

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 603 530**

51 Int. Cl.:

C12N 9/02 (2006.01)

C12N 9/10 (2006.01)

C12N 9/16 (2006.01)

C12N 5/14 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.07.2009 PCT/AU2009/000929**

87 Fecha y número de publicación internacional: **28.01.2010 WO10009499**

96 Fecha de presentación y número de la solicitud europea: **21.07.2009 E 09799857 (9)**

97 Fecha y número de publicación de la concesión europea: **17.08.2016 EP 2315519**

54 Título: **Aceite de semilla de algodón mejorado y usos**

30 Prioridad:

21.07.2008 US 135554 P

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

28.02.2017

73 Titular/es:

**COMMONWEALTH SCIENTIFIC AND INDUSTRIAL
RESEARCH ORGANISATION (100.0%)
Black Mountain Science and Innovation Park,
Clunies Ross Street
Acton ACT 2601, AU**

72 Inventor/es:

**LIU, QING;
GREEN, ALLAN, GRAHAM y
SINGH, SURINDER, PAL**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 603 530 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

Aceite de semilla de algodón mejorado y usos

5 **Campo**

La presente memoria descriptiva se refiere a la producción de plantas y semillas de algodón y aceite preparado a partir de ellas que tienen niveles elevados de ácido oleico y niveles reducidos de ácidos palmítico y linoleico. Además, en el presente documento se describen semillas de algodón que tienen niveles bajos de ácidos grasos de ciclopropano y/o de ciclopropeno y/o niveles reducidos de gossipol. La memoria descriptiva describe también secuencias de nucleótidos y de aminoácidos de FatB y CPA-FAS derivadas del algodón de modo que facilita, entre otras cosas, la modificación directa del contenido y/o la composición de aceite vegetal.

15 **Antecedentes**

El algodón es un cultivo de doble propósito, que produce tanto fibra como semillas como productos agrícolas primarios valiosos. Normalmente, los productos de semilla de algodón, incluidas las cáscaras (26 %), el linter (9 %), el aceite (16 %) y la harina de semilla de algodón (45 %) representan aproximadamente el 15 % del valor de granja de la cosecha de algodón (Cherry y Leffler, Seed. En "Cotton, agronomy monograph No. 24" (eds RJ Kohel, CF Lewis) pp. 511–569. Crop Science Society of America, Madison, WI, USA, 1984), mientras que la hebra proporciona la mayor parte del 85 % restante del valor. El aceite de semilla de algodón es el producto más valioso derivado de la semilla de algodón.

El aceite de semilla de algodón tiene una larga tradición de uso en la elaboración de alimentos. Dado que el aceite de semilla de algodón tiene un sabor blando y neutro que no enmascara el sabor inherente de los alimentos, es un aceite popular y ampliamente utilizado para frituras en el sector de servicios de alimentos para aperitivos y alimentos (Jones y King, Cottonseed Oil. National Cottonseed Products Associations, Inc. and the Cotton Foundation, Memphis, TN, USA, 1993). El aceite de semilla de algodón también se utiliza habitualmente como ingrediente en adobos, aderezos, pasteles, margarinas y mantecas.

El valor nutritivo e industrial del aceite de semilla de algodón, al igual que otros aceites vegetales, se ve afectado por la composición de ácidos grasos en el aceite, es decir, el nivel relativo de cada uno de los ácidos grasos en el aceite y las propiedades conferidas por la longitud de la cadena de carbono y el nivel de insaturación de cada ácido graso.

El aceite de semilla de algodón aislado y purificado se compone principalmente (> 95 %) de triacilglicerol (TAG) que se sintetizan y depositan durante el desarrollo de la semilla. Las moléculas de TAG consisten en tres ácidos grasos esterificados con un esqueleto de glicerol, denominadas las posiciones sn-1, sn-2 y sn-3. En pocas palabras, la biosíntesis *de novo* de ácidos grasos en semillas de algodón, como en otras semillas oleaginosas, se produce en el estroma de plastos durante el desarrollo y el crecimiento de las semillas, es decir, antes de la maduración. A continuación, los ácidos grasos se exportan desde de los plastos en forma de tioésteres de acil-CoA a los sistemas de endomembrana citoplásmicos (retículo endoplásmico, RE) en los que se produce la modificación de los ácidos grasos después de la transferencia de los grupos acilo desde los tioésteres CoA a fosfolípidos mediante aciltransferasas. A esto le sigue el ensamblaje de TAG y el almacenamiento en los oleosotas.

La enzima que contiene biotina acetil-CoA carboxilasa (ACCasa) cataliza la primera etapa comprometida en la vía mediante la activación de acetil-CoA al intermedio de tres carbonos, malonil-CoA, mediante la adición de un grupo carboxilo. A continuación, el grupo malonilo se transfiere desde la CoA a una proteína transportadora de acilo (ACP), que sirve como transportador para la cadena de ácido graso en crecimiento. La malonil-ACP se hace reaccionar con una segunda enzima de condensación de acetil-CoA, cetoacil-ACP sintasa III (KASIII), dando como resultado una cadena de cuatro carbonos. El proceso repetido de la adición de unidades de dos carbonos sobre la cadena de ácidos grasos alargada es catalizado por KASI, lo que conduce a la formación de palmitoil-ACP. La KASII cataliza la elongación de palmitoil-ACP a estearoil-ACP. Una estearoil-ACP $\Delta 9$ -desaturasa introduce el primer doble enlace en el estearoil-ACP para convertirlo en oleoil-ACP en el plasto. La cadena de acilo graso saturada, extendida y el oleato monoinsaturado se escinden del ACP mediante una enzima tioesterasa específica, FatB o FatA, respectivamente, lo que les permite salir del plasto y entrar en el citoplasma. Los ácidos grasos saturados liberados en el citoplasma no se modifican adicionalmente. Sin embargo, el ácido oleico puede modificarse adicionalmente en las membranas del retículo endoplásmico (RE) mediante la acción de las desaturasas unidas a membrana. Las cadenas de acilo unidas a fosfatidilcolina (PC) sirven como sustrato para las enzimas modificadoras de lípidos localizadas en el RE, tales como ácido graso desaturasa 2 (FAD2), que introduce un doble enlace en el ácido oleico en la posición sn-2 de PC para producir ácido linoleico. Todos los grupos acilo grasos modificados y no modificados forman después un grupo mientras están unidos a la CoA. En el algodón, pero no en otras semillas oleaginosas de zonas templadas, se puede utilizar ácido oleico como sustrato para la ciclopropanación catalizada por la ciclopropano ácido graso sintasa para producir ácido dihidroestertérmico. A continuación, este ácido graso se desatura para producir ácido estercúlico y, luego, se oxida en α para producir ácidos malválicos. Finalmente, los grupos acilo grasos se incorporan a lípidos de almacenamiento a través de la vía de Kennedy mediante la esterificación secuencial de glicerol-3-fosfato por la acción de una serie de enzimas de ensamblaje de TAG.

La enzima acil-ACP tioesterasa (FatB) (EC 3.1.2, 14) cataliza la hidrólisis del enlace tioéster entre el resto acilo y ACP en la acil-ACP y la liberación de ácido graso libre en el plasto. Después, el ácido graso libre se vuelve a esterificar a CoA en la envoltura de plasto a medida que se transporta fuera del plasto. FatB pertenece a una clase de enzimas de ácido graso tioesterasa (FAT) solubles codificadas en el núcleo que primero se traducen como proteínas precursoras. Por lo tanto, la especificidad del sustrato de las enzimas FAT en el plasto está implicada en la determinación del espectro de la longitud de cadena y el grado de saturación de los ácidos grasos exportados desde el plasto. Las enzimas FAT pueden clasificarse en dos clases basadas en su especificidad de sustrato y las secuencias de nucleótidos, FatA y FatB (Jones et al., Plant Cell 7: 359–371, 1995). FatA prefiere oleoil-ACP como sustrato, mientras que FatB muestra mayor actividad hacia acil-ACP saturadas de diferentes longitudes de cadena. Los genes que codifican la enzima FatB se aislaron primero de especies de plantas que acumulaban ácidos grasos saturados de longitud de cadena media, tales como ácido láurico (C12:0) del laurel de California (*Umbellularia californica*). La sobreexpresión del gen *FatB1* de la *U. californica* en la canola transgénica condujo a la producción de aceite rico en laurato, comprendiendo el laurato un 50 % de los ácidos grasos totales en el aceite (Voelker et al., Plant Journal 9: 229–241, 1996). Estudios posteriores demostraron que varios ortólogos de *FatB* están presentes en tejidos de plantas, incluso en semillas, con especificidad de sustrato que varía de C8:0-ACP a C18:0-ACP.

El aceite de semilla de algodón producido por el algodón de la altiplanicie (*G. hirsutum*) convencional (silvestre) contiene generalmente aproximadamente 26 % de ácido palmítico (intervalo 22-28 %), 2 % de ácido esteárico, 15 % de ácido oleico (intervalo 13-18 %) y 58 % de ácido linoleico (intervalo 52-60 %) (Cherry, J. Am. Oil Chef. Soc. 60: 360–367, 1983; O'Brien, Cottonseed Oil. En: F.D. Gunstone (Ed.) Vegetable Oils in Food Technology: Composition, Properties and Uses. Blackwell Publishing, Oxford, pp. 203–230, 2002). Además de estos ácidos grasos que también están presentes en la mayoría de los otros cultivos de semillas oleaginosas de zonas templadas, el aceite de semilla de algodón no hidrogenado también contiene niveles bajos (0,5-1 %) de ácidos grasos ciclopropano o ciclopropeno, principalmente ácidos malválico, estercúlico y dihidroestercúlico (Shenstone y Vickery, Nature 190: 68–169, 1961; Cherry, 1983 (citado anteriormente); solicitud de patente japonesa JP S57 208957). Los ácidos grasos de ciclopropano (CPA) y ciclopropeno (CPE) se encuentran en los aceites de semilla de plantas del orden de Malvales y algunas gimnospermas, incluido en algodón. Dos especies de *Sapindaceae*, lichi (*Litchi chinensis*) y ojo de dragón (*Euphoria longan*) contienen hasta un 40 % de ácidos grasos de CPA en sus aceites de semilla, principalmente como ácido dihidroestercúlico (DHS). El aceite de semilla de *Sterculia foetida* contiene hasta un 78 % de ácidos grasos de CPE, principalmente ácidos estercúlicos (STC) y malválicos (MVL). El aceite de semilla de algodón no hidrogenado contiene cantidades relativamente pequeñas (0,5-1,0 %) de ácidos grasos de CPE y de CPA, principalmente en los ejes de los embriones. Asimismo, se observa que las raíces y los hipocótilos de algodón acumula niveles bajos niveles de STC y MVL. CPA y CPE no se encuentran en niveles detectables en los principales cultivos de semillas oleaginosas que no sean algodón, incluyendo en el aceite de palma, de soja, de maíz, de canola, de mostaza, de girasol, de cártamo, de cacahuete, de lino, otras *Brassicás* etc.

La primera etapa comprometida para producir estos ácidos grasos raros es catalizada por una ciclopropano ácido graso sintasa (CPA-FAS) que añade un grupo metileno a través del doble enlace de ácido oleico para producir DHS (Figura 2). En el algodón, la enzima CPA desaturasa desatura la mayor parte del DHS para producir STC, la mayor parte del cual se modifica adicionalmente mediante α -oxidación para formar MVL (Figura 2). El MVL es el ácido graso ciclopropenoide predominante en el aceite de semilla de algodón silvestre.

El nivel relativamente alto de ácidos grasos saturados, principalmente ácido palmítico, en el aceite de semilla de algodón en comparación con los aceites de la mayoría de los otros cultivos de semillas oleaginosas de zonas templadas contribuye a la estabilidad oxidativa del aceite de semilla de algodón, de forma que compensa la mayor inestabilidad de los otros componentes de ácidos grasos insaturados. También imparte el alto punto de fusión necesario para fabricar productos tales como margarina y manteca. Por otra parte, los niveles más altos de ácido palmítico son menos deseables desde el punto de vista nutricional debido a su propiedad de elevar los niveles de colesterol LDL en seres humanos, asociado con un mayor riesgo de cardiopatía cardiovascular (CCV) (Lindsey et al., Exp. Biol. Med. 195: 261–269, 1990). A excepción del aceite de palma, la semilla de algodón contiene el nivel más alto de ácido palmítico (26 %) entre los aceites vegetales básicos principales. Se ha notificado ampliamente que el ácido palmítico y otros ácidos grasos saturados de cadena más corta elevan los niveles plasmáticos totales de colesterol y de colesterol de lipoproteína de baja densidad (Kris–Eherton et al., Nutrition–today (USA). 28: 30–38, 1993). El aceite de semilla de algodón también contiene un nivel elevado de ácido linoleico que es oxidativamente inestable y, por lo tanto, limita la vida útil del aceite y lo hace inadecuado para algunas aplicaciones de alimentos.

Por lo tanto, el aceite de semilla de algodón convencional a menudo se procesa mediante hidrogenación parcial durante la cual el ácido linoleico poliinsaturado se transforma en ácidos grasos monoinsaturados (oleico) y saturados (esteárico) más estables. La hidrogenación parcial da lugar a una serie de cambios estructurales en una fracción de los ácidos grasos, incluyendo el desplazamiento de un doble enlace. Esto puede conducir a la producción de ácidos grasos *trans* (TFA) que son isómeros de los ácidos grasos insaturados de origen natural. En la mayoría de los ácidos grasos insaturados de origen natural, tales como ácido oleico, los dos átomos de hidrógeno unidos a los átomos de carbono de doble enlace están en el mismo lado de los dobles enlaces de la cadena de carbono, y esto se denomina un doble enlace *cis*. Sin embargo, la hidrogenación parcial vuelve a configurar algunos de los dobles enlaces de manera que los dos átomos de hidrógeno están en lados opuestos del doble enlace carbono-carbono, ya que es la forma de energía más baja, y esto se denomina configuración *trans*. Por tanto, la hidrogenación parcial

produce ácidos grasos *trans*, tales como ácido eláidico a partir de ácido oleico. Los ácidos oleico y eláidico contienen el mismo número de átomos (C18:1), con un doble enlace en el mismo lugar, pero es la conformación del doble enlace la que los separa. Los TAG que contienen ácido eláidico, con la configuración de doble enlace *trans*, tiene un punto de fusión mucho más alto que el ácido oleico. En los últimos años se ha reconocido cada vez más que los TFA tienen propiedades a aumento significativo de los niveles de colesterol-LDL y de disminución de los niveles de colesterol-HDL y, por lo tanto, aumentan el riesgo de enfermedades cardiovasculares basándose en pruebas derivadas de estudios epidemiológicos y clínicos (Oomen et al., Lancet 357: 746–751, 2001; Mozaffarian et al., N. Engl. J. Med. 354: 1601–1613, 2006). La hidrogenación parcial convierte también los ácidos grasos ciclopropanoicos o ciclopropenoicos en ácidos grasos de cadena ramificada abriendo el anillo de ciclopropano, de modo que se produce un ácido graso ramificado con un grupo metilo adicional unido al C9 o C10 de la cadena carbonada del ácido graso.

En comparación con los ácidos grasos poliinsaturados, el ácido oleico es más estable frente a la oxidación tanto a temperaturas de almacenamiento ambientales como a temperaturas altas utilizadas en la cocción y fritura de alimentos. Los estudios con una serie de aceites vegetales, tales como aceite de cártamo y aceite de soja, indican que los aceites vegetales con alto contenido de ácido oleico desarrollan rancidez o se descomponen oxidativamente más lentamente durante el almacenamiento o durante la fritura u otro uso, en comparación con los aceites que contienen cantidades altas de ácidos grasos poliinsaturados (Fuller et al., J. Food Sci. 31: 477–480, 1966; Mounts et al., J. Am. Oil Chem. Soc. 65: 624–628, 1998).

Como se ha mencionado anteriormente, el aceite de semilla de algodón se caracteriza también por la presencia de pequeñas cantidades de ácidos malválicos, estercúlico y dihidroestercúlico, grupo de ácidos grasos de propano cíclicos derivado del ácido oleico. Se sabe que los ácidos malválico y estercúlico son potentes inhibidores de la $\Delta 9$ -estearoil-CoA desaturasa animal. Aunque los ácidos grasos de CPA y CPE no son estables y se eliminan en su mayor parte durante el procesamiento del aceite, en particular mediante hidrogenación, el aceite residual en la harina y la semilla de algodón completa utilizada en la industria alimentaria podrían tener efectos negativos sobre la salud animal. Se piensa que la alimentación de animales de granja con cantidades excesivas de semilla de algodón posiblemente cause una serie de problemas de salud para los animales y puede afectar a la calidad de los productos animales, tales como la solidificación de las grasas en la yema de huevo y la leche (Johnson et al., Nature 214: 1244–1245, 1967; Roehm et al., Lipids 5: 80–84, 1970). Se han desarrollado métodos para inactivar los ácidos grasos ciclopropenoicos a través de procesos de hidrogenación parcial especializados. Merker y Mattil, 1965 informaron sobre un proceso de hidrogenación en el que los ácidos malválico y estercúlico se redujeron de forma selectiva en sus derivados dihidro o tetrahidro, mediante un catalizador de níquel, sin reducción significativa de la formación de ácido linoleico o de ácido *trans*. Hutchins et al., Journal of American Oil Chemists Society 45: 397–399, 1968 mostraron hidrogenación selectiva de los grupos ciclopropenoicos en aceite de semilla de algodón por medio de un reactor de lecho relleno y catalizadores de níquel en condiciones más suaves. Sin embargo, estos procesos de hidrogenación añaden costes adicionales para el procesamiento del aceite y no son deseables.

En la década de 1970, el programa de cultivo del algodón de la serie Acala SJ en California (Cherry, 1983 (citado anteriormente)) redujo el ácido palmítico de 23,3 a 22,7 %, aumentó el ácido oleico del 16,6 % al 17,3 % y redujo los ácidos grasos cíclicos totales del 0,9 % al 0,8 % en el aceite de semilla de algodón. Sin embargo, en comparación con los logros obtenidos en otros cultivos de semillas oleaginosas, estos cambios fueron menores, lo que refleja la estrecha base genética de las variedades de algodón de élite como resultado de la selección persistente de rasgos distintos de la calidad del aceite.

Se aislaron cuatro ADNc diferentes que codificaban FAD2 a partir de algodón (Liu et al., Australian Journal of Plant Physiology 26: 101–106, 1999a; Liu et al., Plant Physiol. 120: 339, 1999b; Pirtle et al., Biochim. Biophys. Acta 1522: 122–129, 2001), entre los cuales se determinó que *ghFAD2-1* desempeña un papel principal en la producción de ácido linoleico en el aceite de semilla de algodón. El análisis de la expresión génica sugirió que el gen *ghFAD2-1* se expresaba específicamente en semillas en desarrollo, con expresión máxima durante la etapa de madurez media del desarrollo de semillas (Liu et al., 1999a (citado anteriormente)).

La patente de Estados Unidos n.º 6974898 describe la generación de aceite de semilla de algodón que contiene hasta un 77 % de ácido oleico mediante regulación por disminución de la $\Delta 12$ desaturasa microsomal (FAD2) mediante métodos de ARNi.

Chapman et al., J. Am. Oil Chem. Soc. 78: 941–947, 2001, transformaron plantas de algodón con una construcción que codifica un alelo mutante no funcional de *fad2*, obtenido a partir de semilla de colza y expresado a partir de un promotor de faseolina específico de semilla en orientación de sentido. Aproximadamente la mitad de las 43 líneas transgénicas que se generaron mostraron un ligero aumento del contenido de ácido oleico en el aceite de la semilla, que variaba del 24–30 %. En las generaciones siguientes, se identificaron semillas con niveles más altos de ácido oleico hasta el 40 %.

Los mutantes *fatB* de *Arabidopsis* que tenían aceite de semilla con niveles reducidos de ácido palmítico mostraron un crecimiento vegetativo reducido a temperaturas normales (Bonaventure et al., Plant Cell 15: 1020–1033, 2003). Sin embargo, en estudios anteriores con construcciones antisentido para regular por disminución AtFatB en

Arabidopsis, utilizando el promotor de CaMV35S para expresar el ARN antisentido, se observó una reducción sustancial de los niveles de ácido palmítico solo en las flores y las semillas, no en las hojas ni en las raíces, y no hubo un fenotipo visual obvio (Dormann et al., *Plant Physiol.* 123: 637–644, 2000). La regulación por disminución de FatB también se ha descrito en soja, cáñola o arroz para alterar la composición de ácidos grasos (solicitud de patente internacional WO 2004/000871, patente de Estados Unidos n.º 5.955.650, solicitud de patente internacional WO 2008/006171; solicitud de patente de Estados Unidos n.º 2004/0107460 A1).

En el algodón, como en muchas plantas templadas, las enzimas FatB parecen estar codificadas por una familia multigénica. Yoder et al., *Biochimica et Biophysica Acta* 1446: 403–413, 1999, aislaron una secuencia de ADN genómico que se pensaba que codificaba una acilACP tioesterasa en el algodón.

Wilson et al., *J. Am. Oil Chem. Soc.* 78: 335–340, 2001 y Buhr et al., *Plant J.* 30: 155–163, 2002, utilizaron enfoques antisentido y de ribozimas, respectivamente, para reducir el ácido palmítico en las semillas de soja.

Recientemente se aisló un gen de CPA–FAS de *Sterculia foetida* utilizando un abordaje basado en EST (Bao et al., *Proc. Natl. Acad. Sci. EE.UU.* 99: 7172–7177, 2002; Bao et al., *J. Biol. Chem.* 278(15): 12846–12853, 2003; solicitud de patente de Estados Unidos US 2006/0053512). Se identificaron dos secuencias de EST de *G. hirsutum* que se expresaban de forma diferencial mediante la infección de las raíces/hipocótilos de algodón después de la inoculación de *Fusarium* (Dowd et al., *Molecular Plant–Microbe Interactions.* 17: 654–667, 2004).

Por lo tanto, existe la necesidad de aceite de semilla de algodón mejorado.

Sumario

A lo largo de esta memoria descriptiva, a menos que el contexto requiera lo contrario, se entiende que la palabra "comprenden" o variaciones tales como "comprende" o "que comprende" implica la inclusión de un elemento o número entero o grupo de elementos o de números enteros indicados, pero no la exclusión de cualquier otro elemento, número entero o grupo de elementos o de números enteros. Cada realización debe aplicarse *mutatis mutandis* a cualquier otra realización a menos que se indique expresamente lo contrario.

Como se usa en el presente documento, las formas del singular "un" "uno/una" y "el/la" incluyen los aspectos en plural a menos que el contexto indique claramente lo contrario. Por tanto, por ejemplo, la referencia a "una mutación" incluye una única mutación, así como dos o más mutaciones; la referencia a "un agente" incluye un agente, así como dos o más agentes; etc.

Los detalles bibliográficos de las publicaciones a las que el autor hace referencia en la presente memoria descriptiva se recogen al final de la descripción.

La referencia en esta memoria descriptiva a cualquier publicación previa (o información derivada de ella), o a cualquier asunto que se conozca, no se toma y no debe tomarse como un reconocimiento o admisión o cualquier forma de sugerencia de que esa publicación previa (o información derivada de ella) o materia conocida forma parte del conocimiento general habitual en el campo de esfuerzo al que se refiere la presente memoria descriptiva.

La expresión "derivado de" indica que el número entero especificado se obtiene de una fuente específica aunque no necesariamente directamente de esa fuente.

La designación de secuencias de aminoácidos de ejemplo se muestra en la Tabla 5.

Se hace referencia a las secuencias de nucleótidos y de aminoácidos con un número identificador de secuencia (SEQ ID NO:). Las SEQ ID NO: corresponden numéricamente a los identificadores de secuencia <400> 1 (SEQ ID NO: 1), <400> 2 (SEQ ID NO: 2), etc. En la tabla 4 se proporciona un resumen de identificadores de secuencia. Después de las reivindicaciones se proporciona un listado de secuencias.

En el presente documento, los genes y otro material genético (por ejemplo, ARNm, construcciones de ácidos nucleicos, etc.) se representan en cursiva, mientras que sus productos de expresión proteicos no se representan en cursiva.

En una realización, la presente memoria descriptiva describe un método para producir aceite de semilla de algodón modificado, que comprende las etapas de (i) obtener semilla de algodón, (ii) extraer el aceite de la semilla de algodón, y (iii) recuperar el aceite de semilla de algodón, en el que el aceite de semilla de algodón tiene una composición modificada de ácidos grasos tal que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico. En algunas realizaciones, del 74 % al 86 % del contenido total de ácidos grasos del aceite de semilla de algodón es ácido oleico. En otras realizaciones, del 4 % al 12 % es ácido linoleico. En relación con los ácidos grasos *trans*, en algunas realizaciones, el aceite de semilla de algodón tiene menos del 0,5 %, preferentemente menos del 0,1 %, del contenido total de ácidos grasos es un ácido graso *trans*. En algunas realizaciones, el ácido graso *trans* es ácido

elaídico. En algunas realizaciones, del 0,1 % al 0,5 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido graso de ciclopropano (CPA) o ácido graso de ciclopropeno (CPE). En realizaciones particulares, el ácido graso de ciclopropano o de ciclopropeno es ácido malválico, ácido estercúlico, ácido dihidroestertérmico o cualquier combinación de dos de estos o los tres de estos. En otras realizaciones, menos del 0,5 %, preferentemente menos del 0,1 %, del contenido total de ácidos grasos es un ácido graso ramificado en C9 o C10. En realizaciones preferentes, el contenido total de ácidos grasos del aceite de semilla tiene la composición de 74 % al 86 % de ácido oleico, del 4 % al 10 % de ácido palmítico, del 4 % al 12 % de ácido linoleico y, opcionalmente, del 0,1 % al 0,5 % en total de ácido graso de ciclopropano (CPA) y de ácido graso de ciclopropeno (CPE), y, opcionalmente, menos de 1,0 %, preferentemente menos de 0,5 %, más preferentemente menos de 0,1 %, del contenido total de ácidos grasos es un ácido graso *trans*.

En algunas realizaciones, el método de producción comprende además recoger la pelusa que comprende la semilla de algodón de una planta de algodón y/o desmotar la semilla de algodón a partir de la pelusa que comprende la semilla de algodón. Como alternativa o además de, la etapa (ii) comprende triturar la semilla de algodón y/o extraer el disolvente de la semilla de algodón triturada. En una realización adicional, la etapa (iii) comprende purificar el aceite de semilla de algodón. La purificación del aceite de semilla puede comprender desgomar el aceite, decolorar el aceite, desodorizar el aceite, alterar el pH del aceite, hidrolizar el aceite o cualquier combinación de los mismos.

De forma ventajosa, en algunas realizaciones, el método no comprende una etapa de hidrogenación o hidrogenación parcial del aceite de semilla de algodón. Sin limitarse a ningún modo particular de acción, el bajo nivel de ácido linoleico y el nivel relativamente alto de ácido oleico evitan la necesidad de estabilizar los ácidos grasos, evitando de este modo la producción *de novo* de ácidos grasos *trans* inducida por algunas formas de hidrogenación.

Por consiguiente, en otro aspecto, la memoria descriptiva proporciona aceite de semilla de algodón obtenible mediante extracción de la semilla de algodón como se describe en el presente documento, en el que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, en el que el ácido graso de ciclopropano (CPA) y/o el ácido graso de ciclopropeno (CPE) están presentes en dicho aceite de semilla de algodón, y en el que la composición de ácidos grasos de dicho aceite no se ha modificado después de la extracción. En algunas realizaciones, del 74 % al 86 % del contenido total de ácidos grasos del aceite de semilla de algodón es ácido oleico. En otras realizaciones, del 4 % al 12 % es ácido linoleico. En algunas realizaciones, del 0,1 % al 0,5 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido graso de ciclopropano (CPA) o ácido graso de ciclopropeno (CPE). En realizaciones particulares, el ácido graso de ciclopropano o de ciclopropeno es ácido malválico, ácido estercúlico, ácido dihidroestertérmico o cualquier combinación de dos de estos o los tres de estos. En realizaciones preferentes, el contenido total de ácidos grasos del aceite de semilla tiene la composición de 74 % al 86 % de ácido oleico, del 4 % al 10 % de ácido palmítico, del 4 % al 12 % de ácido linoleico y, opcionalmente, del 0,1 % al 0,5 % en total de ácido graso de ciclopropano (CPA) y de ácido graso de ciclopropeno (CPE), y, opcionalmente, menos de 1,0 %, preferentemente menos de 0,5 %, más preferentemente menos de 0,1 %, del contenido total de ácidos grasos es un ácido graso *trans*.

En otro aspecto, la presente memoria descriptiva proporciona semilla de algodón que tiene una composición de ácido graso modificada en su aceite, en el que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico. En algunas realizaciones, del 74 % al 86 % del contenido total de ácidos grasos del aceite de semilla de algodón es ácido oleico. En otras realizaciones, del 4 % al 12 % es ácido linoleico. En todavía otras realizaciones, del 0,1 % al 0,5 % del contenido total de ácidos grasos en el aceite es un ácido graso de ciclopropano (CPA). En todavía otras realizaciones adicionales, la semilla de algodón tiene un nivel reducido de gossipol, en el que el nivel de gossipol en la semilla de algodón se reduce en al menos un 10 % con respecto al nivel de gossipol en la semilla de algodón de la variedad de algodón Coker. En realizaciones preferentes, el contenido total de ácidos grasos del aceite de semilla tiene la composición de 74 % al 86 % de ácido oleico, del 4 % al 10 % de ácido palmítico, del 4 % al 12 % de ácido linoleico y, opcionalmente, del 0,1 % al 0,5 % en total de ácido graso de ciclopropano (CPA) y de ácido graso de ciclopropeno (CPE), y, opcionalmente, menos de 1,0 %, preferentemente menos de 0,5 %, más preferentemente menos de 0,1 %, del contenido total de ácidos grasos es un ácido graso *trans*.

La memoria descriptiva proporciona semilla de algodón o aceite de semilla de algodón obtenible mediante extracción a partir de la misma, en la que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, y en la que el ácido graso de ciclopropano (CPA) y/o el ácido graso de ciclopropeno (CPE) están presentes en dicho aceite de semilla de algodón, y en la que la composición de ácidos grasos de dicho aceite no se ha modificado después de la extracción y en la que la semilla de algodón tiene una composición de ácidos grasos en su aceite, en la que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico, del 4 al 10 % es ácido palmítico y del 4 al 15 % es ácido linoleico y en la que la semilla de algodón es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son ghFAD2-1, ghFatB-2 y ghCPA-FAS-2, en el que el gen ghFATB-2 es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de la secuencia de ADNc mostrada en la SEQ ID NO: 5 y en la que el gen ghCPA-FAS-2 es una molécula de ácido

nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de secuencia de ADNc mostrada en la SEQ ID NO: 12. En algunas realizaciones, la semilla de algodón es transgénica para una construcción genética que codifica una molécula de ARN que inhibe la expresión de dos o más genes que se seleccionan del grupo que consiste en FAD2, FatB-2 y CPA-FAS-2. En algunas realizaciones los genes son *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2*. En una realización ilustrativa, la construcción genética comprende la construcción MonoCott descrita esquemáticamente en la Figura 12. En otra realización, la semilla de algodón es transgénica para una construcción genética que codifica una molécula de ARN que inhibe la expresión de dos genes que son *ghFAD2-1* y *ghFatB-2*. En otra realización ilustrativa, la semilla de algodón es de la especie *Gossypium hirsutum* o *Gossypium barbadense*.

En otro aspecto, la memoria descriptiva describe un método para producir la semilla de algodón descrita en el presente documento, que comprende recoger la pelusa de una planta de algodón y desmotar la pelusa, produciendo de este modo la semilla de algodón. En otra realización, el método comprende sembrar la semilla de algodón descrita en el presente documento y permitir que la semilla crezca en la planta de algodón.

La presente invención proporciona una planta de algodón que es capaz de producir la semilla de algodón descrita en el presente documento, en la que la planta de algodón es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son *ghFAD2-1*, *ghFatB-2* y *GhCPA-FAS-2* en la que el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:5 y en la que el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteína de la secuencia de ADNc mostrada en la SEQ ID NO:12. La presente memoria descriptiva también describe pelusa de algodón obtenida de dicha planta de algodón.

De acuerdo con la presente invención, la memoria descriptiva proporciona un método para identificar la semilla de algodón que tiene una composición de ácido graso modificada en su aceite, que comprende (i) obtener semillas de algodón transgénicas, (ii) determinar la composición de ácidos grasos del aceite de la semilla de algodón y (iii) si del 72 % al 88 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, seleccionando la semilla de algodón, en la que la semilla de algodón de la etapa (i) es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2*, en la que el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:5 y en la que el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:12. En algunas realizaciones, el método comprende además la transformación de una célula de algodón con una construcción genética y la regeneración de una planta de algodón transgénico antes de la etapa (i). En algunas realizaciones, la semilla de algodón de la etapa (i) es transgénica para una construcción genética que codifica una molécula de ARN que inhibe la expresión de dos o tres genes que son *ghFAD2-1* y *ghFatB-2* o *ghFAD2-1* y *ghFatB-2* y *ghCPA-FAS-2*. En algunas realizaciones, el método comprende además las etapas de producir una planta de algodón a partir de la semilla y cruzar dicha planta de algodón con una segunda planta de algodón, preferentemente de un fondo genético diferente.

Por consiguiente, la memoria descriptiva contempla el uso de la semilla de algodón descrita en el presente documento para producir aceite aislado o harina de semilla de algodón y el uso de la semilla de algodón como alimento para animales o en la producción de un producto alimentario o como combustible. La presente invención también contempla el uso del aceite de semilla de algodón descrito en el presente documento en la producción de un producto alimentario. Se contempla que el aceite se use para aumentar el contenido de ácido oleico o para reducir el contenido de ácido palmítico del producto alimentario. Como alternativa o además de, el aceite se usa para reducir el nivel o la absorción de ácidos grasos saturados o ácidos grasos trans en el producto alimentario o por el animal. Se entenderá que esto se refiere al uso de una cantidad igual de una semilla de algodón, aceite o harina de semilla de algodón sin modificar.

La memoria descriptiva también describe un producto alimentario que comprende un ingrediente alimentario que es el aceite de semilla de algodón como se describe en el presente documento. En otro aspecto, la memoria descriptiva proporciona un método para preparar un producto alimentario, que comprende mezclar el aceite de semilla de algodón tal como se describe en el presente documento con otros ingredientes alimentarios.

En otro aspecto, la invención proporciona un método para preparar un producto alimentario frito, que comprende freír un producto alimentario en el aceite de semilla de algodón tal como se describe de acuerdo con el presente documento.

El producto objeto puede usarse terapéuticamente y la presente memoria descriptiva describe un método para tratar o prevenir una enfermedad en un sujeto, que comprende administrar por vía oral al sujeto, el aceite de semilla de algodón, la semilla de algodón o el producto alimentario como se describe en el presente documento. En algunos

- aspectos, la enfermedad es obesidad, enfermedad cardíaca, alto contenido sérico de TAG, niveles altos de colesterol en suero, niveles altos de LDL en suero, niveles altos de HDL en suero, hipertensión, cáncer, diabetes, o cualquier combinación de los mismos. Se abarca cualquier sujeto que pueda beneficiarse de los presentes métodos o composiciones. El término "sujeto" incluye, sin limitaciones, seres humanos y primates no humanos, animales de ganado, animales de compañía, animales de laboratorio, animales silvestres en cautiverio, reptiles y anfibios, peces, aves y cualquier otro organismo. Se puede hacer referencia a un sujeto, independientemente de si es un organismo humano o no humano, como paciente, individuo, sujeto, animal, huésped o receptor. En un aspecto en concreto, el sujeto es un ser humano.
- En un aspecto en concreto, la memoria descriptiva describe una molécula de ácido nucleico quimérico que codifica un ácido nucleico silenciador de genes que reduce la expresión génica de *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2* de algodón, en el que el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en SEQ ID NO: 5 y en el que el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en la SEQ ID NO: 12. En algunas realizaciones, la molécula de ácido nucleico comprende: (i) una secuencia contigua de nucleótidos expuesta en la SEQ ID NO: 17 o una forma complementaria de la misma, o una secuencia que tiene un 90 % de identidad con la misma o una secuencia que hibrida con un complemento de la misma en condiciones de hibridación de alta rigurosidad o (ii) una variante de la molécula de ácido nucleico de (i) que carece de las secuencias que codifican para el polipéptido *CPA-FAS-2* o sus formas complementarias. En una realización ilustrativa, la molécula de ácido nucleico silenciadora génica es un ARN en horquilla. En algunas realizaciones, la molécula de ácido nucleico se proporciona dentro de una construcción genética.
- En algunas realizaciones, la construcción génica comprende el ácido nucleico y comprende además la estructura expuesta en la figura 12 o las secuencias de los fragmentos A, E, F, G y H expuestas en la figura 13 o una variante funcional de la misma. En otras realizaciones, la molécula de ácido nucleico se proporciona en una célula huésped. En otras realizaciones, se proporciona una planta o tejido o célula de la misma que comprende un ácido nucleico sujeto. En algunas realizaciones, se usan las moléculas de ácido nucleico para reducir la expresión de los genes *FAD2-1* de algodón, *FatB-2* de algodón y/o *CPA-FAS-2* de algodón en una planta de algodón. La memoria descriptiva proporciona, por tanto, una planta de algodón o semilla de la misma modificada que comprende un nivel reducido de uno o más genes endógenos seleccionados de *FAD2* de algodón, *FatB* de algodón y *CPA-FAS* de algodón, preferentemente dos o más de dichos genes o de los tres genes, en los que el aceite de semilla de algodón de la misma muestra niveles elevados de ácido oleico o niveles reducidos de ácido palmítico y ácido linoleico.
- En una realización adicional, el aceite de semilla de algodón exhibe niveles reducidos de ácidos grasos de ciclopropano y/o de ciclopropano o gossipol.
- En otro aspecto, la memoria descriptiva proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de *FatB* de algodón o un fragmento funcional o una variante del mismo, o una forma complementaria del mismo, estando dicha secuencia nucleotídica seleccionada del grupo que consiste en: (i) una secuencia de nucleótidos que tiene al menos un 90 % de identidad con la secuencia de nucleótidos de la región de codificación proteica de la secuencia de nucleótidos expuesta en la SEQ ID NO: 5 o 7; (ii) una secuencia de nucleótidos que codifica un polipéptido que tiene al menos un 80 % de identidad de secuencia con la secuencia de aminoácidos expuesta en la SEQ ID NO: 6 u 8; (iii) una secuencia de nucleótidos que hibrida en condiciones de hibridación de rigurosidad como mínimo moderada con al menos 30 nucleótidos contiguos derivados de la SEQ ID NO: 5 o 7; (iv) una secuencia de nucleótidos que es complementaria a (i), (ii) o (iii); y (v) una secuencia de nucleótidos que comprende los nucleótidos 548-843 de la SEQ ID NO: 5. En algún aspecto, el ácido nucleico, en particular el ARN, o una variante o fragmento funcional del mismo, se proporciona como una molécula sentido, antisentido, ribozima, de cosupresión o ARN bicatenario u otro agente silenciador de genes. Por lo tanto, la memoria descriptiva proporciona una molécula antisentido o de cosupresión que comprende un fragmento de uno cualquiera de (i) - (v) anteriores que se expresa y reduce la expresión de un gen *FatB* de algodón.
- En otro aspecto, se proporciona una sonda o cebador aislado que comprende al menos 15 nucleótidos contiguos derivados de la SEQ ID NO: 5 o la SEC ID N° 7 o una forma complementaria de las mismas.
- En otro aspecto, la memoria descriptiva proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de *CPA-FAS* de algodón o un fragmento funcional o una variante del mismo, o una forma complementaria del mismo, estando dicha secuencia nucleotídica seleccionada del grupo que consiste en: (i) una secuencia de nucleótidos que tiene al menos un 90 % de identidad con la secuencia de nucleótidos de la región de codificación proteica de la secuencia de nucleótidos expuesta en la SEQ ID NO: 10, 12 o 14; (ii) una secuencia de nucleótidos que codifica un polipéptido que tiene al menos un 80 % de identidad de secuencia con la secuencia de aminoácidos expuesta en la SEQ ID NO: 11, 13 o 15; ; (iii) una secuencia de nucleótidos que hibrida en condiciones de hibridación de rigurosidad como mínimo moderada con al menos 30 nucleótidos contiguos derivados de la SEQ ID NO: 10, 12 o 14; (iv) una secuencia de nucleótidos que es

complementaria a (i), (ii) o (iii); y (v) una secuencia de nucleótidos que comprende 442 nucleótidos de las secuencias que codifican CPA-FAS-2 expuestas en el fragmento B de la figura 13. En algunos aspectos, el ácido nucleico, en particular el ARN o una variante o fragmento funcional del mismo se proporciona como una molécula sentido, antisentido, ribozima, de cosupresión o ARN bicatenario u otro agente silenciador de genes. Por lo tanto, la memoria
 5 descriptiva proporciona una molécula antisentido o de cosupresión que comprende un fragmento de uno cualquiera de (i) - (v) anteriores que se expresa y reduce la expresión de un gen CPA-FAS-2 de algodón.

En otro aspecto, se proporciona una sonda o cebador aislado que comprende al menos 15 nucleótidos contiguos derivados de la SEQ ID NO: 10, la SEC ID N° 12 o la SEC ID N° 14 o una forma complementaria de las mismas.

10 El sumario anterior no es ni debe verse de ninguna manera como una lista exhaustiva de todas las realizaciones de la presente invención.

Breve descripción de los dibujos

15 La figura 1 es un diagrama esquemático simplificado de la biosíntesis de ácidos grasos en el desarrollo de semillas de algodón que muestran importantes etapas enzimáticas: 1. ceto-acil sintasa III (KASIII), 2. ceto-acil sintasa I (KASI), 3. ceto-acil sintasa II (KASII), 4. $\Delta 9$ -estearoil-ACP desaturasa (SAD), 5. oleoil-ACP tioesterasa, 6. acil-ACP tioesterasa, 7. $\Delta 12$ -oleoil-lípido desaturasa (FAD2), 8. ciclopropano ácido graso sintasa (CPA-FAS), 9. ciclopropano ácido graso desaturasa (CPA-FAD), 10. α -oxidasa. DHS: ácido dihidroestercúlico; STC, ácido estercúlico; MVL, ácido malvático.

20 La Figura 2 es una representación esquemática de la biosíntesis de los ácidos grasos de ciclopropano (CPA) y de ciclopropeno (CPE) en semillas de algodón en desarrollo. El ácido oleico es el sustrato de ácido graso para la ciclopropano ácido graso sintasa (CPA-FAS). S-adenosil-L-metionina es el donante de metileno y la ciclopropanación de ácido oleico es catalizada por CPA-FAS, produciendo ácido dihidroestercúlico (DHS). El DHS puede modificarse adicionalmente mediante la desaturación a través de la enzima CPA desaturasa para producir ácido estercúlico (STC). Posteriormente, el ácido malvático (MVL) se produce mediante α -oxidación de STC catalizada por una oxidasa.

30 La Figura 3 es una representación de la secuencia de nucleótidos del ADNc correspondiente al gen *ghFAD2-1* y la secuencia de aminoácidos predicha del polipéptido de *ghFAD2-1* precursor codificado. La región que codifica la proteína es desde los nucleótidos 73 a 1227 (codón de terminación TAA).

35 La Figura 4 es una representación de la secuencia de nucleótidos del ADNc de la acil-ACP tioesterasa (*ghFatB-2*) de algodón y la secuencia de aminoácidos predicha del polipéptido precursor *ghFatB-2* codificado por el ADNc. La región que codifica la proteína es desde los nucleótidos 73 a 1472 (codón de terminación TAG).

40 La Figura 5 es una representación de la secuencia de nucleótidos del ADNc de la acil-ACP tioesterasa (*ghFatB-3*) de algodón y la secuencia de aminoácidos predicha del polipéptido precursor *ghFatB-3* codificado por el ADNc. La región que codifica la proteína es desde los nucleótidos 97 a (codón de terminación TAG).

45 La Figura 6 es una representación de la secuencia de ADN de la secuencia EST CD486555 a la que se hace referencia en el Ejemplo 4.

La Figura 7 es una representación de la secuencia de ADN de *ghCPA-FAS-1* y los polipéptidos putativos que codifica.

50 La Figura 8 es una representación de la secuencia de ADN de *ghCPA-FAS-2* y los polipéptidos putativos que codifica.

La Figura 9 es una representación de la secuencia de ADN de *ghCPA-FAS-3* y los polipéptidos putativos que codifica.

55 La Figura 10 es una representación del análisis de transferencia Southern de los genes CPA-FAS en algodón. Los ADN genómicos totales de varias especies diploides o tetraploides de *Gossypium* se digirieron con HindIII. Calle 1, *G. barbadense* (genomas A+D); calle 2, *G. hirsutum* cv. Deltapine-16 (genomas A+D); calle 3, *G. hirsutum* cv Sicala-V4 (genomas A+D); calle 4, *G. herbaceum* (genoma A); calle 5, *G. raimondii* (genoma D).

60 La Figura 11 es una representación del análisis de transferencia Northern de ARN de la expresión de los genes *ghCPA-FAS-1* (A) *ghCPA-FAS-2* (B) en algodón. Las muestras de ARN transferidas a la membrana se obtuvieron de: calle 1, cotiledón; 2, hipocótilo; 3, hoja; 4, raíz; 5-9, embriones de algodón en desarrollo a los 25, 30, 35, 40, 45 días después de la antesis, respectivamente; 10, eje embrionario.

65 La Figura 12 es un diagrama esquemático del transgén quimérico utilizado para generar silenciamiento de tres genes en algodón (construcción MonoCott). Las secuencias de ADN de los fragmentos marcados en la parte

superior de la construcción se detallan en la Figura 11. (Abreviaturas: LB: ADN-T borde izquierdo; LecP: promotor de la lectina de soja; *ghFAD2-1*: ADNc de oleoil $\Delta 12$ desaturasa microsomal de algodón; *ghCPA-FAS-2*: ciclopropano ácido graso sintasa de algodón; *ghFatB-1*: palmitoil-ACP tioesterasa de algodón; lec-T: terminador de lectina de soja; RB2: borde derecho ADN-T 2; S1: promotor del segmento 1 del virus de la atrofia subterránea del trébol; S3: terminador del segmento 3 del virus de la atrofia subterránea del trébol; NPTII: gen de la fosfotransferasa de la neomicina. Los tamaños del ADNc de longitud completa y los fragmentos seleccionados de cada gen diana en la fabricación de la construcción de ARNi también se indicaron).

La Figura 13 es una representación de las secuencias de ADN utilizadas para hacer la construcción del ARNi ilustrada en la Figura 12. El fragmento A es el promotor Lec1 específico de la semilla; el fragmento B es el fragmento quimérico de los tres genes diana, incluyendo *ghFAD2-1* (subrayado), *ghCPA-FAS-2* (fuente en cursiva) y *ghFatB-2* (fuente en negrita); el fragmento C es el intrón **ghFAD2-1** utilizado para separar las repeticiones invertidas. El fragmento D es la repetición invertida del fragmento B; el fragmento E es la secuencia de terminación de la transcripción/poliadenilación del gen de la lectina; el fragmento F es el promotor s1 que controla la expresión del gen marcador seleccionable (F + G + H); el fragmento G es la región de codificación de la proteína *NPTII* que confiere resistencia a la kanamicina y se utiliza como marcador seleccionable; fragmento H, la secuencia de terminación de la transcripción/ poliadenilación s3.

La Figura 14 es una representación gráfica de la acumulación de los ácidos palmítico y oleico en semillas individuales de tres generaciones de DCS9-34: (Semillas T_2 transportadas en la línea transgénica primaria (A); semillas T_3 transportadas en la planta T_2 seleccionada (B) y semillas T_4 transportadas en una planta T_3 seleccionada (C).

La Figura 15 es una representación gráfica de la acumulación de ácidos grasos cíclicos totales en los ejes de embriones extirpados de semillas individuales de algodón. DCS9-34 es la línea MonoCott con rasgos HO, LP y niveles bajos de ácidos grasos cíclicos; LY-2 es la línea con los rasgos HO y LP; y un incremento de los ácidos grasos cíclicos; Coker315 es el algodón convencional o control.

Descripción detallada

La presente memoria descriptiva se basa en la producción de plantas de algodón modificadas en las que las plantas se modifican usando construcciones silenciadoras de genes de ARN en horquilla quimérico. Las progenies estables se caracterizan por comprender aceite que tiene las propiedades altamente deseables, incluyendo niveles significativamente más bajos de ácido palmítico y ácido linoleico, y niveles significativamente más altos de ácido oleico que los controles, y que los niveles que se encuentran normalmente en aceite de semilla de algodón extraído de variedades de cultivos de algodón comercialmente útiles. La expresión génica reducida de *CPA-FAS-2* facilitó una reducción en el nivel de los ácidos grasos ciclopropenoides y la presencia de CPA y/o CPE en el aceite de semilla de algodón. Aunque la presente invención se ilustra utilizando construcciones de silenciamiento de genes de ARN en horquilla, no está limitada y otros métodos conocidos por los expertos en la técnica, tales como los descritos en el presente documento, también se contemplan para su uso en la producción de plantas y productos derivados de la misma con las ventajas divulgadas en el presente documento.

La referencia en el presente documento al aceite de semilla de algodón se refiere a aceite endógeno de semilla de algodón y aceite de semilla de algodón extraído de semillas de algodón derivadas de las plantas de algodón modificadas descritas en el presente documento. Por lo tanto, la presente invención no se refiere a la modificación posterior a la extracción del aceite de semilla de algodón que altera de forma sustancial la composición de ácidos grasos.

La presente invención implica la modificación de la actividad génica y la construcción y uso de genes quiméricos. Tal como se utiliza en el presente documento, el término "gen" incluye cualquier secuencia de desoxirribonucleótidos que incluye una región de codificación de proteínas o que se transcribe en una célula pero no se traduce, así como las regiones reguladoras no codificantes asociadas. Tales regiones asociadas normalmente están situadas adyacentes a la región de codificación o la región transcrita en ambos extremos 5' y 3' para una distancia de aproximadamente 2 kb a cada lado. A este respecto, el gen puede incluir señales de control, tales como promotores, potenciadores, señales de terminación y/o poliadenilación que están asociadas de forma natural con un gen dado o señales de control heterólogas, en cuyo caso el gen se denomina "gen quimérico". Las secuencias que están localizadas en 5' de la región de codificación y que están presentes en el ARNm se denominan secuencias 5' no traducidas. Las secuencias que están localizadas en 3' o aguas abajo de la región de codificación y que están presentes en el ARNm se denominan secuencias 3' no traducidas. El término "gen" abarca tanto ADNc como las formas genómicas de un gen.

Un "gen *FAD2-1* de algodón", "gen *ghFAD2-1*" o similar, tal como se utiliza en el presente documento se refiere a una secuencia de nucleótidos que codifica la oleoil-A12-desaturasa en algodón, que se expresa en semillas de algodón en desarrollo. Un gen *ghFAD2-1* se puede distinguir fácilmente de los genes que codifican otras oleoil-A12-desaturasas u otras proteínas por los expertos en la técnica, en particular de un gen *ghFAD2-2*, por ejemplo la SEQ ID NO: 26. Los genes *FAD2-1* de algodón incluyen los alelos o variantes de origen natural que existen en el

algodón, incluyendo los codificados por los genomas A y D del algodón tetraploide, así como las variantes no naturales que pueden producir los expertos en la técnica de la modificación genética. Un ejemplo de una variante de origen natural de *ghFAD2-1* de algodón es la secuencia mostrada en el presente documento como SEQ ID NO: 24, que tiene aproximadamente un 96 % de identidad a lo largo de su longitud completa con la SEQ ID NO: 1. En una realización preferente, un gen *ghFAD2-1* se refiere a una molécula de ácido nucleico, que puede estar presente en el algodón o aislada del mismo o derivada del mismo, que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en la SEQ ID NO: 1.

10 Un "gen *FatB-2* de algodón", "gen *ghFatB-2*" o similar, tal como se utiliza en el presente documento se refiere a una secuencia de nucleótidos que codifica la acil-ADP tioesterasa en algodón, que se expresa en semillas de algodón en desarrollo. Un gen *ghFatB-2* se puede distinguir fácilmente de los genes que codifican otras acil-ACP tioesterasa u otras proteínas por los expertos en la técnica, en particular de un gen *ghFatB-1*, por ejemplo la SEQ ID NO: 3, o de un gen *ghFatB-3*, por ejemplo la SEQ ID NO: 7. Los genes *FatB-2* de algodón incluyen los alelos o variantes de origen natural que existen en el algodón, incluyendo los codificados por los genomas A y D del algodón tetraploide, así como las variantes no naturales que pueden producir los expertos en la técnica de la modificación genética. En la presente invención, un gen *ghFatB-2* se refiere a una molécula de ácido nucleico, que puede estar presente en el algodón o aislada del mismo o derivada del mismo, que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en la SEQ ID NO: 1. 5.

Un "gen *CPA-FAS-2* de algodón", "gen *ghFatB-2*" o similar, tal como se utiliza en el presente documento se refiere a una secuencia de nucleótidos que codifica la CPA ácido graso sintasa en algodón, que se expresa en semillas de algodón en desarrollo. Un gen *ghCPA-FAS-2* se puede distinguir fácilmente de los genes que codifican otras CPA ácido graso sintasas u otras proteínas por los expertos en la técnica, en particular de un gen *ghCPA-FAS-1*, por ejemplo la SEQ ID NO: 10, o de un gen *ghCPA-FAS-3e*, por ejemplo la SEQ ID NO: 14. Los genes *CPA-FAS* de algodón incluyen los alelos o variantes de origen natural que existen en el algodón, incluyendo los codificados por los genomas A y D del algodón tetraploide, así como las variantes no naturales que pueden producir los expertos en la técnica de la modificación genética. En la presente invención, un gen *ghCPA-FAS-2* se refiere a una molécula de ácido nucleico, que puede estar presente en el algodón o aislada del mismo o derivada del mismo, que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en la SEQ ID NO: 1. 12.

Una forma genómica o clon de un gen que contiene la región transcrita puede estar interrumpida con secuencias no codificantes denominadas "intrones" o "regiones intermedias" o "secuencias intermedias". Un intrón, tal como se usa en el presente documento, es un segmento de un gen que se transcribe como parte de un transcrito de ARN primario pero que no está presente en la molécula de ARNm madura. Los intrones se eliminan o "cortan" del transcrito nuclear o primario; por tanto, los intrones no están en el de ARN mensajero (ARNm). Los intrones pueden contener elementos reguladores tales como potenciadores. "Exones", tal como se utilizan en el presente documento, se refieren a las regiones de ADN correspondientes a las secuencias de ARN que están presentes en el ARNm maduro o la molécula de ARN madura en los casos en los que la molécula de ARN no se traduce. Un ARNm funciona durante la traducción para especificar la secuencia u ordenar los aminoácidos en un polipéptido naciente. El término "gen" incluye una molécula sintética o de fusión que codifica la totalidad o parte de las proteínas de la invención descritas en el presente documento y una secuencia de nucleótidos complementaria a una cualquiera de las anteriores. Se puede introducir un gen en un vector apropiado para el mantenimiento extracromosómico en una célula o para su integración en el genoma del huésped.

Como se usa en el presente documento, un "gen quimérico" se refiere a cualquier gen que no es un gen nativo en su localización nativa. Normalmente, un gen quimérico comprende secuencias reguladoras y transcritas o de codificación de proteínas que no se encuentran juntas en la naturaleza. En consecuencia, un gen quimérico puede comprender secuencias reguladoras y secuencias de codificación derivadas de diferentes fuentes o secuencias reguladoras y secuencias de codificación de la misma fuente, pero dispuestas de un modo distinto a como se encuentran en la naturaleza. El término "endógeno" se utiliza en el presente documento para hacer referencia a una sustancia que está normalmente presente o se produce en una planta no modificada en la misma etapa de desarrollo que la planta bajo investigación. Un "gen endógeno" se refiere a un gen nativo en su localización natural en el genoma de un organismo. Tal como se usa en el presente documento, "molécula de ácido nucleico recombinante" se refiere a una molécula de ácido nucleico que se ha construido o modificada por tecnología de ADN recombinante. Los términos "polinucleótido extraño" o "polinucleótido exógeno" o "polinucleótido heterólogo" y similares se refieren a cualquier ácido nucleico que se introduce en el genoma de una célula mediante manipulaciones experimentales. Estas incluyen secuencias de genes encontradas en esa célula siempre que el gen introducido contenga alguna modificación (por ejemplo, una mutación, la presencia de un gen marcador seleccionable, etc.) en relación con el gen de origen natural. Los genes extraños o exógenos pueden ser genes que se insertan en un organismo no nativo, genes nativos introducidos en una nueva localización dentro del huésped nativo o genes quiméricos. Un "transgén" es un gen que se ha introducido en el genoma mediante un procedimiento de transformación. La expresión "genéticamente modificado" incluye la introducción de genes en las células mediante transformación o transducción, mutación de genes en células y alteración o modulación de la regulación de

un gen en una célula u organismos en los que se han realizado estos actos o su progenie.

Polinucleótidos

5 La presente invención se refiere a diversos polinucleótidos. Como se usa en el presente documento, un "polinucleótido" o "molécula de ácido nucleico" significa un polímero de nucleótidos que puede ser ADN o ARN o una combinación de los mismos e incluye ARNm, ARNc, ADNc, ARNt, ARNsi, ARNhc y ARNhq. Puede ser ADN o ARN de origen celular, genómico o sintético, por ejemplo fabricado en un sintetizador automático y puede combinarse con hidratos de carbono, lípidos, proteínas u otros materiales, marcados con grupos fluorescentes o de otros tipos, o
10 unidos a un soporte sólido para realizar una actividad particular definida en el presente documento, o comprenden uno o más nucleótidos modificados no encontrados en la naturaleza, bien conocidos por los expertos en la técnica. El polímero puede ser monocatenario, esencialmente bicatenario o parcialmente bicatenario. Un ejemplo de una molécula de ARN parcialmente bicatenario es un ARN de horquilla (ARNhq), ARN de horquilla corta (ARNhc) o ARN autocomplementario que incluyen un tallo bicatenario formado por un apareamiento de bases entre una secuencia de nucleótidos y su complementaria y una secuencia de bucle que se une covalentemente a la secuencia de nucleótidos y su complementaria. El apareamiento de bases tal como se utiliza en el presente documento se refiere al apareamiento de bases entre los nucleótidos, incluyendo los pares de bases G:U. "Complementario" significa que dos polinucleótidos que pueden aparearse las bases (hibridar) a lo largo de parte de sus longitudes, o a lo largo de toda la longitud de uno o de ambos. Un "polinucleótido hibridado" significa que el polinucleótido está realmente
15 apareado en sus bases su complementario. El término "polinucleótido" se usa de forma intercambiable en el presente documento con la expresión "ácido nucleico".

Por "aislado" se entiende material que está sustancialmente o esencialmente libre de componentes que normalmente lo acompañan en su estado nativo. Tal como se utiliza en el presente documento, un "polinucleótido
25 aislado" o una molécula de ácido nucleico aislada" significa un polinucleótido que está al menos parcialmente separado de, preferentemente sustancialmente o esencialmente libre de, las secuencias polinucleotídicas del mismo tipo con las que está asociada o unida en su estado nativo. Por ejemplo, un "polinucleótido aislado" incluye un polinucleótido que se ha purificado o separado de las secuencias que lo flanquean en un estado natural, por ejemplo, un fragmento de ADN que se ha eliminado de las secuencias que normalmente están adyacentes al fragmento. Preferentemente, el polinucleótido aislado está también al menos libre del 90 % de otros componentes, tales como proteínas, hidratos de carbono, lípidos, etc. La expresión "polinucleótido recombinante", tal como se utiliza en el presente documento, se refiere a un polinucleótido formado in vitro por la manipulación del ácido nucleico en una forma que normalmente no se encuentra en la naturaleza. Por ejemplo, el polinucleótido recombinante puede estar en forma de un vector de expresión. Generalmente, tales vectores de expresión incluyen
30 ácido nucleico regulador de la transcripción y de la traducción unido operativamente a la secuencia de nucleótidos.

La presente invención se refiere al uso de oligonucleótidos. Como se usa en el presente documento, "oligonucleótidos" son polinucleótidos de hasta 50 nucleótidos de longitud. Pueden ser ARN, ADN, o combinaciones o derivados de cualquiera de ellos. Los oligonucleótidos son, normalmente, moléculas de cadena sencilla relativamente corta de 10 a 30 nucleótidos, normalmente de 15 a 25 nucleótidos de longitud. Cuando se utiliza como sonda o como cebador en una reacción de amplificación, el tamaño mínimo de tal oligonucleótido es el tamaño requerido para la formación de un híbrido estable entre el oligonucleótido y una secuencia complementaria sobre una molécula de ácido nucleico diana. Preferentemente, los oligonucleótidos tienen una longitud de al menos 15 nucleótidos, más preferentemente al menos 18 nucleótidos, más preferentemente al menos 19 nucleótidos, más
45 preferentemente al menos 20 nucleótidos, incluso más preferentemente al menos 25 nucleótidos.

Los polinucleótidos usados como sonda están normalmente conjugados con un marcador detectable, tal como un radioisótopo, una enzima, biotina, una molécula fluorescente o una molécula quimioluminiscente. Los oligonucleótidos de la invención son útiles en métodos de detección de un alelo de un gen *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2* u otro gen unido a un rasgo de interés, por ejemplo composición de aceite modificado. Tales métodos, por ejemplo, emplean hibridación de ácido nucleico y, en muchos casos, incluyen la extensión del cebador de oligonucleótidos mediante una polimerasa adecuada (tal como se utiliza en la PCR).

Una variante de un oligonucleótido de la invención incluye moléculas de diversos tamaños y/o son capaces de hibridar, por ejemplo, con el genoma del algodón próximo al de las moléculas de oligonucleótidos específicas definidas en el presente documento. Por ejemplo, las variantes pueden comprender nucleótidos adicionales (tales como 1, 2, 3, 4, o más) o menos nucleótidos siempre que sigan hibridando con la región diana. Además, algunos nucleótidos pueden estar sustituidos sin influir en la capacidad del oligonucleótido para hibridar con la región diana. Además, pueden diseñarse fácilmente variantes que hibriden cerca de, por ejemplo, en 50 nucleótidos, la región del genoma de la planta en la que hibridan los oligonucleótidos específicos definidos en el presente documento.

Las expresiones "variante polinucleotídica" y "variante" y similares se refieren a polinucleótidos o sus formas complementarias que muestran una identidad de secuencia sustancial con una secuencia polinucleotídica de referencia. Estos términos también abarcan polinucleótidos que se distinguen de un polinucleótido de referencia por la adición, delección o sustitución de al menos un nucleótido. Por consiguiente, las expresiones "variante de polinucleótido" y "variante" incluyen polinucleótidos en los que uno o más nucleótidos se han añadido o eliminado o

se han sustituido con diferentes nucleótidos. A este respecto, se entiende bien en la técnica que se pueden realizar ciertas alteraciones, incluidas mutaciones, adiciones, deleciones y sustituciones, en un polinucleótido de referencia, con lo que el polinucleótido alterado retiene la función o la actividad biológica del polinucleótido de referencia. Por consiguiente, estos términos abarcan polinucleótidos que codifican polipéptidos que exhiben actividad enzimática u otra actividad reguladora, o polinucleótidos capaces de servir como sondas selectivas u otros agentes de hibridación. En particular, esto incluye polinucleótidos que codifican el mismo polipéptido o secuencia de aminoácidos pero que varían en la secuencia de nucleótidos por la redundancia del código genético. Los términos "variante de polinucleótido" y "variante" también incluyen variantes alélicas naturales.

Por "corresponde a" o "correspondiente a" se entiende un polinucleótido (a) que tiene una secuencia de nucleótidos que es sustancialmente idéntica o complementaria a la totalidad o la mayor parte de una secuencia polinucleotídica de referencia o (b) que codifica una secuencia de aminoácidos idéntica a una secuencia de aminoácidos en un péptido o una proteína. Esta frase también incluye dentro de su alcance un péptido o polipéptido que tiene una secuencia de aminoácidos que es sustancialmente idéntica a una secuencia de aminoácidos en un péptido o proteína de referencia. Los términos utilizados para describir las relaciones de secuencia entre dos o más polinucleótidos o polipéptidos incluyen "secuencia de referencia", "ventana de comparación", "identidad de secuencia", "porcentaje de identidad de secuencia", "identidad sustancial" e "idénticos" y se definen con respecto a un número mínimo de nucleótidos o restos de aminoácidos o en toda su longitud. Las expresiones "identidad de secuencia" e "identidad" se usan indistintamente en el presente documento para hacer referencia en la medida en que las secuencias son idénticas base a nucleótido por nucleótido o en base a aminoácido por aminoácido sobre una ventana de comparación. Por tanto, un "porcentaje de identidad de secuencia" se calcula comparando dos secuencias alineadas óptimamente sobre la ventana de comparación, determinando el número de posiciones en las que la base de ácido nucleico idéntica (por ejemplo, A, T, C, G, U) o el resto de aminoácido idéntico (por ejemplo, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys y Met) se produce en ambas secuencias, para dar el número de posiciones coincidentes, dividiendo el número de las posiciones coincidentes por el número total de posiciones en la ventana de comparación (es decir, el tamaño de la ventana) y multiplicando el resultado por 100 para proporcionar el porcentaje de identidad de secuencia.

El % de identidad de un polinucleótido se puede determinar mediante un análisis GAP (Needleman y Wunsch, J. Mol. Biol. 48: 443–453, 1970) (programa GCG) con una penalización por creación de huecos = 5 y una penalización por extensión de huecos = 0,3. A menos que se indique lo contrario, la secuencia problema tiene al menos 45 nucleótidos de longitud, y el análisis GAP alinea las dos secuencias sobre una región de al menos 45 nucleótidos. Preferentemente, la secuencia problema tiene una longitud de al menos 150 nucleótidos y al análisis GAP alinea las dos secuencias sobre una región de al menos 150 nucleótidos. Más preferentemente, la secuencia problema tiene al menos 300 nucleótidos de longitud y el análisis GAP alinea las dos secuencias sobre una región de al menos 300 nucleótidos, o al menos 400, al menos 500 o al menos 600 nucleótidos en cada caso. También se puede hacer referencia a la familia de programas BLAST, como por ejemplo, divulgan Altschul et al., Nucleic Acids Res. 25: 3389, 1997. Una discusión detallada del análisis de secuencias se puede encontrar en la Unidad 19,3 de Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons Inc, 1994–1998, Capítulo 15.

Las secuencias de nucleótidos o aminoácidos se indican como "esencialmente similares" cuando tales secuencias tienen una identidad de secuencia de al menos 90 %, especialmente al menos 95 %, más especialmente son idénticas. Está claro que cuando se describen las secuencias de ARN como esencialmente similares o que corresponden a o que tienen un cierto grado de identidad de secuencia con secuencias de ADN, la timina (T) en la secuencia de ADN se considera igual a uracilo (U) en la secuencia de ARN.

Con respecto a los polinucleótidos definidos, se apreciará que cifras de % de identidad mayores que las proporcionadas anteriormente abarcarán realizaciones preferentes. Por lo tanto, cuando sea aplicable, a la luz de las cifras de porcentaje de identidad mínima, se prefiere que el polinucleótido comprenda una secuencia polinucleotídica que sea al menos 70 %, preferentemente al menos 80 %, más preferentemente al menos 85 %, más preferentemente al menos 90 %, más preferentemente al menos 91 %, más preferentemente al menos 92 %, más preferentemente al menos 93 %, más preferentemente al menos 94 %, más preferentemente al menos 95 %, más preferentemente al menos 96 %, más preferentemente al menos 97 %, más preferentemente al menos 98 %, más preferentemente al menos 99 %, más preferentemente al menos 99,1 %, más preferentemente al menos 99,2 %, más preferentemente al menos 99,3 %, más preferentemente al menos 99,4 %, más preferentemente al menos 99,5 %, más preferentemente al menos 99,6 %, más preferentemente al menos 99,7 %, más preferentemente al menos 99,8 %, e incluso más preferentemente al menos 99,9 % idéntica a la SEQ ID NO citada relevante.

Preferentemente, un polinucleótido de la invención que codifica un polipéptido con actividad ghFAD2-1, ghFatB-2 o ghCPA-FAS-2 es mayor que 800, preferentemente mayor que 900, e incluso más preferentemente mayor que 1.000 nucleótidos de longitud.

Los polinucleótidos de la presente invención pueden poseer, cuando se comparan con moléculas naturales, una o más mutaciones que son deleciones, inserciones o sustituciones de residuos nucleotídicos. Los mutantes pueden ser naturales (es decir, aislados de una fuente natural) o sintéticos (por ejemplo, realizando mutagénesis dirigida a sitio sobre el ácido nucleico).

La presente invención se refiere a la rigurosidad de las condiciones de hibridación para definir el grado de complementariedad de dos polinucleótidos. "Rigurosidad", tal como se utiliza en el presente documento, se refiere a las condiciones de temperatura y de fuerza iónica y a la presencia o ausencia de ciertos disolventes orgánicos, durante la hibridación y el lavado. Cuanto mayor es la rigurosidad, mayor será el grado de complementariedad entre una secuencia de nucleótidos diana y la secuencia polinucleotídica marcada (sonda). "Condiciones rigurosas" hace referencia a la temperatura y a las condiciones iónicas bajo las cuales solo hibridarán las secuencias de nucleótidos que tienen una alta frecuencia de bases complementarias. Como se usa en el presente documento, la expresión "hibrida en condiciones de rigurosidad baja, media, alta o muy alta" describe las condiciones para la hibridación y el lavado. Las directrices para realizar reacciones de hibridación se pueden encontrar en Ausubel et al., (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, NY, 6,3.1–6,3.6., 1989. Los métodos acuosos y no acuosos se describen en dicha referencia y se puede usar cualquiera de ellos. Condiciones de hibridación específicas mencionadas en el presente documento son las siguientes: 1) las condiciones de hibridación de rigurosidad baja son para hibridación en 6x cloruro de sodio/citrato de sodio (SSC) a 45 °C, seguido de dos lavados en 0,2 X SSC, 0,1 % de SDS a 50 – 55 °C; 2) las condiciones de hibridación de rigurosidad media son para la hibridación en 6 X SSC a aproximadamente 45 °C, seguido de uno o más lavados en 0,2 X SSC, 0,1 % de SDS a 60 °C; 3) las condiciones de hibridación de rigurosidad alta son para hibridación en 6 X SSC a 45 °C, seguido de uno o más lavados en 0,2 X SSC, 0,1 % de SDS a 65 °C; y 4) condiciones de hibridación muy rigurosas para hibridación en tampón de fosfato de sodio 0,5 M, 7 % de SDS a 65 °C, seguido de uno o más lavados a 0,2 X SSC, 1 % de SDS a 65 °C.

20 Polipéptidos

Los términos "polipéptido" y "proteína" se usan generalmente de forma intercambiable. Los términos "proteínas" y "polipéptidos", como se usan en el presente documento, también incluyen variantes, mutantes, modificaciones, análogos y/o derivados de los polipéptidos de la invención como se ha descrito en el presente documento. Como se usa en el presente documento, "polipéptido sustancialmente purificado" hace referencia a un polipéptido que se ha separado de los lípidos, ácidos nucleicos, otros péptidos y otras moléculas con la que se asocia en su estado nativo. Preferentemente, el polipéptido sustancialmente purificado está al menos al 90 % libre de otros componentes con los que está naturalmente asociado. Por "polipéptido recombinante" se entiende un polipéptido preparado usando técnicas recombinantes, es decir, a través de la expresión de un polinucleótido recombinante en una célula, preferentemente una célula vegetal y más preferentemente una célula vegetal de cereal.

El % de identidad de un polipéptido con respecto a otro polipéptido puede determinarse mediante un análisis GAP (Needleman y Wunsch, 1970 (citado anteriormente)) (programa GCG) con una penalización de creación de huecos = 5 y una penalización de extensión de hueco = 0,3. La secuencia problema tiene al menos 15 aminoácidos de longitud y el análisis GAP alinea las dos secuencias sobre una región de al menos 15 aminoácidos. Más preferentemente, la secuencia problema tiene una longitud de al menos 50 aminoácidos y al análisis GAP alinea las dos secuencias sobre una región de al menos 50 aminoácidos. Más preferentemente, la secuencia problema tiene una longitud de al menos 100 aminoácidos y al análisis GAP alinea las dos secuencias sobre una región de al menos 100 aminoácidos. Incluso más preferentemente, la secuencia problema tiene una longitud de al menos 250 aminoácidos y al análisis GAP alinea las dos secuencias sobre una región de al menos 250 aminoácidos.

Como se usa en el presente documento, un fragmento "biológicamente activo" de un polipéptido es una porción de un polipéptido de la invención, menor que la longitud total, que mantiene una actividad definida del polipéptido de longitud completa. Los fragmentos biológicamente activos pueden ser de cualquier tamaño siempre y cuando mantengan la actividad definida, pero tienen, preferentemente, al menos 200 o al menos 250 restos de aminoácidos de longitud.

Con respecto a un polipéptido definido, se apreciará que cifras de % de identidad mayores que las proporcionadas anteriormente abarcarán realizaciones preferentes. Por lo tanto, cuando sea aplicable, a la luz de las cifras de % de identidad mínima, se prefiere que el polipéptido comprenda una secuencia de aminoácidos que sea al menos 80 %, más preferentemente al menos 85 %, más preferentemente al menos 90 %, más preferentemente al menos 91 %, más preferentemente al menos 92 %, más preferentemente al menos 93 %, más preferentemente al menos 94 %, más preferentemente al menos 95 %, más preferentemente al menos 96 %, más preferentemente al menos 97 %, más preferentemente al menos 98 %, más preferentemente al menos 99 %, más preferentemente al menos 99,1 %, más preferentemente al menos 99,2 %, más preferentemente al menos 99,3 %, más preferentemente al menos 99,4 %, más preferentemente al menos 99,5 %, más preferentemente al menos 99,6 %, más preferentemente al menos 99,7 %, más preferentemente al menos 99,8 %, e incluso más preferentemente al menos 99,9 % idéntica a la SEQ ID NO citada relevante e incluso más preferentemente al menos 99,9 % idéntica a la SEQ ID NO citada relevante.

Los mutantes de la secuencia de aminoácidos de los polipéptidos de la presente invención se pueden preparar introduciendo cambios nucleotídicos adecuados en el ácido nucleico de la presente invención o mediante síntesis *in vitro* del polipéptido deseado. Dichos mutantes incluyen, por ejemplo, delecciones, inserciones o sustituciones de residuos dentro de la secuencia de aminoácidos. Se puede realizar una combinación de delección, inserción y sustitución para llegar a la construcción final, siempre que el producto peptídico final posea las características deseadas.

Se pueden preparar péptidos mutantes (alterados) usando cualquier técnica conocida en la técnica. Por ejemplo, un polinucleótido de la invención se puede someter a mutagénesis *in vitro*. Dichas técnicas de mutagénesis *in vitro* incluyen a subclonación del polinucleótido en un vector adecuado, transformar el vector en una cepa "mutadora" tal como la *E. coli* XL-1 red (Stratagene) y propagar las bacterias transformadas durante un número adecuado de generaciones. En otro ejemplo, los polinucleótidos de la invención se someten a técnicas de barajado de ADN como describen ampliamente Harayama, Trends Biotechnol. 16: 76-82, 1998. Estas técnicas de barajado de ADN pueden incluir genes relacionados con los de la presente invención, tales como genes de especies de plantas distintas del algodón, y/o incluyen diferentes genes de la misma planta que codifican proteínas similares. Los productos derivados de ADN mutado/alterado pueden someterse fácilmente a detección selectiva usando técnicas descritas en el presente documento para determinar si poseen, por ejemplo, actividad FAD2.

Al diseñar mutantes de secuencias de aminoácidos, la localización del sitio de la mutación y la naturaleza de la mutación dependerán de las características que se van a modificar. Los sitios de la mutación se pueden modificar individualmente o en serie, por ejemplo (1) sustituyendo primero con elecciones de aminoácidos conservadores y, después, como selecciones más radicales en función de los resultados conseguidos, (2) delecionando el residuo diana o (3) insertando otros residuos adyacentes al sitio localizado.

Generalmente, las delecciones en la secuencia de aminoácidos varían de aproximadamente 1 a 15 residuos, más preferentemente de aproximadamente 1 a 10 residuos y normalmente de aproximadamente 1 a 5 residuos contiguos.

Los mutantes de sustitución tienen al menos un residuo de aminoácido en la molécula polipeptídica eliminada y un residuo diferente insertado en su lugar. Los sitios de mayor interés para mutagénesis sustitucional incluyen sitios identificados como el sitio o sitios activos. Otros sitios de interés son aquéllos en los que los residuos concretos obtenidos de varias cepas o especies son idénticos. Estas posiciones pueden ser importantes para la actividad biológica. Estos sitios, especialmente los que entran dentro de una secuencia de al menos otros tres sitios conservados de forma idéntica, están sustituidos preferentemente de un modo relativamente conservador. Dichas sustituciones conservadoras se muestran en la Tabla 5, bajo el encabezado de "sustituciones de ejemplo".

Los polipéptidos de la presente invención se pueden producir de diversos modos, incluyendo producción y recuperación de polipéptidos naturales, producción y recuperación de polipéptidos recombinantes. En una realización, un polipéptido aislado de la presente invención se produce cultivando una célula capaz de expresar el polipéptido en condiciones eficaces para producir el polipéptido y recuperar el polipéptido. Una célula preferente de cultivo es una célula recombinante de la presente invención.

La presente invención se refiere a elementos que están operativamente conectados o unidos. "Conectados operativamente" o "unidos operativamente" y similares se refieren a un enlace de elementos polinucleotídicos en una relación funcional. Normalmente, las secuencias de ácido nucleico conectadas operativamente están unidas de forma contigua y, cuando es necesario, unirse a dos regiones de codificación de proteína, contiguas y en el marco de lectura. Una secuencia de codificación está "conectada operativamente a" otra secuencia de codificación cuando la ARN polimerasa transcriba las dos secuencias de codificación un único ARN, que si se traduce, se traduce a continuación en un único polipéptido que tiene aminoácidos derivados de ambas secuencias de codificación. Las secuencias de codificación no tienen que estar contiguas entre sí, siempre que las secuencias expresadas se procesen en última instancia para producir la proteína deseada.

Como se usa en el presente documento, la expresión "secuencia que actúa en cis", "elemento que actúa en cis" o "región reguladora cis" o "región reguladora" o término similar se entenderá como cualquier secuencia de nucleótidos que, cuando está colocada apropiadamente y conectada con respecto a una secuencia genética expresable, es capaz de regular, al menos en parte, la expresión de la secuencia genética. Los expertos en la técnica sabrán que una región reguladora en cis puede ser capaz de activar, silenciar, mejorar, reprimir o alterar de otro modo el nivel de expresión y/o la especificidad de tipo celular y/o la especificidad de desarrollo de una secuencia génica a nivel transcripcional o postranscripcional. En ciertas realizaciones de la presente invención, la secuencia que actúa en cis es una secuencia activadora que potencia o estimula la expresión de una secuencia genética expresable.

"Unir de forma operativa" un promotor o elemento potenciador a un polinucleótido transcribible significa colocar el polinucleótido transcribible (por ejemplo, el polinucleótido que codifica la proteína u otro transcrito) bajo el control regulador de un promotor, que controla después la transcripción de ese polinucleótido. En la construcción de combinaciones de promotores heterólogos/genes estructurales, generalmente se prefiere posicionar un promotor o una variante del mismo a una distancia del sitio de inicio de la transcripción del polinucleótido transcribible que es aproximadamente igual que la distancia entre ese promotor y la región de codificación de la proteína que controla en su entorno natural, es decir, el gen del que deriva el promotor. Como se conoce en la técnica se puede acomodar alguna variación en esta distancia sin perder la función. De forma similar, el posicionamiento preferente de un elemento de la secuencia reguladora (por ejemplo, un operador, potenciador, etc.) con respecto a un polinucleótido transcribible que se coloca bajo su control se define por el posicionamiento del elemento en su entorno natural; es decir, los genes de los que deriva.

Promotor" o "secuencia promotora" tal como se utiliza en el presente documento se refiere a una región de un gen, generalmente aguas arriba (5') de la región que codifica el ARN, que controla la iniciación y el nivel de transcripción en la célula de interés. Un "promotor" incluye las secuencias reguladoras de la transcripción un gen genómico clásico, incluyendo las secuencias de la caja TATA y de la caja CCAAT, así como elementos reguladores adicionales (es decir, secuencias activadoras aguas arriba, potenciadores y silenciadores) que alteran la expresión génica en respuesta a estímulos de desarrollo y/o ambientales, o de una manera específica de tejidos o de tipo celular. Un promotor normalmente, aunque no necesariamente (por ejemplo, algunos promotores PolIII), está situado aguas arriba de un gen estructural, cuya expresión regula. Además, los elementos reguladores que comprenden un promotor normalmente están colocados dentro de los 2 kb del sitio de iniciación de la transcripción del gen. Los promotores pueden contener elementos reguladores específicos adicionales situados más distales al sitio de iniciación para mejorar adicionalmente la expresión en una célula y/o para alterar el tiempo o la inducibilidad de la expresión de un gen estructural al que está conectado operativamente.

"Promotor constitutivo" hace referencia a un promotor que dirige la expresión de una secuencia transcrita operativamente unida en muchos o todos los tejidos de una planta. El término constitutivo, tal como se usa en el presente documento, no indica necesariamente que un gen se expresa al mismo nivel en todos los tipos de células, sino que el gen se expresa en una amplia gama de tipos de células, aunque algunas variaciones en el nivel son a menudo detectables. "Expresión selectiva", tal como se usa en el presente documento, se refiere a la expresión casi exclusivamente en órganos específicos de la planta, tales como, por ejemplo, endospermo, embrión, hojas, frutos, tubérculos o raíz. En una realización preferente, un promotor se expresa de forma selectiva o preferentemente en semillas en desarrollo de una planta, preferentemente una planta de algodón. El término también puede referirse a la expresión en etapas específicas de desarrollo en un órgano, tal como en la embriogénesis temprana, media o tardía o en diferentes etapas de madurez, o a la expresión que es inducible por ciertas condiciones o tratamientos ambientales. Por lo tanto, la expresión selectiva puede contrastar con la expresión constitutiva, que se refiere a la expresión en muchos o todos los tejidos de una planta en la mayoría o todas las condiciones experimentadas por la planta.

La expresión selectiva también puede dar lugar a la compartimentación de los productos de expresión génica en tejidos, órganos o etapas de desarrollo específicos de la planta. La compartimentación en localizaciones subcelulares específicas, tales como el plasto, el citosol, la vacuola o el espacio apoplástico puede conseguirse mediante la inclusión en la estructura del producto génico de señales apropiadas, por ejemplo, un péptido señal, para el transporte al compartimento celular requerido, o en el caso de los orgánulos semiautónomos (plastos y mitocondrias) mediante la integración del transgén con secuencias reguladoras apropiadas directamente en el genoma del orgánulo.

Un "promotor específico de tejido" o "promotor específico de órgano" es un promotor que se expresa, preferentemente, en un tejido u órgano en relación con muchos otros tejidos u órganos, preferentemente la mayoría si no todos los otros tejidos u órganos de una planta. Normalmente, el promotor se expresa a un nivel 10 veces más alto en el tejido u órgano específico que en otros tejidos u órganos. Un promotor específico de tejido ilustrativo es el promotor de un gen de lectina (SEQ ID NO: 16) que se expresa, preferentemente, en la semilla en desarrollo de plantas dicotiledóneas, tales como algodón. En la técnica se conocen bien otros promotores específicos de la semilla.

Los promotores contemplados por la presente invención pueden ser nativos de la planta huésped que se va a transformar o pueden derivar de una fuente alternativa, en la que la región es funcional en la planta huésped. Otras fuentes incluyen los genes de T-AND de *Agrobacterium*, tales como los promotores de genes para la biosíntesis de nopalina, octapina, manopina u otros promotores de opina, promotores específicos de tejidos (véase, por ejemplo, la patente de Estados Unidos n.º 5.459.252 de Conkling *et al.*; el documento WO 91/13992 de Advanced Technologies); promotores de virus (incluyendo virus específicos del huésped), o promotores parcial o totalmente sintéticos. Numerosos promotores que son funcionales en plantas monocotiledóneas y dicotiledóneas son bien conocidos en la técnica (véase, por ejemplo, Greve, J. Mol. Appl. Genet., 1: 499-511, 1983; Salomon *et al.*, EMBO J., 3: 141-146, 1984; Garfinkel *et al.*, Cell, 27: 143-153, 1983; Barker *et al.*, Plant Mol. Biol., 2: 235-350, 1983; incluyendo varios promotores aislados de plantas y virus, tal como el promotor del virus del mosaico de la coliflor (CaMV 35S, 19S). Se conocen muchas regiones promotoras específicas de tejido. Otras regiones de iniciación de la transcripción que proporcionan, preferentemente, la transcripción en ciertos tejidos o en ciertas condiciones de crecimiento, incluyen las de genes que codifican napina, la semilla ACP, zeína u otras proteínas de almacenamiento de semillas. Los métodos no limitantes para evaluar la actividad promotora se divulgan en Medberry *et al.*, Plant Cell, 4: 185-192, 1992, Medberry *et al.*, Plant J. 3: 619-626, 1993, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (2ª Ed). Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.1989 y McPherson *et al.*, (patente de Estados Unidos n.º 5.164.316).

Como alternativa o además de, el promotor puede ser un promotor inducible o un promotor regulado por el desarrollo que es capaz redirigir la expresión del polinucleótido introducido en una etapa de desarrollo apropiada de la planta. Entre otras secuencias que actúan en cis que pueden usarse se incluyen potenciadores de la transcripción y/o de la traducción. Las regiones potenciadoras son bien conocidas para los expertos en la técnica y pueden incluir un codón de iniciación de la traducción ATG y secuencias adyacentes. Cuando está incluido, el codón de iniciación

debe estar en fase con el marco de lectura de la secuencia de codificación relativa al polinucleótido extraño o exógeno para asegurar la traducción de toda la secuencia si se va a traducir. Las regiones de iniciación de la traducción se pueden proporcionar de la fuente de la región de iniciación de la transcripción o de un polinucleótido extraño o exógeno. La secuencia también puede derivar de la fuente del promotor seleccionado para dirigir la transcripción y puede modificarse específicamente para incrementar la traducción del ARNm.

La construcción de ácido nucleico de la presente invención puede comprender una secuencia 3' no traducida de aproximadamente 50 a 1.000 pares de bases de nucleótidos que puede incluir una secuencia de terminación de la transcripción. Una secuencia 3' no traducida puede contener una señal de terminación de la transcripción que puede o no incluir una señal de poliadenilación y cualquier otra señal reguladora capaz de efectuar el procesamiento del ARNm. Una señal de poliadenilación funciona para la adición de trectos de ácido poliadenílico al extremo 3' de un precursor de ARNm. Las señales de poliadenilación se reconocen habitualmente por la presencia de homología a la forma canónica 5' AATAAA-3' aunque las variaciones no son infrecuentes. Las secuencias de terminación de la transcripción que no incluyen una señal de poliadenilación incluyen terminadores para la ARN polimerasa Poll o PolIII que comprenden una serie de cuatro o más timidinas. Ejemplos de secuencias 3' no traducidas adecuadas son las regiones transcritas no traducidas en 3' que contienen una señal de poliadenilación procedente de un gen de lectina (SEQ ID NO: 20), el gen S3 del virus de la atrofia subterránea del trébol (SEQ ID NO: 23), o el gen de la nopalina sintasa (nos) de *Agrobacterium tumefaciens* (Bevan et al., Nucl. Acid Res., 11: 369, 1983). Las secuencias no traducidas en 3' también pueden derivar de genes vegetales, tales como el extremo 3' de los genes del inhibidor I o II de la proteasa de la patata o del tomate, los genes de la proteína de almacenamiento de soja y el gen de la subunidad pequeña de la ribulosa-1,5-bisfosfato carboxilasa (ssRUBISCO), aunque también pueden usarse otros elementos en 3' conocidos por los expertos en la técnica.

Como la secuencia de ADN insertada entre el sitio de iniciación de la transcripción y el inicio de la secuencia de codificación, es decir, la secuencia líder no traducida en 5' (5'UTR), puede influir en la expresión génica si está traducida y transcrita, también se puede emplear una secuencia líder concreta. Las secuencias líder adecuadas incluyen las que comprenden secuencias seleccionadas para dirigir la expresión óptima de la secuencia de ADN extraña o endógena. Por ejemplo, tales secuencias líder incluyen una secuencia consenso preferente que puede aumentar o mantener la estabilidad del ARNm y evitar el inicio inadecuado de la traducción, como por ejemplo describen Joshi, Nucl Acid Res. 15: 6643, 1987.

Adicionalmente se pueden usar secuencias diana para apuntar a la enzima codificada por el polinucleótido exógeno o extraño a un compartimiento intracelular, por ejemplo al cloroplasto, dentro de las células vegetales o al medio extracelular. Por ejemplo, una secuencia de ácido nucleico que codifica una secuencia peptídica señal o de tránsito puede estar operativamente unida a una secuencia que codifica una enzima elegida de la presente invención de tal manera que, cuando se traduce, el péptido de tránsito o señal puede transportar la enzima a un destino intracelular o extracelular, y, opcionalmente, se puede eliminar postraduccionalmente. Los péptidos de tránsito o señal actúan facilitando el transporte de proteínas a través de membranas intracelulares, por ejemplo, retículo endoplásmico, vacuola, vesícula, plasto, membranas mitocondriales y plasmiales. Por ejemplo, la secuencia diana puede dirigir una proteína deseada a un orgánulo particular, tal como una vacuola o un plasto (por ejemplo, un cloroplasto), en lugar de al citosol. De este modo, la construcción de ácido nucleico de la invención puede comprender además una secuencia de ácido nucleico de codificación de péptido de tránsito de plasto unida operativamente entre una región promotora y el polinucleótido exógeno o extraño. Las secuencias que codifican tales péptidos señal se encuentran a menudo en genes de origen natural y se pueden emplear, tal como, por ejemplo, en los genes *ghFatB* descritos en el presente documento.

Vectores

La presente invención incluye el uso de vectores para la manipulación o transferencia de construcciones genéticas. Por "vector" se entiende una molécula de ácido nucleico, preferentemente una molécula de ADN derivada, por ejemplo, de un plásmido, bacteriófago o virus vegetal, en el que se puede insertar o clonar una secuencia de ácido nucleico. Un vector es, preferentemente, ADN de doble cadena y contiene uno o más sitios de restricción únicos y puede ser capaz de replicación autónoma en una célula huésped definida que incluye una célula o tejido diana o una célula o tejido progenitor de la misma, o capaz de integración en el genoma del huésped definido de manera que la secuencia clonada sea reproducible. Por consiguiente, el vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, por ejemplo, un plásmido circular lineal o cerrado, un elemento extracromosómico, un minicromosoma o un cromosoma artificial. El vector puede contener cualquier medio para asegurar la autorreplicación. Como alternativa, el vector puede ser uno que, cuando se introduce en una célula, está integrado en el genoma de la célula receptora y se replica junto con el cromosoma o cromosomas en el que se ha integrado. Un sistema vectorial puede comprender un único vector o plásmido, dos o más vectores o plásmidos, que juntos contienen el ADN total que se va a introducir en el genoma de la célula huésped, o un transposón. La elección del vector dependerá normalmente de la compatibilidad del vector con la célula en la que se va a introducir el vector. El vector también puede incluir un marcador de selección tal como un gen de resistencia a antibióticos, un gen de resistencia a herbicidas u otro gen que pueda usarse para la selección de transformantes adecuados. Ejemplos de dichos genes son bien conocidos por los expertos en la técnica.

La construcción de ácido nucleico de la invención puede introducirse en un vector, tal como un plásmido. Los vectores plasmídicos incluyen normalmente secuencias de ácido nucleico adicionales que proporcionan una selección, amplificación y transformación fáciles del casete de expresión en células procariotas y eucariotas, por ejemplo, vectores derivados de pUC, vectores derivados de pSK, vectores derivados de pGEM, vectores derivados de pSP, vectores derivados de pBS o vectores binarios que contienen una o más regiones de T-ADN. Entre las secuencias de ácido nucleico adicionales se incluyen orígenes de replicación para proporcionar una replicación autónoma del vector, genes marcadores seleccionables, que codifican, preferentemente, resistencia a antibióticos o a herbicidas, sitios de clonación múltiple únicos que proporcionan sitios múltiples para insertar secuencias de ácido nucleico o genes codificados en la construcción de ácido nucleico, y secuencias que mejoran la transformación de células procariotas y eucariotas (especialmente de plantas).

Por "marcador génico" se quiere decir un gen que imparte un fenotipo distinto a las células que expresan el gen marcador y, por tanto, permite distinguir a dichas células transformadas de las células que no tienen el marcador. Un gen marcador seleccionable confiere un rasgo para el que se puede "seleccionar" según la resistencia a un agente selectivo (por ejemplo, un herbicida, antibiótico, radiación, calor u otro tratamiento perjudicial para las células no transformadas). Un gen marcador detectable (o gen indicador) confiere un rasgo que se puede identificar mediante observación o análisis (es decir, mediante "detección selectiva" (por ejemplo, β -glucuronidasa, luciferasa, GFP u otra actividad enzimática no presente en las células no transformadas). El gen marcador y la secuencia de nucleótidos de interés no tienen que estar unidos. Un ejemplo de un gen marcador seleccionable es el gen terminador *S1-NptII-S3* mostrado esquemáticamente en la Figura 12, que comprende las SEQ ID NOs: 21–23.

Para facilitar la identificación de transformantes, la construcción de ácido nucleico comprende, deseablemente, un gen marcador seleccionable o detectable como, o además, el polinucleótido extraño o exógeno. La elección real de un marcador no es crucial, siempre que sea funcional (es decir, selectivo) en combinación con las células vegetales de elección. El gen marcador y el polinucleótido extraño o exógeno de interés no tienen que estar unidos, ya que la cotransformación de genes no unidos como, por ejemplo, se describe en la patente de Estados Unidos n.º 4.399.216 también es un proceso eficiente en la transformación de plantas.

Entre los ejemplos de marcadores bacterianos seleccionables se encuentran los marcadores que confieren resistencia a antibióticos, tales como resistencia a ampicilina, a eritromicina, a cloranfenicol o a tetraciclina, preferentemente resistencia a kanamicina. Entre los ejemplos de marcadores seleccionables para la selección de transformantes vegetales se incluyen, pero no se limitan a, un gen *hyg* que codifica la resistencia a la higromicina B; un gen de neomicina fosfotransferasa (*nptII*), que confiere resistencia a kanamicina, paromomicina, G418; un gen de la glutatión-S-transferasa de hígado de rata, que confiere resistencia a herbicidas derivados de glutatión como, por ejemplo, se describe en el documento EP-A 256223; un gen de la glutamina sintetasa, que confiere, tras sobreexpresión, resistencia a inhibidores de la glutamina sintetasa, tales como fosfinitricina como, por ejemplo, lo descrito en el documento WO87/05327, un gen de la acetiltransferasa de *Streptomyces viridochromogenes*, que confiere resistencia al agente selectivo fosfinitricina como, por ejemplo, se describe en el documento EP-A-275957, un gen que codifica una 5-enolshikimato-3-fosfato sintasa (EPSPS), que confiere tolerancia a la N-fosfometilglucina como, por ejemplo, describen Hinchee et al. Biotech. 6: 915, 1988, un *bar*, que confiere resistencia a bialafos, como, por ejemplo, se describe en el documento WO91/02071; un gen de la nitrilasa, tal como *bxn* de *Klebsiella ozaenae*, que confiere resistencia al bromoxinilo (Stalker et al., Science, 242: 419, 1988); un gen de dihidrofolato reductasa (DHFR), que confiere resistencia al metotrexato (Thillet et al., J. Biol. Chem. 263: 12500, 1988); un gen de la acetolactato sintasa (ALS) mutante, que confiere resistencia a imidazolinona, sulfonilurea u otros productos químicos inhibidores de la ALS (documento EP-A-154 204); un gen de la antranilato sintasa mutada que confiere resistencia a 5-metilriptófano; o un gen de la dalapon-deshalogenasa, que confiere resistencia al herbicida.

Entre los marcadores detectables preferentes se incluyen, pero no se limitan a, un gen *uidA* que codifica una enzima β -glucuronidasa (GUS) para la cual se conocen varios sustratos cromogénicos, un gen de la β -galactosidasa que codifica una enzima para la cual se conocen sustratos cromogénicos, un gen de la aeuorina (Prasher et al., Biochem. Biophys. Res. Comm. 126: 1259–68, 1985), que puede emplearse en la detección de bioluminiscencia sensible al calcio; un gen de la proteína fluorescente verde (Niedz et al., Plant Cell Reports, 14: 403, 1995.) o derivados de los mismos; un gen de luciferasa (*luc*) (Ow et al., Science, 234: 856, 1986), que permite la detección de bioluminiscencia, y otros conocidos en la técnica. Por "molécula indicadora", tal como se utiliza en la presente memoria descriptiva, se entiende una molécula que, por su naturaleza química, proporciona una señal analíticamente identificable que facilita la determinación de la actividad del promotor por referencia al producto proteico.

Métodos de modificación de la expresión génica

El nivel de una proteína, por ejemplo una enzima implicada en la biosíntesis de aceite en las semillas en desarrollo de una planta, puede modularse aumentando el nivel de expresión de una secuencia de nucleótidos que codifica la proteína en una célula vegetal o disminuyendo el nivel de expresión de un gen que codifica la proteína en la planta, lo que da lugar a una composición de aceite modificado en la semilla madura. El nivel de expresión de un gen puede modularse alterando el número de copias por célula, por ejemplo introduciendo una construcción genética sintética que comprende la secuencia de codificación y un elemento de control transcripcional que está operativamente

conectado a la misma y que es funcional en la célula. Se puede seleccionar y detectar una pluralidad de transformantes para aquellos con un nivel y/o especificidad favorables de expresión de transgenes que surgen de las influencias de secuencias endógenas en la proximidad del sitio de integración de los transgenes. Un nivel y un patrón favorables de la expresión de los transgenes es uno que da como resultado una modificación sustancial de la composición de aceite. Esto puede detectarse mediante pruebas sencillas del aceite de la semilla a partir de los transformantes. Como alternativa, puede seleccionarse un grano mutagenizado o una población de plantas de un programa de cultivos para líneas individuales con un contenido o una composición de aceite alterados.

La reducción de la expresión génica puede lograrse mediante la introducción y la transcripción de un "gen quimérico silenciador de genes" introducido en la célula vegetal. El gen quimérico silenciador de genes puede introducirse de forma estable en el genoma de la célula vegetal, preferentemente en el genoma nuclear, o puede introducirse de forma transitoria, por ejemplo, en un vector viral. Tal como se utiliza en el presente documento, "efecto de silenciamiento de genes" se refiere a la reducción de la expresión de un ácido nucleico diana en una célula vegetal, que puede conseguirse mediante la introducción de un ARN silenciador. Dicha reducción puede ser el resultado de la reducción de la transcripción, incluso mediante la metilación de la remodelación de la cromatina o la modificación postranscripcional de las moléculas de ARN, incluyendo mediante la degradación del ARN, o ambas. El silenciamiento genético incluye la abolición de la expresión del ácido nucleico o gen objetivo y un efecto parcial en su extensión o duración. Es suficiente que el nivel de expresión del ácido nucleico diana en presencia del ARN silenciador sea menor que en ausencia del mismo. El nivel de expresión puede reducirse en al menos aproximadamente 40 %, o al menos aproximadamente 50 %, o al menos aproximadamente 60 %, o al menos aproximadamente 70 %, o al menos aproximadamente 80 %, o al menos aproximadamente 90 %, o al menos aproximadamente 95 %, o al menos aproximadamente 99 %. En realizaciones preferentes, la expresión del gen *ghCPA-FAS-2* de algodón se reduce en la semilla en al menos un 60 %, más preferentemente en al menos un 80 % con respecto a una semilla correspondiente que carece del ADN quimérico silenciador del gen. El ácido nucleico diana puede ser un gen implicado en la biosíntesis de aceite, la acumulación de aceite, tales como los genes implicados en el ensamblaje de TAG, incluyendo, pero sin limitarse a, las aciltransferasas u otras enzimas de la vía de Kennedy, o el metabolismo del aceite, o puede ser cualquier otro gen endógeno, transgén o gen exógeno, tales como genes virales que pueden no estar presentes en la célula vegetal en el momento de la introducción del transgén. En algunas realizaciones, la memoria descriptiva proporciona un método para modificar el aceite endógeno de una planta de algodón que comprende producir una planta de algodón transgénico que tiene una construcción génica que comprende una secuencia de nucleótidos de un gen de biosíntesis de ácidos grasos o un fragmento génico del mismo, en el que dicho gen o fragmento génico está colocado operativamente en conexión con una secuencia promotora capaz de conferir la expresión de dicho gen o fragmento génico en la semilla de una planta de algodón, y en el que dicho gen de la biosíntesis de ácidos grasos es dos o más de *FAD2* desaturasa, *FatB* tioesterasa y *CPA-FAS-2* sintasa de ácido graso. Como se describe en el presente documento, la reducción de la expresión génica mejora el contenido de ácidos grasos del aceite endógeno y extraído.

Moléculas de ARN antisentido

Se pueden usar técnicas antisentido para reducir la expresión génica de acuerdo con la invención. Se entenderá que el término "ARN antisentido" es una molécula de ARN que es complementaria a al menos una porción de una molécula de ARNm específica y capaz de reducir la expresión del gen que codifica el ARNm. Dicha reducción normalmente se produce de una manera dependiente de la secuencia y se piensa que se produce interfiriendo con un acontecimiento postranscripcional, tal como el transporte del ARNm desde el núcleo al citoplasma, la estabilidad del ARNm o la inhibición de la traducción. El uso de métodos antisentido es bien conocido en la técnica (véase, por ejemplo, Hartmann y Endres, *Manual of Antisense Methodology*, Kluwer, 1999). El uso de técnicas antisentido en plantas se ha revisado en Bourque, *Plant Science*, 105: 125–149, 1995 y Senior, *Biotech. Genet. Engin. Revs.* 15: 79–119, 1998. Bourque, 1995 (citado anteriormente) enumera un gran número de ejemplos de cómo se han utilizado secuencias antisentido en sistemas vegetales como método de inactivación génica. También afirma que la obtención de una inhibición del 100 % de cualquier actividad enzimática puede no ser necesaria, ya que muy probablemente la inhibición parcial de lugar a un cambio mensurable en el sistema. Senior, 1998 (citado anteriormente) afirma que los métodos antisentido son ahora una técnica muy bien establecida para manipular la expresión génica.

Como se usa en el presente documento, la expresión "un polinucleótido antisentido que hibrida en condiciones fisiológicas" significa que el polinucleótido (que está total o parcialmente monocatenario) es al menos capaz de formar un polinucleótido bicatenario con un producto de ARN del gen que se va a inhibir, normalmente el ARNm que codifica una proteína, tal como las proporcionadas en el presente documento, en condiciones normales en una célula. Las moléculas antisentido pueden incluir secuencias que corresponden a los genes estructurales o para secuencias que efectúan el control sobre la expresión génica o el acontecimiento de corte y empalme. Por ejemplo, la secuencia antisentido puede corresponder a la región de codificación del gen diana o la región 5' no traducida (UTR) o la 3'-UTR o combinación de estos. Puede ser complementaria en parte con las secuencias intrónicas, que pueden sufrir corte y empalme durante o después de la transcripción, pero, preferentemente, es complementaria a únicamente las secuencias exónicas del gen diana. En vista de la divergencia generalmente mayor de las UTR, el direccionamiento a estas regiones proporciona una mayor especificidad de la inhibición génica.

La longitud de la secuencia antisentido debe ser de al menos 19 nucleótidos contiguos, preferentemente al menos 25 o 30 o 50 nucleótidos, y, más preferentemente, al menos 100, 200, 500 o 1.000 nucleótidos, hasta un máximo de la longitud completa del gen que se va a Inhibir. Se puede usar la secuencia de longitud completa complementaria a todo el transcrito génico. La longitud es, lo más preferentemente, 100-2000 nucleótidos. El grado de identidad de la secuencia antisentido con el transcrito dirigido debe ser de al menos un 90 % y, más preferentemente, de 95-100 %. Las secuencias antisentido preferentes comprenden al menos 30 nucleótidos contiguos que son complementarios de cualquier secuencia de al menos 30 nucleótidos contiguos de SEQ ID NO: 1, 5, o 12. La molécula de ARN antisentido puede comprender, por supuesto, secuencias no relacionadas que pueden funcionar para estabilizar la molécula.

Las construcciones genéticas para expresar un ARN antisentido se pueden preparar fácilmente uniendo una secuencia promotora a una región del gen diana en una orientación "antisentido", que como se usa en el presente documento se refiere a la orientación inversa con respecto a la orientación de la transcripción y la traducción (si se produce) de la secuencia en el gen diana en la célula vegetal. De acuerdo con ello, la presente invención también proporciona una molécula de ácido nucleico, tal como un ADN quimérico que codifica un ARN antisentido de la invención, incluyendo células, tejidos, órganos, plantas, semillas, particularmente plantas de algodón o semilla de algodón y similares, que comprenden la molécula de ácido nucleico.

Ribozimas

El término "ribozima", tal como se utiliza en el presente documento, se refiere a una molécula de ARN que reconoce específicamente un ARN sustrato distinto y cataliza su escisión. Normalmente, la ribozima contiene una región de nucleótidos que es complementaria a una región del ARN diana, lo que permite que la ribozima hibride específicamente con el ARN diana en condiciones fisiológicas, por ejemplo en la célula en la que actúa la ribozima, y una región enzimática a la que en el presente documento se hace referencia como el "dominio catalítico". Los tipos de ribozimas que son particularmente útiles en la presente invención son la ribozima de cabeza de martillo (Haseloff y Gerlach, Nature 334: 585-591, 1988, Perriman et al., Gene, 113: 157-163, 1992) y la ribozima de horquilla (Shippy et al., Mol. Biotech. 12: 117-129, 1999). El ADN que codifica las ribozimas se puede sintetizar usando métodos bien conocidos en la técnica y puede incorporarse en una construcción genética o vector de expresión para la expresión en la célula de interés. De acuerdo con ello, la presente invención también proporciona una molécula de ácido nucleico, tal como un ADN quimérico que codifica una ribozima de la invención, incluyendo células, tejidos, órganos, plantas, granos y similares, que comprenden la molécula de ácido nucleico. Normalmente, el ADN que codifica la ribozima se inserta en un casete de expresión bajo el control de un promotor y una señal de terminación de la transcripción que funciona en la célula. Los sitios específicos de escisión por ribozima dentro de cualquier potencial diana de ARN se pueden identificar mediante barrido de la molécula diana para buscar los sitios de escisión de la ribozima que incluyen las secuencias de tres nucleótidos GUA, GUU y GUC. Una vez identificadas, las secuencias cortas de ARN de entre 5 y 20 ribonucleótidos correspondientes a la región del gen diana 5' y 3' del sitio de escisión se pueden evaluar para determinar las características estructurales previstas, tal como la estructura secundaria, que pueden hacer que la secuencia de oligonucleótidos sea menos adecuada. Cuando se emplean, las ribozimas se pueden seleccionar del grupo que consiste en ribozimas de cabeza de martillo, ribozimas de horquilla, ribozimas de cabeza de hacha, ribozimas satélites newt, ribozimas de Tetrahymena y ribozimas de ARNasa P, y están diseñadas según métodos conocidos en la técnica basados en la secuencia del gen diana (por ejemplo, véase la patente de Estados Unidos n.º 5.741.679). La idoneidad de los candidatos diana también se puede evaluar mediante el análisis de su accesibilidad a la hibridación con oligonucleótidos complementarios usando ensayos de protección de ribonucleasa.

Como con los polinucleótidos antisentido descritos en el presente documento, las ribozimas de la invención deben ser capaces de hibridar con una molécula de ácido nucleico diana (por ejemplo, las SEQ ID NO: 1, 5 o 12) en "condiciones fisiológicas", es decir, las condiciones dentro de una célula, especialmente las condiciones en una célula vegetal, tal como una célula de trigo o cebada.

Interferencia de ARN/ARN dúplex

Como se usa en el presente documento, la "molécula de ARNdc introducida artificialmente" se refiere a la introducción de la molécula de ARNdc, que puede producirse en la célula mediante transcripción de un gen quimérico que codifica dicha molécula de ARNdc, sin embargo no se refiere a la conversión de una molécula de ARN de cadena sencilla en un ARNdc dentro de la célula eucariótica o célula vegetal. La interferencia de ARN (ARNi) es particularmente útil para reducir específicamente la expresión de un gen o inhibir la producción de una proteína particular. Aunque sin desear estar limitados por la teoría, Waterhouse et al., Proc. Natl. Acad. Sci. USA.95: 13959-13964, 1998 han proporcionado un modelo para el mecanismo mediante el cual se puede usar ARNdc para reducir la expresión génica o la producción de proteínas. Esta tecnología se basa en la presencia de moléculas de ARNdc que contienen una secuencia que es esencialmente idéntica al ARNm del gen de interés o parte del mismo. Convenientemente, el ARNdc puede producirse a partir de un único promotor en un vector recombinante o célula huésped, en el que las secuencias sentido y antisentido se transcriben para producir un ARN de horquilla en el que las secuencias sentido y antisentido hibridan para formar la región de ARNdc con una secuencia intermedia o región espaciadora que forma una estructura de bucle, de manera que el ARN de horquilla comprende una estructura de

tallo-bucle. El diseño y la producción de moléculas de ARNdc adecuadas para la presente invención está dentro de la capacidad de un experto en la técnica, particularmente considerando Waterhouse *et al.*, 1998 (citado anteriormente), Smith *et al.*, *Nature*, 407: 319–320, 2000, el documento WO 99/53050, el documento WO 99/49029 y el documento WO 01/34815. De acuerdo con ello, la presente invención también proporciona una molécula de ácido nucleico, tal como un ADN quimérico que codifica un ARN dúplex, tal como un ARN de horquilla de la invención, incluyendo células, tejidos, órganos, plantas, semillas, particularmente plantas de algodón o semilla de algodón y similares, que comprenden la molécula de ácido nucleico. En una realización preferente, el ADN quimérico tiene la estructura mostrada esquemáticamente en la Figura 12, que comprende la secuencia mostrada en la Figura 13, a la que se hace referencia en el presente documento como la construcción MonoCott.

En un ejemplo, se introduce un ADN que dirige la síntesis de un producto o productos de ARN al menos parcialmente bicatenario con homología con el gen diana que se va a inactivar. Por lo tanto, el ADN comprende tanto secuencias sentido como antisentido que, cuando se transcriben a ARN, pueden hibridar para formar la región de ARN de doble cadena. En una realización preferente, las secuencias sentido y antisentido están separadas por una región espaciadora que comprende un intrón que, cuando se transcribe a ARN, se elimina mediante corte y empalme. Un intrón preferente es un intrón del gen *ghFAD2-1*, por ejemplo la SEQ ID NO: 18. Se ha demostrado que esta disposición da como resultado una mayor eficacia del silenciamiento génico (Smith *et al.*, 2000 (citado anteriormente)). La región bicatenaria puede comprender una o dos moléculas de ARN, transcritas desde una región de ADN o dos. El ARNdc puede clasificarse como ARNhq largo que tiene regiones largas sentido y antisentido que pueden ser en gran parte complementarias, pero no necesitan ser totalmente complementarias (que normalmente forman una región de pares de bases mayor de aproximadamente 100 pb, que preferentemente varía de 200-1.000 pb). El ARNhq también puede ser más pequeño, en el que la porción bicatenaria varía de tamaño de aproximadamente 30 a aproximadamente 50 pb, o de 30 a aproximadamente 100 pb (véase el documento WO04/073390). Se cree que la presencia de la región de ARN de doble cadena desencadena una respuesta de un sistema de planta endógeno que procesa el ARN de doble cadena a oligonucleótidos de 21-24 nucleótidos de longitud, y también usa estos oligonucleótidos para la escisión específica de secuencia del transcrito de ARN homólogo del gen de la planta diana, reduciendo o eliminando eficazmente la actividad del gen diana.

La longitud de las secuencias sentido y antisentido que hibridan debe ser de al menos 19 nucleótidos contiguos, preferentemente de al menos 27 o 30 o 50 nucleótidos, y, más preferentemente, de al menos 100, 200 o 500 nucleótidos, hasta la secuencia de longitud completa correspondiente a la totalidad del transcrito génico. Las longitudes son, lo más preferentemente, 100-2.000 nucleótidos. El grado de identidad de las secuencias sentido y antisentido con el transcrito dirigido debe ser de al menos un 85 %, preferentemente de al menos de 90 % y, más preferentemente, 95-100 %. Cuanto más larga sea la secuencia, menor será el requisito de identidad de secuencia global. La molécula de ARN puede comprender, por supuesto, secuencias no relacionadas que pueden funcionar para estabilizar la molécula. La molécula de ARN puede ser un híbrido entre diferentes secuencias dirigidas a diferentes ARN diana, lo que permite la reducción de la expresión de más de un gen diana, o puede ser una secuencia que corresponde a una familia de genes diana relacionados, tales como una familia multigénica. En una realización preferente, la molécula de ARN se dirige a al menos tres genes diana diferentes, más preferentemente los genes *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2* en algodón. Las secuencias utilizadas en el ARNdc corresponden preferentemente a las secuencias exónicas del gen diana y pueden corresponder a las secuencias 5' y/o 3' no traducidas o las secuencias de codificación de proteínas o cualquier combinación de las mismas. El promotor utilizado para expresar la construcción que forma el ARNdc puede ser cualquier tipo de promotor si el ARNdc resultante es específico para un producto génico en el linaje celular dirigido para la destrucción. Como alternativa, el promotor puede ser específico de linaje en cuanto a que solo se expresa en las células de un linaje de desarrollo en particular. Esto podría ser ventajoso cuando se observa cierta superposición en la homología con un gen que se expresa en un linaje celular no diana. El promotor también puede ser inducible por factores controlados externamente o por factores ambientales intracelulares. Normalmente, la molécula de ARN se expresa bajo el control de un promotor de la ARN polimerasa II o de la ARN polimerasa III. Entre los ejemplos de estas últimas se incluyen los promotores de ARNt o ARNsn.

En el ejemplo 5 se proporcionan ejemplos de construcciones genéricas que codifican moléculas de ARNdc que pueden usarse para regular por disminución la producción de polipéptidos para la modificación de la composición de aceite en semilla de algodón.

Otro ARN silenciador puede ser "ARN no poliadenilado" que comprende al menos 20 nucleótidos consecutivos que tienen al menos un 95 % de identidad de secuencia con la complementaria de una secuencia de nucleótidos de un transcrito de ARN del gen diana, tal como se describe en los documentos WO01/12824 o US6423885. Todavía otro tipo de ARN silenciador es una molécula de ARN como se describe en el documento WO03/076619 que comprende al menos 20 nucleótidos consecutivos que tienen al menos un 95 % de identidad de secuencia con la secuencia del ácido nucleico diana o la complementaria de la misma y que comprende además una región en su mayor parte bicatenaria como se describe en el documento WO03/076619.

La regulación del microARN es una rama especializada de la vía de silenciamiento del ARN que evolucionó hacia la regulación génica, divergiendo de los ARNi/PTGS convencionales. Los microARN son una clase específica de ARN pequeños que están codificados en elementos similares a genes organizados en una repetición invertida parcial

característica. Cuando se transcriben, los genes de los microARN dan lugar a ARN precursores tallo-bucle con apareamientos parciales de bases, a partir de los cuales se procesan a continuación los microARN. Los microARN procesados tienen, normalmente, 21 nucleótidos contiguos de longitud o 21-23 nucleótidos de longitud con una secuencia al menos un 95 % idéntica con 21 nucleótidos contiguos de la complementaria del transcrito de ARN del gen diana. Los ARNmi liberados se incorporan en complejos similares a RISC que contienen un subconjunto particular de proteínas Argonauta que ejercen una represión génica específica de la secuencia (véase, por ejemplo, Millar y Waterhouse, *Funct Integr Genomics*, 5: 129–135, 2005; Pasquinelli et al., *Curr Opin Genet Develop* 15: 200–205, 2005; Almeida y Allshire, *Trends Cell Biol.* 15: 251–258, 2005).

10 Cosupresión

Otro enfoque biológico molecular que puede usarse para reducir específicamente la expresión génica es la cosupresión. El mecanismo de la cosupresión no se entiende bien, pero se piensa que implica silenciamiento génico postranscripcional (PTGS) y, en ese sentido, puede ser muy similar a muchos ejemplos de supresión antisentido. Implica la introducción de una copia adicional de un gen o un fragmento del mismo en una planta en la "orientación sentido" con respecto a un promotor para su expresión, que, como se usa en el presente documento, se refiere a la misma orientación que la transcripción y la traducción (si se produce) de la secuencia con respecto a la secuencia en el gen diana. El tamaño del fragmento sentido, su correspondencia con las regiones génicas diana y su grado de homología con el gen diana son los mismos que para las secuencias antisentido descritas anteriormente. En algunos casos, la copia adicional de la secuencia del gen interfiere con la expresión del gen de la planta diana. Se hace referencia a la memoria descriptiva de la patente WO 97/20936 y a la memoria descriptiva de la patente europea 0465572 para métodos de implementación de enfoques de cosupresión. Las moléculas de ARN antisentido, de cosupresión o de doble cadena pueden comprender también una región de ARN en gran parte bicatenaria, que comprende, preferentemente, una señal de localización nuclear, como se describe en el documento WO 03/076619.

Cualquiera de estas tecnologías para reducir la expresión génica puede utilizarse para reducir de forma coordinada la actividad de múltiples genes. Por ejemplo, una molécula de ARN puede ser dirigida a una familia de genes relacionados usando como diana una región de los genes que es común. Como alternativa, los genes no relacionados pueden utilizarse como dianas a través de la inclusión de múltiples regiones en una molécula de ARN, estando cada región dirigida a un gen diferente. Esto se puede hacer fácilmente fusionando las múltiples regiones bajo el control de un solo promotor, tal como en el Ejemplo 5.

Métodos de introducción de ácidos nucleicos en células vegetales/transformación

Existen varias técnicas para la introducción de moléculas de ácido nucleico en una célula huésped vegetal, bien conocidas por los trabajadores en la técnica. El término "transformación" significa la alteración del genotipo de un organismo, por ejemplo una bacteria o una planta, mediante la introducción de un ácido nucleico extraño o exógeno. Por "transformante" se entiende un organismo alterado de este modo. Tal como se utiliza en el presente documento, el término "transgénico" se refiere a una planta genéticamente modificada en la que el genoma endógeno se complementa o modifica mediante la integración o mantenimiento estable en una forma no replicable de un gen o secuencia exógena o exógena. Por "transgén" se entiende un gen exógeno o extraño o secuencia exógena o extraña que se introduce en el genoma de una planta. La molécula de ácido nucleico puede integrarse de forma estable en el genoma de la planta o puede replicarse como un elemento extracromosómico. Por "genoma" se entiende el complemento genético total heredado de la célula, planta o parte de la planta, e incluye moléculas de ADN cromosómico, ADN plastídico, ADN mitocondrial y de ADN extracromosómico. El término "regeneración", tal como se usa en el presente documento en relación con materiales vegetales, significa el cultivo de una planta completa diferenciada de una célula vegetal, un grupo de células vegetales, una parte de la planta, tal como, por ejemplo, de un embrión, un escutelo, un protoplasto, un callo u otro tejido, pero sin incluir el crecimiento de una planta a partir de una semilla.

La elección particular de una tecnología de transformación estará determinada por su eficiencia para transformar ciertas especies vegetales, así como la experiencia y la preferencia de la persona que practica la invención con una metodología de elección particular. Será evidente para el experto en la técnica que la elección particular de un sistema de transformación para introducir una construcción de ácido nucleico en las células vegetales no es esencial ni una limitación de la invención, siempre que logre un nivel aceptable de transferencia de ácido nucleico. La guía en la implementación práctica de sistemas de transformación para la mejora de las plantas se proporciona en Birch, *Ann Rev Plant Physiol Plant Mol Biol.* 48: 297–326, 1997.

En principio, tanto las plantas dicotiledóneas como las monocotiledóneas que son susceptibles de transformación pueden modificarse introduciendo una construcción de ácido nucleico de acuerdo con la invención en una célula receptora y cultivando una nueva planta que alberga y expresa un polinucleótido de acuerdo con la invención.

Se ha demostrado que la introducción y la expresión de polinucleótidos extraños o exógenos en plantas dicotiledóneas, tales como algodón, tabaco, patata y legumbres, o plantas monocotiledóneas, tales como cereales, incluyendo trigo, cebada, arroz, maíz, avena, centeno y sorgo, es posible usando el T-AND del plásmido inductor de tumores (Ti) de *Agrobacterium tumefaciens* (véase, por ejemplo, Umbeck, patente de Estados Unidos n.º 5.004.863,

y la solicitud internacional PCT/US93/02480). Se puede introducir una construcción de la invención en una célula vegetal utilizando *A. tumefaciens* que contiene el plásmido Ti. Al utilizar un cultivo de *A. tumefaciens* como vehículo de transformación, es muy ventajoso utilizar una cepa no oncogénica de *Agrobacterium* como portador de vectores de manera que sea posible la diferenciación normal no oncogénica de los tejidos transformados. Se prefiere que el *Agrobacterium* albergue un sistema de plásmido Ti binario. Dicho sistema binario comprende (1) un primer plásmido Ti que tiene una región de virulencia esencial para la introducción del ADN de transferencia (T-ADN) en plantas, y (2) un plásmido quimérico. El plásmido quimérico contiene al menos una región de borde de la región de T-ADN de un plásmido Ti de tipo silvestre que flanquea el ácido nucleico que se va a transferir. Se ha demostrado que los sistemas de plásmidos Ti binarios son eficaces para transformar las células vegetales como, por ejemplo, describen De Framond, *Biotechnology*, 1: 262, 1983 y Hoekema et al., *Nature*, 303: 179, 1983. Dicho sistema binario se prefiere entre otros porque no requiere la integración en el plásmido Ti en *Agrobacterium*.

Los métodos que implican el uso de *Agrobacterium* incluyen, pero no se limitan a, la transformación de células o tejidos vegetales con *Agrobacterium*, tal como la transformación de semillas, ápices o meristemas con *Agrobacterium*, o la inoculación en la planta, tal como el método de inmersión floral descrito por Bechtold et al, *CR Acad. Sci. Paris*, 316: 1194, 1993. Este abordaje se basa en la infiltración de una suspensión de células de *Agrobacterium*. Como alternativa, la construcción quimérica puede introducirse usando plásmidos inductores de raíces (Ri) de *Agrobacterium* como vectores.

Los métodos para la transformación de plantas de algodón mediante la introducción de un ácido nucleico exógeno y para la regeneración de plantas a partir de células mediante embriogénesis somática son bien conocidos en la técnica.

Otros métodos para introducir la construcción de ácido nucleico en una célula vegetal son por electroporación, o penetración balística a alta velocidad por partículas pequeñas (también conocido como bombardeo de partículas o bombardeo con microproyectiles) con el ácido nucleico que se va a introducir dentro de la matriz de pequeñas perlas o partículas, o en su superficie como, por ejemplo, describen Klein et al., *Nature*, 327: 70, 1987.

Tal como se utiliza en el presente documento, una "mutación inducida" es una variación genética inducida artificialmente que puede ser el resultado de radiación química o mutagénesis basada biológicamente, por ejemplo transposón o inserción de T-ADN. Las mutaciones preferentes son mutaciones nulas, tales como mutaciones sin sentido, mutaciones de cambio de marco, mutaciones de inserción o variantes de sitio de corte y empalme que inactivan completamente el gen de interés. Entre los derivados de inserción de nucleótidos se incluyen fusiones en los extremos 5' y 3', así como inserciones dentro de la secuencia de uno o más nucleótidos. Las variantes de la secuencia de nucleótidos de inserción son aquellas en las que uno o más nucleótidos se introducen en la secuencia de nucleótidos, que puede obtenerse por inserción aleatoria con una detección selectiva adecuada de los productos resultantes. Las variantes de delección se caracterizan por la eliminación de uno o más nucleótidos de la secuencia. Preferentemente, un gen mutante tiene solo una sola inserción o delección de una secuencia de nucleótidos con relación al gen de tipo silvestre. Las variantes por sustitución de nucleótidos son aquellas en las que se ha eliminado al menos un nucleótido en la secuencia y se ha insertado un nucleótido diferente en su lugar. El número preferente de nucleótidos afectados por sustituciones en un gen mutante con relación al gen de tipo Silvestre es un máximo de diez nucleótidos, más preferentemente un máximo de 9, 8, 7, 6, 5, 4, 3 o 2, o solo un nucleótido. Dicha sustitución puede ser "silenciosa" en el sentido en que la sustitución no cambia el aminoácido definido por el codón. Como alternativa, los sustituyentes conservadores están diseñados para alterar un aminoácido para otro aminoácido de acción similar.

El término "mutación", tal como se utiliza en el presente documento, no incluye sustituciones silenciosas de nucleótidos que no afectan a la actividad del gen, y, por lo tanto, incluye solo alteraciones en la secuencia génica que afectan a la actividad génica. El término "polimorfismo" se refiere a cualquier cambio en la secuencia de nucleótidos que incluye dichas sustituciones de nucleótidos silenciosas.

La mutagénesis puede lograrse mediante medios químicos o de radiación, por ejemplo tratamiento de las semillas con EMS o azida sódica (Zwar y Chandler, *Planta* 197: 39–48, 1995), o irradiación gamma, bien conocidos en la técnica. El aislamiento de mutantes puede conseguirse mediante detección selectiva de plantas o semillas mutagenizadas. En una planta poliploide, tal como algodón, la detección selectiva se realiza, preferentemente, en un genotipo que ya carece de una de las actividades enzimáticas, de modo que se busca un mutante que carezca completamente de la actividad funcional. Como alternativa, la mutación puede identificarse usando técnicas tales como "TILLING" en una población mutagenizada con un agente tal como EMS (Slade y Knauf, *Transgenic Res.* 14: 109–115, 2005). Dichas mutaciones pueden introducirse después en fondos genéticos deseables cruzando el mutante con una planta del fondo genético deseado y realizando un número adecuado de retrocruzamientos para cruzar el fondo parental originalmente no deseado. La invención se extiende claramente a métodos para producir o identificar tales plantas o la semilla producida por tales plantas.

El proceso de producción de plantas de la invención puede incluir etapas de mutagénesis y/o detección selectiva tales como TILLING (direccionamiento de lesiones locales inducidas en genomas). En un primer paso, las mutaciones introducidas, tales como nuevos cambios de pares de bases simples se inducen en una población de

plantas tratando las células, semillas, polen u otras partes de plantas con un mutágeno químico o radiación química y, después, haciendo avanzar las plantas a una generación en la que se las mutaciones se heredarán de manera estable. El ADN se extrae y las semillas se almacenan de todos los miembros de la población para crear un recurso al que se puede acceder repetidamente con el tiempo. Para un ensayo de TILLING, los cebadores de PCR están
 5 diseñados para amplificar específicamente una sola Diana génica de interés. La especificidad es especialmente importante si una diana es un miembro de una familia de genes o parte de un genoma poliploide. A continuación, se pueden usar cebadores marcados con colorante para amplificar productos de PCR a partir de ADN agrupado de múltiples individuos. Estos productos de PCR se desnaturalizan y se rehibridan para permitir la formación de pares de bases apareados erróneamente. Los apareamientos erróneos, o heterodúplex, representan tanto los
 10 polimorfismos de nucleótidos individuales (SNP) que ose producen de forma natural (es decir, es probable que varias plantas de la población porten el mismo polimorfismo) como los SNP inducidos (es decir, solo las plantas individuales raras es probable que muestren la mutación). Después de la formación de heterodúplex, el uso de una endonucleasa, tal como Cell, que reconoce y escinde el ADN apareado erróneamente, es la clave para descubrir nuevos SNP dentro de una población de TILLING.

Utilizando este enfoque, se pueden someter a detección selectiva miles de plantas para identificar cualquier individuo con un solo cambio de base, así como pequeñas inserciones o deleciones (1-30 pb) en cualquier gen o región específica del genoma. Los fragmentos genómicos que se analizan pueden variar de tamaño desde 0,3 hasta 1,6 kb. En una agrupación multiplicada por 8 veces, los fragmentos de 1,4 kb (descontando los extremos de los
 20 fragmentos en los que la detección de SNP es problemática debido al ruido) y 96 carriles por ensayo, esta combinación permite detectar hasta un millón de pares de bases de ADN genómico por ensayo único, lo que convierte al TILLING en una técnica de alto rendimiento. El TILLING se describe adicionalmente en Slade y Knauft, 2005 (citado anteriormente), y Henikoff et al., *Plant Physiol.* 135: 630–636, 2004.

Además de permitir la detección eficiente de mutaciones, la tecnología de TILLING de alto rendimiento es ideal para la detección de polimorfismos naturales. Por lo tanto, la interrogación de un ADN homólogo desconocido mediante formación de heterodúplex a una secuencia conocida revela el número y la posición de sitios polimórficos. Se identifican tanto los cambios de nucleótidos como las pequeñas inserciones y deleciones, incluyendo al menos algunos polimorfismos de número de repetición. Esto se ha denominado Ecotiling (Comai et al., *Plant J.* 37: 778–
 30 786, 2004).

Tal como se utiliza en el presente documento, la expresión "unido genéticamente" se refiere a un locus marcador y a un segundo locus que está suficientemente cerca en un cromosoma que se heredarán juntos en más del 50 % de las meiosis, por ejemplo, no aleatoriamente. Esta definición incluye la situación en la que el locus marcador y el segundo locus forman parte del mismo gen. Además, esta definición incluye la situación en la que el locus marcador comprende un polimorfismo que es responsable del rasgo de interés (en otras palabras, el locus marcador está directamente "ligado" al fenotipo). Por lo tanto, el porcentaje de recombinación observado entre los loci por generación (centimorgans (cM)) será menor que 50. En realizaciones particulares de la invención, los loci unidos genéticamente pueden ser 45, 35, 25, 15, 10, 5, 4, 3, 2, o 1 o menos cM aparte en un cromosoma. Preferentemente, los marcadores están separados por menos de 5 cM o 2 cM, y lo más preferentemente, separados por aproximadamente 0 cM.

Como se usa en el presente documento, los "otros marcadores genéticos" pueden ser cualquier molécula que esté ligada a un rasgo deseado de una planta de algodón. Dichos marcadores son bien conocidos por los expertos en la técnica e incluyen marcadores moleculares ligados a genes que determinan rasgos tales como la resistencia a enfermedades, el rendimiento, la morfología de la planta, la calidad de la pelusa. Cualquier técnica biológica molecular conocida en la técnica que sea capaz de detectar alelos de genes de interés puede usarse en los métodos de la presente invención. Tales métodos incluyen, pero no se limitan a, el uso de amplificación de ácidos nucleicos, secuenciación de ácidos nucleicos, hibridación de ácidos nucleicos con sondas adecuadamente marcadas, análisis conformacional de una sola hebra (SSCA), electroforesis en gel de gradiente desnaturalizante (DGGE), análisis heterodúplex (HET), Análisis de escisión química (CCM), escisión catalítica de ácido nucleico o una combinación de los mismos. La invención también incluye el uso de técnicas de marcadores moleculares para detectar polimorfismos ligados a alelos de. Tales métodos incluyen la detección o análisis de polimorfismos de longitud de fragmentos de restricción (RFLP), RAPD, polimorfismos de longitud de fragmentos amplificados (AFLP) y polimorfismos de microsatélites (repetición de secuencia simple, SSR). Los marcadores estrechamente ligados pueden obtenerse fácilmente por métodos bien conocidos en la técnica, tal como el análisis de segregantes agrupados.

La "reacción en cadena de la polimerasa" ("PCR") es una reacción en la que se hacen copias duplicadas de un polinucleótido diana usando un "par de cebadores" o "conjunto de cebadores" que consiste en un cebador "aguas arriba" y "aguas abajo", un catalizador de la polimerización, tal como una ADN polimerasa, y, normalmente, una enzima polimerasa térmicamente estable. Los métodos para la PCR son conocidos en la técnica y se enseñan, por ejemplo, en "PCR" (McPherson y Moller (Ed), BIOS Scientific Publishers Ltd, Oxford, 2000). La PCR puede realizarse sobre ADNc obtenido a partir de transcripción inversa del ARNm aislado a partir de células de plantas que expresan un gen SSII o sobre ADN genómico aislado de una planta. Un cebador en este contexto es una secuencia oligonucleotídica que es capaz de hibridar de una manera específica de la secuencia con la secuencia diana y que se extiende durante la PCR. Los amplicones o productos de la PCR o fragmentos de PCR o productos de

amplificación son productos de extensión que comprenden el cebador y las copias recién sintetizadas de las secuencias diana. Los cebadores pueden coincidir perfectamente con la secuencia diana o pueden contener bases internas apareadas erróneamente que pueden dar lugar a la introducción de sitios de reconocimiento/escisión de enzimas de restricción o de ácido nucleico catalítico en secuencias diana específicas. Los cebadores también pueden contener secuencias adicionales y/o contener nucleótidos modificados o marcados para facilitar la captura o detección de amplicones. Ciclos repetidos de desnaturalización por calor del ADN, hibridación de los cebadores con sus secuencias complementarias y extensión de los cebadores hibridados con la polimerasa dan como resultado una amplificación exponencial de la secuencia Diana. Los términos objetivo o secuencia diana o molde se refieren a secuencias de ácido nucleico que se amplifican.

Los métodos para la secuenciación directa de las secuencias de nucleótidos son bien conocidos por los expertos en la técnica y pueden encontrarse, por ejemplo, en Ausubel *et al.* (citado anteriormente) y Sambrook *et al.*, 1989 (citado anteriormente). La secuenciación puede llevarse a cabo mediante cualquier método adecuado, por ejemplo, secuenciación dideoxi, secuenciación química o variaciones de las mismas. La secuenciación directa tiene la ventaja de determinar la variación en cualquier par de bases de una secuencia particular.

Plantas

El término "planta", tal como se utiliza en el presente documento como sustantivo, se refiere a plantas enteras y se refiere a cualquier miembro del Reino Plantae, pero cuando se usa como adjetivo se refiere a cualquier sustancia que está presente, se obtiene, deriva o está relacionada con una planta, tales como, por ejemplo, órganos de las plantas (por ejemplo, hojas, tallos, raíces, flores), células individuales (por ejemplo, polen), semillas, células vegetales y similares. Las plántulas y las semillas germinadas de las que han surgido raíces y brotes también se incluyen dentro del significado de "planta". La expresión "partes de plantas" tal como se usa en el presente documento se refiere a uno o más tejidos u órganos vegetales que se obtienen de una planta y que comprenden ADN genómico de la planta. Partes de planta incluyen estructuras vegetativas (por ejemplo, hojas, tallos), raíces, órganos/estructuras florales, semillas (incluidos embriones, cotiledones y cubierta de la semilla), tejido vegetal (por ejemplo, tejido vascular, tejido fundamental y similares), células y progenie de las mismas. El término "célula vegetal", tal como se usa en el presente documento, se refiere a una célula obtenida de una planta o en una planta e incluye protoplastos u otras células derivadas de plantas, células productoras de gametos y células que se regeneran en plantas enteras. Las células vegetales pueden ser células en cultivo. Por "tejido vegetal" se entiende tejido diferenciado en una planta u obtenido de una planta ("explante") o tejido indiferenciado derivado de embriones inmaduros o maduros, semillas, raíces, brotes, frutos, tubérculos, polen, tejido tumoral tal como agallas de corona y varias formas de agregaciones de células vegetales en cultivo, tales como callos. Ejemplos de tejidos vegetales en o de las semillas son cotiledón, embrión y eje embrionario. Por consiguiente, la invención incluye plantas y partes de plantas y productos que las comprenden, particularmente semillas que comprenden la composición de aceite modificada.

Tal como se utiliza en el presente documento, el término "semilla" se refiere a la "semilla madura" de una planta, que está lista para la recolección o que se ha recolectado de la planta, tal como normalmente se recolecta comercialmente en el campo o como "semilla en desarrollo" que se produce en una planta después de la fertilización y antes de que se establezca la latencia de las semillas y antes de la recolección.

Una "planta transgénica", como se usa en el presente documento, se refiere a una planta que contiene una construcción génica que no se encuentra en una planta silvestre de la misma especie, variedad o cultivo. Es decir, las plantas transgénicas (plantas transformadas) contienen material genético (un transgén) que no contenían antes de la transformación. El transgén puede incluir secuencias genéticas obtenidas o derivadas de una célula vegetal, u otra célula vegetal, o una fuente no vegetal, o una secuencia sintética. Normalmente, el transgén se ha introducido en la planta mediante manipulación genética tal como, por ejemplo, mediante transformación, pero se puede usar cualquier procedimiento que reconozca un experto en la técnica. El material genético está integrado, preferentemente, de forma estable en el genoma de la planta. El material genético introducido puede comprender secuencias que se producen de forma natural en la misma especie pero en un orden reordenado o en una disposición diferente de elementos, por ejemplo una secuencia antisentido. Las plantas que contienen tales secuencias se incluyen en el presente documento en "plantas transgénicas". Una "planta no transgénica" es una que no se ha modificado genéticamente mediante la introducción de material genético por técnicas de ADN recombinante. En una realización preferente, las plantas transgénicas son homocigotas para cada gen que se ha introducido (transgén) de modo que su progenie no segregue el fenotipo deseado.

Tal como se utiliza en el presente documento, la expresión "planta no transgénica correspondiente" se refiere a una planta que es isogénica con relación a la planta transgénica pero sin el transgén de interés. Preferentemente, la planta no transgénica correspondiente es del mismo cultivo o variedad que el progenitor de la planta transgénica de interés, o una línea de plantas hermanas que carece de la construcción, a menudo denominado "segregante", o una planta del mismo cultivo o variedad transformado con una construcción de "vector vacío" y puede ser una planta no transgénica. "Tipo silvestre", tal como se usa en el presente documento, se refiere a una célula, tejido o planta que no se ha modificado de acuerdo con la invención. Se pueden usar células, tejidos o plantas de tipo silvestre como controles para comparar los niveles de expresión de un ácido nucleico exógeno o la extensión y la naturaleza de la

modificación del rasgo con células, tejidos o plantas modificadas como se describe en el presente documento. Una planta de tipo silvestre típica con respecto al algodón es de la variedad "Coker".

Las plantas transgénicas, como se definen en el contexto de la presente invención, incluyen la progenie de las plantas que se han modificado genéticamente usando técnicas recombinantes, en las que la progenie comprende el transgén de interés. Dicha progenie puede obtenerse por autofecundación de la planta transgénica primaria o mediante cruzamiento de dichas plantas con otra planta de la misma especie. Esto generalmente sería modular la producción de al menos una proteína/enzima definida en el presente documento en la planta u órgano de la planta deseado. Las partes de plantas transgénicas incluyen todas las partes y células de dichas plantas que comprenden el transgén, tal como, por ejemplo, tejidos cultivados, callos y protoplastos.

Se puede emplear cualquiera de varios métodos para determinar la presencia de un transgén en una planta transformada. Por ejemplo, se puede usar la reacción en cadena de polimerasa (PCR) para amplificar secuencias que son únicas para la planta transformada, con detección de los productos amplificados mediante electroforesis en gel u otros métodos. El ADN se puede extraer de las plantas usando métodos convencionales y la reacción de PCR se lleva a cabo utilizando cebadores para amplificar un ADN específico, cuya presencia distinguirá las plantas transformadas de las no transformadas. Por ejemplo, se pueden diseñar cebadores que amplifiquen una región de ADN desde la lectura del vector de transformación en la construcción y el cebador inverso diseñado a partir del gen de interés. Estos cebadores solo amplificarán un fragmento si la planta se ha transformado con éxito. Un método alternativo para confirmar un transformante positivo es mediante hibridación de transferencia Southern, bien conocida en la técnica. Las plantas que se transforman también pueden identificarse, es decir, distinguirse de las plantas no transformadas o de tipo silvestre por su fenotipo, por ejemplo conferido por la presencia de un gen marcador seleccionable, o conferido por el fenotipo de la composición de aceite modificada de la semilla de la planta, o fenotipo relacionado, tal como la actividad enzimática alterada.

Como se usa en el presente documento, "germinación" se refiere a la aparición de la punta de la raíz de la cubierta de la semilla después de la imbibición. La "tasa de germinación" se refiere al porcentaje de semillas en una población que ha germinado durante un período de tiempo, por ejemplo 14 o 21 días, después de la imbibición. Una población de semillas puede evaluarse diariamente durante varios días para determinar el porcentaje de germinación con el tiempo. Con respecto a las semillas de la presente invención, tal como se usa en el presente documento, la expresión "tasa de germinación que es sustancialmente el mismo" significa que la tasa de germinación de las semillas transgénicas es al menos el 80 % de la de las semillas silvestres isogénicas.

Las plantas proporcionadas o contempladas para su uso en la práctica de la presente invención incluyen angiospermas, incluyendo tanto monocotiledóneas como dicotiledóneas. En realizaciones preferentes, las plantas de la presente invención son plantas de cultivo (por ejemplo, cereales y legumbres, maíz, trigo, patatas, tapioca, arroz, sorgo, mijo, yuca, cebada o guisantes) u otras legumbres. Las plantas se pueden cultivar para la producción de raíces, tubérculos, hojas, tallos, flores o frutos comestibles. Preferentemente, la planta es una planta de algodón. Ejemplos de plantas de cereales incluyen, pero no se limitan a, trigo, cebada, arroz, maíz (maíz), sorgo, avena y centeno. En una realización, la planta de algodón es una planta progenie de la línea DCS9 (Ejemplo 7) y contiene el mismo transgén situado en la misma posición en el genoma.

La referencia a una "planta de algodón" se refiere a cualquier planta perteneciente al género *Gossypium* y a cualquier pariente silvestre, especie progenitora, germoplasma, cultivo o variedad de los mismos. Por ejemplo, las plantas de algodón incluyen especies tales como: *G. anapoides*, *G. anomalum*, *G. arboretum*, *G. areysianum*, *G. aridum*, *G. armourianum*, *G. australe*, *G. barbadense*, *G. barbosanum*, *G. benadireense*, *G. bickii*, *G. bricetti*, *G. capitiviridis*, *G. costulatum*, *G. cunninghamii*, *G. darwinii*, *G. davidsonii*, *G. enthyale*, *G. exiguum*, *G. gossypoides*, *G. harknessii*, *G. herbaceum*, *G. hirsutum*, *G. incanum*, *G. klotzschianum*, *G. laxum*, *G. lobatum*, *G. londonderriense*, *G. longicalyx*, *G. marchantii*, *G. mustelinum*, *G. nandewarense*, *G. nelsonii*, *G. nobile*, *G. pilosum*, *G. populifolium*, *G. pulchellum*, *G. raimondii*, *G. robinsonii*, *G. rotundifolium*, *G. schwendimaantii*, *G. somalense*, *G. soudanense*, *G. stocksii*, *G. sturtianum*, *G. thurberi*, *G. timorense*, *G. tomentosum*, *G. trilobum*, *G. triphyllum*, y *G. vindis*. Especies concretas son *Gossypium hirsutum* (gh) y *Gossypium barbadense* (gb). Como apreciarán los expertos en la técnica, una variedad, cultivo o mutante convenientes puede estar modificado genéticamente y el fenotipo producido en variedades o especies relacionadas mediante técnicas de reproducción estándar. La referencia a "progenitor" se refiere a cualquiera de las especies, variedades, cultivos o germoplasmas, de los cuales deriva una planta. Como conocerán los expertos en la técnica, el algodón más comercialmente útil es un alotetraploide de origen natural derivado mediante hibridación sexual entre padres progenitores diploides antiguos. Basándose en la enseñanza proporcionada en el presente documento, los expertos en la técnica pueden realizar fácilmente el método de la invención en uno o más algodones diploides o alotetraploides y producir híbridos sexuales entre las plantas transgénicas producidas a partir de los mismos.

Como se usa en el presente documento, se entenderá que la expresión "especies derivadas, germoplasma o variedad" significa cualquier especie vegetal, germoplasma o variedad que se produce usando una especie, variedad, cultivo o germoplasma de algodón indicado, utilizando procedimientos estándar de hibridación sexual, tecnología de ADN recombinante, cultivo de tejidos, mutagénesis, o una combinación de uno o más de dichos procedimientos. En particular, se han producido híbridos interespecíficos entre diversas especies importantes de

algodón, tales como *G. Barbadosense* Y *G. hirsutum*, y entre ciertas especies diploides, y la producción de tales híbridos interespecíficos es rutinaria para los expertos en la técnica.

5 Tal como se utiliza en el presente documento, el término "algodón" se refiere a cualquier especie del género *Gossypium*, preferentemente de la especie *Gossypium hirsutum* o *Gossypium barbadense*. En algunas realizaciones, una planta control o de referencia adecuada es una planta de algodón sustancialmente isogénica no transformada tratada de la misma manera que la planta de "ensayo". El aceite de semilla de algodón se puede distinguir por la presencia en el mismo de ácidos grasos de CPE y/o de CPA.

10 Producción de alimentos

En otro aspecto, la invención proporciona plantas de algodón y semilla de algodón, y productos obtenidos a partir de ellas, que comprenden aceite de la semilla, particularmente aceite de semilla de algodón, que es útil para la producción de alimentos o piensos, teniendo la semilla una composición del aceite de la semilla modificada en comparación con la semilla de tipo silvestre correspondiente. Preferentemente, la planta a partir de la cual se obtiene la semilla tiene un nivel reducido de actividades de ghFAD2-1, ghFatB-2 y ghCPA-FAS-2 en la semilla durante el desarrollo. El aceite de semilla de algodón de la presente invención es útil para la producción de alimentos y, en particular, para la producción comercial de alimentos. Tal producción de alimentos podría incluir la mezcla del aceite de semilla de algodón como un ingrediente con otros ingredientes en la producción comercial de alimentos. En realizaciones preferentes que son deseables para su uso en la producción de alimentos, el aceite de semilla de algodón tiene una composición modificada como se especifica en el presente documento.

El aceite se aísla fácilmente a partir de semillas limpias, deshinchadas y descascaradas de la invención usando métodos estándar, por ejemplo cocción a altas temperaturas, prensado, triturado usando una prensa de tornillo (alta presión) y/o procedimientos para la extracción de aceite usando disolventes, vapor y/o presión alta. El contenido de aceite de la semilla de algodón, o el contenido de cualquier ácido graso en el aceite de la semilla de algodón se determina convenientemente como se describe en el presente documento o es bien entendido en la técnica. Como alternativa o además se pueden usar los procedimientos de Folch et al., J. Biol. Chem. 226: 497, 1957 o variaciones de los mismos como se describe en otra parte (véase, por ejemplo, Liu et al., 2002 (citado anteriormente)). El contenido de ácidos grasos y/o la composición del aceite de semilla de algodón pueden determinarse convenientemente utilizando cromatografía de gases líquidos contra ácidos grasos estándar conocidos, comparando los picos del éster metílico de ácidos grasos y los tiempos de retención de los patrones de referencia con la muestra que se está analizando y mediante la integración estándar de los picos obtenidos. Sin embargo, la presente invención no está limitada por el método para determinar el contenido y/o la composición del aceite de semilla de algodón, en particular los medios para determinar los ácidos grasos u otros componentes lipídicos. El aceite está compuesto casi totalmente por triacilglicerol (TAG) que comprenden tres ácidos grasos esterificados con una cadena principal de glicerol. Para evaluar el contenido de TAG del aceite, el aceite puede purificarse tal como mediante extracción en fase sólida (SPE) en cartuchos de gel de sílice. La composición de TAG puede evaluarse cualitativamente mediante cromatografía líquida de alta resolución de fase inversa (HPLC) usando un detector del índice de refracción y propionitrilo como fase móvil. A partir de aceite purificado, se preparan ésteres metílicos de ácidos grasos (FAME), tal como mediante metilación con solución fría de KOH en metanol y los ésteres analizados por cromatografía de gases capilar (GC) usando columnas polares altas. A partir de la composición de ácidos grasos, la composición teórica de TAG se puede calcular mediante un programa informático que emplea una distribución típica de ácidos grasos en el triacilglicerol para el aceite de semilla de algodón. Los algoritmos matemáticos se pueden calcular a partir de composiciones teóricas y experimentales (HPLC) de triacilglicerol y los valores resultantes se comparan con los contenidos en una base de datos que comprende conjuntos de datos determinados realizando el análisis sobre diferentes patrones de aceites.

50 Productos alimentarios

La memoria descriptiva también describe alimentos, bebidas o preparaciones farmacéuticas producidas con productos que comprenden el aceite de semilla de algodón. La planta de la invención o productos derivados de la misma que contienen aceite o pelusa se pueden utilizar en diversas aplicaciones para uso o consumo humano. Tal como se usa en el presente documento, "seres humanos" hace referencia a *Homo sapiens*. El aceite de semilla de algodón se puede utilizar fácilmente en los procedimientos de procesamiento de alimentos, en particular en la fritura de productos alimentarios. El aceite puede incorporarse en productos, tales como margarina, manteca, aderezo para ensalada, mayonesa, productos lácteos tales como helado, yogur o queso, o añadirse como ingrediente a otros alimentos o materiales alimentarios, tales como pan, tartas, galletas, pasteles, cereales de desayuno, pasta, fideos o salsas.

Otras partes de las plantas de la invención que son comestibles pueden usarse como alimentos para consumo humano o como pienso para uso animal. Por ejemplo, pueden utilizarse para consumo humano o animal hojas, tallos, raíces, tubérculos, frutos, vainas o extractos o partes de estos que comprenden células, como se describe en el presente documento de cualquiera de estos. El contenido y la composición de aceite modificado de las plantas de la invención y partes de las mismas pueden proporcionar ventajas para el uso de estos materiales como alimento animal tal como, por ejemplo, alimento para cerdos, ganado vacuno, caballos, aves de corral, tales como pollos, y

otros animales.

El alimento o la bebida o la preparación farmacéutica pueden envasarse listos para la venta o en forma a granel. La memoria descriptiva describe también métodos para preparar alimentos, bebidas o preparados farmacéuticos de la invención, y recetas o instrucciones para preparar dichos alimentos o bebidas. Los métodos pueden comprender las etapas de trituración, extracción, molienda, cocción, fritura, enlatado, envasado u otras etapas de procesamiento conocidas en la técnica. Los métodos o recetas o instrucciones pueden incluir las etapas de procesar el aceite de la invención y/o mezclarlo con otros ingredientes alimentarios, tales como calentar u hornear la mezcla o el producto a, por ejemplo, al menos 100 °C. El método puede incluir la etapa de empaquetar el producto para que esté listo para la venta.

Métodos

Los productos de la presente invención pueden formularse en composiciones farmacéuticas que se preparan de acuerdo con técnicas convencionales de composición farmacéutica. Véase, por ejemplo, "Remington's Pharmaceutical Sciences", 18ª ed. (1990, Mack Publishing, Company, Easton, PA, EE.UU.). La composición puede contener el agente activo o un agente activo derivado farmacéuticamente aceptable. Estas composiciones pueden comprender, además de una de las sustancias activas, un excipiente, vehículo, tampón, estabilizante u otros materiales farmacéuticamente aceptables bien conocidos en la técnica. Tales materiales deben ser no tóxicos y no deben interferir con la eficacia del ingrediente activo. El vehículo puede tomar una amplia variedad de formas según la forma de preparación deseada para la administración.

Para la administración oral, los compuestos se pueden formular en preparaciones líquidas tales como cápsulas. En la preparación de las composiciones en forma de dosificación oral, se pueden emplear cualquiera de los medios farmacéuticos usuales, tales como, por ejemplo, agua, glicoles, aceites, alcoholes, agentes aromatizantes, conservantes, agentes colorantes, agentes de suspensión y similares en el caso de preparaciones líquidas orales (tales como, por ejemplo, suspensiones, elixires y soluciones); o vehículos tales como almidones, azúcares, diluyentes, agentes de granulación, lubricantes, aglutinantes.

El agente activo se administra, preferentemente, en una cantidad terapéuticamente eficaz. La cantidad real administrada y la velocidad y evolución en el tiempo de la administración dependerán de la naturaleza y la gravedad de la afección que se esté tratando. La prescripción del tratamiento, por ejemplo, las decisiones sobre la dosificación, el momento, etc., están bajo la responsabilidad de médicos generales o especialistas y, por lo general, tienen en cuenta el trastorno a tratar, la afección del paciente individual, el lugar de administración, el método de administración y otros factores conocidos por los profesionales. Ejemplos de técnicas y protocolos se pueden encontrar en Remington's Pharmaceutical Sciences, (citado anteriormente).

Uso industrial

Los productos vegetales, preferentemente aceite, se pueden utilizar en la producción de productos industriales tales como, por ejemplo, etanol.

La presente invención se describe de forma adicional mediante los siguientes ejemplos no limitantes.

45 Ejemplo 1: Métodos y materiales ilustrativos

Aislamiento de ARN

Se trituraron dos gramos de embriones de algodón o tejido de hojas congelado con nitrógeno líquido hasta obtener un polvo fino usando un mortero y una mano de mortero y se transfirieron a un vaso de precipitados que contenía 22 ml de tampón de extracción en frío y se agitaron constantemente. El tampón de extracción contenía Tris-HCl 200 mM pH 8,5, dodecilsulfato de litio al 1,5 %, LiCl 300 mM, Na₂EDTA 10 mM, desoxicolato sódico al 1 %, Nonidet P-40 al 1 %. A esto le siguió la adición de PVP insoluble al 5 %, mercaptoetanol 90 mM, DTT 10 mM (ditiotreitól), DEPC al 0,1 %, y se agitó durante 10 minutos antes de transferirse a un tubo Corex. Después se añadieron 18,4 ml de acetato de amonio 3 M y se mezclaron bien. Se centrifugó a 6.000 x rpm durante 20 minutos a 4 °C. El sobrenadante se transfirió a un nuevo tubo y se precipitó añadiendo 1/10 del volumen de NaAc 3 M, pH 5,2 y 1/2 del volumen final de isopropanol frío y se almacenó a -20 °C durante 1 hora antes de la centrifugación a 6.000 x rpm durante 30 minutos utilizando un rotor giratorio. El sedimento se resuspendió en 1 ml de dH₂O y se transfirió a dos tubos Eppendorf (500 µl en cada tubo). La suspensión se extrajo con un volumen igual de solución de fenol/cloroformo/alcohol isoamílico (25:24:1) y las fases se separaron mediante centrifugación durante 5 minutos a 4 °C. La capa superior acuosa se transfirió cuidadosamente a un nuevo tubo Eppendorf y se extrajo de nuevo con cloroformo como anteriormente. Se añadió medio volumen de LiCl 5M a la muestra acuosa, se mezcló bien y se dejó en hielo durante 3 horas antes de la centrifugación a 12.000 x rpm durante 15 minutos a 4 °C. El sedimento se resuspendió en 50 µl de dH₂O. Finalmente, la muestra de ARN se precipitó añadiendo 5 µl de NaAc y 138 µl de etanol frío y se incubó en hielo seco durante 30 minutos antes de la centrifugación durante 15 minutos a 4 °C. El sedimento de ARN se secó al vacío y después se disolvió en 30 µl de H₂O libre de ARNasa.

Construcción de una biblioteca de ADNc de semilla de algodón

Se aisló ARN poli(A)⁺ de algodón a partir de ARN total preparado como se ha descrito anteriormente, usando un kit de purificación de ARNm (Pharmacia), esencialmente como describe el fabricante. Para preparar el ADNc se utilizó un kit de síntesis de ADNc (Pharmacia), esencialmente como describe el fabricante, usando 1-5 µg de ARN poli (A)⁺ como material de partida. El producto de ADNc de doble cadena se cortó en extremos romos y se ligó a adaptadores de *EcoRI/NotI* usando procedimientos estándar. Después de la eliminación del exceso de adaptadores no ligados, el ADNc se clonó en el vector bacteriófago Lambda ZAPII (Stratagene, EE.UU.) y se empaquetó usando un sistema de envasado comercialmente disponible (Stratagene, EE.UU.), de acuerdo con las instrucciones del fabricante.

Las bibliotecas de ADNc descritas en el presente documento contenían generalmente aproximadamente 92 % de partículas de bacteriófago recombinante, en un total de aproximadamente $1,5 \times 10^7$ unidades formadoras de placas (ufp) por ml de biblioteca no amplificada.

Después de la purificación de las placas de hibridación positivas, se utilizó el sistema ExAssist/SOLR (Stratagene, EE.UU.) para eliminar el fagémido pBluescript SK (-) del vector Lambda ZAPII, según lo descrito por el fabricante.

Todos los otros métodos, tales como electroforesis en gel, transferencia de ácidos nucleicos a membranas para hibridación, preparación de sondas de ADN marcadas y selección de bibliotecas de ADNc fueron por métodos estándar (Ausubel et al., (citado anteriormente)). A menos que se indique lo contrario, las condiciones de hibridación fueron como las descritas por Khandjian, *Bio/Technology*, 5: 165–167, 1987) usando Tris-HCl 50 mM, pH 7,5, NaCl 1M, formamida al 50 %, 10x solución de Denhardt, sulfato de dextrano al 10 %, SDS al 1 %, pirofosfato sódico al 0,1 %, 0,1 mg/ml de ADN de esperma de arenque a 42 °C. Después, se lavaron las membranas se lavaron brevemente en 2x SSC, SDS al 0,1 % a 65 °C, seguido de dos lavados adicionales en SSC 0,2x, SDS al 0,1 % a 65 °C durante 15 minutos cada uno (alta rigurosidad).

Ejemplo 2: Aislamiento de genes *fad2* de algodón

Se aislaron cuatro genes DIFERENTES de FAD2 a partir de algodón, cada uno de los cuales codifica posibles oleoil-Δ12-desaturasas microsomales que podrían desaturar el ácido oleico en ácido linoleico (Liu *et al.*, 1999a (citado anteriormente), Liu *et al.*, 1999b (citado anteriormente); Kargiotidou et al., *Journal of Experimental Botany* 2008 59(8): 2043–2056, 2008). primer gen, designado *ghFAD2-1*, se expresó específicamente en semillas en desarrollo aproximadamente al mismo tiempo que la biosíntesis de aceite activo (Liu *et al.*, 1999a (citado anteriormente)). Dos secuencias de nucleótidos para *ghFAD2-1* se presentan en el presente documento (SEQ ID Nos: 1 y 24) que son idénticos al 96 % a lo largo de sus longitudes completas y, por lo tanto, probablemente representan ADNc correspondientes a alelos diferentes o, más probablemente, a los genes FAD2-1 homólogos en el algodón tetraploide. Un segundo gen, *ghFAD2-2*, (número de acceso Y10112) tenía una expresión constitutiva de bajo nivel y se expresó a un nivel bajo a lo largo del desarrollo de la semilla (Pirtle *et al.*, 2001 (citado anteriormente)). La secuencia nucleotídica de *ghFAD2-2* (SEQ ID NO: 26) es aproximadamente un 72 % idéntica a la mitad central de *ghFAD2-1* (SEQ ID NO: 1). El tercer y cuarto miembros *ghFAD2-3* (número de acceso AF331163) y *ghFAD2-4* (número de acceso AY279314) parecían expresarse más altamente en hojas y otros tejidos de algodón en lugar de semillas (Kargiotidou et al., 2008 (citado anteriormente)). *ghFAD2-3* y *ghFAD2-4* tienen secuencias de nucleótidos que son aproximadamente el 72 % idénticas a *ghFAD2-1* (SEQ ID NO: 1).

Basándose en estas observaciones y en la producción previa de aceite de semilla de algodón con alto contenido de ácido oleico mediante la regulación negativa por ARNi de *ghFAD2-1* (Liu et al., *Plant Physiol.* 129: 1732–1743, 2002; patente de Estados Unidos n.º 6974898), se pensó que este gen codificaba la principal actividad de FAD2 en las semillas de algodón. Por lo tanto, se seleccionó una región del gen *ghFAD2-1* que tenía la secuencia de nucleótidos 5-354 de SEQ ID NO: 1) para fabricar una construcción de ARNi para la regulación por disminución del gen. Esta era la misma región descrita por Liu et al., 2002 (citado anteriormente) y se esperaba que fuera específica para *ghFAD2-1*. Otras regiones del gen podrían haberse seleccionado igualmente. Por ejemplo, la patente de Estados Unidos n.º 6974898 usó una región de 92 pb a partir de la 5'-UTR de *ghFAD2-1*, y Liu et al., 2002 (citado anteriormente) usaron una región de 540 pb desde el extremo 5' de la parte transcrita del gen. La secuencia de nucleótidos y aminoácidos de FAD-2-1 se muestra en la Figura 3.

Ejemplo 3: Aislamiento de genes *fatb* de algodón

Se identificaron dos clones de ADNc que codificaban proteínas con homología con FatB en algodón y se aislaron a partir de una biblioteca de ADNc de semilla de algodón en desarrollo (Ejemplo 1) como sigue. Se amplificó un fragmento de PCR correspondiente a aproximadamente 480 pb del primer exón del gen *ghFatB-1* (Yoder et al., 1999 (citado anteriormente), número de acceso: AF076535) correspondientes en secuencia a los nucleótidos 210 a 713 de la SEQ ID NO: 3 usando cebadores directo (sentido) e inverso (antisentido): 5'-atggttgctactgctgtgac-3' (SEQ ID NO: 29 y 5'-ctgtaaagattcattagtg-3' (SEQ ID NO: 30) y el ARN molde obtenido de embriones de algodón en desarrollo (Ejemplo 1). El fragmento de la PCR se utilizó para sondear la biblioteca de ADNc en condiciones de rigurosidad alta y se aislaron dos ADNc similares a FatB. Estos se designaron *ghFatB-2* y *ghFatB-3*. Ambas

secuencias parecían ser ADNc de longitud completa, ya que tenían regiones no traducidas en ambos extremos 5' y 3'. El ADNc de *ghFatB-2* tenía 1647 pb de longitud (SEQ ID NO: 5) y codificaba una proteína predicha de 421 aminoácidos (SEQ ID NO: 6), que incluía la secuencia señal usada para la translocación de la proteína al plasto. El ADNc de *ghFatB-3* tenía 1498 pb de longitud (SEQ ID NO: 7) y codificaba una proteína predicha de 419 aminoácidos (SEQ ID NO: 8).

Las secuencias de aminoácidos predichas de *ghFatB-2*, y *-3* tuvieron 60-75 % de identidad a lo largo de su longitud completa con varias tioesterasas de FatB descritas, tales como las de *Arabidopsis* y soja (n.º de acceso DQ418764) y *Brassica juncea* (DQ856315), mientras que solo un 30-40 % de homología con la FatA tioesterasa de plantas dicotiledóneas, tales como *Arabidopsis* (AK176105). Los polipéptidos putativos codificados por la región de codificación de *ghFatB-2* y *-3* mostraron un 89 % de identidad entre sí. Debido a que sus secuencias 3'UTR eran muy diferentes, se pensó que derivaban de dos genes diferentes pertenecientes a una familia de genes de FatB en algodón. Un ADNc que codifica FatB y su correspondiente gen estructural fue aislado por Pirtle et al., Plant Cell Physiology, 40: 155-163, 1999 (ADNc: N.º de acceso AF034266; gen estructural: AF076535). En comparación con el *ghFatB-1* previamente aislado (Yoder et al., 1999 (citado anteriormente), número de acceso: AF076535), *ghFatB-2* tiene aproximadamente un 87 % de identidad y *ghFatB-3* tiene aproximadamente un 84 % de identidad. Como todas las acil-ACP tioesterasas caracterizadas hasta la fecha, las FatB tioesterasas de algodón tienen secuencias peptídicas de tránsito en el extremo N-terminal para dirigir las proteínas a los plastos. El sitio de escisión exacto del péptido de tránsito no se conoce, pero se ha sugerido que está en algún lugar al principio de una región hidrofóbica altamente conservada única para las secuencias FatB de la planta (Jones Jones et al., 1995 (citado anteriormente)). Las Figuras 4 y 5 muestran las secuencias de nucleótidos y polipéptidos putativos de *ghFatB-2* y *-3* respectivamente.

Se ha sugerido previamente que los genes *FatB* de la planta se expresan de forma constitutiva (Dormann et al., Arch. Biochem. Biophys. 316: 612-618, 1995). En consonancia con esto, el análisis de la expresión mostró que el gen *ghFatB-1* se expresaba en embriones, plántulas y hojas de plantas de algodón maduras (Pirtle et al., 1999 (citado anteriormente)). Se pensó que tanto *ghFatB-2* como *-3* se expresaban en semillas en desarrollo, ya que se aislaron de la biblioteca de ADNc de semilla de algodón como se ha descrito anteriormente, y posiblemente de otros tejidos. El nivel de expresión de *ghFatB-2* en las semillas fue mayor que el de *ghFatB-3*, por lo que se sugirió que el primero codificaba la mayor actividad de FatB en semillas de algodón. Sin embargo, para establecer que este gen en lugar de los otros dos codificaba la actividad principal de FatB en semilla de algodón, era necesario demostrar que su inactivación daría como resultado una actividad de tioesterasa reducida y una liberación reducida de ácido palmítico del plasto. También era posible que dos o tres de los genes en combinación tuvieran que ser silenciados para reducir los niveles de ácido palmítico. Por lo tanto, se hizo una construcción de silenciamiento de genes usando una región de *ghFatB-2*, que tenía la secuencia de nucleótidos 458-843 de SEQ ID NO: 5.

Ejemplo 4: Clonación de genes que codifican la ciclopropano ácido graso sintasa (cpa-fas) en algodón

La CPA-FAS cataliza la primera etapa comprometida en la producción de ácidos grasos de ciclopropano, la conversión de ácido oleico en DHS. Se identificaron dos secuencias de EST de *G. hirsutum* que se expresaban de forma diferencial tras la infección de las raíces e hipocótilos de algodón con *Fusarium oxysporum* (Dowd et al., 2004 (citado anteriormente)). Se utilizó una de las secuencias EST, CD486555, como sonda de ADN para cribar la biblioteca de ADNc hecha de semilla de algodón en desarrollo (Ejemplo 1) y una segunda biblioteca de ADNc hecha de ARN obtenido a partir de raíces de algodón. La secuencia de ADN del ADN de la sonda se muestra en la Figura 6 (SEQ ID NO: 9).

Después de la hibridación de alta rigurosidad con la sonda, se aislaron dos ADNc de longitud completa diferentes con secuencias UTR 5' y 3' únicas a partir de la biblioteca de ADNc hecha de ARN de raíz de algodón. Se aisló un tercer clon de ADNc de la biblioteca de ADNc de semilla de algodón. Se designaron como *ghCPA-FAS-1*, *2* y *-3*, respectivamente. Las secuencias de ADN y las secuencias de aminoácidos predichas de los polipéptidos codificados se muestran en las figuras 7-9, (SEC ID N.º: 10-15).

Los polipéptidos de *ghCPA-FAS* deducidos comprendían cada uno 865-873 aminoácidos y tenían masas moleculares calculadas de aproximadamente 99 kDa. Esto probablemente representaba la proteína madura, ya que la enzima CPA-FAS, que estaba activa en el RE, no se esperaba que tuviera un péptido señal en 5'. De forma similar a la CPA-FAS de *Sterculia foetida* y las proteínas homólogas de *Arabidopsis* (n.º de acceso AT23510) y arroz (AK069115) predichas a partir de secuencias genómicas, las proteínas *ghCPA-FAS* codificadas tenían un dominio en N-terminal de unión a FAD fusionado a un dominio en C-terminal que tenía homología con varias metiltransferasas. Esto era consistente con la actividad predicha de CPA-FAS de las proteínas. En el extremo N, los primeros 20 aminoácidos de las proteínas *ghCPA-FAS* parecían ser hidrófobos y se pensaba que estaban implicados en el anclaje a la membrana.

Cuando se compararon las secuencias de aminoácidos deducidas de los tres ADNc diferentes de *ghCPA-FAS* con secuencias de ADN homólogas de *Sterculia foetida*, *Arabidopsis* y arroz, se observó que *ghCPA-FAS-1* (AY574036) y *ghCPA-FAS-2* (AY574037) compartían una identidad de aminoácidos del 97 %, pero solo un 64-65 % de identidad con *ghCPA-FAS-3* (AY574038). Por el contrario, *ghCPA-FAS-3* mostró mayor homología de secuencia con cada uno

de los genes de *Arabidopsis*, 74 % de identidad de aminoácidos con At23510, y 75 % de identidad de aminoácidos con At23530. Esto sugirió que *ghCPA-FAS-3* evolucionaba por separado antes de la especiación del algodón. No se sabe si *Arabidopsis* y arroz acumulan ácidos grasos de CPA o CPE y, por lo tanto, se determina la funcionalidad de los genes de CPA-FAS en estas plantas.

5 La organización genómica de los genes *CPA-FAS* en algodón se investigó mediante análisis de hibridación de transferencia Southern utilizando la región codificante de proteínas de *ghCPA-FAS-1* como sonda. Se detectaron al menos tres bandas de hibridación en ADN digerido con *HindIII* de algodones diploides, mientras que había dos veces más bandas de hibridación en los algodones tetraploides (Figura 10). Se concluyó que cada uno de los tres genes *ghCPA-FAS* estaban representados por un único locus en *Gossypium* diploide y dos loci homólogos en algodones tetraploides. Esto demostró que el algodón alotetraploide contenía dos copias de cada uno de los tres genes presentes en los progenitores diploides A y D del genoma.

15 Se analizó la expresión de los genes CPA-FAS de algodón, buscando la expresión diferencial en tejidos de algodón. Como se muestra en la Figura 11, el análisis del ARN total extraído de diversos tejidos de algodón, incluyendo raíces, hipocótilos, hojas y embriones en desarrollo, en varios momentos después de la antesis reveló que el nivel del transcrito de *ghCPA-FAS-1* era alto en las raíces y los hipocótilos, pero no se detectó en las hojas ni en los embriones en desarrollo, mientras que se detectó expresión de *ghCPA-FAS-2* en todos los tejidos analizados, a excepción de las hojas. En los embriones en desarrollo, los niveles máximos de transcrito de *ghCPA-FAS-2* se encontraron en los embriones durante la parte media del desarrollo de la semilla, aproximadamente de 20 a 40 días después de la antesis. Este fue el período de tiempo principal para la producción de aceite en los embriones de algodón en desarrollo. Por lo tanto, se predijo que *ghCPA-FAS-2* desempeñaba un papel clave en la determinación de la biosíntesis de los ácidos grasos ciclopropenoides en la semilla de algodón y, por lo tanto, era un gen diana candidato para la regulación por disminución para reducir los ácidos grasos de CPA y CPE en aceite de semilla de algodón.

Ejemplo 5: Construcción genética para regular por disminución simultáneamente *ghfad2-1*, *ghcpa-fas-2* y *ghfatb-2* en semilla de algodón

30 Una construcción genética para expresar una molécula de ARN de horquilla quimérico para la reducción mediada por ARNi de la expresión génica se realizó como sigue. La construcción de silenciamiento génico de ARNi se diseñó para dirigirse simultáneamente a tres genes diferentes con el objetivo de lograr reducciones significativas de los niveles de ácido palmítico y ácidos grasos de ciclopropano en combinación con ácido oleico significativamente mayor en aceite de semilla de algodón. La construcción se muestra esquemáticamente en la Figura 12 y contenía una repetición invertida de una secuencia quimérica hecha de 350 pb de *ghFAD2-1*, 442 pb de *ghCPA-FAS-2* y 358 pb de *ghFatB-1* fusionados entre sí. Las unidades repetidas de esta secuencia quimérica estaban separadas por el intrón 1 del gen *ghFad2-1*, que tenía 1.120 pb de longitud, con límites exón/intrón 5' y 3' intactos (Figura 13, fragmento C). El intrón actuó como espaciador para estabilizar la repetición invertida y, de este modo, facilitar la clonación de la repetición invertida en el vector plasmídico en *E. coli*. Para conseguir la expresión específica de la semilla, la secuencia repetida invertida se colocó bajo el control de un promotor de lectina de soja (Lec-P) y una secuencia de terminación de la transcripción/poliadenilación (Lec-T) (Cho y col., Plant Molecular Biology Reporter 13: 255–269, 1995). El gen que expresaba ARN de horquilla se colocó adyacente al extremo 3' de un gen marcador seleccionable comprendido por una región codificante de la proteína NPTII dirigida por el promotor del virus de la atrofia subterránea del trébol (Sc1-P) y el terminador (Sc3-T). Las secuencias específicas usadas se muestran en la Figura 13 (SEQ ID NO: 16–23). Esta construcción genética se denominó construcción MonoCott.

También se fabricó una segunda construcción genética, denominada LY-2. Esta era igual que la construcción MonoCott, excepto que los segmentos de 442 pb del gen *ghCPA-FAS-2* se reemplazaron por una región de un gen de algodón que codifica la cadineno hidroxilasa, una enzima requerida para la síntesis de gopipol en semillas en desarrollo. Para preparar LY-2, un fragmento de ADN de 401 pb derivado del gen *ghCAD-H* que codifica (+)-delta-cadineno-8-hidroxilasa (número de acceso: AF332974), correspondiente a nucleótidos que tienen la secuencia 1420 a 1821 de SEQ ID NO: 28 se amplificaron en reacciones de PCR usando los cebadores: S1: 5'-cctgagaggttctgactgatcatg-3' (SEQ ID NO: 31) y A1: 5'-gcttatgatgtataataaacacattat -3' (SEQ ID NO: 32). Este fragmento se substituyó por las porciones de *ghCPA-FAS-2* en la construcción MonoCott para producir LY-2.

Ejemplo 6: Transformación de algodón con la construcción MonoCott

60 Las construcciones genéticas MonoCott y LY-2 se insertaron cada una en un vector binario y se introdujeron en la cepa AGL1 de *Agrobacterium tumefaciens*. Las bacterias transformadas se usaron para transformar la variedad de algodón Coker315 como describen Liu et al., 2002 (citado anteriormente). En pocas palabras, los cotiledones escindidos de plántulas de algodón cultivadas asépticamente de 10 días de edad se usaron como explantes y se infectaron y cocultivaron con los transformantes de *A. tumefaciens* durante un período de dos días. A esto siguió un período de seis semanas de selección en medio MS (Murashige y Skoog, Physiologia Plantarum. 15: 473–497, 1962) que contenía 0,1 mg/l de 2,4-D, 0,1 mg/l de cinetina, 50 mg/l de sulfato de canamicina y 25 mg/l de cefotaxima. Los callos sanos derivados de los explantes de cotiledón se transfirieron después a medio MS que contenía 5 mg/l de 6-(γ , γ -dimetilalilamino) -purina (2ip), 0,1 mg/l de ácido naftalenoacético (NAA), 25 mg/l de

kanamicina y 250 mg/l de cefotaxima durante un segundo período de seis semanas a 28 °C. Los embriones somáticos comenzaron a formarse después de aproximadamente seis a diez semanas de incubación y se transfirieron a placas frescas, pero sin añadir fitohormona o antibióticos, hasta que germinaron. Las plántulas que se desarrollaron a partir de los embriones somáticos se transfirieron posteriormente al suelo y se mantuvieron en un invernadero una vez que se desarrollaron las hojas y las raíces, con una temperatura de crecimiento de 28/20 °C (día/noche).

Seis líneas de algodón independientes transformadas con la construcción de ARNi de MonoCott se regeneraron a partir de callos y se dejaron crecer hasta su maduración en el invernadero, floreciendo y produciendo semilla como lo normal. No se observaron diferencias fenotípicas evidentes entre las plantas transgénicas y las plantas parentales no transgénicas (de tipo silvestre) de la variedad Coker315. Diez líneas de algodón independientes transformadas con la construcción de ARNi de LY-2 también se regeneraron a partir de callos y se trataron de manera similar.

Ejemplo 7: Análisis fenotípico de los transformantes de algodón

Una porción consistente en aproximadamente 1/8 de los cotiledones de las semillas maduras se escindió de las semillas T₂ obtenidas de las plantas transgénicas primarias T₁ y cada porción se sometió a análisis de la composición de ácidos grasos mediante FAME y GC-MS. Los lípidos totales se aislaron usando el método de Bligh et al., Canadian Journal of Biochemistry and Physiology 37: 911–917, 1959 y se prepararon ésteres metílicos de ácidos grasos (FAME) usando métodos estándar como describen Liu et al., 2002 (citado anteriormente). Los FAME se separaron posteriormente mediante cromatografía de gases (CG), utilizando un GC Agilent 6890 equipado con una columna capilar GC fuerte (30m x 0,25 mm) Los ácidos grasos fueron identificados por referencia a los patrones de FAME.

Las semillas de la línea transgénica DCS9-34 tenían una composición de ácido graso que tenía un nivel marcadamente aumentado de ácido oleico, en combinación con niveles reducidos de ácido palmítico y niveles reducidos de ácidos grasos de ciclopropano. Esta línea transgénica se seleccionó y se cultivo hasta homocigosis. Los datos que muestran la herencia de la composición de ácidos grasos en el aceite de semilla de la progenie se dan en la Tabla 1. Como se muestra en la Tabla 1, el contenido de ácido oleico en el aceite de la semilla de la línea DCS9-34 aumentó a aproximadamente un 75-80 % y el contenido de ácido palmítico se redujo a menos del 10 %. En las líneas con alto contenido de ácido oleico desarrolladas previamente (Liu et al., 2002 (citado anteriormente)), los niveles de ácido palmítico se redujeron a 16 % de los ácidos grasos totales en el aceite de semilla como un efecto pleiotrópico de la regulación por disminución mediada por el ARNi de *ghFAD2-1* (Liu et al., 2002 (citado anteriormente)). Se supuso que la reducción adicional del ácido palmítico hasta el 7 % obtenida con la construcción MonoCott se debía al efecto combinado de la regulación por disminución tanto de *ghFAD2-1* como de *ghFatB-2*. Se observaron fenotipos similares en las otras plantas transformadas con la construcción MonoCott, pero en menor grado.

En las semillas T₂ obtenidas de las plantas transgénicas DCS9-34 primarias, tres de cada diez semillas eran segregantes nulos. Por lo tanto, parece que el transgén se segregaba como un solo gen dominante con patrón de herencia mendeliana. En las semillas T₃ posteriores, tres de las 15 semillas eran nulas/segregantes que carecían del transgén. No se encontraron semillas nulas en treinta semillas T₄ analizadas; Por lo tanto, éstos representaban una población homocigótica. Dado que sólo una porción aproximadamente 1/8 de los cotiledones de las semillas T₂ y T₃ se utilizaron en el análisis de ácidos grasos, los ácidos grasos cíclicos no se detectaron debido a su presencia predominante en radículas de semillas. Como se muestra también en la Figura 14, la composición de ácidos grasos a través de cada una de las tres generaciones se caracterizó por un aumento del ácido oleico y unos niveles reducidos de ácido palmítico. Estos niveles permanecieron constantes y se heredaron de forma estable en plantas que eran heterocigóticas u homocigotas para el transgén como se muestra en la Tabla 2. Esta correlación demostró claramente que la herencia estable de la alteración genética de la composición de ácidos grasos se lograba mediante expresión transgénica de la construcción de ARNi de MonoCott dirigida simultáneamente a los tres genes *ghFAD2-1*, *ghCPA-FAS-2* y *ghFatB-2*.

En la semilla transgénica LY-2, los niveles de ácido oleico también se elevaron, por encima del 80 %, y en algunas semillas más del 85 %. El nivel de ácido palmítico en muchas semillas fue de aproximadamente el 7 %, y el nivel de ácido linoleico aproximadamente del 6-10 % (Tabla 2). Estos niveles alterados se heredaron de forma estable como un locus mendeliano junto con el transgén (Tabla 2).

Los ácidos grasos de ciclopropano, incluyendo DHS, STC y MVL, no están presentes a niveles significativos en los cotiledones de algodón en la semilla de algodón madura, sino que se concentran en los ejes de embriones de las semillas maduras. Con el fin de analizar los niveles de ácidos grasos de ciclopropano en las líneas transgénicas, se escindieron los ejes embrionarios y se sometieron a un análisis de la composición de ácidos grasos. El perfil completo de ácidos grasos de los ejes embrionarios DCS9-34, junto con el control no transformado, Coker315 y una línea transgénica transformada con la construcción LY-2 se tabula en la Tabla 3 y se representa en la Figura 15. La semilla transformada LY-2 también se analizó. Como se muestra anteriormente, la semilla transformada con LY-2 mostró niveles modificados de forma similar, niveles bajos de ácido palmítico y altos de ácido, como semilla transformada con DCS9-34. Como se muestra en la Tabla 3, los ejes de embrión de algodón no transformados,

representados por el Coker315 no transformado, contenían ácidos grasos totales de ciclopropano a un nivel de 11,5 % de ácidos grasos totales (promedio de 23 semillas tomadas al azar). En la semilla LY-2, como resultado del aumento sustancial del ácido oleico que es el sustrato para la enzima CPA-FAS, cada uno de los tres ácidos grasos de ciclopropano DHS, STC y MVL, así como el porcentaje total de ácidos grasos de ciclopropano se incrementaron sustancialmente en comparación con el control no transformado. Esto contrastaba claramente con la línea MonoCott, DCS9-34, que mostró una reducción sustancial en los tres ácidos grasos de ciclopropano con el nivel total de ácidos grasos cíclicos, promediando sólo el 3,8 % en 26 semillas seleccionadas al azar. Esto representó una reducción de los ácidos grasos totales de ciclopropano más del 60 % en comparación con el tipo silvestre y más del 80 % en comparación con la semilla transgénica LY-2.

TABLA 1

Composición de ácidos grasos del aceite de semilla en la semilla de algodón de la línea DCS9-34 transgénica, en tres generaciones. Las semillas con más de 20 % de contenido de ácido palmítico eran segregantes nulos y eran esencialmente al nivel de tipo silvestre.						
Generación	n.º de semilla	% de palmítico (16:0)	% de esteárico (18:0)	% de oleico (18:1)	% de linoleico (18:2)	Total % CPFA
T ₂	1	24,1	1,8	13,3	58,5	ND
	2	22,0	1,8	13,4	60,4	ND
	3	11,8	1,6	63,7	21,6	ND
	4	9,3	1,6	73,4	14,6	ND
	5	9,8	1,7	70,6	16,7	ND
	6	22,5	1,8	17,0	56,6	ND
	7	11,3	1,6	64,8	20,9	ND
	8	9,5	1,8	70,0	17,6	ND
	9	8,5	1,6	75,4	13,3	ND
	10	10,9	1,6	72,3	14,1	ND
T ₃	1	8,6	2,2	76,6	12,0	ND
	2	9,5	2,2	73,4	14,5	ND
	3	8,9	2,6	73,6	14,5	ND
	4	7,4	2,4	80,0	9,8	ND
	5	22,8	2,4	15,0	58,8	ND
	6	7,6	2,3	81,0	8,7	ND
	7	9,3	2,4	73,4	14,4	ND
	8	8,0	2,5	76,7	12,3	ND
	9	8,9	2,6	78,7	9,2	ND
	10	9,0	2,6	74,3	13,7	ND
	11	7,9	2,9	78,6	10,2	ND
	12	23,2	2,8	15,6	57,4	ND
	13	9,0	2,6	74,2	13,7	ND
	14	7,8	2,2	79,5	10,1	ND
	15	23,2	2,7	14,9	58,4	ND
T ₄	1	8,3	2,3	74,2	12,6	0,2
	2	8,2	2,5	75,6	11,2	0,4
	3	8,1	2,5	74,9	11,9	0,2
	4	8,7	2,3	72,9	13,6	0,2
	5	8,0	2,5	75,3	11,6	0,3
	6	8,2	2,3	75,2	11,8	0,2
	7	7,8	2,6	76,3	10,7	0,3
	8	7,6	2,4	76,6	11,0	0,2
	9	8,2	2,7	75,1	11,4	0,3
	10	8,3	2,5	74,4	12,4	0,2
	11	8,0	2,5	75,6	11,5	0,2

ES 2 603 530 T3

	12	8,2	2,4	75,5	11,2	0,2
	13	7,4	2,5	76,5	11,3	0,2
	14	7,8	2,4	76,3	11,3	0,0
	15	7,8	2,6	75,2	12,1	0,3
	16	8,3	2,2	74,5	12,4	0,3
	17	8,2	2,3	74,6	12,3	0,3
	18	8,3	2,3	74,1	12,8	0,2
	19	8,5	2,3	73,9	12,6	0,3
	20	7,8	2,6	75,5	11,7	0,2
	21	7,7	2,5	76,3	11,1	0,2
	22	8,3	2,6	74,4	12,2	0,3
	23	7,9	2,6	73,8	13,2	0,2
	24	7,5	2,5	76,7	11,0	0,1
	25	8,0	2,5	75,5	11,5	0,2
	26	8,2	2,3	74,7	12,0	0,3
	27	7,9	2,4	75,6	11,6	0,2
	28	7,9	2,6	74,9	12,0	0,3
	29	8,1	2,4	75,3	11,8	0,2
	30	8,2	2,4	73,8	12,9	0,4
ND: no determinado						

TABLA 2

Composición de ácidos grasos de tres generaciones de algodón transgénico que expresa la construcción de ARNi de LY-2 dirigido a tres genes, a saber *ghFad2-1*, *ghCAD-H* y *ghFatB-2*. LY2-1, LY2-8 y LY2-9 parecen mostrar una composición de ácidos grasos de tipo silvestre y pueden ser "escapes" no transformados.

Muestra	Mirístico	Palmitico	Estearico	Oleico	Linoleico	Linolénico	Araquidico
Semillas T2							
LY2-1	0,88	21,26	1,47	11,34	65,50	0,00	0,00
LY2-2	0,13	8,22	0,87	78,27	12,37	0,21	0,00
LY2-3	0,20	9,65	0,66	74,23	15,21	0,15	0,00
LY2-4	0,00	9,42	1,46	76,76	12,20	0,16	0,00
LY2-5	0,00	8,88	0,60	76,82	13,70	0,00	0,00
LY2-6	0,16	8,10	1,52	80,67	9,36	0,18	0,00
LY2-7	0,22	8,41	1,68	79,97	9,36	0,23	0,14
LY2-8	0,85	21,37	1,73	12,22	63,58	0,24	0,00
LY2-9	0,67	21,45	1,65	11,50	64,45	0,28	0,00
LY2-10	0,21	8,02	1,46	80,51	9,67	0,13	0,00

Semillas T3							
LY2,2-1	0,35	6,97	2,07	83,35	6,94	0,13	0,19
LY2,2-2	0,43	6,88	2,03	83,10	7,25	0,14	0,17
LY2,2-3	0,34	7,16	1,71	83,37	7,15	0,12	0,16
LY2,2-4	0,12	7,13	2,10	82,05	8,28	0,18	0,14
LY2,2-5	0,14	7,13	1,91	83,54	7,00	0,12	0,17
LY2,2-6	0,11	6,82	2,07	83,79	6,87	0,14	0,19
LY2,2-7	0,12	6,85	1,69	85,53	7,48	0,13	0,20
LY2,2-8	0,15	7,18	2,03	83,11	7,22	0,16	0,16
LY2,2-9	0,12	6,69	1,93	84,04	6,87	0,13	0,23
LY2,2-10	0,32	6,45	1,91	85,02	5,94	0,17	0,19
LY2,2-11	0,24	7,85	1,74	83,46	6,39	0,11	0,18
LY2,2-12	0,22	7,38	1,24	85,48	5,39	0,13	0,17
LY2,2-13	0,35	8,57	1,27	83,34	6,15	0,20	0,12
LY2,2-14	0,19	6,83	1,68	85,62	5,37	0,13	0,19
LY2,2-15	0,23	7,47	1,94	83,22	6,79	0,14	0,21

ES 2 603 530 T3

Semillas T4							
LY2,2.1-1	0,20	8,96	1,93	77,80	10,41	0,17	0,29
LY2,2.1-2	0,16	7,88	2,15	81,14	8,06	0,18	0,26
LY2,2.1-3	0,17	8,08	2,00	81,33	7,87	0,16	0,26
LY2,2.1-4	0,17	8,96	2,17	78,29	9,61	0,25	0,26
LY2,2.1-5	0,16	8,34	2,39	79,31	8,94	0,21	0,30
LY2,2.1-6	0,17	7,81	2,09	81,32	7,96	0,16	0,30
LY2,2.1-7	0,15	7,77	2,01	81,42	8,00	0,17	0,27
LY2,2.1-8	0,19	8,66	2,03	79,96	8,61	0,15	0,26
LY2,2.1-9	0,18	8,31	2,08	80,60	8,21	0,17	0,27
LY2,2.1-10	0,19	8,77	1,76	80,39	8,28	0,16	0,24
LY2,2.1-11	0,16	8,26	2,03	80,52	8,41	0,18	0,25
LY2,2.1-12	0,26	10,60	2,45	74,43	11,41	0,47	0,30
LY2,2.1-13	0,17	8,02	2,14	81,09	8,01	0,18	0,25
LY2,2.1-14	0,13	7,28	2,24	82,01	7,72	0,18	0,28
LY2,2.1-15	0,17	8,16	1,83	81,27	8,03	0,15	0,26
LY2,2.1-16	0,17	8,04	1,95	81,25	8,03	0,16	0,25
LY2,2.1-17	0,15	7,19	1,83	83,02	7,31	0,14	0,25
LY2,2.1-18	0,20	8,88	2,01	79,40	8,89	0,17	0,27
LY2,2.1-19	0,15	7,70	2,16	81,75	7,69	0,15	0,27
LY2,2.1-20	0,19	8,64	1,89	79,62	9,04	0,16	0,26
LY2,2.1-21	0,18	8,19	2,01	81,25	7,84	0,15	0,25
LY2,2.1-22	0,19	8,56	2,06	80,19	8,35	0,16	0,27
LY2,2.1-23	0,18	8,67	2,03	80,45	8,00	0,20	0,25
LY2,2.1-24	0,16	7,95	1,95	81,53	7,89	0,15	0,26
LY2,2.1-25	0,16	7,92	2,33	81,02	8,02	0,17	0,27
LY2,2.1-26	0,18	8,49	2,08	80,43	8,15	0,17	0,27
LY2,2.1-27	0,18	8,58	2,05	80,19	8,32	0,18	0,27
LY2,2.1-28	0,17	8,26	2,04	80,08	8,88	0,18	0,26
LY2,2.1-29	0,16	7,77	2,04	81,44	7,99	0,17	0,27
LY2,2.1-30	0,18	8,14	1,76	81,16	8,16	0,19	0,24

Semillas T4							
LY2,2.4-1	0,17	9,14	2,76	79,72	7,65	0,13	0,31
LY2,2.4-2	0,18	9,57	2,82	78,68	8,08	0,17	0,31
LY2,2.4-3	0,15	9,16	3,18	78,75	8,05	0,17	0,31
LY2,2.4-4	0,17	9,28	2,97	79,31	7,62	0,15	0,32
LY2,2.4-5	0,17	9,33	2,56	79,58	7,79	0,15	0,28
LY2,2.4-6	0,17	9,29	2,79	79,09	8,03	0,16	0,29
LY2,2.4-7	0,16	9,42	2,66	78,09	9,00	0,17	0,29
LY2,2.4-8	0,21	9,82	2,59	77,69	9,03	0,18	0,30
LY2,2.4-9	0,19	9,15	2,43	80,02	7,66	0,14	0,28
LY2,2.4-10	0,18	9,00	2,72	79,48	8,10	0,13	0,31
LY2,2.4-11	0,18	9,63	2,99	78,71	7,84	0,17	0,30
LY2,2.4-12	0,22	9,88	2,42	78,79	8,05	0,15	0,31
LY2,2.4-13	0,21	9,62	2,37	79,01	8,22	0,13	0,30
LY2,2.4-14	0,16	9,01	2,92	79,82	7,53	0,15	0,30
LY2,2.4-15	0,18	9,52	2,76	79,06	7,85	0,16	0,29
LY2,2.4-16	0,19	9,75	2,76	78,63	8,05	0,16	0,29
LY2,2.4-17	0,18	9,24	2,38	80,03	7,62	0,14	0,27
LY2,2.4-18	0,20	9,82	2,73	77,41	9,15	0,18	0,31
LY2,2.4-19	0,23	10,18	2,16	78,61	8,36	0,13	0,25
LY2,2.4-20	0,16	9,06	2,74	80,11	7,43	0,14	0,27
LY2,2.4-21	0,17	9,31	2,77	79,66	7,47	0,15	0,31
LY2,2.4-22	0,18	9,26	2,73	79,77	7,44	0,18	0,28
LY2,2.4-23	0,19	8,94	2,38	79,98	7,99	0,13	0,27
LY2,2.4-24	0,19	9,41	2,38	79,71	7,70	0,17	0,27
LY2,2.4-25	0,19	9,47	2,43	79,60	7,74	0,14	0,28
LY2,2.4-26	0,20	10,08	2,84	76,62	9,54	0,17	0,31
LY2,2.4-27	0,21	9,90	2,49	77,69	9,06	0,17	0,29
LY2,2.4-28	0,20	9,84	2,65	78,72	7,91	0,19	0,30
LY2,2.4-29	0,18	9,08	2,90	79,08	8,19	0,14	0,32
LY2,2.4-30	0,16	8,89	2,65	79,82	7,91	0,15	0,27

Reducción de los ácidos grasos de ciclopropano en los ejes de embriones de algodón de la semilla DCS9-34 homocigota T4

TABLA 3

Composición de ácidos grasos de los ejes de embriones en tres tipos de algodón, incluyendo DCS9-34 (MonoCott) con rasgos combinados de HO, LP y niveles bajos de ácidos grasos cíclicos, LY-2 con rasgos HO y LP, pero con niveles altos de ácidos grasos cíclicos y Coker315 como el control convencional del algodón.

Ejes embrionarios de DCS9

N.º de semilla									CPFA			Total de CPF A
	14: 0	16: 0	16: 1	18: 0	18: 1 n = -9	18: 1 n-7	18: 2	18: 3	DHS	MVA	STC	
1	0,5	16,6	0,8	2,5	48,5	1,0	25,8	0,8	0,3	2,1	0,5	3,0
2	0,5	15,1	0,8	2,2	47,4	1,0	27,7	0,9	0,4	2,8	0,7	4,0
3	0,6	16,0	0,8	1,5	43,8	1,1	29,3	1,2	0,4	4,1	0,8	5,3
4	0,5	14,7	0,8	2,1	49,0	1,0	26,7	0,8	0,4	2,9	0,7	4,0
5	0,4	13,0	0,8	2,2	54,5	1,0	25,0	0,7	0,2	1,4	0,4	2,1
6	0,6	16,3	0,8	1,5	42,0	1,1	31,5	1,1	0,4	3,4	0,8	4,6
7	0,6	17,3	0,9	1,4	39,5	1,1	32,6	1,1	0,5	3,8	0,9	5,1
8	0,5	14,6	0,8	1,7	47,9	1,0	28,8	0,8	0,2	2,5	0,6	3,4
9	0,5	14,0	0,8	2,3	51,3	1,0	25,5	0,9	0,3	2,4	0,5	3,3
10	0,6	16,0	0,8	1,5	43,7	1,1	30,4	1,0	0,4	3,2	0,7	4,4
11	0,4	14,1	0,7	2,2	49,4	1,0	27,0	0,8	0,3	2,8	0,7	3,8
12	0,4	14,5	0,8	1,7	48,9	1,0	28,3	0,9	0,2	2,2	0,6	3,0
13	0,4	14,8	0,8	2,0	48,5	1,1	27,6	0,7	0,3	2,5	0,7	3,6
14	0,6	16,8	0,9	1,3	40,6	1,1	32,7	1,2	0,3	3,2	0,7	4,2
15	0,6	16,9	0,8	1,5	40,3	1,0	31,2	1,2	0,6	4,3	1,1	5,9
16	0,4	12,4	0,8	2,1	57,5	1,0	23,4	0,6	0,2	0,8	0,3	1,3
17	0,5	15,8	0,8	2,1	43,9	1,0	29,7	0,9	0,5	3,3	0,9	4,7
18	0,6	16,1	0,9	1,4	43,5	1,1	30,7	1,1	0,3	3,1	0,7	4,2
19	0,4	13,1	0,7	2,4	54,6	0,9	23,9	0,8	0,2	2,0	0,5	2,7
20	0,5	15,2	0,9	1,3	45,4	1,2	30,9	1,1	0,2	2,3	0,5	2,9
21	0,6	16,0	0,9	1,4	43,1	1,1	31,1	1,1	0,4	3,2	0,8	4,3
22	0,5	15,2	0,7	1,7	46,8	1,0	28,6	0,9	0,4	2,9	0,7	4,0
23	0,5	14,7	0,9	1,4	46,5	1,2	29,8	1,2	0,2	2,6	0,5	3,3
24	0,5	16,6	0,8	1,7	43,9	1,0	29,1	0,9	0,6	3,5	0,9	5,0
25	0,6	16,8	0,9	1,5	41,0	1,1	31,4	1,2	0,4	3,8	0,9	5,1
26	0,5	15,2	0,8	1,8	47,9	1,1	28,0	1,0	0,2	2,3	0,5	3,0
Promedio									0,3	2,8	0,7	3,8

Ejes embrionarios de LY2

N.º de semilla												Total de CPF A
	14: 0	16: 0	16: 1	18: 0	18: 1 n-9	18: 1 n-7	18: 2	18: 3	DHS	MVA	STC	
1	0,4	18,1	0,6	5,0	27,5	0,6	20,9	0,8	9,2	9,5	6,3	25,0
2	0,4	18,6	0,5	5,1	22,2	0,5	20,5	1,1	10,3	12,9	6,8	30,0
3	0,4	17,1	0,7	3,6	41,3	0,7	20,0	0,9	4,4	6,2	4,0	14,6
4	0,4	16,8	0,7	4,3	34,2	0,7	19,3	1,1	6,9	8,7	6,0	21,6
5	0,3	14,7	0,6	4,5	42,4	0,6	15,6	0,8	7,1	7,4	5,1	19,6
6	0,3	15,5	0,6	4,3	37,4	0,6	18,4	0,6	7,8	7,8	5,7	21,3
7	0,3	14,3	0,6	4,1	46,1	0,6	17,2	0,7	4,9	5,8	4,5	15,2
8	0,5	18,5	0,6	4,1	29,2	0,7	21,9	1,0	7,0	9,0	6,6	22,6
9	0,3	16,4	0,6	4,5	33,3	0,6	18,6	1,2	7,5	9,5	6,5	23,5
10	0,3	13,4	0,6	4,5	44,9	0,6	15,2	0,7	6,9	7,1	5,0	19,0
11	0,4	15,5	0,6	4,6	42,0	0,6	17,0	0,7	6,9	5,9	5,0	17,7
12	0,4	16,5	0,6	4,9	32,6	0,6	18,7	1,1	8,4	9,2	5,9	23,5
13	0,4	16,1	0,7	3,9	33,0	0,7	20,3	1,0	7,3	9,3	6,4	23,1
14	0,4	16,1	0,6	4,5	43,0	0,6	17,2	0,7	5,6	6,2	4,1	15,9
15	0,4	16,7	0,6	4,2	38,9	0,6	19,0	0,8	5,7	7,1	5,0	17,8
16	0,4	18,1	0,5	4,7	37,4	0,6	18,7	0,8	6,5	6,6	4,6	17,6
17	0,3	14,3	0,6	4,4	46,0	0,6	15,2	0,6	6,7	6,3	4,2	17,1
18	0,3	16,2	0,6	5,1	35,7	0,6	18,0	0,8	8,6	7,2	5,8	21,6
19	0,3	14,0	0,6	4,2	44,6	0,6	15,5	0,7	6,4	7,2	5,1	18,7
20	0,7	21,7	0,6	4,2	19,9	0,6	26,1	1,2	7,0	11,4	5,6	24,1
21	0,3	13,7	0,6	4,3	47,3	0,6	15,4	0,7	5,5	6,3	4,4	16,1
22	0,3	13,5	0,6	3,8	49,0	0,6	16,3	0,6	5,1	5,2	4,2	14,4
23	0,3	14,2	0,6	4,1	45,5	0,6	16,4	0,7	5,8	6,5	4,4	16,7
24 25	0,3	15,8	0,7	4,3 4,5	37,1	0,6	18,5	0,8	7,4	7,0	6,6	20,9 24,9
	0,5	19,5	0,6		25,5		21,8	1,2	7,9	10,6	6,4	
Promedio									6,9	7,8	5,4	20,1

Ejes embrionarios de Coker

N.º de semilla	14: 0	16: 0	16: 1	18: 0	18: 1 n=9	18: 1 n=7	18: 2	18: 3	DHS	MVA	STC	Total de CPFA
B.1	0,6	25,1	0,5	3,4	12,0	0,6	44,5	0,7	2,2	7,0	3,0	12,2
B.2	0,6	26,3	0,5	3,5	11,2	0,6	44,3	0,6	2,4	6,5	2,7	11,6
B.3	0,7	26,1	0,4	4,0	10,6	0,6	41,6	0,7	3,1	8,1	3,2	14,4
B.4	0,7	26,7	0,4	3,9	11,5	0,6	42,0	0,6	2,8	7,1	3,0	12,9
B.5	0,6	27,2	0,5	3,7	12,0	0,6	45,0	0,4	2,2	4,4	2,7	9,3
B.6	0,6	27,3	0,5	3,6	11,4	0,6	42,4	0,6	2,7	6,7	2,9	12,3
B.7	0,7	27,2	0,5	3,5	11,6	0,6	43,9	0,6	2,4	5,7	2,7	10,8
B.8	0,6	26,6	0,4	4,1	11,5	0,6	42,8	0,5	2,7	6,4	2,9	12,0
B.9	0,7	26,2	0,4	3,8	11,3	0,6	44,6	0,5	2,4	6,0	2,7	11,2
B.10	0,7	26,8	0,5	3,8	11,6	0,6	42,4	0,6	2,7	6,6	3,0	12,3
B.11	0,7	26,4	0,5	3,3	10,8	0,7	42,8	0,8	2,4	7,9	3,1	13,4
B.12	0,7	27,0	0,4	3,9	11,5	0,6	43,1	0,6	2,4	6,3	2,7	11,4
B.13	0,7	26,5	0,5	3,9	12,0	0,6	42,9	0,6	2,6	6,3	2,8	11,6
B.14	0,7	26,5	0,5	3,7	11,7	0,6	43,7	0,6	2,5	6,1	2,7	11,3
B.15	0,6	26,9	0,5	3,5	12,2	0,6	44,1	0,5	2,3	5,5	2,6	10,3
B.16	0,6	26,4	0,5	3,3	12,2	0,6	48,0	0,4	1,8	4,2	1,9	8,0
B.17	0,6	26,6	0,5	3,7	11,5	0,6	43,4	0,6	2,6	6,4	2,7	11,8
B.18	0,6	28,8	0,5	4,1	11,2	0,7	39,2	0,7	2,3	7,9	3,2	13,5
B.19	0,7	26,0	0,5	3,2	11,2	0,7	45,3	0,7	2,0	6,5	2,7	11,1
B.20	0,6	26,8	0,5	3,8	10,7	0,6	43,0	0,7	2,1	7,6	3,0	12,7
B.21	0,7	27,2	0,5	3,6	11,3	0,6	40,7	0,7	3,0	7,3	3,6	13,9
B.22	0,7	26,4	0,5	3,2	11,8	0,7	44,5	0,6	2,0	6,2	2,6	10,9
B.23	0,5	24,6	0,5	2,9	12,7	0,6	51,4	0,4	1,0	3,3	1,5	5,8
Promedio									2,4	6,3	2,8	11,5

TABLA 4

Resumen de los identificadores de secuencia	
SEQ ID NO:	DESCRIPCIÓN
1	Secuencia de ADNc que codifica FAD2 de algodón del gen <i>ghFAD2</i> - 1. 1362 nucleótidos, región codificante de proteína: nucleótidos 73-1227
2	Secuencia de aminoácidos codificada por SEQ ID NO: 1: 384 aminoácidos
3	Secuencia de nucleótidos del ADN genómico de <i>ghFatB-1</i> que codifica la tioesterasa de la proteína transportadora de acilo-acilo (FatB-1) de <i>Gossypium hirsutum</i> (Yoder et al., 1999 (citado anteriormente)), n.º de acceso AF076535, 5201 pb. La secuencia completa del ADNc corresponde a los nucleótidos 915-1397, 1530-1663, 2352-2465, 2716-2887, 3001-3069 y 3322-3591, todos unidos entre sí
4	Secuencia de aminoácidos codificada por SEQ ID NO: 3: 413 aminoácidos
5	Secuencia de ADNc del gen <i>ghFatB</i> - 2, 1647 nucleótidos, La región que codifica la proteína es de los nucleótidos 210 a 1472 (codón de terminación TAG)
6	Secuencia de aminoácidos codificada por SEQ ID NO: 5: 420 aminoácidos
7	Secuencia de ADNc del gen <i>ghFatB-3</i> , 1498 nt
8	Secuencia de aminoácidos codificada por SEQ ID NO: 7: 418 aminoácidos
9	Secuencia EST de <i>G. hirsutum</i> (Dowd et al., 2004 (citado anteriormente)). N.º de acceso CD486555
10	Secuencia de nucleótidos de <i>ghCPA-FAS-1</i> que codifica la ciclopropano ácido graso sintasa (CPA-FAS-1), de <i>Gossypium hirsutum</i> , número de acceso AY574036, secuencia completa de ADNc, 2884 pb, región codificante de proteína: nucleótidos 33-2654
11	Secuencia de aminoácidos codificada por SEQ ID NO: 10: 873 aminoácidos
12	Secuencia de nucleótidos de <i>ghCPA-FAS-2</i> que codifica la ciclopropano ácido graso sintasa (CPA-FAS-2), de <i>Gossypium hirsutum</i> , número de acceso AY574037, secuencia completa de ADNc, 2827 pb, región codificante de proteína: nucleótidos 16-2613
13	Secuencia de aminoácidos codificada por SEQ ID NO: 12: 865 aminoácidos
14	Secuencia de nucleótidos de <i>ghCPA-FAS-3</i> que codifica la ciclopropano ácido graso sintasa (CPA-FAS-3), de <i>Gossypium hirsutum</i> , número de acceso AY574038, secuencia completa de ADNc, 2912 pb, región codificante de proteína: nucleótidos 109-2706
15	Secuencia de aminoácidos codificada por SEQ ID NO: 4: 865 aminoácidos
16	Fragmento A de MonoCot, que comprende el promotor de lectina específico de la semilla derivado del gen <i>lec1</i> de la soja: 1771 nucleótidos
17	Fragmento B de MonoCot, ADN quimérico de tres fragmentos del gen diana
18	Fragmento C de MonoCot: Intrón de <i>ghFAD2-1</i>
19	Fragmento D de MonoCot, complementario del ADN quimérico

Resumen de los identificadores de secuencia	
SEQ ID NO:	DESCRIPCIÓN
20	Fragmento E de MonoCot, región de terminación de la transcripción/ poliadenilación del gen de la lectina
21	Fragmento F de MonoCot, promotor S1
22	Fragmento G de MonoCot, gen <i>NPTII</i>
23	Fragmento H de MonoCot, región de terminación de la transcripción/ poliadenilación del gen S3
24	ARNm de <i>Gossypium hirsutum</i> para la omega-6 desaturasa (FAD2-1) también conocida como oleoil-delta 12 desaturasa. N.º de acceso X97016 región codificante de la proteína 1386 pb: nucleótidos 54–1211. Homología con la SEQ ID NO: 1 fue del 96 % a lo largo de toda la longitud de los ADNc
25	Secuencia de aminoácidos codificada por SEQ ID NO: 24: 385 aminoácidos
26	Secuencia de ADNc del gen <i>ghFAD2 - 2</i> , N.º de acceso Y10112 (Liu et al., 1999), 1422 nucleótidos, región codificante de proteína: nucleótidos 98-1249
27	Secuencia de aminoácidos codificada por SEQ ID NO: 26: 383 aminoácidos
28	Secuencia completa del ADNc del gen de la monooxigenasa P450 de <i>Gossypium arboreum</i> , N.º de acceso AF332974, 1933 nt, nucleótidos de la región codificante de proteína 91-1701
29	cebador sentido para la amplificación de los nucleótidos 210 a 713 de la SEQ ID NO: 3
30	cebador antisentido para la amplificación de los nucleótidos 210 a 713 de la SEQ ID NO: 3
31	cebador sentido para la amplificación de los nucleótidos 1420 a 1821 de la SEQ ID NO: 28
32	cebador antisentido para la amplificación de los nucleótidos 1420 a 1821 de la SEQ ID NO: 28

TABLA 5

Sustituciones de aminoácidos de ejemplo preferentes		
Resto original	SUSTITUCIONES DE EJEMPLO	SUSTITUCIONES PREFERENTES
Ala	Val, Leu, He	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln, His, Lys, Arg	Gln
Asp	Glu	Glu
Cys	Ser	Ser
Gln	Asn, His, Lys,	Asn
Glu	Asp, Lys	Asp
Gly	Pro	Pro
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleu	Leu
Leu	Norleu, He, Val, Met, Ala, Phe	Ile
Lys	Arg, Gln, Asn	Arg
Met	Leu, He, Phe	Leu
Phe	Leu, Val, He, Ala	Leu
Pro	Gly	Gly
Ser	Thr	Thr
Thr	Ser	Ser
Trp	Tyr	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	He, Leu, Met, Phe, Ala, Norleu	Leu

Bibliografía

5 Almeida y Allshire, Trends Cell Biol. 15: 251–258, 2005.
 Altschul et al., Nucleic Acids Res. 25: 3389, 1997.
 Ausubel et al., (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, NY, 6,3.1–6,3.6., 1989.
 Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons Inc, 1994–1998, Capitulo 15.
 10 Bao et al., Proc. Natl. Acad. Sci. U.S.A. 99: 7172–7177, 2002.
 Barker et al., Plant Mol. Biol., 2: 235–350, 1983.
 Bechtold et al., C.R. Acad. Sci. Paris, 316: 1194, 1993.
 Bevan et al., Nucl. Acid Res., 11: 369, 1983.
 Birch, Ann Rev Plant Physiol Plant Mol Biol. 48: 297–326, 1997.
 15 Bligh et al., Canadian Journal of Biochemistry and Physiology 37: 911–917, 1959.
 Bonaventure et al., Plant Cell 15: 1020–1033, 2003.
 Bourque, Plant Science, 105: 125–149, 1995.
 Buhr et al., Plant J. 30: 155–163, 2002.
 Chapman et al., J. Am. Oil Chem. Soc. 78: 941–947, 2001.
 20 Cherry y Leffler, Seed. In "Cotton, agronomy monograph No. 24" (eds RJ Kohel, CF Lewis) pp. 511–569. Crop Science Society of America, Madison, WI, USA, 1984.

- Cherry, J. *Am. Oil Chem. Soc.* 60: 360–367, 1983.
 Cho et al., *Plant Molecular Biology Reporter* 13: 255–269, 1995.
 Comai et al., *Plant J.* 37: 778–786, 2004.
 De Framond, *Biotechnology*, 1: 262, 1983.
 5 Dormann et al., *Plant Physiol.* 123: 637–644, 2000.
 Dormann et al., *Arch. Biochem. Biophys.* 316: 612–618, 1995.
 Dowd et al., *Molecular Plant–Microbe Interactions*. 17: 654–667, 2004.
 Folch et al., *J. Biol. Chem.* 226: 497, 1957.
 Fuller et al., *J. Food Sci.* 31: 477–480, 1966.
 10 Garfinkel et al., *Cell*, 27: 143–153, 1983.
 Greve, J. *Mol. Appl. Genet.*, 1: 499–511, 1983.
 Harayama, *Trends Biotechnol.* 16: 76–82, 1998.
 Hartmann y Endres, *Manual of Antisense Methodology*, Kluwer, 1999.
 Haseloff y Gerlach, *Nature* 334: 585–591, 1988.
 15 Henikoff et al., *Plant Physiol.* 135: 630–636, 2004.
 Hinchee et al., *Biotech.* 6: 915, 1988.
 Hoekema et al., *Nature*, 303: 179, 1983.
 Hutchins et al., *Journal of American Oil Chemists Society* 45: 397–399, 1968.
 Johnson et al., *Nature* 214: 1244–1245, 1967.
 20 Jones y King, *Cottonseed Oil*. National Cottonseed Products Associations, Inc. and the Cotton Foundation, Memphis, TN, USA, 1993.
 Jones et al., *Plant Cell* 7: 359–371, 1995.
 Joshi, *Nucl. Acid Res.* 15: 6643, 1987.
 Kargiotidou et al., *Journal of Experimental Botany* 2008 59(8): 2043–2056, 2008.
 25 Khandjian, *Bio/Technology*, 5: 165–167, 1987.
 Klein et al., *Nature*, 327: 70, 1987.
 Kris–Etherton et al., *Nutrition–today (USA)*. 28: 30–38, 1993.
 Lindsey et al., *Exp. Biol. Med.* 195: 261–269, 1990.
 Liu et al., *Plant Physiol.* 120: 339, 1999b.
 30 Liu et al., *Plant Physiol.* 129: 1732–1743, 2002.
 Liu et al., *Australian Journal of Plant Physiology* 26: 101–106, 1999a.
 McPherson y Moller (Ed), *BIOS Scientific Publishers Ltd*, Oxford, 2000.
 Medberry et al., *Plant Cell*, 4: 185–192, 1992.
 Medberry et al., *Plant J.* 3: 619–626, 1993.
 35 Merker y Mattil, 1965. Patente de Estados Unidos 3.201.431.
 Millar y Waterhouse, *Funct Integr Genomics*, 5: 129–135, 2005.
 Mounts et al., *J. Am. Oil Chem. Soc.* 65: 624–628, 1998.
 Mozaffarian et al., *N. Engl. J. Med.* 354: 1601–1613, 2006.
 Murashige y Skoog, *Physiologia Plantarum*. 15: 473–497, 1962.
 40 Needleman y Wunsch, *J. Mol. Biol.* 48: 443–453, 1970.
 Niedz et al., *Plant Cell Reports*, 14: 403, 1995.
 O'Brien, *Cottonseed Oil*. In: F.D. Gunstone (Ed.) *Vegetable Oils in Food Technology: Composition, Properties and Uses*. Blackwell Publishing, Oxford, pp. 203–230, 2002.
 OGTR DIR–039/2003 Field evaluation of genetically modified high–oleic (HO) cotton
 45 Oomen et al., *Lancet* 357: 746–751, 2001.
 Ow et al., *Science*, 234: 856, 1986.
 Pasquinelli et al., *Curr Opin Genet Develop* 15: 200–205, 2005.
 Perriman et al., *Gene*, 113: 157–163, 1992.
 Pirtle et al., *Plant Cell Physiology*, 40: 155–163, 1999.
 50 Pirtle et al., *Biochim. Biophys. Acta* 1522: 122–129, 2001.
 Prasher et al., *Biochem. Biophys. Res. Comm.* 126: 1259–68, 1985.
 Roehm et al., *Lipids* 5: 80–84, 1970.
 Salomon et al., *EMBO J.*, 3: 141–146, 1984.
 Sambrook et al., *Molecular Cloning: A Laboratory Manual (2nd Ed)*. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y. 1989.
 55 Senior, *Biotech. Genet. Engin. Revs.* 15: 79–119, 1998.
 Shenstone y Vickery, *Nature* 190: 68–169, 1961.
 Shippy et al., *Mol. Biotech.* 12: 117–129, 1999.
 Slade y Knauf, *Transgenic Res.* 14: 109–115, 2005.
 60 Smith et al., *Nature*, 407: 319–320, 2000.
 Stalker et al., *Science*, 242: 419, 1988.
 Thillet et al., *J. Biol. Chem.* 263: 12500, 1988.
 Voelker et al., *Plant Journal* 9: 229–241, 1996.
 Waterhouse et al., *Proc. Natl. Acad. Sci. U.S.A.* 95: 13959–13964, 1998.
 65 Wilson et al., *J. Am. Oil Chem. Soc.* 78: 335–340, 2001.
 Yoder et al., *Biochimica et Biophysica Acta* 1446: 403–413, 1999.

Zwar y Chandler, Planta 197: 39–48, 1995.

LISTADO DE SECUENCIAS

- 5 <110> Commonwealth Scientific and Industrial Research Organisation
<120> Aceite de semilla de algodón mejorado y usos
<130> 30812011/JEH/DXT
10 <160> 32
<170> PatentIn versión 3.5
15 <210> 1
<211> 1362
<212> ADN
<213> *Gossypium hirsutum*
20 <400> 1

ES 2 603 530 T3

```

caaaaccaac acgccttctt tgccctcgtgt ttcacacact ggcgtaaac tgctttcttt      60
aaaaccaaca aaatgggtgc cgggtgggta ggatgccaat tgacgggtat aaaggaggaa      120
aatcgaggct cggccaatcg agttccgacg gagaagcctc cgtttacgct cggtcagatc      180
aagcaagcca ttccgccccca ctgttttcgc cgctccctcc ttcgatcctt ctctacgtg      240
gtocatgacc tatgcttagc ctctctcttt tactacattg caacatcata ttttcacttt      300
ctcccacaac ccttttctta cattgcttgg cctgtctatt gggttctcca aggttgcac      360
ctcaccggtg tttgggtcat cgcacacgaa tgcggtcacc acgctttcag tgactaccaa      420
tgggttgacg acaccgtcgg gttgatcctt cactccgcc ttttagtccc gtacttctcg      480
tggaaaatca gtcaccgccc tcaccactcg aacaccggtt ccatggagcg tgacgaagta      540
ttcgtgcccc aaccaagtc taaattatca tgctttgcga aatacttcaa caatccacc      600
ggtcgagttc tctctcttgt agtcacattg actcttgggt ggcttatgta cttagccttc      660
aacgtttcgg gtcgatacta tgatcgatta gcttcccact ataaccctta cggccccatt      720
tactccgaac gcgagaggct acaagtttac atctccgatg ctggtatagt tgcggtaatt      780
tatgtacttt ataagattgc tgcaacaaaa gggctggctt ggcttttatg cacttatggg      840
gtacctctac ttattgtgaa tgccctcctt gtgttgatca cctacttgca acatactcac      900
tcggcattgc cgcattacga ctcgcttgaa tgggattggt ttcgaggagc attgtcgacg      960
attgatcgag attacggggt gttgaacaaa gtgttcata acatcaccga tacgcatgtg     1020
gctcatcacc tcttotcaac gatgccacat tatcatgcaa tggaggccac taaagcaatc     1080
aaaccgatac tcggcaagta ttatcctttc gacgggacac cgatttataa ggcaatgtgg     1140
agggaggcaa aagagtgcct ttacgtcgag gctgacgttg gtggtggtgg tagcaaaggt     1200
gttttttggg atcgtaacaa gttctaaaga cagaccaact gcctgatagc tggccggcaa     1260
aatcgacgta aaacgtactt attagactag tgtaactag ggaagttaat aatggtagga     1320

aaatgtggaa tagctgccta gtagttttat gtattaagtg tt                               1362

```

<210> 2
 <211> 384
 <212> PRT
 <213> *Gossypium hirsutum*

 <400> 2

5

ES 2 603 530 T3

Met Gly Ala Gly Trp Val Gly Cys Gln Leu Thr Gly Ile Lys Glu Glu
 1 5 10 15

Asn Arg Gly Ser Val Asn Arg Val Pro Ile Glu Lys Pro Pro Phe Thr
 20 25 30

Leu Gly Gln Ile Lys Gln Ala Ile Pro Pro His Cys Phe Arg Arg Ser
 35 40 45

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

Leu Phe Tyr Tyr Ile Ala Thr Ser Tyr Phe His Phe Leu Pro Gln Pro
 65 70 75 80

Phe Ser Tyr Ile Ala Trp Pro Val Tyr Trp Val Leu Gln Gly Cys Ile
 85 90 95

Leu Thr Gly Val Trp Val Ile Ala His Glu Cys Gly His His Ala Phe
 100 105 110

Ser Asp Tyr Gln Trp Val Asp Asp Thr Val Gly Leu Ile Leu His Ser
 115 120 125

Ala Leu Leu Val Pro Tyr Phe Ser Trp Lys Ile Ser His Arg Arg His
 130 135 140

His Ser Asn Thr Gly Ser Met Glu Arg Asp Glu Val Phe Val Pro Lys
 145 150 155 160

Pro Lys Ser Lys Leu Ser Cys Phe Ala Lys Tyr Phe Asn Asn Pro Pro
 165 170 175

Gly Arg Val Leu Ser Leu Val Val Thr Leu Thr Leu Gly Trp Pro Met
 180 185 190

Tyr Leu Ala Phe Asn Val Ser Gly Arg Tyr Tyr Asp Arg Leu Ala Ser
 195 200 205

ES 2 603 530 T3

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

Val Tyr Ile Ser Asp Ala Gly Ile Val Ala Val Ile Tyr Val Leu Tyr
 225 230 235 240

Lys Ile Ala Ala Thr Lys Gly Leu Ala Trp Leu Leu Cys Thr Tyr Gly
 245 250 255

Val Pro Leu Leu Ile Val Asn Ala Phe Leu Val Leu Ile Thr Tyr Leu
 260 265 270

Gln His Thr His Ser Ala Leu Pro His Tyr Asp Ser Ser Glu Trp Asp
 275 280 285

Trp Phe Arg Gly Ala Leu Ser Thr Ile Asp Arg Asp Tyr Gly Val Leu
 290 295 300

Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Ala His His Leu
 305 310 315 320

Phe Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr Lys Ala Ile
 325 330 335

Lys Pro Ile Leu Gly Lys Tyr Tyr Pro Phe Asp Gly Thr Pro Ile Tyr
 340 345 350

Lys Ala Met Trp Arg Glu Ala Lys Glu Cys Leu Tyr Val Glu Ala Asp
 355 360 365

Val Gly Gly Gly Gly Ser Lys Gly Val Phe Trp Tyr Arg Asn Lys Phe
 370 375 380

<210> 3
 <211> 5201
 <212> ADN
 <213> *Gossypium hirsutum*

<400> 3

ctcgagaaaa tttctcatca tttgattcgg tttttcttc gttaaggtagt ggtcgtagtt 60

tttcgtttct cagtttttagt ttgatctgag cttttttggt tgaaatgatg ctgatctctt 120

tagttaccga gaaaataaat gagaaggatg tattgttata ttgttttttt ttctcgaaag 180

gtttcagagtg attgtggcgg ccataggaga tctaagtgtt tctcgcattt tctcggcaat 240

caaaggcagc ttaaatttat aatttatatg aaaaagtcca gcttattggt tattttatcg 300

tatttttttt ctcatgaaac acagaaagac caaagctcca ttctgtattg tttgctattt 360

tctctttcat tttttaagtt tattttccgg tttttcgtct tgtaaacttt ttaagcgtat 420

5

10

ES 2 603 530 T3

tactgttttt gtttttagcg attctacatc ctttttgcaa catatgttct tcaacgcoga	480
tttaagattc tttatgcgac gctctttggt taagtactgt gatttttagt acataacaat	540
ttgctttgag ctactgggta atcctatctt cctttttgta aatagatccc tataaatcta	600
gaactttgaa atataaacia acacgtgat attatgttct ttagattatg gctgcttatt	660
aaataaagta gaaagaaatt atttacctat tagattggaa cctgctttta aacatataga	720
atatgtaatc ttctatatag tcaaaatctg aagatttatg tgattatatt tattgttagt	780
gogtcttttg gatatgttct gattctttaa tacatgtgat actgttagct catgcaccag	840
ctgcttgat aaaaagcttt agattttgca aaagaagggg tttcagcacg aaattgaagt	900
tgtttttaaa aaccatgggt gctactgctg tgacatcggc gtttttccca gtcacttctt	960
cacctgactc ctctgactcg aaaaacaaga agctcggaag catcaagtcg aagccatcgg	1020
tttcttctgg aagtttgcaa gtcaaggcaa atgctcaagc acctccgaaa ataaacggca	1080
ctgtggcgtc gacgactccc gtggaagggt ccaagaacga tgacggtgca agtccccctc	1140
ctcctaggac gtttatcaac cagttacctg attggagcat gcttcttgct gctatcacia	1200
ccattttctt ggctgctgag aagcagtgga tgatgcttga ttggaagccg aggcggcctg	1260
acatggtcac tgatccgttt ggcatagggg agattgttca ggatggctct gttttcagtc	1320
agaacttctc gattagatca tatgagatag gcgctgatca aacagcatcc atagagacac	1380
taatgaatca tttacaggta gagttacagt ttttggtca gtatgtttga acaatgaaca	1440
ttgggaaaca ggatttacat ttattggttt cttttgtaga gatatggcat aagcttgagt	1500
tttagttaat acatctcact ttttggcagg aaacagctat aaatcattgt cgaagtgctg	1560
gactgcttgg agaaggtttt ggtgcaacac ctgagatgtg caagaagaac ctaatatggg	1620
ttgtcacacg gatgcaagtt gtggttgatc gctatcctac ttggtaagac atgctttttt	1680
gctcatgatt atagcaacia ttcatgataa gccacttttg ctctacagta tggtgtggc	1740
atatcttttg atactaacta gttcagttct tgaattccag caatattctg tattatacaa	1800
atgatctgta tcacatctgg cggacttggt tttgtttcat taaaacttgg attgatgta	1860
ttgtttaagc ttttaaaggt taagatatga agtcgaagac aattaaggct agccccagc	1920
aatgaataac ataagaaaga taaacctgat acgcttcttt gtttaatgag attccctggt	1980
taatactaag aacgggacct taacttgta tttttgtaa tgtaataact tccccctaa	2040
ttgtgattta ggtacctatg cactatattg ttttaagcac ttggcaagtt attgggtgtg	2100
aaaataacat tcaacttaact gatattagct tggcattagg gcttgtttca aaaataataa	2160
taaaaacgaa gaagttggcc taaaatagtt acttttagca aggtatgtac ttgttggagc	2220
accgttcttt tttgtatgtc acaaaattag tagttctgga ggtattaaga gtatatagaa	2280

ES 2 603 530 T3

atattattat tatttttcggt aggtgtatgt ctatattgat gaacattgaa cttattttttg 2340
ttgtccttgca ggggtgatgt tgttcaagtc gacacttggg tcagtgcacg ggggaagaat 2400
ggcatgcgaa gagattggct tgtcagcaat agtgaaactg gtgaaatfff aacacgagcc 2460
acaaggtcgg tcattgttta tggaaggatc atgaccatat gttttttttc gtgatgtaac 2520
ctggtgaact ctaaaatatt tcaaaatttg ttttgctagc atttaatatg ttttcaatcg 2580
atacacgaat cgatattttg gtttccagg ccacatctaa tgactttttc ctgacctgtg 2640
gtttgcactt taatgaacag tgtttcatga gtgactaatc ccagtctect ctgtttttgt 2700
tttgttgatc tgcagtgtat ggggtgatgat gaataaactg actagaaggt tatctaaaat 2760
cccagaagag gttcgagggg aatagaacc tttttttatg aattcagatc ctgtttctggc 2820
tgaggatagc cagaaactag tgaaactcga tgacagcaca gctgaacacg tgtgcaaagg 2880
tttaactgta agtccccgct tcccctgttt ctctttcata tacttacagc gtctgtcact 2940
tgtaattgct gttatgttca ttcgttgcaa tttgtaataa agtttatatc taatttgcag 3000
cctaaatgga gcgacttggg tgtcaaccag catgtcaata atgtgaagta cattggctgg 3060
atccttgagg tagactcact ccggttgtat ttcaaggatt ttcttttgaa cattctcacc 3120
attacctctt cgtatccgaa gataacaatt aaatggaaat cataactgat ttttatttcg 3180
tataccttaa tttttacctg tgatgtcaag ctaatatgaa attggaattt tggtagaagt 3240
ctgcttcaga tcttcacgag ttaagtcatt tatagtctgt tggttacatg cattttaacc 3300
ggatggtagt acttgttgca gagtgctcca ttaccaatct tggagagtca cgagctttcc 3360
gccttgactc tggaaatatag gagggagtgc gggagggaca gcgtgctgca gtcactgacc 3420
actgtgtctg attccaatac ggaaaatgca gtaaagtgtg gtgaatttaa ttgccaacat 3480
ttgctccgac tcgacgatgg agctgagatt gtgagaggca ggacccgatg gaggcctaaa 3540
catgccaaaa gttccgctaa catggatcaa attaccgcaa aaagggcata gaaatccaag 3600
taatctcatt gctgtgtgta gtatctatcg tgctcttttc ggatttatat acatatattc 3660
cttatgatta ttagtcttcc tttgagaaaa aaaaaggggg ttgtaattag gcttgtttag 3720
gagtcggggt ttcgtacata gccttgtaag gctcagctcg tatgaccga gcctcggaca 3780
cggattttgt gaagttgggc ccgtgccta accagcatag gctctttcca tggaaagggt 3840
ggtctgcttt tgaaaaattg aatagccatg tgagatggct ctctccctac attatgggct 3900
tttaaccagt tagagaccgg gtagtttagg ataaaattta tctttaattt gggaggattt 3960
gtatattttt tttgccttta ttttaacctg aatttgctta taattatttg gttttatatt 4020
taggtattga atcaatgaag tttttaaatt ttaaaatggt ttgattggct tacattgaca 4080
aagcattgta tcaaccaatt tttttagatc aaatgtaagc ccttgtgtaa ctattgcatg 4140
catgtagga aaaaaaaaaac aaaaattcac ttaaatttag gtgaggtggt tattttttta 4200

ES 2 603 530 T3

```

aggtatatat gcgtttataa attgatttat gtacgtaaatt gtatttcatg tgcacgggtg      4260
cattogaaat caaataaagg gggttaaatta atttttaagt tcatcgtttc gaattggtaa      4320
aaaaatagaa aattgagata aaacagtgat tgaatttggt tttattattt ttaacatttt      4380
attaatattt taattgttta tttaatgatt gttggattga tgcacgaact aattaaatta      4440
agaattaatg gtttgataag tttgatcatg acgatgaact cataatatct gtatcttgtc      4500
tgaaacacag tcttctgtac atcttgcctc ttcactctta attggtatcc agatctcaag      4560
tcaatcttag aaaatacaat ggctcccttt aactaatcaa acaggtcacc aaagacgata      4620
ttactacagg cagactcatt tggccaact tgaagaacat tcccgctctt acatttcaac      4680
tcaattatct ttttcccaca gtcgactata aactatgac cagtcaacca atctatccca      4740
aggattatat cgaactcatt aaatgacaat aacatcaagt tagtcgaaaa acaatgacct      4800
ctaattgtca aagggcaatt tctacatact tgatccacta gtgcatgttt gcctaagggg      4860
ttggacaact ttattacaaa ctacgtggac tctgctagca tatccatag aggtatcaat      4920
tccatacaaaa tataagagtg ggtagatctg gggcaacta aagcaatgac agatatttca      4980
tggatagaaa aggtaccctg gatcacgtca ggagattctg cctcttcccg agcttgaata      5040
gcataagtcc ttgcagttgt tctaccctcg gacctactg cagcatctct tggatgcct      5100
ctactgctag cccacttct ggggttcttt tgtgatctac cctcaaggg agcactgctt      5160
gctttcacat ctgttttct ctctctctca ttcaactcga g                               5201

```

<210> 4
 <211> 413
 <212> PRT
 <213> *Gossypium hirsutum*

5

<400> 4

```

Met Val Ala Thr Ala Val Thr Ser Ala Phe Phe Pro Val Thr Ser Ser
 1                5                10                15

Pro Asp Ser Ser Asp Ser Lys Asn Lys Lys Leu Gly Ser Ile Lys Ser
      20                25                30

Lys Pro Ser Val Ser Ser Gly Ser Leu Gln Val Lys Ala Asn Ala Gln
      35                40                45

Ala Pro Pro Lys Ile Asn Gly Thr Val Ala Ser Thr Thr Pro Val Glu
      50                55                60

Gly Ser Lys Asn Asp Asp Gly Ala Ser Ser Pro Pro Pro Arg Thr Phe
      65                70                75                80

```

10

ES 2 603 530 T3

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

Ile Phe Leu Ala Ala Glu Lys Gln Trp Met Met Leu Asp Trp Lys Pro
100 105 110

Arg Arg Pro Asp Met Val Ile Asp Pro Phe Gly Ile Gly Lys Ile Val
115 120 125

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

Ile Gly Ala Asp Gln Thr Ala Ser Ile Glu Thr Leu Met Asn His Leu
145 150 155 160

Gln Glu Thr Ala Ile Asn His Cys Arg Ser Ala Gly Leu Leu Gly Glu
165 170 175

Gly Phe Gly Ala Thr Pro Glu Met Cys Lys Lys Asn Leu Ile Trp Val
180 185 190

Val Thr Arg Met Gln Val Val Val Asp Arg Tyr Pro Thr Trp Gly Asp
195 200 205

Val Val Gln Val Asp Thr Trp Val Ser Ala Ser Gly Lys Asn Gly Met
210 215 220

Arg Arg Asp Trp Leu Val Ser Asn Ser Glu Thr Gly Glu Ile Leu Thr
225 230 235 240

Arg Ala Thr Ser Val Trp Val Met Met Asn Lys Leu Thr Arg Arg Leu
245 250 255

Ser Lys Ile Pro Glu Glu Val Arg Gly Glu Ile Glu Pro Phe Phe Met
260 265 270

Asn Ser Asp Pro Val Leu Ala Glu Asp Ser Gln Lys Leu Val Lys Leu
275 280 285

Asp Asp Ser Thr Ala Glu His Val Cys Lys Gly Leu Thr Pro Lys Trp
290 295 300

Ser Asp Leu Asp Val Asn Gln His Val Asn Asn Val Lys Tyr Ile Gly
305 310 315 320

Trp Ile Leu Glu Ser Ala Pro Leu Pro Ile Leu Glu Ser His Glu Leu
325 330 335

ES 2 603 530 T3

Ser Ala Leu Thr Leu Glu Tyr Arg Arg Glu Cys Gly Arg Asp Ser Val
 340 345 350

Leu Gln Ser Leu Thr Thr Val Ser Asp Ser Asn Thr Glu Asn Ala Val
 355 360 365

Asn Val Gly Glu Phe Asn Cys Gln His Leu Leu Arg Leu Asp Asp Gly
 370 375 380

Ala Glu Ile Val Arg Gly Arg Thr Arg Trp Arg Pro Lys His Ala Lys
 385 390 395 400

Ser Ser Ala Asn Met Asp Gln Ile Thr Ala Lys Arg Ala
 405 410

<210> 5
 <211> 1647
 <212> ADN
 <213> *Gossypium hirsutum*

5

<400> 5

ttttgtagat cgcaagccgt agaagtaaac aaagaaggat ttgcagatct taaatcatgc	60
tgtaattttc tcgaggaaat tttttacctt cattgattcg tttctatttt cgcttgagtt	120
ggagaatgct tcagctgcct ttctaaaaga ttgtttccaa aagaggggat tttagtggat	180
aatagaagtt ctttttaatt ctcaaaatca tggttgccac tgctgtaca tcctcattct	240
ttcccgtaac ttcttccctt gactcctctg actcaaaaaa caagaagctt ggaagtggat	300
ctactaacct cggaggcatc aagtogaaac catctgcttc ttctggaagt ttgcaagtca	360
aggcaaatgc tcaagcccct ccaaagataa atggtaccac tgttgtaact tctccggttg	420
aaggtttcaa gaacgaagat ggtgcagggt cccctcatcc tcggacctt atcaatcaat	480
tacctgattg gagcatgctt cttgccgcta tcacaacat tttcctggct getgagaagc	540
agtggatgat gcttgattgg aagccaaggc ggctgacat gctcattgat ccttttggtg	600
tagggaggat tgttcaggat ggtcttgttt tccgtcaaaa cttctcgatt aggtcttatg	660
agataggtgc tgatcgtagc gcatccatag agacgctaataaatcattta caggaaaccg	720
cgattaatca ttgtaaaagt gctggactgc ttggagaagg ttttgggtgct acccctgaga	780
tgtgcaagaa gaacctaat tgggtggtca ctccgatgca agttgtgttt gatcggtatc	840
ctacttgggg tgatgttgtt caagtagaca cttgggtcag tgcatcagga aagaatggca	900
tgcaagaga ttggcttgtc agtgatagta aaactggtga agttttaaca agagcctcaa	960
gtgtgtgggt gatgatgaat aaattgacta gaaggctatc taaaattcct gaggaggtcc	1020
gaggagaaat agaaccttat tttatgaatt ccgatcctgt tgtggcagaa gatagccgga	1080

10

ES 2 603 530 T3

```

aattagtgaa gctcgataaa agcatggetg agcacgtgcg taaaggttta actcctagat      1140
ggagtgactt ggatgtcaac caacatgtca ataacgtgaa gtacattggc tggatcctcg      1200
agagtgctcc attgccggtg ttggaaactc acgagctttc ttccatgaca ctggagtata      1260
ggagggagtg tgggagggag agcatactgc agtcgctaac aaccgtgtcc gactccagtg      1320
taggagactt ggtgaatgtg ggtgaaatcg agtgccagca cctgctgcaa ctcgaggaag      1380
ggtccgagat tgtgagaggg agaactcaat ggaggcccaa gtatgccaaa agttttggta      1440
atgtgggtca aattccagca gaaagtgcac agaaggaaaa aatcccaaaa tttctcttat      1500
tgtgacctaa gtggggcaat agtctgattg ccgggtgtca caatgattta tgtagaatct      1560
aatcatgtgt tctatggata tatatatata tatatttatg cttctttttt atatataaat      1620
aatattatat attcctttta aaaaaaa      1647

```

<210> 6
 <211> 420
 <212> PRT
 <213> *Gossypium hirsutum*
 <400> 6

5

```

Met Val Ala Thr Ala Ala Thr Ser Ser Phe Phe Pro Val Thr Ser Ser
 1                5                10                15

Pro Asp Ser Ser Asp Ser Lys Asn Lys Lys Leu Gly Ser Gly Ser Thr
          20                25                30

Asn Leu Gly Gly Ile Lys Ser Lys Pro Ser Ala Ser Ser Gly Ser Leu
          35                40                45

Gln Val Lys Ala Asn Ala Gln Ala Pro Pro Lys Ile Asn Gly Thr Thr
          50                55                60

Val Val Thr Ser Pro Val Glu Gly Phe Lys Asn Glu Asp Gly Ala Gly
 65                70                75                80

Ser Pro His Pro Arg Thr Phe Ile Asn Gln Leu Pro Asp Trp Ser Met
          85                90                95

Leu Leu Ala Ala Ile Thr Thr Ile Phe Leu Ala Ala Glu Lys Gln Trp
          100                105                110

Met Met Leu Asp Trp Lys Pro Arg Arg Pro Asp Met Leu Ile Asp Pro
          115                120                125

Phe Gly Ile Gly Arg Ile Val Gln Asp Gly Leu Val Phe Arg Gln Asn
 130                135                140

```

10

ES 2 603 530 T3

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

Glu Thr Leu Met Asn His Leu Gln Glu Thr Ala Ile Asn His Cys Lys
 165 170 175

Ser Ala Gly Leu Leu Gly Glu Gly Phe Gly Ala Thr Pro Glu Met Cys
 180 185 190

Lys Lys Asn Leu Ile Trp Val Val Thr Arg Met Gln Val Val Phe Asp
 195 200 205

Arg Tyr Pro Thr Trp Gly Asp Val Val Gln Val Asp Thr Trp Val Ser
 210 215 220

Ala Ser Gly Lys Asn Gly Met Arg Arg Asp Trp Leu Val Ser Asp Ser
 225 230 235 240

Lys Thr Gly Glu Val Leu Thr Arg Ala Ser Ser Val Trp Val Met Met
 245 250 255

Asn Lys Leu Thr Arg Arg Leu Ser Lys Ile Pro Glu Glu Val Arg Gly
 260 265 270

Glu Ile Glu Pro Tyr Phe Met Asn Ser Asp Pro Val Val Ala Glu Asp
 275 280 285

Ser Arg Lys Leu Val Lys Leu Asp Lys Ser Met Ala Glu His Val Arg
 290 295 300

Lys Gly Leu Thr Pro Arg Trp Ser Asp Leu Asp Val Asn Gln His Val
 305 310 315 320

Asn Asn Val Lys Tyr Ile Gly Trp Ile Leu Glu Ser Ala Pro Leu Pro
 325 330 335

Val Leu Glu Thr His Glu Leu Ser Ser Met Thr Leu Glu Tyr Arg Arg
 340 345 350

Glu Cys Gly Arg Glu Ser Ile Leu Gln Ser Leu Thr Thr Val Ser Asp
 355 360 365

Ser Ser Val Gly Asp Leu Val Asn Val Gly Glu Ile Glu Cys Gln His
 370 375 380

Leu Leu Gln Leu Glu Glu Gly Ser Glu Ile Val Arg Gly Arg Thr Gln
 385 390 395 400

ES 2 603 530 T3

Trp Arg Pro Lys Tyr Ala Lys Ser Phe Gly Asn Val Gly Gln Ile Pro
 405 410 415

Ala Glu Ser Ala
 420

5 <210> 7
 <211> 1498
 <212> ADN
 <213> *Gossypium hirsutum*

<400> 7

```

tagcgctaga agttaccgag aagagtttag agatccctat tatcggaaag aggggatttc      60
agcggataac agaagttcat tttaatttat aaaatcatgg ttgccactgc tgctacatcc      120
tcattctttc caatcacttc ttccccggac tccattgact caaaaaacaa gaagcttggga      180
aatggatcta ctaaccttgg aggtataaag ttgaaacat ctgcttcttc tgggaagtttg      240
caagttaagg caaatgcaca agcccccca aagataaatg gtaccacagt tgtgatgact      300
ccagtagaag gtttccccgag cgaagatgct gcaagttccc tacctcccag gacgtttatc      360
aatcagctac ctgattggag catgcttctt gctgctatga caaccatttt cctggctgct      420
gagaagcagt ggatgatgct tgattggaag ccaaagcggc ctgacatgct cattgacca      480
tttgggatag ggaggattgt tcaggatggt cttgtttttc gtcagaactt ctcaattagg      540
tcttatgaga taggtgctga tcgtacagca tccatagaga cgctaatgaa tcatttacag      600
gaaacagcga ttaatcattg taaaagtgct ggactgctag gagatggttt tgggtgctacc      660
cctgggatgt gcaagaaaaa cctaatatgg gtagtcaccc ggatgcaagt tgtggttgat      720
tgttatccaa cttgggggtga tgttgttcaa gtagacactt gggtcagtgc atcaggaaag      780
aatggcatgc gaagggattg gcttgtcagc aatagtaaaa ctggtgaaat ttaactaga      840
gcctcaagtg tgtgggtgat gatgaataaa ttgaccagaa ggttatctaa aattccagaa      900
gaggtcagag gagaaataga acctcatttt atgaattcag atccagtggg ggctgaggat      960
aaccggaaat tagtgaaact tgacgacagc acagcccaat atgtgcgcaa gggtttaact     1020
cctcgatgga gcgacctgga tgtgaatcag catgtcaaca atgtgaagta cgttggttgg     1080
atccttgaga gtacaccatt ggggaattgtg gagagtcatg agctttgttc catgacactg     1140
gagtatagga gggagtgtgg gagggacagc gtgctgcagt cactaactgc ggtgtctggt     1200
gtgggcaacc tcgggaatat gggggaaatt gagtgccagc acttgetcca acttgaagag     1260
gggtctgaga ttgtgagagg gaggacacag tggaggccaa agaatgcaa gagttttggt     1320
aaaatggatc aagttcccgc acaaagtgca tagatccgaa gtctctttgc tgcgtgtcaa     1380
aactagcagt caacgcattg tgtagaatct ttcttttggg ctttgaatcc ataatatata     1440

tatatgatat tagcttghaa gctttcaaag cttgctghaa ttagctctaa aaaaaaaaa     1498
    
```

10

ES 2 603 530 T3

<210> 8
 <211> 418
 <212> PRT
 <213> *Gossypium hirsutum*

5

<400> 8

Met Val Ala Thr Ala Ala Thr Ser Ser Phe Phe Pro Ile Thr Ser Ser
 1 5 10 15

Pro Asp Ser Ile Asp Ser Lys Asn Lys Lys Leu Gly Asn Gly Ser Thr
 20 25 30

Asn Leu Gly Gly Ile Lys Leu Lys Pro Ser Ala Ser Ser Gly Ser Leu
 35 40 45

Gln Val Lys Ala Asn Ala Gln Ala Pro Pro Lys Ile Asn Gly Thr Thr
 50 55 60

Val Val Met Thr Pro Val Glu Gly Phe Pro Ser Glu Asp Ala Ala Ser
 65 70 75 80

Ser Leu Pro Pro Arg Thr Phe Ile Asn Gln Leu Pro Asp Trp Ser Met
 85 90 95

Leu Leu Ala Ala Met Thr Thr Ile Phe Leu Ala Ala Glu Lys Gln Trp
 100 105 110

Met Met Leu Asp Trp Lys Pro Lys Arg Pro Asp Met Leu Ile Asp Pro
 115 120 125

Phe Gly Ile Gly Arg Ile Val Gln Asp Gly Leu Val Phe Arg Gln Asn
 130 135 140

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

Glu Thr Leu Met Asn His Leu Gln Glu Thr Ala Ile Asn His Cys Lys
 165 170 175

Ser Ala Gly Leu Leu Gly Asp Gly Phe Gly Ala Thr Pro Gly Met Cys
 180 185 190

Lys Lys Asn Leu Ile Trp Val Val Thr Arg Met Gln Val Val Val Asp
 195 200 205

ES 2 603 530 T3

Cys Tyr Pro Thr Trp Gly Asp Val Val Gln Val Asp Thr Trp Val Ser
 210 215 220

Ala Ser Gly Lys Asn Gly Met Arg Arg Asp Trp Leu Val Ser Asn Ser
 225 230 235 240

Lys Thr Gly Glu Ile Leu Thr Arg Ala Ser Ser Val Trp Val Met Met
 245 250 255

Asn Lys Leu Thr Arg Arg Leu Ser Lys Ile Pro Glu Glu Val Arg Gly
 260 265 270

Glu Ile Glu Pro His Phe Met Asn Ser Asp Pro Val Val Ala Glu Asp
 275 280 285

Asn Arg Lys Leu Val Lys Leu Asp Asp Ser Thr Ala Gln Tyr Val Arg
 290 295 300

Lys Gly Leu Thr Pro Arg Trp Ser Asp Leu Asp Val Asn Gln His Val
 305 310 315 320

Asn Asn Val Lys Tyr Val Gly Trp Ile Leu Glu Ser Thr Pro Leu Gly
 325 330 335

Ile Val Glu Ser His Glu Leu Cys Ser Met Thr Leu Glu Tyr Arg Arg
 340 345 350

Glu Cys Gly Arg Asp Ser Val Leu Gln Ser Leu Thr Ala Val Ser Gly
 355 360 365

Val Gly Asn Leu Gly Asn Met Gly Glu Ile Glu Cys Gln His Leu Leu
 370 375 380

Gln Leu Glu Glu Gly Ser Glu Ile Val Arg Gly Arg Thr Gln Trp Arg
 385 390 395 400

Pro Lys Asn Ala Lys Ser Phe Gly Lys Met Asp Gln Val Pro Ala Gln
 405 410 415

Ser Ala

- <210> 9
- <211> 743
- 5 <212> ADN
- <213> *Gossypium hirsutum*
- <220>
- <221> misc_feature
- 10 <222> (571)..(571)
- <223> n es a, c, g o t
- <400> 9

ES 2 603 530 T3

atggtttacc cacgcgtccg gtagaacatg ttggtgaaga atatattgag gagttttaca 60
gatgctgtga ccaattactg aaagaagatg gactttttgt tcttcagttc atatctatcc 120
cagaagagct ttccaaagaa atccagcaaa cagcaggttt tctaaaggaa tatatattcc 180
ccggtggaac cctgctttct ttggatagga atttatcagc catggctgct gcaacaagat 240
tcagtgtgga gcatgtggaa aatataggaa tgagttatta ccacacactg agatggtgga 300
gaaaactttt cctggaaaac acaagcaaag ttctagctct gggattcgac gagaagttca 360
tgaggacatg ggaatactat ttcgattact gcgctgccgg ttttaagaca ggaaccctta 420
tagattacca ggttgtattt tcgcgggccg gaaatttcgg taccactcgg gatccataca 480
aaggtttccc ttctgcatac tccttcatgg atgattgaac aaagtgtggt tgaacattga 540
tccaaagaag caaacaaaat tatcaccaca ntgccagtgt taagaacaac ctatctccct 600
agtcocctact tttctttatt atggctatgt ttgcaatgca agaataagca aacattgtaa 660
tgtcaataaa gtttgcactt ttgtagactg gatgggatgt tatcaatgaa gtacctagtt 720
tataagtaaa aaaaaaaaaa aga 743

<210> 10
<211> 2884
<212> ADN
<213> *Gossypium hirsutum*

5

<400> 10

gcacaaggta aagcagtgta ccggcggcag tgatggaagt ggccgtgatc ggaggtggga 60
taaaaggggt gctttcggcc tacgtactgg tcaaagccgg cgtggacgtg gtggtttacg 120
agaaagaaga acaattaggc ggccatgcaa agactgttaa cttcgacgcc gttgatttag 180
accttggtct cttgtttctc aatccagcaa gatatgcaac actattgcat atgttcgaca 240
gccttgggtg tgatgtagaa acatccgatg tttcattctc tataagccat gacaaaggca 300
acaatggcta tgaatggtgc agccaatatg gattttccaa ttactttgct caaaagaaga 360
aactgttgaa ccctttcaat tggcaaagcc tcagagagat catcaaattc ggcaatgatg 420
tcgaaagtta ccttggatca cttgagaaca acccagacat tgatcgtact gagaccttgg 480
gacagtttat aaactcaaag ggctactctg aaaattttca aaacacttat ctggctccta 540
tatgtggttc aatgtggtca agctccaagg aagatgttac gagcttttca gctttttcca 600
tcctttcatt ttgccgtact catcatttgt accagctatt tgggcagtca cagtggttga 660
ctatcaaagg gcactcacat tttgttaaaa gggttagggg agtgcctggag actaaagggt 720
gtcaatttaa actcggttgt gaagtacaat ctgttttgcc cgttgataat ggtaccgcca 780

10

ES 2 603 530 T3

tggctctgtgg agatggtttc caagaaactt acaatggatg cataatggct gttgatgctc 840
 ccaactgcctt aaaattatta ggaaaccaag caacatttga agaaacaaga gtactgggtg 900
 ctttccaata tgctaccagt gatattttcc ttcaccagga cagtacttta atgccacaaa 960
 acaaatcagc ttggagtgca ttgaattttc tcaatagtag caaaaataat gcattcttaa 1020
 catactggct caatgcacta cagaatattg ggaaaacaag tgagccattt tttgtgactg 1080
 tcaatccaga ccataccccg aagaatacct tacttaagtg gtcaaccggc catgcaatts 1140
 cctctgttgc tgcatcaaaa gcttcacttg agcttggca gattcagggg aagagaggaa 1200
 tctggttctg tggctatgac ttcaatcagg atgaactaaa ggctggtatg gatgctgcac 1260
 atggtatctt gggaaagcat tcttctgttc cgccagctcc aaagaatatg tcaccctctt 1320
 taccaaagaa tatgtcacc tctttcatgg aaacaacggc acgcctcttt gttaccaaat 1380
 tctttcaaca atatatatct atgggctgcg taattttttt agaggaagga ggcagaattt 1440
 tcactttcaa aggaaacatg gaaaagtgtc ctcttaaac agttctgaaa gtgcataatc 1500
 ctcagtttta ctggaggatc atgaaagaag ctgatatagg ccttgcagac gcatatatcc 1560
 atggagattt tcttttctt gatgaaaatg aaggccttct taatcttttc eggattcttg 1620
 ttgccaataa agagaactca gctgcctcag ggtcgactaa aagaaggact tgggtgctgc 1680
 ctgctctgtt aacagctagt atatcatctg ccaagtattt tgtgaagcat ctcttaagac 1740
 aaaatactat tacacaagct cgtaggaaca tttctcgtca ttatgatctg agtaatgaac 1800
 tttctctct atacttgggc aaaatgatgc aatactcttc tggagtcttt aggacaggag 1860
 aagaacattt ggacgttgca cagcgaagaa aatcagttc tctaattgag aaaacaagga 1920
 tagagaaatg gcatgaagtt ctagacattg ggtgcggttg gggaaagctta gctattgaaa 1980
 ctgtgaaaag aacaggatgc aatatactg gcatcactct atcagaacag caactgaaat 2040
 atgctcaaga aaaagtgaag gaagctggac tcgaggataa catcaaaata cttctctgtg 2100
 actatogcca gttacctaag gaacaccaat ttgacagaat catatctgta gagatggtag 2160
 aacatgttgg tgaagaatat attgaggaat tttacagatg ctgtgatcaa ttactgaaag 2220
 aagatggact tttcgttctt cagttcatat ctatcccaga ggagctttcc aaagaaatcc 2280
 agcaaacagc tggttttctt aaggaatata tattccctgg tggaaacctg ctttctttgg 2340
 ataggaattt atcagccatg gctgctgcaa caagattcag tgtggagcat gtggaaaaca 2400
 taggaatgag ttattaccac aactgagat ggtggagaaa acttttctctg aaaaacacaa 2460
 gcaaagttct ggctttgggg ttcgacgaga agttcatgcg gacatgggaa tactatctcg 2520
 attactgtgc tgctggtttt aagacaggaa cccttataga ttaccagggt gtattttctc 2580
 gagccggtaa tttcgttaca cttggagatc catacaaagg tttcccttct gcatattcct 2640

ES 2 603 530 T3

```

tcattgatga ttgaacaaag tgtttgaata tatgatcacc atacaatgat tcaaccagct      2700
ggatcaaact ggtaccagtg tttacctagt ccctgcttt tgtttagtta tggttttcgt      2760
ttcgttgcca aaaagaaaaa agcaaataat gtatgttaat aatgaaatgt ttgtatctgg      2820
tatatctata ctggttggat tttatgtatg gagatctggt tctttttaa aaaaaaaaaa      2880
aaaa                                                                    2884

```

<210> 11
 <211> 873
 <212> PRT
 <213> *Gossypium hirsutum*

<220>
 <221> misc_feature
 <222> (370)..(370)
 <223> Xaa puede ser cualquier aminoácido de origen natural

<400> 11

```

Met Glu Val Ala Val Ile Gly Gly Gly Ile Lys Gly Leu Leu Ser Ala
 1                               5                               10           15

Tyr Val Leu Val Lys Ala Gly Val Asp Val Val Val Tyr Glu Lys Glu
                20                               25           30

Glu Gln Leu Gly Gly His Ala Lys Thr Val Asn Phe Asp Ala Val Asp
                35                               40           45

Leu Asp Leu Gly Phe Leu Phe Leu Asn Pro Ala Arg Tyr Ala Thr Leu
 50                               55           60

Leu His Met Phe Asp Ser Leu Gly Val Asp Val Glu Thr Ser Asp Val
 65                               70           75           80

Ser Phe Ser Ile Ser His Asp Lys Gly Asn Asn Gly Tyr Glu Trp Cys
                85                               90           95

Ser Gln Tyr Gly Phe Ser Asn Tyr Phe Ala Gln Lys Lys Lys Leu Leu
                100                              105           110

Asn Pro Phe Asn Trp Gln Ser Leu Arg Glu Ile Ile Lys Phe Gly Asn
                115                              120           125

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

Arg Thr Glu Thr Leu Gly Gln Phe Ile Asn Ser Lys Gly Tyr Ser Glu
 145                              150           155           160

```

5

10

15

ES 2 603 530 T3

Asn Phe Gln Asn Thr Tyr Leu Ala Pro Ile Cys Gly Ser Met Trp Ser
 165 170 175
 Ser Ser Lys Glu Asp Val Thr Ser Phe Ser Ala Phe Ser Ile Leu Ser
 180 185 190
 Phe Cys Arg Thr His His Leu Tyr Gln Leu Phe Gly Gln Ser Gln Trp
 195 200 205
 Leu Thr Ile Lys Gly His Ser His Phe Val Lys Arg Val Arg Glu Val
 210 215 220
 Leu Glu Thr Lys Gly Cys Gln Phe Lys Leu Gly Cys Glu Val Gln Ser
 225 230 235 240
 Val Leu Pro Val Asp Asn Gly Thr Ala Met Val Cys Gly Asp Gly Phe
 245 250 255
 Gln Glu Thr Tyr Asn Gly Cys Ile Met Ala Val Asp Ala Pro Thr Ala
 260 265 270
 Leu Lys Leu Leu Gly Asn Gln Ala Thr Phe Glu Glu Thr Arg Val Leu
 275 280 285
 Gly Ala Phe Gln Tyr Ala Thr Ser Asp Ile Phe Leu His Gln Asp Ser
 290 295 300
 Thr Leu Met Pro Gln Asn Lys Ser Ala Trp Ser Ala Leu Asn Phe Leu
 305 310 315 320
 Asn Ser Ser Lys Asn Asn Ala Phe Leu Thr Tyr Trp Leu Asn Ala Leu
 325 330 335
 Gln Asn Ile Gly Lys Thr Ser Glu Pro Phe Phe Val Thr Val Asn Pro
 340 345 350
 Asp His Thr Pro Lys Asn Thr Leu Leu Lys Trp Ser Thr Gly His Ala
 355 360 365
 Ile Xaa Ser Val Ala Ala Ser Lys Ala Ser Leu Glu Leu Gly Gln Ile
 370 375 380
 Gln Gly Lys Arg Gly Ile Trp Phe Cys Gly Tyr Asp Phe Asn Gln Asp
 385 390 395 400
 Glu Leu Lys Ala Gly Met Asp Ala Ala His Gly Ile Leu Gly Lys His
 405 410 415

ES 2 603 530 T3

Ser Ser Val Pro Pro Ser Pro Lys Asn Met Ser Pro Ser Leu Pro Lys
420 425 430

Asn Met Ser Pro Ser Phe Met Glu Thr Thr Ala Arg Leu Phe Val Thr
435 440 445

Lys Phe Phe Gln Gln Tyr Ile Ser Met Gly Cys Val Ile Phe Leu Glu
450 455 460

Glu Gly Gly Arg Ile Phe Thr Phe Lys Gly Asn Met Glu Lys Cys Pro
465 470 475 480

Leu Lys Thr Val Leu Lys Val His Asn Pro Gln Phe Tyr Trp Arg Ile
485 490 495

Met Lys Glu Ala Asp Ile Gly Leu Ala Asp Ala Tyr Ile His Gly Asp
500 505 510

Phe Ser Phe Leu Asp Glu Asn Glu Gly Leu Leu Asn Leu Phe Arg Ile
515 520 525

Leu Val Ala Asn Lys Glu Asn Ser Ala Ala Ser Gly Ser Thr Lys Arg
530 535 540

Arg Thr Trp Trp Ser Pro Ala Leu Leu Thr Ala Ser Ile Ser Ser Ala
545 550 555 560

Lys Tyr Phe Val Lys His Leu Leu Arg Gln Asn Thr Ile Thr Gln Ala
565 570 575

Arg Arg Asn Ile Ser Arg His Tyr Asp Leu Ser Asn Glu Leu Phe Ser
580 585 590

Leu Tyr Leu Gly Lys Met Met Gln Tyr Ser Ser Gly Val Phe Arg Thr
595 600 605

Gly Glu Glu His Leu Asp Val Ala Gln Arg Arg Lys Ile Ser Ser Leu
610 615 620

Ile Glu Lys Thr Arg Ile Glu Lys Trp His Glu Val Leu Asp Ile Gly
625 630 635 640

Cys Gly Trp Gly Ser Leu Ala Ile Glu Thr Val Lys Arg Thr Gly Cys
645 650 655

ES 2 603 530 T3

Lys Tyr Thr Gly Ile Thr Leu Ser Glu Gln Gln Leu Lys Tyr Ala Gln
660 665 670

Glu Lys Val Lys Glu Ala Gly Leu Glu Asp Asn Ile Lys Ile Leu Leu
675 680 685

Cys Asp Tyr Arg Gln Leu Pro Lys Glu His Gln Phe Asp Arg Ile Ile
690 695 700

Ser Val Glu Met Val Glu His Val Gly Glu Glu Tyr Ile Glu Glu Phe
705 710 715 720

Tyr Arg Cys Cys Asp Gln Leu Leu Lys Glu Asp Gly Leu Phe Val Leu
725 730 735

Gln Phe Ile Ser Ile Pro Glu Glu Leu Ser Lys Glu Ile Gln Gln Thr
740 745 750

Ala Gly Phe Leu Lys Glu Tyr Ile Phe Pro Gly Gly Thr Leu Leu Ser
755 760 765

Leu Asp Arg Asn Leu Ser Ala Met Ala Ala Ala Thr Arg Phe Ser Val
770 775 780

Glu His Val Glu Asn Ile Gly Met Ser Tyr Tyr His Thr Leu Arg Trp
785 790 795 800

Trp Arg Lys Leu Phe Leu Lys Asn Thr Ser Lys Val Leu Ala Leu Gly
805 810 815

Phe Asp Glu Lys Phe Met Arg Thr Trp Glu Tyr Tyr Phe Asp Tyr Cys
820 825 830

Ala Ala Gly Phe Lys Thr Gly Thr Leu Ile Asp Tyr Gln Val Val Phe
835 840 845

Ser Arg Ala Gly Asn Phe Gly Thr Leu Gly Asp Pro Tyr Lys Gly Phe
850 855 860

Pro Ser Ala Tyr Ser Phe Met Asp Asp
865 870

<210> 12
<211> 2827
5 <212> ADN
<213> *Gossypium hirsutum*

<400> 12

gtcacggcgg cagtgatgga agtggcggtg atcggagggtg ggataaaagg gttggtttcg

60

10

ES 2 603 530 T3

gcctacgtac tggctaaagc cggcgtggac gtgggtggtt acgagaaaga agagcaatta 120
ggcggccatg cgaagactgt taacttcgac gccgttgact tagaccttg cttcttgttt 180
cttaatcctg caagatatgc aacctgttg gatataatcg acagccttg tgttgatgta 240
gaaacatccg atgtttcatt ctctataagc catgacaaaag gcaacaatgg ctatgaatgg 300
tgcagtcaat atggattttc caattacttt gcacaaaaga agaaactggt gaaccctttc 360
aattggcaaa accttagaga gatcatcaga ttcagcaacg atgtcgaaag ttaccttgga 420
tcacttgaga acaaccgaga cattgatcgt actgagacct tgggacagtt tataaaatca 480
aagggctact ctgaaaattt tcaaaacact tacctggctc ctatatgtgg ttcaatgtgg 540
tcaagctcca aggaagatgt tatgagcttt tcagcatttt ccatcctttc attttgccgt 600
actcatcatt tgtaccagca atttgggcag ccacagtggg tgactatcaa agggcactca 660
cattttgtta aaagggttag ggaagtgtg gagactaaag gttgtcaatt taaactcggg 720
tgtgaagtac aatctgtttt gcctgctgat aatggtacca ccatggctg tggagatggg 780
ttccaagaaa cttacaatgg atgcataatg gctgttgatg ctcccactgc cctaaaatta 840
ttaggaaacc aagcaacatt tgaagaaaca agagtactgg gtgctttcca atatgctacc 900
agtgatattt tccttcaccg ggacagtact ttaatgccac aaaacaaatc agcttggagt 960
gcattgaatt ttctcaatag tagcaaaaat aatgcattct taacatactg gctcaatgca 1020
ctacagaata ttgggaaaac aagtgaacca tttttgtga ctgtcaatcc agaccatacc 1080
ccgaagaata ccttgcttaa gtggctgact ggccatgcaa ttccctctgt tgctgcatca 1140
aaagctcac ttgagcttg tcagattcag gggagagag gaatctggtt ctgtggctat 1200
gacttcaatc aggatgaact aaaggctgg atggatgctg cacatggtat cttgggaaag 1260
cattctctg ttctgcatag tccaaagagt atgtcacctt ctttcatgga aacaacggca 1320
cgectcttg ttactaaatt ctttcaacaa tatatatcta tgggctgtgt aattttctta 1380
gaggaaggag gcagaatttt cactttcaa ggaacatgg aaaagtgtcc tcttaaaca 1440
gttctgaaag tacataatcc tcagttttac tggaggatca tgaaagaagc tgatataggc 1500
cttgcagatg catatatcca tggagatttt tctttcttg atgaaactga aggccttctt 1560
aatcttttcc ggattcttgt tgccaataaa gagaactcag ctgcctcagg gtcgaataaa 1620
agaaggactt ggtggtcacc tgctctgtta acagctagta tatcatctgc aaagtatttt 1680
gtgaagcatc tcttgagaca aaatactatt acacaagctc gtaggaacat ttctcgtcat 1740
tatgatctga gtaatgaact tttcactcta tacttgggca aatgatgca atactcttct 1800
ggagtcttta ggacgggaga agaacatttg gacgttgca acgctagaaa aatcagttct 1860
ctaattgaga aagcaaggat agagaaacgg cacgaagttc tcgacattgg gtgctggttg 1920

ES 2 603 530 T3

```

ggaagcttag ctattgaaac tgtgaaaaga acaggatgca aatatactgg catcactcta      1980
tcagaacagc aactgaaata tgctcaagaa aaagtgaagg aagctggact ccaggataac      2040
atcaaaatac ttctctgtga ctatcgccag ttacctaagg aacaccaatt tgacagaatc      2100
atatctgtag agatggtaga acatgttggg gaagaatata ttgaggagtt ttacagatgc      2160
tgtgaccaat tactgaaaga agatgggctt tttgttcttc agttcatatc tatcccagaa      2220
gagctttcca aagaaatcca gcaaacagca ggttttctaa aggaatatat attccctgga      2280
ggaaccctgc tttctttgga taggaattta tcagccatgg ctgctgcaac aagattcagt      2340
gtggagcatg tggaaaatat aggaatgagt tattaccaca cactgagatg gtggagaaaa      2400
cttttctcgg aaaacacaag caaagttcta gctctggggg tcgacgagaa gttcatgagg      2460
acatgggaat actatttcga ttactgogct gccggtttta agacaggaac tcttatagat      2520
taccaggttg tttttcaag ggccggaaat ttcggtacac tcggagatcc atacaaaggt      2580
ttccctctcg catattcctt catggatgat tgaacaaagt gtggttgaac attgatccaa      2640
agaagcaaac aaaattatca ccacatgcca gtgtaagaa caacctatct ccctagtccc      2700
tacttttggt tattatggct atgtttgcaa tgcaagaata agcaaacatt gtaatgttaa      2760
taaagtttgc actttttag actggatgga tgttatcaat gaagtaccta gtttataaaa      2820
aaaaaaa                                     2827

```

<210> 13
 <211> 865
 <212> PRT
 <213> *Gossypium hirsutum*
 <400> 13

5

```

Met Glu Val Ala Val Ile Gly Gly Gly Ile Lys Gly Leu Val Ser Ala
 1                               5                               10                               15

Tyr Val Leu Val Lys Ala Gly Val Asp Val Val Val Tyr Glu Lys Glu
                               20                               25                               30

Glu Gln Leu Gly Gly His Ala Lys Thr Val Asn Phe Asp Ala Val Asp
                               35                               40                               45

Leu Asp Leu Gly Phe Leu Phe Leu Asn Pro Ala Arg Tyr Ala Thr Leu
 50                               55                               60

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

Ser Phe Ser Ile Ser His Asp Lys Gly Asn Asn Gly Tyr Glu Trp Cys
                               85                               90                               95

```

10

ES 2 603 530 T3

Ser Gln Tyr Gly Phe Ser Asn Tyr Phe Ala Gln Lys Lys Lys Leu Leu
100 105 110

Asn Pro Phe Asn Trp Gln Asn Leu Arg Glu Ile Ile Arg Phe Ser Asn
115 120 125

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

Arg Thr Glu Thr Leu Gly Gln Phe Ile Lys Ser Lys Gly Tyr Ser Glu
145 150 155 160

Asn Phe Gln Asn Thr Tyr Leu Ala Pro Ile Cys Gly Ser Met Trp Ser
165 170 175

Ser Ser Lys Glu Asp Val Met Ser Phe Ser Ala Phe Ser Ile Leu Ser
180 185 190

Phe Cys Arg Thr His His Leu Tyr Gln Gln Phe Gly Gln Pro Gln Trp
195 200 205

Leu Thr Ile Lys Gly His Ser His Phe Val Lys Arg Val Arg Glu Val
210 215 220

Leu Glu Thr Lys Gly Cys Gln Phe Lys Leu Gly Cys Glu Val Gln Ser
225 230 235 240

Val Leu Pro Ala Asp Asn Gly Thr Thr Met Val Cys Gly Asp Gly Phe
245 250 255

Gln Glu Thr Tyr Asn Gly Cys Ile Met Ala Val Asp Ala Pro Thr Ala
260 265 270

Leu Lys Leu Leu Gly Asn Gln Ala Thr Phe Glu Glu Thr Arg Val Leu
275 280 285

Gly Ala Phe Gln Tyr Ala Thr Ser Asp Ile Phe Leu His Arg Asp Ser
290 295 300

Thr Leu Met Pro Gln Asn Lys Ser Ala Trp Ser Ala Leu Asn Phe Leu
305 310 315 320

Asn Ser Ser Lys Asn Asn Ala Phe Leu Thr Tyr Trp Leu Asn Ala Leu
325 330 335

Gln Asn Ile Gly Lys Thr Ser Glu Pro Phe Phe Val Thr Val Asn Pro
340 345 350

ES 2 603 530 T3

Asp His Thr Pro Lys Asn Thr Leu Leu Lys Trp Ser Thr Gly His Ala
 355 360 365

Ile Pro Ser Val Ala Ala Ser Lys Ala Ser Leu Glu Leu Gly Gln Ile
 370 375 380

Gln Gly Lys Arg Gly Ile Trp Phe Cys Gly Tyr Asp Phe Asn Gln Asp
 385 390 395 400

Glu Leu Lys Ala Gly Met Asp Ala Ala His Gly Ile Leu Gly Lys His
 405 410 415

Ser Ser Val Leu His Ser Pro Lys Ser Met Ser Pro Ser Phe Met Glu
 420 425 430

Thr Thr Ala Arg Leu Phe Val Thr Lys Phe Phe Gln Gln Tyr Ile Ser
 435 440 445

Met Gly Cys Val Ile Phe Leu Glu Glu Gly Gly Arg Ile Phe Thr Phe
 450 455 460

Lys Gly Asn Met Glu Lys Cys Pro Leu Lys Thr Val Leu Lys Val His
 465 470 475 480

Asn Pro Gln Phe Tyr Trp Arg Ile Met Lys Glu Ala Asp Ile Gly Leu
 485 490 495

Ala Asp Ala Tyr Ile His Gly Asp Phe Ser Phe Leu Asp Glu Thr Glu
 500 505 510

Gly Leu Leu Asn Leu Phe Arg Ile Leu Val Ala Asn Lys Glu Asn Ser
 515 520 525

Ala Ala Ser Gly Ser Asn Lys Arg Arg Thr Trp Trp Ser Pro Ala Leu
 530 535 540

Leu Thr Ala Ser Ile Ser Ser Ala Lys Tyr Phe Val Lys His Leu Leu
 545 550 555 560

Arg Gln Asn Thr Ile Thr Gln Ala Arg Arg Asn Ile Ser Arg His Tyr
 565 570 575

Asp Leu Ser Asn Glu Leu Phe Thr Leu Tyr Leu Gly Lys Met Met Gln
 580 585 590

ES 2 603 530 T3

Tyr Ser Ser Gly Val Phe Arg Thr Gly Glu Glu His Leu Asp Val Ala
 595 600 605
 Gln Arg Arg Lys Ile Ser Ser Leu Ile Glu Lys Ala Arg Ile Glu Lys
 610 615 620
 Arg His Glu Val Leu Asp Ile Gly Cys Gly Trp Gly Ser Leu Ala Ile
 625 630 635 640
 Glu Thr Val Lys Arg Thr Gly Cys Lys Tyr Thr Gly Ile Thr Leu Ser
 645 650 655
 Glu Gln Gln Leu Lys Tyr Ala Gln Glu Lys Val Lys Glu Ala Gly Leu
 660 665 670
 Gln Asp Asn Ile Lys Ile Leu Leu Cys Asp Tyr Arg Gln Leu Pro Lys
 675 680 685
 Glu His Gln Phe Asp Arg Ile Ile Ser Val Glu Met Val Glu His Val
 690 695 700
 Gly Glu Glu Tyr Ile Glu Glu Phe Tyr Arg Cys Cys Asp Gln Leu Leu
 705 710 715 720
 Lys Glu Asp Gly Leu Phe Val Leu Gln Phe Ile Ser Ile Pro Glu Glu
 725 730 735
 Leu Ser Lys Glu Ile Gln Gln Thr Ala Gly Phe Leu Lys Glu Tyr Ile
 740 745 750
 Phe Pro Gly Gly Thr Leu Leu Ser Leu Asp Arg Asn Leu Ser Ala Met
 755 760 765
 Ala Ala Ala Thr Arg Phe Ser Val Glu His Val Glu Asn Ile Gly Met
 770 775 780
 Ser Tyr Tyr His Thr Leu Arg Trp Trp Arg Lys Leu Phe Leu Glu Asn
 785 790 795 800
 Thr Ser Lys Val Leu Ala Leu Gly Phe Asp Glu Lys Phe Met Arg Thr
 805 810 815
 Trp Glu Tyr Tyr Phe Asp Tyr Cys Ala Ala Gly Phe Lys Thr Gly Thr
 820 825 830
 Leu Ile Asp Tyr Gln Val Val Phe Ser Arg Ala Gly Asn Phe Gly Thr
 835 840 845

ES 2 603 530 T3

Leu Gly Asp Pro Tyr Lys Gly Phe Pro Ser Ala Tyr Ser Phe Met Asp
 850 855 860

Asp
 865

5 <210> 14
 <211> 2912
 <212> ADN
 <213> *Gossypium hirsutum*

<400> 14

```

    tccctatctc cattaactat tttcttctct cttcttcttc tttcgaacca ttttcagagt      60
    tcataaattc agggttttgt tttttttttg ggtgtagtga aataaaggat gaaaatagca      120
    gtgataggag gagggataag tggggtggtg tcagcctata ctttagccaa agccggtgca      180
    aatgtagtgc tttacgagaa agaagagtat ttgggaggcc attccaagac cgttcacttc      240
    gatgggtgtg atttagacct tggtttcatg gtttttaatc gcgttacata tccaaatag      300
    atggagttgt ttgagagcct tgggattgat atggaacat ttgatatgtc actctcagtg      360
    agccttaatg aaggcaaagg ctgtgaatgg ggcagccgta atggcctttc ggcttgttt      420
    gcccaaaaat ccaacctctt caatccttac ttttgcaaaa tgcttagaga aattctcaaa      480
    ttcaagaatg atgttattag ttatcttgaa ttgctcgaaa acaaccgga tattgaccgt      540
    aatgaaacat tgggacagtt cataaaatca aagggttact ctgatttatt tcagaaggct      600
    tatctgggtg ctgtatgtgg ttcaatatgg tcatgcccta cagaaagagt tatggatttt      660
    tcagctttct ctattcttct attttgccgc aatcatcacc tacttcagat ctttgacga      720
    ccacagtgga tgaccgttcg atggcgttca catcgttacg tcaataaggt tagagaagag      780
    ctggagagta caggttgtca aataagaact ggttgcgagg tgcatctctg tttgagtgat      840
    gctgaagggt gcactgtatt atgtggagat gactctcacg agttatatca agggcgcata      900
    atggctgttc atgcaccata tgctttgaga ttgttaggga atcaagcaac atatgatgaa      960
    tcaacagtgc ttggcgcttt ccaatatgtc tatagtgata tttatcttca tcgtgacaaa     1020
    aatttaatgc caaaaaacc agcagcatgg agtgcattga attttcttgg aagtacagac     1080
    aagaatgtat ctttgacata ctggcttaat gtgcttcaga atctaggaga aacaagccta     1140
    cccttttttg tactctcaa tccagattat acacaaaaac acaccttgot taagtggaga     1200
    acaggccatc cagtaccatc tgttctgca aaaaagctt ctcttgagct tgatcggatt     1260
    caaggaaga gaggaatttg gttttgtgga gcataacctg gctatggctt ccatgaagat     1320
    ggattaaagg ctgggatgat tgctgcaaac ggtctgctgg gaaaaagttg taatattctg     1380
    agcaatccaa agcatatggt gcctctctg atggaacag gggcacgtct tttgttact     1440
    
```

10

ES 2 603 530 T3

agattcctca gtcattttat atcaaccggc tgtgtgattt tattggaaga aggtggcact 1500
 atgtttacct ttgaaggaac tagcaataag tgttctctaa aaactgtaat taaagttcac 1560
 agtccacatt tttattggaa ggttatgaca gaggcagatt taggccttgc agattcatat 1620
 atcaatgggg atttttcttt tgttgataaa aaagacggtc tgctgaacct tgtaatgatt 1680
 cttattgcca acagagattt gatttcttcc aactcaaaac ttagtaagaa aaggggttgg 1740
 tggacaccat tgttgtttac agctggctca acatcagcaa agtatttctt caagcatgtc 1800
 ttaagacaaa atactcttac acaagctcgt aggaacattt ctcgccatta cgacytgagt 1860
 aatgaccttt ttgcaactct cttggatgag acaatgacat actcttgtgc agtatttaag 1920
 acagaagatg aggatttgaa agatgcacaa cacagaaaga tctctctttt gattgaaaaa 1980
 gcaagaattg atagcaagca tgaatttctt gagattggat gtggttgggk aagcttagct 2040
 attgagggtg tcaaacgaac tggatgcaaa tataccggca ttactttatc cgaagagcaa 2100
 ctcaaacttg cagaaaaaag agtgaaggaa gctggacttc aggaaaaat aagatttcaa 2160
 ctctgtgact atcgacaact acctagcacc tacaagtatg acagaattat atcgtgtgag 2220
 atgatagaag ctgttggcca tgaatacatg gaggacttct tcggttgctg tgaatcagtg 2280
 ttagcagatg atggacttct tgttttacag ttcatatcaa taccagagga acggtacaat 2340
 gaatacaggc gaagctcgga tttcatcaag gaatacatct tccttgggtg atgettacct 2400
 tctctggcta ggataacaac agccatgaat gctgcgtcca aactctgtgt ggagcatgtg 2460
 gaaaacatcg gacttcatta ctaccaaacy cttagatatt ggagaaagaa tttcttggag 2520
 aaacagagca aaatccatgc cttgggattc aatgacaagt tcatccggac atgggaatac 2580
 tattttgatt attgtgctgc tggtttcaag tccaatactc ttggtaatta ccaggttga 2640
 ttttctcggc ctggaaatgt agttgcactt ggcaaccat acaaagactt ccctcagct 2700
 tcttaattat ttattttctc cttatttcaa tcgtaccata gccataattt gagcttgttg 2760
 aaaactgatg ctacacgttt ggtttcatc aaatatggta tttgagtgca tatctataca 2820
 ttgatgaatg taattctggc ttgcctcgtg ggaacttgcc agcaggatta tctttttaca 2880
 tggacattta ttttaattct ctgttcaaat tt 2912

<210> 15
 <211> 865
 <212> PRT
 <213> *Gossypium hirsutum*

5

<220>
 <221> misc_feature
 <222> (641)..(641)
 <223> Xaa puede ser cualquier aminoácido de origen natural

10

<400> 15

ES 2 603 530 T3

Met Lys Ile Ala Val Ile Gly Gly Gly Ile Ser Gly Val Val Ser Ala
1 5 10 15

Tyr Thr Leu Ala Lys Ala Gly Ala Asn Val Val Leu Tyr Glu Lys Glu
20 25 30

Glu Tyr Leu Gly Gly His Ser Lys Thr Val His Phe Asp Gly Val Asp
35 40 45

Leu Asp Leu Gly Phe Met Val Phe Asn Arg Val Thr Tyr Pro Asn Met
50 55 60

Met Glu Leu Phe Glu Ser Leu Gly Ile Asp Met Glu Pro Phe Asp Met
65 70 75 80

Ser Leu Ser Val Ser Leu Asn Glu Gly Lys Gly Cys Glu Trp Gly Ser
85 90 95

Arg Asn Gly Leu Ser Ala Leu Phe Ala Gln Lys Ser Asn Leu Phe Asn
100 105 110

Pro Tyr Phe Trp Gln Met Leu Arg Glu Ile Leu Lys Phe Lys Asn Asp
115 120 125

Val Ile Ser Tyr Leu Glu Leu Leu Glu Asn Asn Pro Asp Ile Asp Arg
130 135 140

Asn Glu Thr Leu Gly Gln Phe Ile Lys Ser Lys Gly Tyr Ser Asp Leu
145 150 155 160

Phe Gln Lys Ala Tyr Leu Val Pro Val Cys Gly Ser Ile Trp Ser Cys
165 170 175

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

Cys Arg Asn His His Leu Leu Gln Ile Phe Gly Arg Pro Gln Trp Met
195 200 205

Thr Val Arg Trp Arg Ser His Arg Tyr Val Asn Lys Val Arg Glu Glu
210 215 220

Leu Glu Ser Thr Gly Cys Gln Ile Arg Thr Gly Cys Glu Val His Ser
225 230 235 240

Val Leu Ser Asp Ala Glu Gly Cys Thr Val Leu Cys Gly Asp Asp Ser
245 250 255

ES 2 603 530 T3

His Glu Leu Tyr Gln Gly Cys Ile Met Ala Val His Ala Pro Tyr Ala
 260 265 270

Leu Arg Leu Leu Gly Asn Gln Ala Thr Tyr Asp Glu Ser Thr Val Leu
 275 280 285

Gly Ala Phe Gln Tyr Val Tyr Ser Asp Ile Tyr Leu His Arg Asp Lys
 290 295 300

Asn Leu Met Pro Lys Asn Pro Ala Ala Trp Ser Ala Trp Asn Phe Leu
 305 310 315 320

Gly Ser Thr Asp Lys Asn Val Ser Leu Thr Tyr Trp Leu Asn Val Leu
 325 330 335

Gln Asn Leu Gly Glu Thr Ser Leu Pro Phe Leu Val Thr Leu Asn Pro
 340 345 350

Asp Tyr Thr Pro Lys His Thr Leu Leu Lys Trp Arg Thr Gly His Pro
 355 360 365

Val Pro Ser Val Ala Ala Thr Lys Ala Ser Leu Glu Leu Asp Arg Ile
 370 375 380

Gln Gly Lys Arg Gly Ile Trp Phe Cys Gly Ala Tyr Leu Gly Tyr Gly
 385 390 395 400

Phe His Glu Asp Gly Leu Lys Ala Gly Met Ile Ala Ala Asn Gly Leu
 405 410 415

Leu Gly Lys Ser Cys Asn Ile Leu Ser Asn Pro Lys His Met Val Pro
 420 425 430

Ser Leu Met Glu Thr Gly Ala Arg Leu Phe Val Thr Arg Phe Leu Ser
 435 440 445

His Phe Ile Ser Thr Gly Cys Val Ile Leu Leu Glu Glu Gly Gly Thr
 450 455 460

Met Phe Thr Phe Glu Gly Thr Ser Asn Lys Cys Ser Leu Lys Thr Val
 465 470 475 480

Ile Lys Val His Ser Pro His Phe Tyr Trp Lys Val Met Thr Glu Ala
 485 490 495

ES 2 603 530 T3

Asp Leu Gly Leu Ala Asp Ser Tyr Ile Asn Gly Asp Phe Ser Phe Val
 500 505 510
 Asp Lys Lys Asp Gly Leu Leu Asn Leu Val Met Ile Leu Ile Ala Asn
 515 520 525
 Arg Asp Leu Ile Ser Ser Asn Ser Lys Leu Ser Lys Lys Arg Gly Trp
 530 535 540
 Trp Thr Pro Leu Leu Phe Thr Ala Gly Leu Thr Ser Ala Lys Tyr Phe
 545 550 555 560
 Phe Lys His Val Leu Arg Gln Asn Thr Leu Thr Gln Ala Arg Arg Asn
 565 570 575
 Ile Ser Arg His Tyr Asp Leu Ser Asn Asp Leu Phe Ala Leu Phe Leu
 580 585 590
 Asp Glu Thr Met Thr Tyr Ser Cys Ala Val Phe Lys Thr Glu Asp Glu
 595 600 605
 Asp Leu Lys Asp Ala Gln His Arg Lys Ile Ser Leu Leu Ile Glu Lys
 610 615 620
 Ala Arg Ile Asp Ser Lys His Glu Ile Leu Glu Ile Gly Cys Gly Trp
 625 630 635 640
 Xaa Ser Leu Ala Ile Glu Val Val Lys Arg Thr Gly Cys Lys Tyr Thr
 645 650 655
 Gly Ile Thr Leu Ser Glu Glu Gln Leu Lys Leu Ala Glu Lys Arg Val
 660 665 670
 Lys Glu Ala Gly Leu Gln Glu Asn Ile Arg Phe Gln Leu Cys Asp Tyr
 675 680 685
 Arg Gln Leu Pro Ser Thr Tyr Lys Tyr Asp Arg Ile Ile Ser Cys Glu
 690 695 700
 Met Ile Glu Ala Val Gly His Glu Tyr Met Glu Asp Phe Phe Gly Cys
 705 710 715 720
 Cys Glu Ser Val Leu Ala Asp Asp Gly Leu Leu Val Leu Gln Phe Ile
 725 730 735
 Ser Ile Pro Glu Glu Arg Tyr Asn Glu Tyr Arg Arg Ser Ser Asp Phe
 740 745 750

ES 2 603 530 T3

Ile Lys Glu Tyr Ile Phe Pro Gly Gly Cys Leu Pro Ser Leu Ala Arg
 755 760 765

Ile Thr Thr Ala Met Asn Ala Ala Ser Lys Leu Cys Val Glu His Val
 770 775 780

Glu Asn Ile Gly Leu His Tyr Tyr Gln Thr Leu Arg Tyr Trp Arg Lys
 785 790 795 800

Asn Phe Leu Glu Lys Gln Ser Lys Ile His Ala Leu Gly Phe Asn Asp
 805 810 815

Lys Phe Ile Arg Thr Trp Glu Tyr Tyr Phe Asp Tyr Cys Ala Ala Gly
 820 825 830

Phe Lys Ser Asn Thr Leu Gly Asn Tyr Gln Val Val Phe Ser Arg Pro
 835 840 845

Gly Asn Val Val Ala Leu Gly Asn Pro Tyr Lys Asp Phe Pro Ser Ala
 850 855 860

Ser
 865

<210> 16
 <211> 1771
 <212> ADN
 <213> Artificial

<220>
 <223> Artificial

<400> 16

gaattctcta gaaaagttaa cccttcgaag atgatactga cattaacacc attttttaat 60
 attgtttttc tatatcgta ttgatctcag cacattctta gaaagatatt taaattagat 120
 aaaagtaa atatatatat atatatatatat atatatatatat ataaatgtaa 180
 cataaatcta tgggtcaatta caatatttaa ttaaataaaa tagaaatata aacaccactt 240
 taatttgact cggatacatg catccataaa gactacaaaa ggcaaaaaga gaaggaaatg 300
 agatacgaat atatgtcata agtatatata ggtgacaagg gcaaattaa taggttggtg 360
 tttaaatgca aaatcctatg tttgataaag aatgggatga aaaacaggca aagttaattg 420
 caattcaaag gtgaacaaag catttctttg tctacactaa tggcatgtct aagtaaatta 480
 ttagtcttgt atctatatgt ccacaagtta ttaattagtc ttatactatc aaaaacaagt 540
 taagttgcaa atcaaacatg aacaaagcat ttgtgttgta acctacgaaa aaataccta 600
 acatactgat acgaataatg tggcctaaat tgatcgttta ccaaattacg gtgctggaaa 660

ES 2 603 530 T3

aaaaaattgc tcctttacca acaaaattaa gaactgatac atcttgtttt ttgtcactga 720
 agataaacac gtgatctttg gcaaaacata aaggccaaca aaacaaactt gtctcatccc 780
 tgaatgatto gaatgccatc gtatgctgtt cacaaagtgg aatacagcaa tgaacaaatg 840
 ctatcctctt gagaaaagtg aatgcagcag cagcagcaga ctagagtgct acaaatgctt 900
 atcctcttga gaaaagtgaa tgcagcggca gcagacctga gtgctatata caattagaca 960
 cagggctctat taattgaaat tgtcttatta ttaaataatt cgttttatat taatttttta 1020
 aattttaatt aaatttatat atattatatt taagacagat atatttattt gtgattataa 1080
 atgtgtcact tttcttttta gtccatgtat tctctatttt tttcaattta actttttatt 1140
 tttattttta agtcaactctt gatcaagaaa acattggtga cataaaacta ttaacataaa 1200
 attatgttaa catgtgataa catcatattt tactaatata acgtcgcatt ttaacgtttt 1260
 tttaacaaat atcgactgta agagtaaaaa tgaatgttt gaaaaggtta attgcatact 1320
 aactattttt tttcctataa gtaatctttt ttgggatcaa ttgtatatca ttgagatacg 1380
 atattaaata tgggtacctt ttcacaaaac ctaacccttg ttagtcaaac cacacataag 1440
 agaggatgga tttaaaccag tcagcaccgt aagtatatag tgaagaaggc tgataacaca 1500
 ctctattatt gttagtacgt acgtatttcc ttttttgttt agtttttgaa ttaattaat 1560
 taaaatata atgctaacaa cattaaattt taaatttacg tctaattata tattgtgatg 1620
 tataataaat tgtcaacctt taaaaattat aaaagaata ttaattttga taaacaactt 1680
 ttgaaaagta cccaataatg ctagtataaa taggggcatg actccccatg catcacagtg 1740
 caatttaact gaagcaaage agcggccccg g 1771

<210> 17
 <211> 1188
 <212> ADN
 <213> Artificial

5

<220>
 <223> Artificial

10

<400> 17

accaacacgc cttctttgcc tcgtgtttca tcacctggcg ttaaactgct ttctttaaaa 60
 ccaacaaaat ggggtgccggg tgggtaggat gccaatgac gggatataag gaggaaaatc 120
 gaggetcggg caatcgagtt ccgatcgaga agcctccggt tacgctcggg cagatcaage 180
 aagccattcc gccccactgt tttgcgcgct cctctcttcg atccttctcc tacgtggctc 240
 atgacctatg cttagcctct ctcttttact acattgcaac atcatatttt cactttctcc 300
 cacaacctt ttcctacatt gcttggcctg tctattgggt tctccaagggt tggtagcatg 360
 gtctgtggag atggtttcca agaaacttac aatggatgca taatggctgt tgatgctccc 420

ES 2 603 530 T3

actgccctaa aattattagg aaaccaagca acatttgaag aaacaagagt actgggtgct 480
 ttccaatatg ctaccagtga tattttcctt caccgggaca gtactttaat gccacaaaac 540
 aatcagctt ggagtgcatt gaattttctc aatagtagca aaaataatgc attcttaaca 600
 tactggctca atgcaactaca gaatattggg aaaacaagtg agccattttt tgtgactgtc 660
 aatccagacc ataccccgaa gaataccttg ctttaagtggc cgactggcca tgcaattccc 720
 tctgttgctg catcaaaaagc ttcacttgag cttggctcaga ttcaggggaa gagaggaatc 780
 tggttctgtg gctatgagag ctccctcggac cttfatcaat caattacctg attggagcat 840
 gcttcttgcc gctatcacia ccattttcct ggctgctgag aagcagtgga tgatgcttga 900
 ttggaagcca agggggcctg acatgctcat tgatcctttt ggtatagggg ggattgttca 960
 ggatggctct gttttccgtc aaaacttctc gattaggtct tatgagatag gtgctgatcg 1020
 tacggcatcc atagagacgc taatgaatca tttacaggaa accgagatta atcattgtaa 1080
 aagtgctgga ctgcttgag aaggttttgg tgctaccctc gagatgtgca agaagaacct 1140
 aatttgggtg gtcactcggg tgcaagttgt gtttgatcgg taccctac 1188

<210> 18
 <211> 1110
 <212> ADN
 <213> Artificial

5

<220>
 <223> Artificial

10

<400> 18

gtacatttct ctttaatttc cttttttttt catttcatgt ttttcatggt aatggtgcat 60
 tgaagtgata aatttgagtg aatgatgttt ggtatatctt cttagtaact gaccttttga 120
 aaatactagc atttttttta atatcaagtg aaagaagaag aagaatttcg ccatgcaaaa 180
 gctttttaag gctttttctt ttccttagat caaaatttat ttgtttactt atactgttct 240
 ttttaagccc aagaaagaag ccatggtttc aatttttgag agttttaaat cccaaatacc 300
 agagagcttc atcgtttatt catatatttt taaacatttt ttaaagcaag aacttgatgat 360
 ttgtttttaa taaaatatgc aataaatttt tatatttttc gtaaatttaa aatttaattt 420
 ttctactttt aaaatttaaa aaagtaaatt ttaaaatata cctttcatta aattaaatta 480
 ttataagtaa ttgagtattt ttaattttta aatttcacac atcaaattaa aaaaaagtt 540
 aacacttgca cttgattttg aaaagtaaaa ggattaaatt tcaaattttc agtaaaagga 600
 ctaaatttca aatttttaaa gagtatagag actcctctac atttttagatt ttaaaattta 660
 aatctaacag ttaacacttt ctttaattact ttacgataaa ttaactaaa aaattacaat 720
 attaatggtt aaaattaaat tttgaaaagt ataaagatta aattgtaaatt tttcaaaaag 780
 cataggaagt tatagtatat tttaaccttt atttatttta tatctggtga ggttccctgca 840

ES 2 603 530 T3

tgcaccgaag atgtcacctt ttgccagtat tttccagtgg cttgtttctc tcaaaactac 900
 cttgaatctt gagacagaat taaatatatt tttggccttc tcttcatttt ctctctctct 960
 attttctttt aaaaattgct ttagagaatt cagaaaaaat actttccaac acgaaaattt 1020
 cttcaaattt attgtttata tctaataaat ggttgcttaa ttttgaaaaa caaaagttat 1080
 tgtagttagt tttgcttctt gcgtgtccag 1110

5 <210> 19
 <211> 1188
 <212> ADN
 <213> Artificial

10 <220>
 <223> Artificial
 <400> 19

gtaggatacc gatcaaacac aacttgcatc cgagtgacca cccaaattag gttcttcttg 60
 cacatctcag gggtagcacc aaaaccttct ccaagcagtc cagcactttt acaatgatta 120
 atcgcggttt cctgtaaagt attcattagc gtctctatgg atgccgtacg atcagcacct 180
 atctcataag acctaatacg gaagttttga cggaaaacaa gaccatcctg aacaatcctc 240
 cctataccaa aaggatcaat gagcatgtca ggccgccttg gcttccaatc aagcatcatc 300
 cactgcttct cagcagccag gaaaatgggt gtgatagcgg caagaagcat gctccaatca 360
 ggtaattgat tgataaaggc ccgaggagct ctcatagcca cagaaccaga ttcctctctt 420
 cccctgaatc tgaccaagct caagtgaagc ttttgatgca gcaacagagg gaattgcatg 480
 gccagtcgac cacttaagca aggtattctt cgggggatgg tctggattga cagtcacaaa 540
 aatggctca cttgttttcc caatattctg tagtgattg agccagtatg ttaagaatgc 600
 attatttttg ctactattga gaaaattcaa tgcactcaa gctgatttgt tttgtggcat 660
 taaagtactg tcccggtgaa ggaaaatata actggtagca tattggaaag caccagctac 720
 tctgtttct tcaaatgttg cttggtttcc taataattt agggcagtgg gagcatcaac 780
 agccattatg catccattgt aagtttcttg gaaaccatct ccacagacca tggtagcaac 840
 cttggagaac ccaatagaca ggccaagcaa tgtaggaaaa gggttgtggg agaaagtga 900
 aatatgatgt tgcaatgtag taaaagagag aggctaagca taggtcatgg accacgtagg 960
 agaaggatcg aaggagggag cggcgaaaac agtggggcgg aatggcttgc ttgatctgac 1020
 cgagcgtaaa cggaggcttc tcgatcggaa ctcgattgac cgagcctcga ttttctctct 1080
 ttatacccgct caattggcat cctaccacc cggcacccat tttgttggt ttaaagaaag 1140
 cagtttaacg ccaggtgatg aaacacgagg caaagaaggc gtgttggt 1188

15 <210> 20
 <211> 358
 <212> ADN
 <213> Artificial

ES 2 603 530 T3

<220>
<223> Artificial

<400> 20

5

```

aggggccgcc atgtgacaga togaaggaag aaagtgtaat aagacgactc tcactactcg      60
atcgctagtg attgtcattg ttatatataa taatgttatc tttcacaact tatcgtaatg      120
catgtgaaac tataacacat taatcctact tgtcatatga taacactctc cccatttaaa      180
actcttgca atttaaagat ataagattct ttaaatgatt aaaaaaata tattataaat      240
tcaatcactc ctactaataa attattaatt aatatttatt gattaaaaaa atacttatac      300
taatttagtc tgaatagaat aattagattc tagagtcgac ctgcaggcat gcaagctt      358
    
```

<210> 21
<211> 620
<212> ADN
<213> Artificial

10

<220>
<223> Artificial

15

<400> 21

```

tctagaggat cctatgttgt aattttatat ggattaatga gaattattat tattctgttc      60
ttcgtctgtg tttttaagc tagcttctc gaagcagcgt ataactttaa tttgaatttg      120
gttttgccgc gttagtgaag ttgcccgtgt aaacgtgtca agttgtgagt ggctgaaata      180
agataataga tatattatta ttgttttaat ttaattccgc gaagcgatat gttaagtgat      240
aatgaaacg aagcgttttg atgacgtcat atgtctccgt gcctacgtca gcacggggct      300
tagtattacc cccgtgccg gatcagagac atttgaccaa tagttgacta gtataatagc      360
ccttgatta aatgacacgt ggacgctcag gatctgtgat gctagtgaag cgcttaagct      420
gaacgaatct gacggaagag cgttcacact tagatctagt tagcgtactt agtacgcgtt      480
gtcttgggtc tataaataga gtgcttctga acagattggt cagaatttca tagcggcgat      540
aacaatgatt gaacaagatg gattgcacgc aggttctccg gccgcttggg tggagaggct      600
attcggctat gactgggcac      620
    
```

20 <210> 22
<211> 709
<212> ADN
<213> Artificial

25 <220>
<223> Artificial

<400> 22

ES 2 603 530 T3

aacagacaat cggctgctct gatgccgccc tgttccgget gtcagcgcag gggcgcccgg 60
 ttctttttgt caagaccgac ctgtccggtg ccctgaatga actgcaggta aatttctagt 120
 ttttctcett cattttcttg gttaggaccc ttttctcttt ttattttttt gagctttgat 180
 ctttctttaa actgatctat tttttaattg attggttatg gtgtaaatac tacatagctt 240
 taactgataa tctgattact ttatttcgtg tgtctatgat gatgatgata actgcaggac 300
 gaggcagcgc ggctatcgtg gctggccacg acgggegttc cttgcgcagc tgtgctcgac 360
 gttgtcactg aagcgggaag ggactggctg ctattgggag aagtgccggg gcaggatctc 420
 ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg 480
 ctgcatacgc ttgatccggc tacctgccca ttcgaccacc aagcgaacaac tcgcatcgag 540
 cgagcacgta ctccgatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat 600
 caggggctcg cggcagccga actgttcgcc aggctcaagg cgcgcatgcc cgacggcgag 660
 gatctcgtcg tgacccatgg cgatgcctgc ttgccgaata tcatggtgg 709

5 <210> 23
 <211> 549
 <212> ADN
 <213> Artificial

10 <220>
 <223> Artificial
 <400> 23

aaaatggccc cttttctgga ttcacgact gtggccggct ggggtgtggcg gaccgctatc 60
 aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa tgggctgacc 120
 gcttctcgtg gctttacggg atcgcgcctc ccgattcgca gcgcatcgcc ttctatcgcc 180
 ttcttgaaga gttcttctga gcgggactct ggggttcgaa atgaccgacc aagcgaagcc 240
 caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggc tgggcttcgg 300
 aatcgttttc cgggacgccc gctggatgat cctccagcgc ggggatctca tgctggagtt 360
 cttcgcccac cccgggtcga cggattgag tttatgtcta ttgtaattga tagaggttct 420
 attaagatag aattatgaga tgtaattgtg attaataaat aaagagttgt tattattctt 480
 tgaattactc cgcgaagcgg tgtgttatgt ttttgttggg agctttctag actcgactag 540
 agcggcccgc 549

15 <210> 24
 <211> 1386
 <212> ADN
 <213> *Gossypium hirsutum*

20 <400> 24

ES 2 603 530 T3

tgcttcgtgt ttcacccaacc tggcgttaaa ctgctttctt taaagccagc aaaatgggtg	60
ccgggtgtag gatgcccaatt gacgggataa aggaggaaaa tcgaggctcg gtcaatcgag	120
ttccgatcga gaagcctccg tttacgctcg gtcagatcaa gcaagccatt ccgccccact	180
gttttcgccc ctccctcctt cgatccttct cctacgtggt ccatgaccta tgcttagcct	240
ctttctttta ctacattgca acatcatatt ttcactttct cccacaacc ttttcctaca	300
ttgcttgccc tgtctattgg gttctccaag gttgcatcct caccgggtgt tgggtcatcg	360
cacacgagtg gggtcaccac gctttcagag actaccaatg gggtgacgac accgctcgggt	420
tgatcctca ttccgcccct ttagtcccgt acttctcgtg gaaaatcagt caccgccctc	480
accactcgaa caccgggtcc atggagcgtg acgaagtatt cgtgccc aaa cccaagteta	540
aattatcatg ctttgcgaaa tacttaaaaca atccaccgg tcgagttcta tctctttag	600
tcacattgac tcttggttg cctatgtact tagcctcaa cgtttcgggt cgatactatg	660
atcgattagc ttcccactat aacccttatg gccccattta ctccgatcgc gagaggctac	720
aagtttacat ctccgatact ggtatatttg cggtaattta tgtactttat aagattgctg	780
caacaaaagg gctggcttg cttttatgca cttatgggg gectctactt attgtgaatg	840
ccttccttgt gttgatcacc tacttgcaac atactcactc ggcattgccg cattatgact	900
cgccgaatg ggattggtg cgaggagcat tgtcagcagat ggatcgagat ttcgggggtg	960
tgaacaaagt gttccataac atcaccgata cgcattgtgc tcatcacctc ttctcaacga	1020
tgccacatta tcatgcaatg gaggccacta aagcaatcaa accaatactc ggcaagtatt	1080
atcctttcga cgggacaccg atttacaagg caatgtggag ggaggcaaaa gagtgccctt	1140
acgttgagcc tgacgttggg ggtggtggtg gtggtagcaa aggtgttttt tggtatcgta	1200
acaagttcta aagaccgacc aactgcctga tagctggccg gcgaaatcaa cgtaaaacgt	1260
acttattaga ctagtgttaa ctagggaagt taataattaa tggtaggaaa atgtggaata	1320
gttgcctagt agttttatgt attaagtgtt gtattaataa actatatggt agaaaaaaaa	1380
aaaaaa	1386

- <210> 25
- <211> 385
- <212> PRT
- <213> *Gossypium hirsutum*
- <400> 25

ES 2 603 530 T3

Met Gly Ala Gly Gly Arg Met Pro Ile Asp Gly Ile Lys Glu Glu Asn
1 5 10 15

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

Gly Gln Ile Lys Gln Ala Ile Pro Pro His Cys Phe Arg Arg Ser Leu
35 40 45

ES 2 603 530 T3

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

Phe Tyr Tyr Ile Ala Thr Ser Tyr Phe His Phe Leu Pro Gln Pro Phe
 65 70 75 80

Ser Tyr Ile Ala Trp Pro Val Tyr Trp Val Leu Gln Gly Cys Ile Leu
 85 90 95

Thr Gly Val Trp Val Ile Ala His Glu Trp Gly His His Ala Phe Arg
 100 105 110

Asp Tyr Gln Trp Val Asp Asp Thr Val Gly Leu Ile Leu His Ser Ala
 115 120 125

Leu Leu Val Pro Tyr Phe Ser Trp Lys Ile Ser His Arg Arg His His
 130 135 140

Ser Asn Thr Gly Ser Met Glu Arg Asp Glu Val Phe Val Pro Lys Pro
 145 150 155 160

Lys Ser Lys Leu Ser Cys Phe Ala Lys Tyr Leu Asn Asn Pro Pro Gly
 165 170 175

Arg Val Leu Ser Leu Val Val Thr Leu Thr Leu Gly Trp Pro Met Tyr
 180 185 190

Leu Ala Phe Asn Val Ser Gly Arg Tyr Tyr Asp Arg Leu Ala Ser His
 195 200 205

Tyr Asn Pro Tyr Gly Pro Ile Tyr Ser Asp Arg Glu Arg Leu Gln Val
 210 215 220

Tyr Ile Ser Asp Thr Gly Ile Phe Ala Val Ile Tyr Val Leu Tyr Lys
 225 230 235 240

Ile Ala Ala Thr Lys Gly Leu Ala Trp Leu Leu Cys Thr Tyr Gly Val
 245 250 255

Pro Leu Leu Ile Val Asn Ala Phe Leu Val Leu Ile Thr Tyr Leu Gln
 260 265 270

His Thr His Ser Ala Leu Pro His Tyr Asp Ser Ser Glu Trp Asp Trp
 275 280 285

ES 2 603 530 T3

Leu Arg Gly Ala Leu Ser Thr Met Asp Arg Asp Phe Gly Val Leu Asn
 290 295 300

Lys Val Phe His Asn Ile Thr Asp Thr His Val Ala His His Leu Phe
 305 310 315 320

Ser Thr Met Pro His Tyr His Ala Met Glu Ala Thr Lys Ala Ile Lys
 325 330 335

Pro Ile Leu Gly Lys Tyr Tyr Pro Phe Asp Gly Thr Pro Ile Tyr Lys
 340 345 350

Ala Met Trp Arg Glu Ala Lys Glu Cys Leu Tyr Val Glu Pro Asp Val
 355 360 365

Gly Gly Gly Gly Gly Gly Ser Lys Gly Val Phe Trp Tyr Arg Asn Lys
 370 375 380

Phe
 385

<210> 26
 <211> 1422
 <212> ADN
 <213> *Gossypium hirsutum*
 <400> 26

5

taaaaaaaaa aggcatttct ttcattcttaa agagacagcg aggaagccac gaagataata 60
 gagtgatttt caatctccat ttttaagggtg tggaacaatg ggtgctggag gcagaatgct 120
 ggttccaacg agtccaaaaa aaccggaatt caactcactg aagcgagttc cataactcaaa 180
 gccacccttc actctgagtg aaatcaagaa agccatccca ccacactggt tccagcgctc 240
 cgttttacgc tcattctcat atctccttta cgactttata ttggcctctc ttttttacca 300
 tgtggccacc aattacttcc ctaaccttcc tcaggctctc tccaacgtgg cttggcctct 360
 ttattgggcc atgcaagggt gcattttgac cggcgtttgg gtcatagcc atgaatgtgg 420
 ccaccatgct ttcagtgatt atcaatggct tgacgacacc gtgggcctta tcttccactc 480
 ttctctctta gttccatatt tctcttgga atatagccac cggcgtcacc attctaacac 540
 cggttccctc gaaagggatg aagtgttcgt tccaagaaa aaatctggt taagatgggt 600
 ggccaaacac ttcaacaatc caccgggtcg gtttctgtca atcaccattc aacttacct 660
 tggttggccg ctttacttag ctttcaacgt tgccggccgg ccttacgaca ggttcgcttg 720
 ccactatgac ccttacggcc ccatatttcc cgaccgggaa cgactccaaa tctatatctc 780
 tgacgccggc gtctctgctg tgcctatgc gctctaccgt ctctgtgttg ccaaaggggt 840
 aggttggggt attagcgttt atgggggtgc attattggtg gttaacgcct tcttagtaat 900

10

ES 2 603 530 T3

gatcacgtat ttgcaacaca ctcacccatc tttgccgcac tatgattcct cggagtggga 960
 ctggatgaga ggagctttat caactgtgga cagagattat gggattttaa acaaggtttt 1020
 ccataacata accgacactc atgtggctca tcatttgttt tcgacaatgc ctcactatca 1080
 tgccatggtg gccaccaagg cgataaagcc catattgggg gaatactatc agttcgatgg 1140
 gatgcctgtc tataaggcga tatggagôga ggcgaaggag tgtctctacg ttgaaccaga 1200
 tgagggcgac aaggataaag gtgtgttttg gtttagaaac aagctttaaa tatttgcatt 1260
 ttaccttagg catgttctag tcgttgatgt ttttaaggata ttttagccga cataactggg 1320
 tttccttttt gggacttttt agctttgtat ttgcagacaa taatcttggt cactattaa 1380
 taatggtaga aataaataca cagcatggat tggcaataaa aa 1422

<210> 27
 <211> 383
 <212> PRT
 <213> *Gossypium hirsutum*
 <400> 27

5

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

10

ES 2 603 530 T3

His Ser Asn Thr Gly Ser Leu Glu Arg Asp Glu Val Phe Val Pro Lys
 145 150 155 160

Lys Lys Ser Gly Leu Arg Trp Trp Ala Lys His Phe Asn Asn Pro Pro
 165 170 175

Gly Arg Phe Leu Ser Ile Thr Ile Gln Leu Thr Leu Gly Trp Pro Leu
 180 185 190

Tyr Leu Ala Phe Asn Val Ala Gly Arg Pro Tyr Asp Arg Phe Ala Cys
 195 200 205

His Tyr Asp Pro Tyr Gly Pro Ile Phe Ser Asp Arg Glu Arg Leu Gln
 210 215 220

Ile Tyr Ile Ser Asp Ala Gly Val Leu Ala Val Ala Tyr Ala Leu Tyr
 225 230 235 240

Arg Leu Val Leu Ala Lys Gly Val Gly Trp Val Ile Ser Val Tyr Gly
 245 250 255

Val Pro Leu Leu Val Val Asn Ala Phe Leu Val Met Ile Thr Tyr Leu
 260 265 270

Gln His Thr His Pro Ser Leu Pro His Tyr Asp Ser Ser Glu Trp Asp
 275 280 285

Trp Met Arg Gly Ala Leu Ser Thr Val Asp Arg Asp Tyr Gly Ile Leu
 290 295 300

Asn Lys Val Phe His Asn Ile Thr Asp Thr His Val Ala His His Leu
 305 310 315 320

Phe Ser Thr Met Pro His Tyr His Ala Met Val Ala Thr Lys Ala Ile
 325 330 335

Lys Pro Ile Leu Gly Glu Tyr Tyr Gln Phe Asp Gly Met Pro Val Tyr
 340 345 350

Lys Ala Ile Trp Arg Glu Ala Lys Glu Cys Leu Tyr Val Glu Pro Asp
 355 360 365

Glu Gly Asp Lys Asp Lys Gly Val Phe Trp Phe Arg Asn Lys Leu
 370 375 380

<210> 28
 <211> 1933
 <212> ADN
 <213> *Gossypium arboreum*

5

<400> 28

ES 2 603 530 T3

ccacttcgca gcaatattat tgcagttcct ggttggttac ctctgagttt tcaacttaaa 60
 atttcttggt tttcctcaag aaggaagaag atgttgcaaa tagctttcag ctctgattca 120
 tggctgttga ctgctagcaa ccagaaagat ggaatgttgt tcccagtagc tttgtcattt 180
 ttggtagcca tattgggaat ttcactgtgg cacgtatgga ccataaggaa gccaaagaaa 240
 gacatcgccc cattaccgcc ggtccccgt gggttgcaa tagtgggata tcttccatat 300
 cttggaactg ataactttca cttggtgttt acagatttggt ctgcagctta cggccccatc 360
 tacaagcttt ggctaggaaa caaattatgc gtagtcatta gctcggcacc actggcgaaa 420
 gaagtggttc gtgacaacga catcacattt tctgaaaggg atcctcccgt ttgtgcaaag 480
 attattacct ttggcctcaa tgatattgta tttgattctt acagtagtcc agattggaga 540
 atgaagagaa aagtgcctgg acgtgaaatg cttagccata gtagcattaa agcttgttat 600
 ggtctaagga ggaacaagt gcttaaaggc gtacaaaatg ttgctcaaag tgctggcaag 660
 ccaattgatt ttggtgaaac ggcattttta acatcaatca atgcgatgat gagcatgctg 720
 tggggtggca aacagggagg agagcggaaa gggcccgacg tttggggcca atttcgagat 780
 ctcataaccg aactaatggt gatacttggg aaaccaaacg tttctgatat tttcccggtg 840
 cttgcaaggt ttgacataca gggattggag aaggaaatga ctaaaatcgt taattctttc 900
 gataagcttt tcaactccat gattgaagaa agagagaact ttagcaacaa attgagcaaa 960
 gaagatggaa aactgaaac aaaagacttc ttgcagcttc tgttggaact caagcagaag 1020
 aacgatagcg gaatatcgat aacaatgaat caagtcaagg ccttgctcat ggacattgtg 1080
 gtcggtggaa ctgatacaac atcaaccatg atggaatgga caatggctga actaattgca 1140
 aatcctgaag caatgaaaaa ggtgaagcaa gaaatagacg atgttgtcgg ttcggatggc 1200
 gccgtcgatg agactcactt gcctaagttg cgctatctag atgctgcagt aaaggagacc 1260
 ttccgattgc acccaccgat gccactcctt gtaccccggt gcccgggcga ctcaagcaac 1320
 gttggtggct atagcgtacc aaagggcacc aggtcttctt taaacatttg gtgtattcag 1380
 aggatccac agctttggga aaatccttta gaattcaagc ctgagaggtt cttgactgat 1440
 catgagaagc tcgattattt aggaaaagat tcccgtaca tgccgtttgg ttctggaagg 1500
 agaatgtgtg ccggagtatc tctcggtgaa aagatgttgt attcctcctt ggcagcaatg 1560
 atccatgctt atgattggaa cttggccgac ggtgaagaaa atgacttgat tggcttattt 1620
 ggaattatta tgaagaaaaa gaagccttta attcttgttc ctacaccaag accatcaaat 1680
 ctccagcact atatgaagta actttactat tgtatttctt ttataccact ttattgcctc 1740
 tttgtcatgt ttaggcaaca attctaagta ataagtttgg ctatatggtg aacaataatg 1800
 tgtttattat acatcataag caatgagctc tttccgacc tagggcaata caatgatact 1860
 gtgtattaag tgaatcaac aaatctttta ttctaaaaaa aaaaaaaaaa aaaaaaaaaa 1920
 aaaaaaaaaa aaa 1933

ES 2 603 530 T3

<210> 29
<211> 20
<212> ADN
<213> Artificial
5
<220>
<223> Artificial

<400> 29
10 atggttgcta ctgctgtgac 20

<210> 30
<211> 21
<212> ADN
15 <213> Artificial

<220>
<223> Artificial
20
<400> 30
ctgtaaatga tcattagtg t 21

<210> 31
<211> 25
25 <212> ADN
<213> Artificial

<220>
<223> Artificial
30
<400> 31
cctgagaggt tcttgactga tcatg 25

<210> 32
<211> 27
35 <212> ADN
<213> Artificial

<220>
40 <223> Artificial

<400> 32
gcttatgatg tataataaac acattat 27

REIVINDICACIONES

1. Semilla de algodón que tiene una composición de ácido graso en su aceite, en la que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico, del 4 al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, y en la que la semilla de algodón es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son *ghFAD2-1*, *ghFatB-2* y *ghCPA-FAS-2*, en donde el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región codificante de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO: 5 y en donde el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región codificante de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO: 12.
2. La semilla de algodón de la reivindicación 1, en la que del 74 % al 86 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico o en la que del 4 % al 12 % es ácido linoleico.
3. La semilla de algodón de acuerdo con las reivindicaciones 1 o 2, en la que del 0,1 % al 0,5 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido graso de ciclopropano (CPA) o ácido graso de ciclopropeno (CPE) o una combinación de estos.
4. La semilla de algodón de la reivindicación 3, en la que el ácido graso de ciclopropano o de ciclopropeno es ácido malválico, ácido estercúlico, ácido dihidroestercúlico o cualquier combinación de dos de estos o de los tres.
5. Un método para producir aceite de semilla de algodón, que comprende las etapas de (i) obtener semilla de algodón de acuerdo con una cualquiera de las reivindicaciones 1 a 4, (ii) extraer el aceite de la semilla de algodón, y (iii) recuperar el aceite de semilla de algodón, en donde el aceite de semilla de algodón tiene una composición modificada de ácidos grasos tal que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico.
6. El método de la reivindicación 5, en el que la etapa (i) comprende además recoger la pelusa que comprende la semilla de algodón de una planta de algodón y/o desmotar la semilla de algodón de la pelusa que comprende la semilla de algodón, y/o en el que la etapa (ii) comprende triturar la semilla de algodón y en el que la etapa (iii) comprende purificar el aceite de semilla de algodón, y/o en donde el método no comprende una etapa de hidrogenación del aceite de semilla de algodón.
7. El método de la reivindicación 6, en el que la purificación del aceite comprende desgomar el aceite, decolorar el aceite, desodorizar el aceite, alterar el pH del aceite, hidrolizar el aceite o cualquier combinación de los mismos.
8. Aceite de semilla de algodón obtenible mediante extracción de la semilla de algodón de una cualquiera de las reivindicaciones 1 a 4, en el que del 72 % al 88 % del contenido total de ácidos grasos en el aceite de semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, en donde el ácido graso de ciclopropano (CPA) y/o el ácido graso de ciclopropeno (CPE) están presentes en dicho aceite de semilla de algodón y en donde la composición de ácidos grasos de dicho aceite no se ha modificado después de la extracción.
9. El aceite de semilla de algodón de la reivindicación 8, en el que menos del 0,5 % del contenido total de ácidos grasos es un ácido graso *trans* o en el que menos de 0,5 % del contenido total de ácidos grasos es un ácido graso ramificado en C9 o C10.
10. Una planta de algodón que es capaz de producir la semilla de algodón de una cualquiera de las reivindicaciones 1 a 4, en donde la planta de algodón es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son *ghFAD2-1*, *ghFatB-2* y *GhCPA-FAS-2* en donde el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:5 y en donde el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteína de la secuencia de ADNc mostrada en la SEQ ID NO:12.
11. Un método para producir la semilla de algodón de un cualquiera de las reivindicaciones 1 a 4, que comprende recolectar la pelusa de una planta de algodón de acuerdo con la reivindicación 10 y desmotar la pelusa, produciendo de este modo la semilla de algodón.
12. Un método para identificar la semilla de algodón que tiene una composición de ácido graso modificada en su aceite, que comprende (i) obtener semillas de algodón transgénicas, (ii) determinar la composición de ácidos grasos del aceite de la semilla de algodón y (iii) si del 72 % al 88 % del contenido total de ácidos grasos en el aceite de la semilla de algodón es ácido oleico, del 4 % al 10 % es ácido palmítico y del 4 % al 15 % es ácido linoleico, seleccionando la semilla de algodón, en donde la semilla de algodón de la etapa (i) es transgénica para una construcción genética que codifica una o más moléculas de ARN que inhiben la expresión de tres genes que son

- 5 ghFAD2-1, ghFatB-2 y ghCPA-FAS-2, en donde el gen ghFATB-2 es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:5 y en donde el gen ghCPA-FAS-2 es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de la proteína de la secuencia de ADNc mostrada en la SEQ ID NO:12.
13. Uso de la semilla de algodón de una cualquiera de las reivindicaciones 1 a 4, para producir aceite aislado o harina de semilla de algodón o pienso para animales.
- 10 14. Uso del aceite de semilla de algodón de las reivindicaciones 8 o 9 en la producción de un producto alimentario, preferentemente, para aumentar el contenido de ácido oleico o para reducir el contenido de ácido palmítico del producto alimentario o como combustible.
- 15 15. El uso de cualquiera de las reivindicaciones 13 a 14 para aumentar el nivel o la absorción de ácidos grasos monoinsaturados o reducir el nivel o la absorción de ácidos grasos saturados o ácidos grasos *trans* en el producto alimentario o por el animal.
- 20 16. Un método para preparar un producto alimentario, que comprende mezclar el aceite de semilla de algodón de las reivindicaciones 8 o 9 con otros ingredientes alimentarios.
- 25 17. Un método para preparar un producto alimentario frito, que comprende freír un producto alimentario en el aceite de semilla de algodón de las reivindicaciones 8 o 9.
- 30 18. Una molécula de ácido nucleico quimérico que codifica un ácido nucleico silenciador de genes que reduce la expresión génica de ghFAD2-1, ghFatB-2 y ghCPA-FAS-2 de algodón, en donde el gen *ghFATB-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteínas de la secuencia de ADNc mostrada en SEQ ID NO: 5 y en donde el gen *ghCPA-FAS-2* es una molécula de ácido nucleico que comprende nucleótidos que tienen una secuencia que tiene al menos un 90 % de identidad con la región de codificación de proteína de la secuencia de ADNc mostrada en la SEQ ID NO: 12.
19. Una célula huésped, una planta o un tejido o una célula de la misma que comprende la molécula de ácido nucleico quimérico de la reivindicación 18.

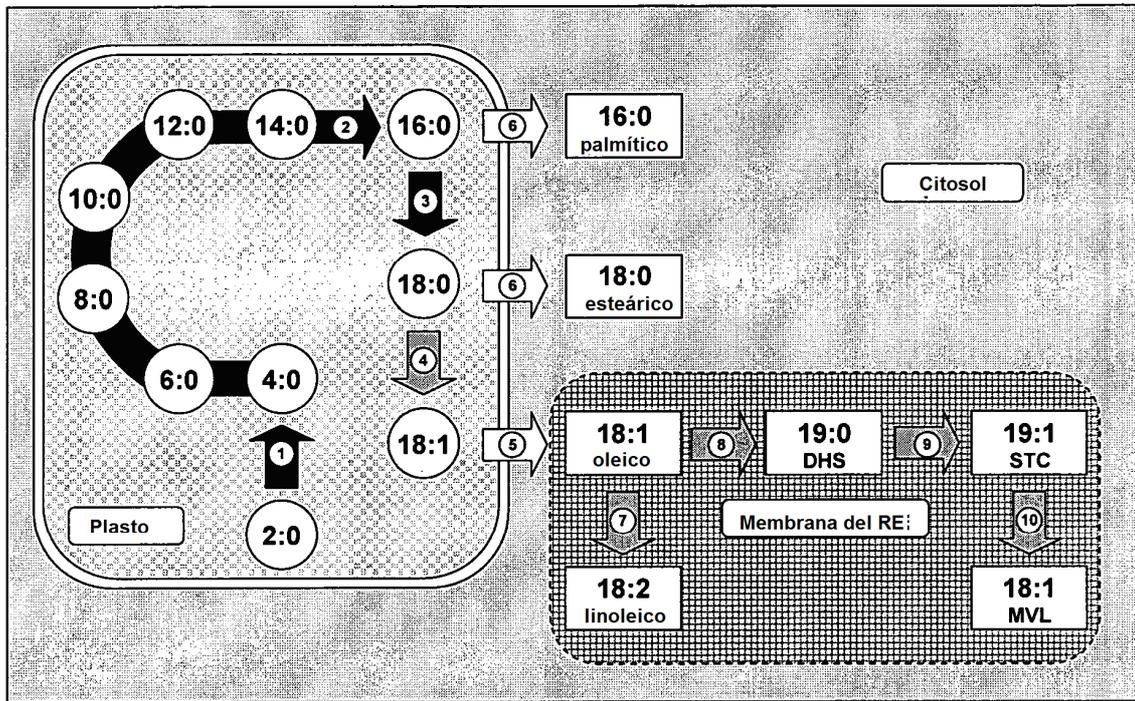


Figura 1

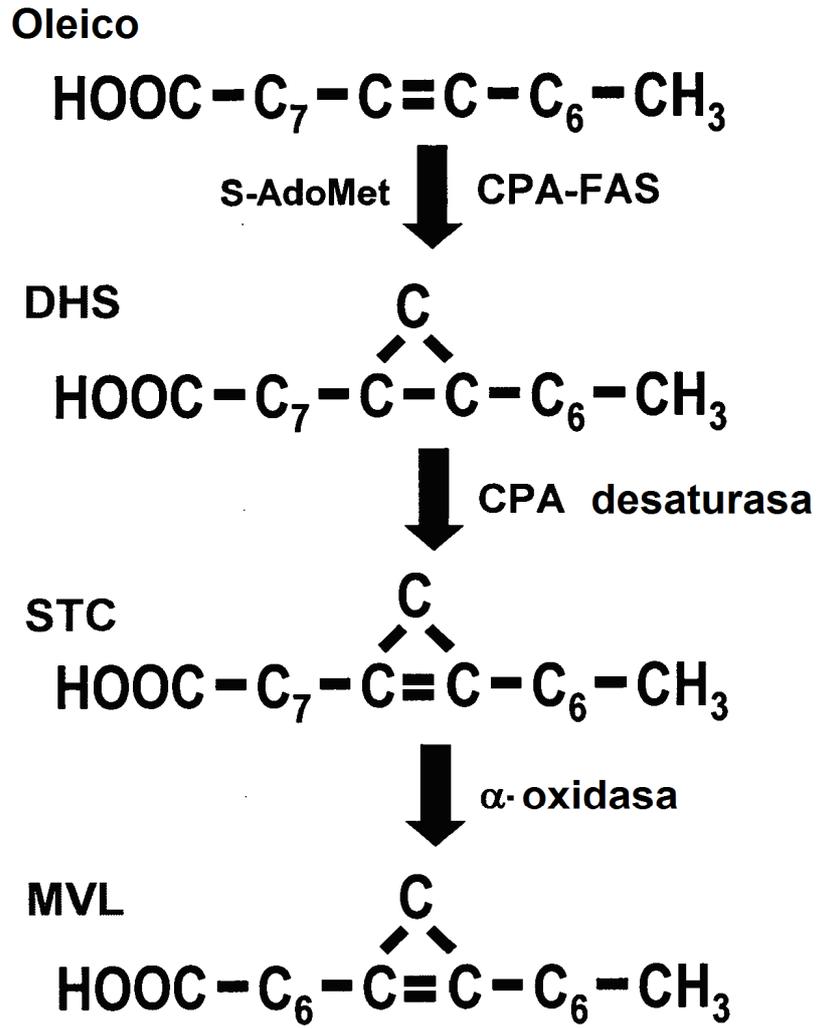


Figura 2

ES 2 603 530 T3

10 30 50
CAAACCAACACGCCTTCTTTGCCTCGTGTTCATCACCTGGCGTTAAACTGCTTTCTTT
70 90 110
AAAACCAACAAAATGGGTGCCGGGTGGGTAGGATGCCAATTGACGGGTATAAAGGAGGAA
M G A G W V G C Q L T G I K E E
130 150 170
AATCGAGGCTCGGTCAATCGAGTTCGATCGAGAAGCCTCCGTTTACGCTCGGTCAGATC
N R G S V N R V P I E K P P F T L G Q I
190 210 230
AAGCAAGCCATTCCGCCCCACTGTTTTCGCCGCTCCCTCCTTCGATCCTTCTCCTACGTG
K Q A I P P H C F R R S L L R S F S Y V
250 270 290
GTCCATGACCTATGCTTAGCCTCTCTCTTTACTACATTGCAACATCATATTTCACTTT
V H D L C L A S L F Y Y I A T S Y F H F
310 330 350
CTCCACAACCCTTTTCCTACATGCTTGGCCTGTCTATTGGGTTCTCCAAGGTTGCATC
L P Q P F S Y I A W P V Y W V L Q G C I
370 390 410
CTCACCGGTGTTTGGGTCATCGCACACGAATGCGGTCACCACGCTTTCAGTGACTACCAA
L T G V W V I A H E C G H H A F S D Y Q
430 450 470
TGGGTTGACGACACCGTTCGGGTTGATCCTTCACTCCGCCCTTTTAGTCCCCTACTTCTCG
W V D D T V G L I L H S A L L V P Y F S
490 510 530
TGGAAAATCAGTCACCGCGTCCACTCGAACACCGGTTCCATGGAGCGTGACGAAGTA
W K I S H R R H H S N T G S M E R D E V
550 570 590
TTCGTGCCCAAACCAAGTCTAAATTATCATGCTTTGCGAAATACTTCAACAATCCACCC
F V P K P K S K L S C F A K Y F N N P P
610 630 650
GGTCGAGTCTCTCTTGTAGTCACATTGACTCTTGGTTGGCCTATGTACTTAGCCTTC
G R V L S L V V T L T L G W P M Y L A F
670 690 710
AACGTTTCGGGTCGATACTATGATCGATTAGCTTCCCACTATAACCCTTACGGCCCCATT
N V S G R Y Y D R L A S H Y N P Y G P I

Figura 3

ES 2 603 530 T3

730 750 770
TACTCCGAACGCGAGAGGCTACAAGTTTACATCTCCGATGCTGGTATAGTTGCGGTAATT
Y S E R E R L Q V Y I S D A G I V A V I
790 810 830
TATGTACTTTATAAGATTGCTGCAACAAAAGGGCTGGCTTGGCTTTTATGCACTTATGGG
Y V L Y K I A A T K G L A W L L C T Y G
850 870 890
GTACCTCTACTTATTGTGAATGCCTTCCTTGTGTTGATCACCTACTTGCAACATACTCAC
V P L L I V N A F L V L I T Y L Q H T H
910 930 950
TCGGCATTGCCGATTACGACTCGTCTGAATGGGATTGGTTTCGAGGAGCATTGTGCGACG
S A L P H Y D S S E W D W F R G A L S T
970 990 1010
ATTGATCGAGATTACGGGGTGTGAACAAAAGTGTCCATAACATCACCGATACGCATGTG
I D R D Y G V L N K V F H N I T D T H V
1030 1050 1070
GCTCATCACCTCTTCTCAACGATGCCACATTATCATGCAATGGAGGCCACTAAAGCAATC
A H H L F S T M P H Y H A M E A T K A I
1090 1110 1130
AAACCGATACTCGGCAAGTATTATCCTTTCGACGGGACACCGATTTATAAGGCAATGTGG
K P I L G K Y Y P F D G T P I Y K A M W
1150 1170 1190
AGGGAGGCAAAAGAGTGCCTTTACGTGAGGCTGACGTTGGTGGTGGTGGTAGCAAAGGT
R E A K E C L Y V E A D V G G G G S K G
1210 1230 1250
GTTTTTTGGTATCGTAACAAGTTCTAAAGACAGACCAACTGCCTGATAGCTGGCCGGCAA
V F W Y R N K F *
1270 1290 1310
AATCGACGTAAAACGTACTTATTAGACTAGTGTTAACTAGGAAGTTAATAATGGTAGGA
1330 1350
AAATGTGGAATAGCTGCCTAGTAGTTTTATGTATTAAGTGTT

Figura 3 continuación

ES 2 603 530 T3

```

      10              30              50
TTTTGTAGATCGCAAGCCGTAGAAGTAAACAAAGAAGGATTTGCAGATCTTAAATCATGC
      70              90              110
TGTAATTTTCTCGAGGAAATTTTTACCTTCATTGATTTCGTTTCTATTTTCGCTTGAGTT
      130             150             170
GGAGAATGCTTCAGCTGCCTTTCTAAAAGATTGTTTCCAAAAGAGGGAATTTTAGTGGAT
      190             210             230
AATAGAAGTTCTTTTTAATTCTCAAATCATGGTTGCCACTGCTGCTACATCCTCATTCT
                                M V A T A A T S S F F
      250             270             290
TTCCCGTAACTTCTTCCCCTGACTCCTcTGA CTCAAAAAaCAAGAAGCTTGAAGTGGAT
      P V T S S P D S S D S K N K K L G S G S
      310             330             350
CTACTAACCTCGGAGGCATCAAGTCGAAACCATCTGCTTCTTCTGGAAGTTTGCAAGTCA
      T N L G G I K S K P S A S S G S L Q V K
      370             390             410
AGGCAAATGCTCAAGCCCTCCAAGATAAATGGTACCCTGTTGTAACCTCTCCGTTG
      A N A Q A P P K I N G T T V V T S P V E
      430             450             470
AAGGTTTCAAGAACGAAGATGGTGCAGGTTCCCCTCATCCTCGGACCTTTATCAATCAAT
      G F K N E D G A G S P H P R T F I N Q L
      490             510             530
TACCTGATTGGAGCATGCTTCTTGCCGCTATCACAACCATTTTCTGGCTGCTGAGAAGC
      P D W S M L L A A I T T I F L A A E K Q
      550             570             590
AGTGGATGATGCTTGATTGGAAGCCAAGGCGCCTGACATGCTCATTGATCCTTTTGGTA
      W M M L D W K P R R P D M L I D P F G I
      610             630             650
TAGGGAGGATGTTCAGGATGGTCTTGT TTTCCGTCAAAACTTCTCGATTAGGTCTTATG
      G R I V Q D G L V F R Q N F S I R S Y E
      670             690             710
AgATAGGTGCTGATCGTACGGCATCCATAGAGACGCTAATGAATCATTTACAGGAAACCG
      I G A D R T A S I E T L M N H L Q E T A
      730             750             770
CGATTAATCATTTGTA AAAAGTCTGGACTGcTTGGAgAAgGtTTTGGTGTACCCCTGAGA
      I N H C K S A G L L G E G F G A T P E M

```

Figura 4

ES 2 603 530 T3

790 810 830
TGTGCAAGAAGAACCTAATTTGGGTGGTCACTCGGATGCAAGTTGTGTTTTGATCGGTATC
C K K N L I W V V T R M Q V V F D R Y P
850 870 890
CTACTTGGGGTGATGTTGTTCAAGTAGACACTTGGGTGAGTGCATCAGGAAAGAATGGCA
T W G D V V Q V D T W V S A S G K N G M
910 930 950
TGCGAAGAGATTGGCTTGTCTAGTATAGTAAAAGTGGTGAAGTTTTAACAAGAGCCTCAA
R R D W L V S D S K T G E V L T R A S S
970 990 1010
GTGTGTGGGTGATGATGAATAAATGACTAGAAAGCTATCTAAAATTCCTGAGGAGGTCC
V W V M M N K L T R R L S K I P E E V R
1030 1050 1070
GAGGAGAAATAGAACCTTATTTTATGAATTCCGATCCTGTTGTGGCAGAAGATAGCCGGA
G E I E P Y F M N S D P V V A E D S R K
1090 1110 1130
AATTAGTGAAGCTCGATAAAAAGCATGGCTGAGCACGTGcGTAAAGGTTTTAACTCCTAGAT
L V K L D K S M A E H V R K G L T P R W
1150 1170 1190
GGAGTGACTTGGATGTCAACCAACATGTCAATAACGTGAAGTACATTGGCTGGATCCTCG
S D L D V N Q H V N N V K Y I G W I L E
1210 1230 1250
AGAGTGCTCCATTGCCGGTGTGGAAACTCACGAGCTTTCTTCCATGACACTGGAGTATA
S A P L P V L E T H E L S S M T L E Y R
1270 1290 1310
GGAGGGAGTGTGGGAGGGAGAGCATACTGCAGTCGCTAACAACCGTGTCCGACTCCAGTG
R E C G R E S I L Q S L T T V S D S S V
1330 1350 1370
TAGGAGACTTGGTGAATGTGGGTGAAATCGAgTGCCAGCACCTGCTGCAACTCGAGGAAG
G D L V N V G E I E C Q H L L Q L E E G
1390 1410 1430
GGTCCGAGATTGTGAGAGGGGAGAACTCAATGGAGGCCCAAGTATGCCAAAAGTTTTGGTA
S E I V R G R T Q W R P K Y A K S F G N
1450 1470 1490
ATGTGGGTCAAATTCAGCAGAAAGTGCATAGAAGGAAAAAATCCCAAATTTCTCTTAT
V G Q I P A E S A *

Figura 4 continuación

1510 1530 1550
TGTGACCTAAGTGGGGCAATAGTCTGATTGCCGGGTGTCACAATGATTTATGTAGAATCT
1570 1590 1610
AATCATGTGTTCTATGGATATATATATATATATATATTTATGCTTCTTTTTTATATATAAAT
1630
AATATTATATATTCCTTTTAAAAAAA

Figura 4 continuación

TAGCGCTAGAAGTTACCGAGAAGAGTTTAGAGATCCCTATTATCGGAAAGAGGGGATTTC
 AGCGGATAACAGAAGTTCATTTTAAATTTATAAAAATCATGGTTGCCACTGCTGCTACATCC
 M V A T A A T S
 TCATTCTTTCCAATCACTTCTTCCCCGACTCCATTGACTCAAAAAACAAGAAGCTTGGA
 S F F P I T S S P D S I D S K N K K L G
 AATGGATCTACTAACCTTGGAGGTATAAAGTTGAAACCATctGCTTCTTCTGGAAGTTTG
 N G S T N L G G I K L K P S A S S G S L
 CAAGTTAAGGCAAATGCACAAGCCCCCAAAGATAAATGGTACCACAGTTGTGATGACT
 Q V K A N A Q A P P K I N G T T V V M T
 CCAGTAGAAGGTTTCCCGAGCGAAGATGCTGCAAGTTCCTTACCTCCAGGACGTTTATC
 P V E G F P S E D A A S S L P P R T F I
 AATCAGCTACCTGATTGGAGCATGCTTCTTCTGCTATGACAACCATTTTCTGGCTGCT
 N Q L P D W S M L L A A M T T I F L A A
 GAGAAGCAGTGGATGATGCTTGGATTGGAAGCCAAAGCGGCCTGACATGCTCATTGACCCA
 E K Q W M M L D W K P K R P D M L I D P
 TTTGGGATAGGGAGGATTGTTTCAAGATGGTCTTGTGTTTTTCGTCAGAACTTCTCAATTAGG
 F G I G R I V Q D G L V F R Q N F S I R
 TCTTATGAGATAGGTGCTGATCGTACAGCATCCATAGAGACGCTAATGAATCATTTACAG
 S Y E I G A D R T A S I E T L M N H L Q
 GAAACAGCGATTAATCATTGTAAAAGTGCTGGACTGCTAGGAGATGGTTTTGGTGCTACC
 E T A I N H C K S A G L L G D G F G A T
 CCTGGGATGTGCAAGAAAACCTAATATGGGTAGTCACCCGGATGCAAGTTGTGGTTGAT
 P G M C K K N L I W V V T R M Q V V V D
 TGTTATCCAACCTGGGGTGATGTTGTTCAAGTAGACACTTGGGTGAGTGCATCAGGAAAG
 C Y P T W G D V V Q V D T W V S A S G K
 AATGGCATGCGAAGGGATTGGCTTGTGCAAGTAAAGTAAACTGGTGAATTTTAACTAGA
 N G M R R D W L V S N S K T G E I L T R
 GCCTCAAGTGTGTGGGTGATGATGAATAAATTGACCAGAAGGTTATCTAAAATTCCAGAA
 A S S V W V M M N K L T R R L S K I P E
 GAGGTCCGAGGAGAAATAGAACCTCATTTTATGAATTCAGATCCAGTGGTGGCTGAGGAT
 E V R G E I E P H F M N S D P V V A E D
 AACCGGAAATTAGTGAAACTTGACGACAGCACAGCCCAATATGTGCGCAAGGGTTTAACT
 N R K L V K L D D S T A Q Y V R K G L T
 CCTCGATGGAGCGACCTGGATGTGAATCAGCATGTCAACAATGTGAAGTACGTTGGTTGG
 P R W S D L D V N Q H V N N V K Y V G W
 ATCCTTGAGAGTACACCATTGGGAATTGTGGAGAGTCATGAGCTTTGTTCCATGACACTG
 I L E S T P L G I V E S H E L C S M T L

Figura 5

GAGTATAGGAGGGAGTGTGGGAgGGACAGCGTGCTGCAGTCACTAACTGCGGTGTCTGGT
E Y R R E C G R D S V L Q S L T A V S G
GTGGGCAACCTCGGGAATATGGGGAAATTGAGTGCCAGCACTTGCTCCAACCTGAAGAG
V G N L G N M G E I E C Q H L L Q L E E
GGGTCTGAGATTGTGAGAGGGAGGACACAGTGGAGGCCAAAGAATGCCAAGAGTTTTGGT
G S E I V R G R T Q W R P K N A K S F G
AAAATGGATCAAGTTCCCGCACAAAGTGCATAGATCCGAAGTCTCTTTGCTGCGTGTCAA
K M D Q V P A Q S A *
AACTAGCAGTCAACGCATTGTGTAGAATCTTCTTTTGTCTTTGAATCCATAATATATA
TATATGATATTAGCTTGTAAGCTTTCAAAGCTTGCTGTAATTAGCTCTAAAAAAAAA

Figura 5 continuación

```

1  ATGGTTTACC CACGCGTCCG GTAGAACATG TTGGTGAAGA ATATATTGAG
51  GAGTTTTACA GATGCTGTGA CCAATTACTG AAAGAAGATG GACTTTTTGT
101 TCTTCAGTTC ATATCTATCC CAGAAGAGCT TTCCAAAGAA ATCCAGCAAA
151 CAGCAGGTTT TCTAAAGGAA TATATATTCC CCGGTGGAAC CCTGCTTTCT
201 TTGGATAGGA ATTTATCAGC CATGGCTGCT GCAACAAGAT TCAGTGTGGA
251 GCATGTGGAA AATATAGGAA TGAGTTATTA CCACACACTG AGATGGTGGA
301 GAAAACTTTT CCTGGAAAAC ACAAGCAAAG TTCTAGCTCT GGGATTTCGAC
351 GAGAAGTTCA TGAGGACATG GGAATACTAT TTCGATTACT GCGCTGCCGG
401 TTTTAAGACA GGAACCCTTA TAGATTACCA GGTGTATTT TCGCGGGCCG
451 GAAATTTCCG TACTCTCGGA GATCCATACA AAGGTTTCCC TTCTGCATAC
501 TCCTTCATGG ATGATTGAAC AAAGTGTGGT TGAACATTGA TCCAAAGAAG
551 CAAACAAAAT TATCACCACA CTGCCAGTGT TAAGAACAAC CTATCTCCCT
601 AGTCCCTACT TTTCTTTATT ATGGCTATGT TTGCAATGCA AGAATAAGCA
651 AACATTGTAA TGTCAATAAA GTTTGCACCT TTGTAGACTG GATGGGATGT
701 TATCAATGAA GTACCTAGTT TATAAGTAAA AAAAAAAAAA AGA

```

Figura 6

GCACAAGGTAAAGCAGTGTACCGCGCGCAGTGTGGAAGTGGCCGTGATCGGAGGTGGGA
M E V A V I G G G I
TAAAAGGGTTGCTTTTCGGCCTACGTACTGGTCAAAGCCGGCGTGGACGTGGTGGTTTACG
K G L L S A Y V L V K A G V D V V V Y E
AGAAAGAAGAACAATTAGGCGGCCATGCAAAGACTGTAACTTCGACGCCGTTGATTTAG
K E E Q L G G H A K T V N F D A V D L D
ACCTTGGCTTCTTGTCTCAATCCAGCAAGATATGCAACACTATTGCATATGTTTCGACA
L G F L F L N P A R Y A T L L H M F D S
GCCTTGGTGTGTAGTGTAGAAACATCCGATGTTTCATTCTCTATAAGCCATGACAAAGGCA
L G V D V E T S D V S F S I S H D K G N
ACAATGGCTATGAATGGTGCAGCCAATATGGATTTTCCAATTACTTTGCTCAAAGAAGA
N G Y E W C S Q Y G F S N Y F A Q K K K
AACTGTTGAACCCTTTCAATTGGCAAAGCCTCAGAGAGATCATCAAATTCGGCAATGATG
L L N P F N W Q S L R E I I K F G N D V
TCGAAAGTTACCTTGGATCACTTGAGAACAACCCAGACATTGATCGTACTGAGACCTTGG
E S Y L G S L E N N P D I D R T E T L G
GACAGTTTATAAACTCAAAGGGCTACTCTGAAAATTTTCAAACACTTATCTGGCTCCCTA
Q F I N S K G Y S E N F Q N T Y L A P I
TATGTGGTTCAATGTGGTCAAGCTCCAAGGAAGATGTTACGAGCTTTTTCAGCTTTTTTCCA
C G S M W S S S K E D V T S F S A F S I
TCCTTTCATTTTGCGTACTCATCATTTGTACCAGCTATTTGGGCAGTCACAGTGGTTGA
L S F C R T H H L Y Q L F G Q S Q W L T
CTATCAAAGGGCACTCACATTTTGTAAAAGGGTTAGGGAAGTGTGGAGACTAAAGGTT
I K G H S H F V K R V R E V L E T K G C
GTCAATTTAAACTCGGTTGTGAAGTACAATCTGTTTTGCCCGTTGATAATGGTACCGCCA
Q F K L G C E V Q S V L P V D N G T A M
TGGTCTGTGGAGATGGTTTCCAAGAAACTTACAATGGATGCATAATGGCTGTTGATGCTC
V C G D G F Q E T Y N G C I M A V D A P
CCACTGCCCTAAAATTATTAGGAAACCAAGCAACATTTGAAGAAAACAAGACTGCGGTG
T A L K L L G N Q A T F E E T R V L G A
CTTTCCAATATGCTACCAGTGATATTTTCTTACCAGGACAGTACTTTAATGCCACAAA
F Q Y A T S D I F L H Q D S T L M P Q N
ACAATCAGCTTGGAGTGCATTGAATTTTCTCAATAGTAGCAAAAATAATGCATTCTTAA
K A W S A L N F L N S S K N N A F L T
CATACTGGCTCAATGCACTACAGAATATTGGGAAAAAAGTGAGCCATTTTTTGTGACTG
Y W L N A L Q N I G K T S E P F F V T V
TCAATCCAGACCATACCCGAAGAATACCTTACTTAAGTGGTCAACCGGCCATGCAATTS
N P D H T P K N T L L K W S T G H A I X
CCTCTGTGCTGCATCAAAGCTTCACTTGAGCTTGGTCAGATTGAGGAAAGAGAGGAA
S V A A S K A S L E L G Q I Q G K R G I
TCTGGTTCTGTGGCTATGACTTCAATCAGGATGAACTAAAGGCTGGTATGGATGCTGCAC
W F C G Y D F N Q D E L K A G M D A A H
ATGGTATCTTGGGAAAGCATTCTTCTGTTCCGCCAGTCCAAAGAATATGTCACCCTCTT
G I L G K H S S V P P S P K N M S P S L
TACCAAAGAATATGTCACCCCTTTTCATGGAACAACGGCAGCCTCTTTGTTACCAAAT
P K N M S P S F M E T T A R L F V T K F
TCTTTCAACAATATATATCTATGGGCTGCGTAATTTTTTTAGAGGAAGGAGGCAGAATTT
F Q Q Y I S M G C V I F L E E G G R I F
TCACTTTCAAAGGAAACATGGAAAAGTGTCTCTTAAAACAGTTCTGAAAAGTGCATAATC
T F K G N M E K C P L K T V L K V H N P
CTCAGTTTTACTGGAGGATCATGAAAGAAGCTGATATAGGCCTTGCAGACGCATATATCC
Q F Y W R I M K E A D I G L A D A Y I H
ATGGAGATTTTTCTTTCTTGATGAAAATGAAGGCTTCTTAATCTTTTTCCGGATTCTTG
G D F S F L D E N E G L L N L F R I L V
TTGCCAATAAAGAGAACTCAGCTGCCTCAGGGTCGACTAAAAGAAGGACTTGGTGGTCGC
A N K E N S A A S G S T K R R T W W S P
CTGCTCTGTTAACAGCTAGTATATCATCTGCCAAGTATTTTTGTGAAGCATCTCTTAAGAC
A L L T A S I S S A K Y F V K H L L R Q

Figura 7

AAAATACTATTACACAAGCTCGTAGGAACATTTCTCGTCATTATGATCTGAGTAATGAAC
 N T I T Q A R R N I S R H Y D L S N E L
 TTTTCTCTATACTTGGGCAAATGATGCAATACTCTTCTGGAGTCTTTAGGACAGGAG
 F S L Y L G K M M Q Y S S G V F R T G E
 AAGAACATTTGGACGTTGCACAGCGAAGAAAATCAGTTCTCTAATTGAGAAAAACAAGGA
 E H L D V A Q R R K I S S L I E K T R I
 TAGAGAAATGGCATGAAGTTCTAGACATTGGGTGCGGTTGGGGAAGCTTAGCTATTGAAA
 E K W H E V L D I G C G W G S L A I E T
 CTGTGAAAAGAACAGGATGCAAATATACTGGCATCACTCTATCAGAACAGCAACTGAAAT
 V K R T G C K Y T G I T L S E Q Q L K Y
 ATGCTCAAGAAAAAGTGAAGGAAGCTGGACTCGAGGATAACATCAAATACTTCTCTGTG
 A Q E K V K E A G L E D N I K I L L C D
 ACTATCGCCAGTTACCTAAGGAACACCAATTTGACAGAATCATATCTGTAGAGATGGTAG
 Y R O L P K E H Q F D R I I S V E M V E
 AACATGTTGGTGAAGAATATATTGAGGAATTTACAGATGCTGTGATCAATTACTGAAAG
 H V G E E Y I E E F Y R C C D Q L L K E
 AAGATGGACTTTTCGTTCTTCAGTTCATATCTATCCCAGAGGAGCTTTCCAAAGAAATCC
 D G L F V L Q F I S I P E E L S K E I Q
 AGCAAACAGCTGGTTTTCTTAAGGAATATATATTCCTGGTGAACCCTGCTTTCTTTGG
 Q T A G F L K E Y I F P G G T L L S L D
 ATAGGAATTTATCAGCCATGGCTGCTGCAACAAGATTCAGTGTGGAGCATGGGAAAACA
 R N L S A M A A A T R F S V E H V E N I
 TAGGAATGAGTTATTACCACACACTGAGATGGTGGAGAAAACCTTTTCTGAAAAACACAA
 G M S Y Y H T L R W W R K L F L K N T S
 GCAAAGTTCTGGCTTTGGGGTTCGACGAGAAGTTCATGCGGACATGGGAATACTATTTTCG
 K V L A L G F D E K F M R T W E Y Y F D
 ATTACTGTGCTGCTGGTTTTAAGACAGGAACCCTTATAGATTACCAGGTTGATTTTCTC
 Y C A A G F K T G T L I D Y Q V V F S R
 GAGCCGGTAATTCGGTACACTTGGAGATCCATACAAAGGTTTCCCTTCTGCATATTCCT
 A G N F G T L G D P Y K G F P S A Y S F
 TCATGGATGATTGAACAAAGTGTGTTGAATATATGATCACCATACAATGATTCACCAGCT
 M D D *
 GGATCAAACGGTACCAGTGTACCTAGTCCCTGCTTTTGTGTTAGTTATGGTTTTTCGT
 TTCGTTGCGAAAAAGAAAAAGCAAATAATGTATGTTAATAATGAAATGTTTGTATCTGG
 TATATCTATACTGGTTGGATTTTATGTATGGAGATCTGTTTCTTTTTAAAAAAAAAAAAA
 AAAA

Figura 7 continuación

ES 2 603 530 T3

10 30 50
 GTCACGGCGGCAGTGATGGAAGTGGCGGTGATCGGAGGTGGGATAAAAAGGGTTGGTTTCG
 M E V A V I G G G I K G L V S
 70 90 110
 GCCTACGTACTGGTCAAAGCCGGCGTGGACGTGGTGGTTTACGAGAAAGAAGAGCAATTA
 A Y V L V K A G V D V V V Y E K E E Q L
 130 150 170
 GGCGCCATGCGAAGACTGTAACTTCGACGCCGTTGACTTAGACCTTGGCTTCTTGTTT
 G G H A K T V N F D A V D L D L G F L F
 190 210 230
 CTTAATCCTGCAAGATATGCAACACTGTTGGATATAATCGACAGCCTTGGTGTGATGTA
 L N P A R Y A T L L D I I D S L G V D V
 250 270 290
 GAAACATCCGATGTTTTCATTCTCTATAAGCCATGACAAAGGCAACAATGGCTATGAATGG
 E T S D V S F S I S H D K G N N G Y E W
 310 330 350
 TGCAGTCAATATGGATTTTCCAATTACTTTGCACAAAAGAAGAACTGTTGAACCCCTTTC
 C S Q Y G F S N Y F A Q K K K L L N P F
 370 390 410
 AATTGGCAAAAACCTTAGAGAGATCATCAGATTGACCAACGATGTCGAAAGTTACCTTGGA
 N W Q N L R E I I R F S N D V E S Y L G
 430 450 470
 TCACTTGAGAACAACCCAGACATTGATCGTACTGAGACCTTGGGACAGTTATAAAAATCA
 S L E N N P D I D R T E T L G Q F I K S
 490 510 530
 AAGGGCTACTCTGAAAATTTTCAAACACTTACCTGGCTCCTATATGTGGTTCATGTGG
 K G Y S E N F Q N T Y L A P I C G S M W
 550 570 590
 TCAAGCTCCAAGGAAGATGTTATGAGCTTTTCAGCATTTCATCCTTTTCATTTTGGCCGT
 S S S K E D V M S F S A F S I L S F C R
 610 630 650
 ACTCATCATTTGTACCAGCAATTTGGGCAGCCACAGTGGTGTGACTATCAAAGGGCACTCA
 T H H L Y Q Q F G Q P Q W L T I K G H S
 670 690 710
 CATTTTGTTAAAAGGGTTAGGGAAGTGTGAGACTAAAGGTTGTCAATTTAAACTCGGT
 H F V K R V R E V L E T K G C Q F K L G
 730 750 770
 TGTGAAGTACAATCTGTTTTGCCTGCTGATAATGGTACCACCATGGTCTGTGGAGATGGT
 C E V Q S V L P A D N G T T M V C G D G
 790 810 830
 TTCCAAGAACTTACAATGGATGCATAATGGCTGTTGATGCTCCCACTGCCCTAAAATTA
 F Q E T Y N G C I M A V D A P T A L K L
 850 870 890
 TTAGGAAACCAAGCAACATTTGAAGAAACAAGAGTACTGGGTGCTTTCCAATATGCTACC
 L G N Q A T F E E T R V L G A F Q Y A T
 910 930 950
 AGTGATATTTTCTTACCAGGACAGTACTTTAATGCCACAAAACAATCAGCTTGGAGT
 S D I F L H R D S T L M P Q N K S A W S
 970 990 1010
 GCATTGAATTTTCTCAATAGTAGCAAAAATAATGCATTCTTAACATACTGGCTCAATGCA
 A L N F L N S S K N N A F L T Y W L N A
 1030 1050 1070
 CTACAGAATATTGGGAAAACAAGTGAGCCATTTTGTGACTGTCAATCCAGACCATAACC
 L Q N I G K T S E P F F V T V N P D H T
 1090 1110 1130
 CCGAAGAATACCTTGCTTAAAGTGGTTCGACTGGCCATGCAATCCCTCTGTGCTGCATCA
 P K N T L L K W S T G H A I P S V A A S
 1150 1170 1190
 AAAGCTTCACTTGGCTTGGTTCAGATTCAGGGGAAGAGAGGAATCTGGTTCTGTGGCTAT
 K A S L E L G Q I Q G K R G I W F C G Y

Figura 8

ES 2 603 530 T3

1210 1230 1250
 GACTTCAATCAGGATGAACTAAAGGCTGGTATGGATGCTGCACATGGTATCTTGGGAAAG
 D F N Q D E L K A G M D A A H G I L G K
 1270 1290 1310
 CATTCTTCTGTTCTGCATAGTCCAAAGAGTATGTCACCCTCTTTCATGGAAACAACGGCA
 H S S V L H S P K S M S P S F M E T T A
 1330 1350 1370
 CGCCTCTTGTACTAAATTCTTCAACAATATATATCTATGGGCTGTGTAATTTTCTTA
 R L F V T K F F Q Q Y I S M G C V I F L
 1390 1410 1430
 GAGGAAGGAGGCAGAATTTTCACTTTCAAAGGAAACATGGAAAAGTGTCTCTTAAACA
 E E G G R I F T F K G N M E K C P L K T
 1450 1470 1490
 GTTCTGAAAGTACATAATCCTCAGTTTTACTGGAGGATCATGAAAGAAGCTGATATAGGC
 V L K V H N P Q F Y W R I M K E A D I G
 1510 1530 1550
 CTTGCAGATGCATATATCCATGGAGATTTTTCTTTCTTGTGAAACTGAAGGCCTTCTT
 L A D A Y I H G D F S F L D E T E G L L
 1570 1590 1610
 AATCTTTCCGGATTCTTGTGGCCAATAAAGAGAAGTCACTGCCTCAGGGTCGAATAAA
 N L F R I L V A N K E N S A A S G S N K
 1630 1650 1670
 AGAAGGACTTGGTGGTCACCTGCTCTGTTAACAGCTAGTATATCATCTGCAAAGTATTTT
 R R T W W S P A L L T A S I S S A K Y F
 1690 1710 1730
 GTGAAGCATCTCTTGAGACAAAATACTATTACACAAGCTCGTAGGAACATTTCTCGTCAT
 V K H L L R Q N T I T Q A R R N I S R H
 1750 1770 1790
 TATGATCTGAGTAATGAACTTTTCACTCTATACTTGGGCAAAATGATGCAATACTCTTCT
 Y D L S N E L F T L Y L G K M M Q Y S S
 1810 1830 1850
 GGAGTCTTTAGGACGGGAGAAGAACATTTGGACGTTGCACAGCGTAGAAAAATCAGTTCT
 G V F R T G E E H L D V A Q R R K I S S
 1870 1890 1910
 CTAATTGAGAAAGCAAGGATAGAGAAACGGCACGAAGTCTCGACATTTGGGTGCGGTTGG
 L I E K A R I E K R H E V L D I G C G W
 1930 1950 1970
 GGAAGCTTAGCTATTGAAACTGTGAAAAGAACAGGATGCAAATATACTGGCATCACTCTA
 G S L A I E T V K R T G C K Y T G I T L
 1990 2010 2030
 TCAGAACAGCAACTGAAATATGCTCAAGAAAAAGTGAAGGAAGCTGGACTCCAGGATAAC
 S E Q Q L K Y A Q E K V K E A G L Q D N
 2050 2070 2090
 ATCAAAATACTTCTCTGTGACTATCGCCAGTTACCTAAGGAACACCAATTTGACAGAATC
 I K I L L C D Y R Q L P K E H Q F D R I
 2110 2130 2150
 ATATCTGTAGAGATGGTAGAACATGTTGGTGAAGAATATATGAGGAGTTTACAGATGC
 I S V E M V E H V G E E Y I E E F Y R C
 2170 2190 2210
 TGTGACCAATTACTGAAAGAAGATGGGCTTTTGTTCCTCAGTTCATATCTATCCCAGAA
 C D Q L L K E D G L F V L Q F I S I P E
 2230 2250 2270
 GAGCTTTCAAAGAAATCCAGCAACAGCAGGTTTTCTAAAGGAATATATATTCCCTGGA
 E L S K E I Q Q T A G F L K E Y I F P G
 2290 2310 2330
 GGAACCTGCTTTCTTTGGATAGGAATTTATCAGCCATGGCTGCTGCAACAAGATTCACT
 G T L L S L D R N L S A M A A A T R F S
 2350 2370 2390
 GTGGAGCATGTGAAAATATAGGAATGAGTTATTACCACACACTGAGATGGTGGAGAAAA
 V E H V E N I G M S Y Y H T L R W W R K

Figura 8 continuación

ES 2 603 530 T3

2410 2430 2450
CTTTTCCTGGAAAACACAAGCAAAGTTCTAGCTCTGGGGTTCGACGAGAAGTTCATGAGG
L F L E N T S K V L A L G F D E K F M R
2470 2490 2510
ACATGGGAATACTATTTTCGATTACTGCGCTGCCGGTTTTAAGACAGGAAGTCTTATAGAT
T W E Y Y F D Y C A A G F K T G T L I D
2530 2550 2570
TACCAGGTTGTATTTTCAAGGGCCGGAAATTTTCGGTACACTCGGAGATCCATACAAAGGT
Y Q V V F S R A G N F G T L G D P Y K G
2590 2610 2630
TTCCCTTCTGCATATTCCTTCATGGATGATTGAACAAAGTGTGGTTGAACATTGATCCAA
F P S A Y S F M D D *
2650 2670 2690
AGAAGCAAACAAAATTATCACCACATGCCAGTGTTAAGAACAACCTATCTCCCTAGTCCC
2710 2730 2750
TACTTTTGTATTATTATGGCTATGTTTGCAATGCAAGAATAAGCAAACATTGTAATGTTAA
2770 2790 2810
TAAAGTTTGCACCTTTTGTAGACTGGATGGATGTTATCAATGAAGTACCTAGTTTATAAAA

AAAAAAA

Figura 8 continuación

ES 2 603 530 T3

10 30 50
 TCCCTATCTCCATTTACTATTTTTCTTCTCTCTTCTTCTTTCTTTCGAACCATTTTCAGAGT
 70 90 110
 TCATAAAATTCAGGGTTTTGTTTTTTTTTTGGGTGTAGTCAAATAAAGGATGAAAATAGCA
 M K I A
 130 150 170
 GTGATAGGAGGAGGATAAGTGGGGTGGTATCAGCCTATACTTTAGCCAAAGCCGGTGCA
 V I G G G I S G V V S A Y T L A K A G A
 190 210 230
 AATGTAGTGCTTTACGAGAAAGAAGAGTATTTGGGAGGCCATTCCAAGACCGTTCACTTC
 N V V L Y E K E E Y L G G H S K T V H F
 250 270 290
 GATGGTGTGATTTAGACCTTGGTTTCATGGTTTTTAATCGCGTTACATATCCAAATATG
 D G V D L D L G F M V F N R V T Y P N M
 310 330 350
 ATGGAGTTGTTGAGAGCCTTGGGATTGATATGGAACCATTTGATATGTCACCTCTCAGTG
 M E L F E S L G I D M E P F D M S L S V
 370 390 410
 AGCCTTAATGAAGGCAAAGGCTGTGAATGGGGCAGCCGTAATGGCCTTTCGGCCTTGT
 S L N E G K G C E W G S R N G L S A L F
 430 450 470
 GCCCAAAAATCCAACCTCTTCAATCCTTACTTTTGGCAAATGCTTAGAGAAAATCTCAA
 A Q K S N L F N P Y F W Q M L R E I L K
 490 510 530
 TTCAAGAAATGATGTTATTAGTTATCTTGAATTGCTCGAAAACAACCCGGATATTGACCGT
 F K N D V I S Y L E L L E N N P D I D R
 550 570 590
 AATGAAACATTGGGACAGTTCATAAAAATCAAAGGGTTACTCTGATTTATTTTCAGAAGGCT
 N E T L G Q F I K S K G Y S D L F Q K A
 610 630 650
 TATCTGGTGCCTGTATGTGGTTCAATATGGTCATGCCCTACAGAAAGAGTTATGGATTTT
 Y L V P V C G S I W S C P T E R V M D F
 670 690 710
 TCAGCTTTCTCTATTCTTTTCATTTTGCCGCAATCATCATCTACTTCAGATCTTTGGACGA
 S A F S I L S F C R N H H L L Q I F G R
 730 750 770
 CCACAGTGGATGACCGTTCGATGGCGTTCACATCGTTACGTCAATAAGGTTAGAGAAGAG
 P Q W M T V R W R S H R Y V N K V R E E
 790 810 830
 CTGGAGAGTACAGGTTGTCAAATAAGAAGTGGTTGCGAGGTGCATTCTGTTTTGAGTGT
 L E S T G C Q I R T G C E V H S V L S D
 850 870 890
 GCTGAAGGTTGCACTGTATTATGTGGAGATGACTCTCACGAGTTATATCAAGGGTGCATA
 A E G C T V L C G D D S H E L Y Q G C I
 910 930 950
 ATGGCTGTTTCATGCACCAATATGCTTTGAGATTGTTAGGGAATCAAGCAACATATGATGAA
 M A V H A P Y A L R L L G N Q A T Y D E
 970 990 1010
 TCAACAGTGCCTTGGCGCTTTCCAATATGTCTATAGTGATATTTATCTTCATCGTGACAAA
 S T V L G A F Q Y V Y S D I Y L H R D K
 1030 1050 1070
 AATTTAATGCCCAAAAACCCAGCAGCATGGAGTGCATGGAATTTTCTTGGAAGTACAGAC
 N L M P K N P A A W S A W N F L G S T D
 1090 1110 1130
 AAGAATGTATCTTTGACATACTGGCTTAATGTGCTTCAGAATCTAGGAGAAAACAAGCCTA
 K N V S L T Y W L N V L Q N L G E T S L
 1150 1170 1190
 CCCTTTTGGTCACTCTCAATCCAGATTATACACCAAAAACACACCTTGGCTTAAGTGGAGA
 P F L V T L N P D Y T P K H T L L K W R

Figura 9

ES 2 603 530 T3

1210 ACAGGCCATCCAGTACCATCTGTTGCTGCAACAAAAGCTTCTCTTGAGCTTGATCGGATT
 T G H P V P S V A A T K A S L E L D R I
 1270 1290 1310
 CAAGGGAAGAGAGGAATTTGGTTTTGTGGAGCATACTGGGCTATGGCTTCCATGAAGAT
 Q G K R G I W F C G A Y L G Y G F H E D
 1330 1350 1370
 GGATTAAGGCTGGGATGATTGCTGCAAACGGTCTGCTGGGAAAAGTTGTAATATTCTG
 G L K A G M I A A N G L L G K S C N I L
 1390 1410 1430
 AGCAATCAAAGCATATGGTGCCCTCTCTGATGGAAACAGGGGCACGTCTTTTGTACT
 S N P K H M V P S L M E T G A R L F V T
 1450 1470 1490
 AGATTCTCAGTCATTTTATATCAACCGCTGTGTGATTTTATGGAAGAAGGTGGCACT
 R F L S H F I S T G C V I L L E E G G T
 1510 1530 1550
 ATGTTACCTTTGAAGGAAGTCAATAAGTGTCTCTAAAACTGTAATTAAGTTCAC
 M F T F E G T S N K C S L K T V I K V H
 1570 1590 1610
 AGTCCACATTTTTATTGGAAGGTTATGACAGAGGCAGATTTAGGCCTTGCAGATTCATAT
 S P H F Y W K V M T E A D L G L A D S Y
 1630 1650 1670
 ATCAATGGGGATTTTCTTTTGTGATAAAAAAGACGGTCTGCTGAACCTTGTATGATT
 I N G D F S F V D K K D G L L N L V M I
 1690 1710 1730
 CTTATTGCCAACAGAGATTTGATTTCTTCCAACCTCAAACCTTAGTAAGAAAAGGGTGG
 L I A N R D L I S S N S K L S K K R G W
 1750 1770 1790
 TGGACACATTGTTGTTTACAGCTGGTCTAACATCAGCAAAGTATTTCTTCAAGCATGTC
 W T P L L F T A G L T S A K Y F F K H V
 1810 1830 1850
 TTAAGACAAAATACTCTTACACAAGCTCGTAGGAACATTTCTCGCCATTACGACyTGAGT
 L R Q N T L T Q A R R N I S R H Y D L S
 1870 1890 1910
 AATGACCTTTTTGCACTCTTCTGGATGAGACAATGACATACTTGTGTCAGTATTTAAG
 N D L F A L F L D E T M T Y S C A V F K
 1930 1950 1970
 ACAGAAGATGAGGATTTGAAAGATGCACAACACAGAAAGATCTCTTTTGTGAAAAA
 T E D E D L K D A Q H R K I S L L I E K
 1990 2010 2030
 GCAAGAATTGATAGCAAGCATGAAATCTTGAGATTGGATGTGGTTGGGkAAGCTTAGCT
 A R I D S K H E I L E I G C G W X S L A
 2050 2070 2090
 ATTGAGTTGTCAAACGAACCTGGATGCAAATATACCGGCATTACTTTATCCGAAGAGCAA
 I E V V K R T G C K Y T G I T L S E E Q
 2110 2130 2150
 CTCAACTTGCAGAAAAAAGAGTGAAGGAAGCTGGACTTCAGGAAAATATAAGATTTCAA
 L K L A E K R V K E A G L Q E N I R F Q
 2170 2190 2210
 CTCTGTGACTATCGACAACCTACCTAGCACCTACAAGTATGACAGAATTATATCGTGTGAG
 L C D Y R Q L P S T Y K Y D R I I S C E
 2230 2250 2270
 ATGATAGAAGCTGTTGGCCATGAATACATGGAGGACTTCTTCGGTTGCTGTGAATCAGTG
 M I E A V G H E Y M E D F F G C C E S V
 2290 2310 2330
 TTAGCAGATGATGGACTTCTGTTTTACAGTTCATATCAATACCAGAGGAACGGTACAAT
 L A D D G L L V L Q F I S I P E E R Y N
 2350 2370 2390
 GAATCAGGCGAAGCTCGGATTCATCAAGGAATACATCTTCCCTGGTGGATGCTTACCT
 E Y R R S S D F I K E Y I F P G G C L P

Figura 9 continuación

ES 2 603 530 T3

2410 2430 2450
TCTCTGGCTAGGATAACAACAGCCATGAATGCTGCGTCCAACTCTGTGTGGAGCATGTG
S L A R I T T A M N A A S K L C V E H V
2470 2490 2510
GAAAACATCGGACTTCATTACTACCAAACGCTTAGATATTGGAGAAAGAATTTCTGGAG
E N I G L H Y Y Q T L R Y W R K N F L E
2530 2550 2570
AAACAGAGCAAAATCCATGCCTTGGGATTCAATGACAAGTTCATCCGGACATGGGAATAC
K Q S K I H A L G F N D K F I R T W E Y
2590 2610 2630
TATTTTGATTATTGTGCTGCTGGTTTCAAGTCCAATACTCTTGGTAATTACCAGGTGTGA
Y F D Y C A A G F K S N T L G N Y Q V V
2650 2670 2690
TTTTCTCGGCCTGGAAATGTAGTTGCCTTGGCAACCCATACAAAGACTTCCCCTCAGCT
F S R P G N V V A L G N P Y K D F P S A
2710 2730 2750
TCTTAATTATTTATTTCTCCTTATTTCAATCGTACCATAGCCATAATTTGAGCTTGTTG
S *
2770 2790 2810
AAAACATGATGCTACACGTTTGGTTTCATTCAAATATGGTATTTGAGTGCATATCTATACA
2830 2850 2870
TTGATGAATGTAATTCTGGCTTGCCTCGTAGGAACTTGCCAGCAGGATTATCTTTTTACA
2890 2910
TGGACATTTATTTAATTCTCTGTTCAAATTT

Figura 9 continuación

1 2 3 4 5

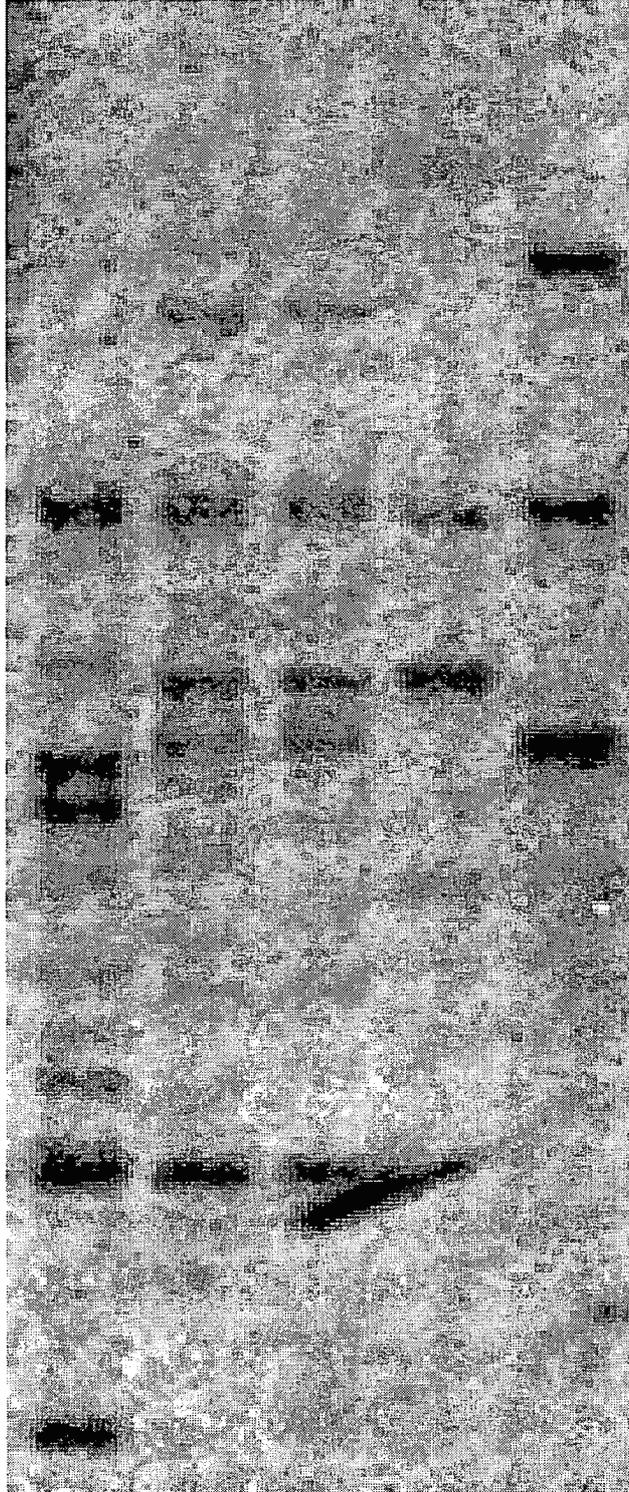


Figura 10

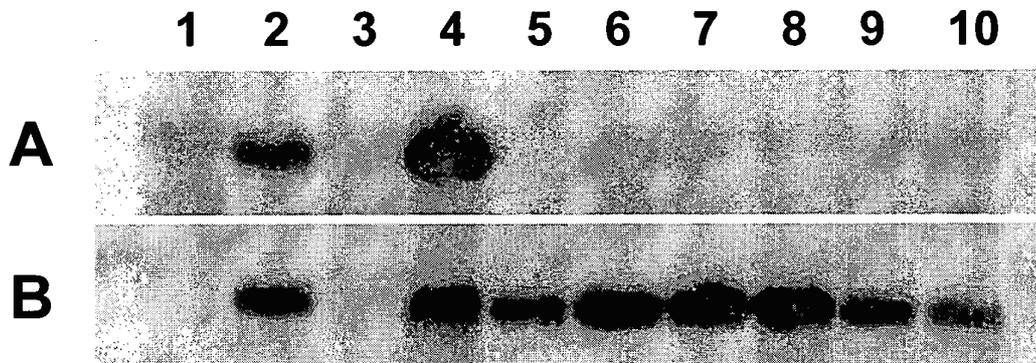


Figura 11

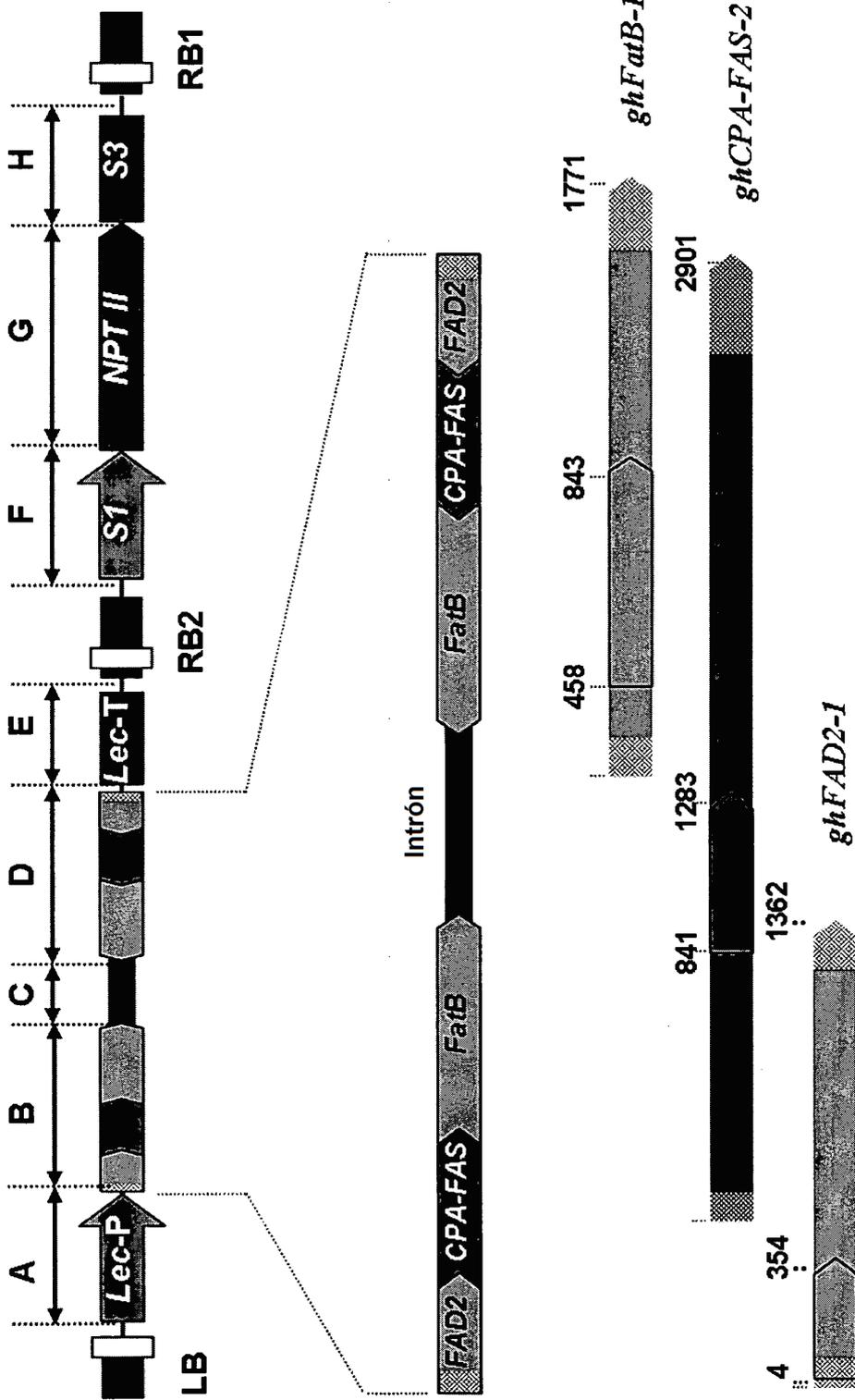


Figura 12

Fragmento A, el promotor de la lectina específico de semilla derivado del gen *lec1* de soja:

```

1   GAATTCTCTA  GAAAAGTTAA  CCCTTCGAAG  ATGATACTGA  CATTAAACACC
51  ATTTTTTAAT  ATTGTTTTTC  TATATCGTTA  TTGATCTCAG  CACATTCTTA
101 GAAAGATATT  TAAATTAGAT  AAAAGTAAAT  TTATATATAT  ATATATATAT
151 ATATATATAT  ATATATATAT  ATAAATGTAA  CATAAATCTA  TGGTCAATTA
201 CAATATTTAA  TTAAATAAAA  TAGAAATATA  AACACCACTT  TAATTTGACT
251 CGGATACATG  CATCCATAAA  GACTACAAAA  GGCAAAAAGA  GAAGGAAATG
301 AGATACGAAT  ATATGTCATA  AGTATATATA  GGTGACAAGG  GCAAATTTAA
351 TAGGTTGGTA  TTTAAATGCA  AAATCCTATG  TTTGATAAAG  AATGGTATGA
401 AAAACAGGCA  AAGTTAATTG  CAATTCAAAG  GTGAACAAAG  CATTTCCTTG
451 TCTACACTAA  TGGCATGTCT  AAGTAAATTA  TTAGTCTTGT  ATCTATATGT
501 CCACAAGTTA  TTAATTAGTC  TTATACTATC  AAAAAACAAGT  TAAGTTGCAA
551 ATCAAACATG  AACAAAGCAT  TTGTGTTGTA  ACCTACGAAA  AAATACCCTA
601 ACATACTGAT  ACGAATAATG  TGGCCTAAAT  TGATCGTTTA  CCAAATACG
651 GTGCTGGAAA  AAAAAATTGC  TCCTTTACCA  ACAAATTTAA  GAACTGATAC
701 ATCTTGTTTT  TTGTCACTGA  AGATAAACAC  GTGATCTTTG  GCAAAACATA
751 AAGGCCAACA  AAACAACTT  GTCTCATCCC  TGAATGATTC  GAATGCCATC
801 GTATGCGTGT  CACAAAGTGG  AATACAGCAA  TGAACAAATG  CTATCCTCTT
851 GAGAAAAGTG  AATGCAGCAG  CAGCAGCAGA  CTAGAGTGCT  ACAAATGCTT
901 ATCCTCTTGA  GAAAAGTGAA  TGCAGCGGCA  GCAGACCCTGA  GTGCTATATA
951 CAATTAGACA  CAGGGTCTAT  TAATTGAAAT  TGTCTTATTA  TTAAATATTT
1001 CGTTTTATAT  TAATTTTTTA  AATTTAATT  AAATTTATAT  ATATTATATT
1051 TAAGACAGAT  ATATTTATTT  GTGATTATAA  ATGTGTCACT  TTTTCTTTTA
1101 GTCCATGTAT  TCTTCTATTT  TTTCAATTTA  ACTTTTTATT  TTTATTTTTA
1151 AGTCACTCTT  GATCAAGAAA  ACATTGTTGA  CATAAAACTA  TTAACATAAA
1201 ATTATGTTAA  CATGTGATAA  CATCATATTT  TACTAATATA  ACGTCGCATT
1251 TTAACGTTTT  TTTAACAAAT  ATCGACTGTA  AGAGTAAAAA  TGAAATGTTT
1301 GAAAAGGTTA  ATTGCATACT  AACTATTTTT  TTTCTTATAA  GTAATCTTTT
1351 TTGGGATCAA  TTGTATATCA  TTGAGATACG  ATATTTAATA  TGGGTACCTT
1401 TTCACAAAAC  CTAACCCTTG  TTAGTCAAAC  CACACATAAG  AGAGGATGGA
1451 TTTAAACCAG  TCAGCACCGT  AAGTATATAG  TGAAGAAGGC  TGATAACACA
1501 CTCTATTATT  GTTAGTACGT  ACGTATTTCC  TTTTTTGTTT  AGTTTTTGAA
1551 TTTAATTAAT  TAAAATATAT  ATGCTAACAA  CATTAATTTT  TAAATTTACG
1601 TCTAATTATA  TATTGTGATG  TATAATAAAT  TGTCAACCTT  TAAAAATTAT
1651 AAAAGAAATA  TTAATTTTGA  TAAACAACCT  TTGAAAAGTA  CCCAATAATG
1701 CTAGTATAAA  TAGGGGCATG  ACTCCCCTATG  CATCACAGTG  CAATTTAACT
1751 GAAGCAAAGC  AGCGGCCCCG  G

```

Figura 13

Fragmento B, fragmento quimérico de tres genes diana:

```

1   ACCAACACGC CTTCTTTGCC TCGTGTTTCA TCACCTGGCG TTAAACTGCT
51  TTCTTTAAAA CCAACAAAAT GGGTGCCGGG TGGGTAGGAT GCCAATTGAC
101 GGGTATAAAG GAGGAAAATC GAGGCTCGGT CAATCGAGTT CCGATCGAGA
151 AGCCTCCGTT TACGCTCGGT CAGATCAAGC AAGCCATTCC GCCCCACTGT
201 TTTGCGCGCT CCCTCCTTCG ATCCTTCTCC TACGTGGTCC ATGACCTATG
251 CTTAGCCTCT CTCTTTTACT ACATTGCAAC ATCATATTTT CACTTTCTCC
301 CACAACCCTT TTCCTACATT GCTTGGCCTG TCTATTGGGT TCTCCAAGGT
351 TGGTACCATG GTCTGTGGAG ATGGTTTCCA AGAAACTTAC AATGGATGCA
401 TAATGGCTGT TGATGCTCCC ACTGCCCTAA AATTATTAGG AAACCAAGCA
451 ACATTTGAAG AAACAAGAGT ACTGGGTGCT TTCCAATATG CTACCAGTGA
501 TATTTTCCTT CACCGGGACA GTACTTTAAT GCCACAAAAC AAATCAGCTT
551 GGAGTGCATT GAATTTTCTC AATAGTAGCA AAAATAATGC ATTCTTAAAC
601 TACTGGCTCA ATGCACTACA GAATATTGGG AAAACAAGTG AGCCATTTTT
651 TGTGACTGTC AATCCAGACC ATACCCCGAA GAATACCTTG CTTAAGTGGT
701 CGACTGGCCA TGCAATTCCC TCTGTTGCTG CATCAAAAGC CATCAAAAGC
751 CTTGGTCAGA TTCAGGGGAA GAGAGGAATC TGGTTCTGTG GCTATGAGAG
801 CTCCTCGGAC CTTTATCAAT CAATTACCTG ATTGGAGCAT GCTTCTTGCC
851 GCTATCACAA CCATTTTCCT GGCTGCTGAG AAGCAGTGA TGATGCTTGA
901 TTGGAAGCCA AGGCGGCCTG ACATGCTCAT TGATCCTTTT GGTATAGGGA
951 GGATTGTTCA GGATGGTCTT GTTTTCCGTC AAAACTTCTC GATTAGGTCT
1001 TATGAGATAG GTGCTGATCG TACGGCATCC ATAGAGACGC TAATGAATCA
1051 TTTACAGGAA ACCGCGATTA ATCATTGTAA AAGTGTGGA CTGCTTGGAG
1101 AAGGTTTTGG TGCTACCCCT GAGATGTGCA AGAAGAACCT AATTTGGGTG
1151 GTCACTCGGA TGCAAGTTGT GTTTGATCGG TATCCTAC

```

Fragmento C: intrón de *ghFAD2-1*:

```

1   GTACATTTCT CTTTAATTTT CTTTTTTTTT CATTTCATGT TTTTCATGTT
51  AATGTTGCAT TGAAGTGATA AATTTGAGTG AATGATGTTT GGTATATCTT
101 CTTAGTAACT GACTTTTTGA AAATACTAGC ATTTTTTTTA ATATCAAGTG
151 AAAGAAGAAG AAGAATTTTCG CCATGCAAAA GCTTTTTAAG CTTTTTCTT
201 TTCCTTAGAT CAAAATTTAT TTGTTTACTT ATACTGTTCT TTTAAGCCCG
251 AAGAAAGAAG CCATGGTTTC AATTTTTGAG AGTTTTAAAT CCCAAATACC
301 AGAGAGCTTC ATCGTTTATT CATATATTTT TAAACATTTT TTAAAGCAAG
351 AACTTGTGAT TTGTTTTTAA TAAAATATGC AATAAATTTT TATATTTTTC
401 GTAAATTTAA AATTTAATTT TTCTACTTTT AAAATTTAAA AAAGTAAATT
451 TTAAATATA CTTTTCATT AATTTAAATTA TTATAAGTAA TTGAGTATTT
501 TTAATTTTAA AATTTACAC ATCAAATTA AAAAAAAGTT AACACTTGCA
551 CTTGATTTTG AAAAGTAAA GGATTAATTT TCAAATTTT AGTAAAAGGA
601 CTAATTTTCA AATTTTAAA GAGTATAGAG ACTCCTCTAC ATTTTAGATT
651 TTAATTTTAA AATCTAACAG TTAACACTTT CTTAATTAAT TTACGATAAAA
701 TTTAACTAAA AAATTACAAT ATTAATGGTT AAAATTTAAAT TTTGAAAAGT
751 ATAAAGATTA AATTGTAAT TTTCAAAAAG CATAGGAAGT TATAGTATAT
801 TTTAACCTTT ATTTATTTTA TATCTGGTGA GGTTCCTGCA TGCACCGAAG
851 ATGTCACCTT TTGCCAGTAT TTTCCAGTGG CTGTTTCTC TCAAACTAC
901 CTTGAATCTT GAGACAGAAT TAAATATATT TTGGCCTTT TCTTCATTTT
951 CTCTCTCTCT ATTTTCTTTT AAAAATTGCT TTAGAGAATT CAGAAAAAAT
1001 ACTTTCCAAC ACGAAAATTT CTTCAAATTT ATTGTTTATA TCTAATAAAT
1051 GGTGCTTAA TTTTGAAAA CAAAAGTTAT TGTAGTTAGT TTTGCTTCTT
1101 GCGTGTCCAG

```

Figura 13 continuación

Fragmento D, repetición invertida de los genes quiméricos:

```

1  GTAGGATACC GATCAAACAC AACTTGCATC CGAGTGACCA CCCAAATTAG
51  GTTCTTCTTG CACATCTCAG GGGTAGCACC AAAACCTTCT CCAAGCAGTC
101 CAGCACTTTT ACAATGATTA ATCGCGGTTT CCTGTAAATG ATTCATTAGC
151 GTCTCTATGG ATGCCGTACG ATCAGCACCT ATCTCATAAG ACCTAATCGA
201 GAAGTTTTGA CGGAAAACAA GACCATCCTG AACAATCCTC CCTATACCAA
251 AAGGATCAAT GAGCATGTCA GGCCGCCTTG GCTTCCAATC AAGCATCATC
301 CACTGCTTCT CAGCAGCCAG GAAAATGGTT GTGATAGCGG CAAGAAGCAT
351 GCTCCAATCA GGTAATTGAT TGATAAAGGT CCGAGGAGCT CTCATAGCCA
401 CAGAACCAGA TTCCTCTCTT CCCCTGAATC TGACCAAGCT CAAGTGAAGC
451 TTTTGATGCA GCAACAGAGG GAATTGCATG GCCAGTCGAC CACTTAAGCA
501 AGGTATTCTT CGGGGTATGG TCTGGATTGA CAGTCACAAA AAATGGCTCA
551 CTTGTTTTCC CAATATTCTG TAGTGCATTG AGCCAGTATG TTAAGAAATGC
601 ATTATTTTTG CTACTATTGA GAAAATTCAA TGCCTCCAA GCTGATTGT
651 TTTGTGGCAT TAAAGTACTG TCCCAGTGAA GGAAAATATC ACTGGTAGCA
701 TATTGGAAAG CACCCAGTAC TCTTGTCTTCT TCAAATGTTG CTTGGTTTCC
751 TAATAATTTT AGGGCAGTGG GAGCATCAAC AGCCATTATG CATCCATTGT
801 AAGTTTCTTG GAAACCATCT CCACAGACCA TGGTACCAAC CTTGGAGAAC
851 CCAATAGACA GGCCAAGCAA TGTAGGAAAA GGGTTGTGGG AGAAAGTGAA
901 AATATGATGT TGCAATGTAG TAAAAGAGAG AGGCTAAGCA TAGGTCATGG
951 ACCACGTAGG AGAAGGATCG AAGGAGGGAG CGGCGAAAAC AGTGGGGCGG
1001 AATGGCTTGC TTGATCTGAC CGAGCGTAAA CGGAGGCTTC TCGATCGGAA
1051 CTCGATTGAC CGAGCCTCGA TTTTCTCCTT TTATACCCGT CAATTGGCAT
1101 CCTACCCACC CGGCACCCAT TTTGTTGGTT TTAAGAAAAG CAGTTTAAACG
1151 CCAGGTGATG AAACACGAGG CAAAGAAGGC GTGTTGGT

```

Fragmento E, terminador de lectina:

```

AGGGGCCGCC ATGTGACAGA TCGAAGGAAG AAAGTGTAAT AAGACGACTC
TCACTACTCG ATCGCTAGTG ATTGTCATTG TTATATATAA TAATGTTATC
TTTCACAAC TATCGTAATG CATGTGAAAC TATAACACAT TAATCCTACT
TGTCATATGA TAACACTCTC CCCATTTAAA ACTCTTGTCA ATTTAAAGAT
ATAAGATTCT TTAATGATT AAAAAAATA TATTATAAAT TCAATCACTC
CTACTAATAA ATTATTAATT AATATTTATT GATTAATAAA ATACTTATAC
TAATTTAGTC TGAATAGAAT AATTAGATTC TAGAGTCGAC CTGCAGGCAT
GCAAGCTT

```

Fragmento F, promotor S1:

```

1  TCTAGAGGAT CCTATGTTGT AATTTTATAT GGATTAATGA GAATTATTAT
51  TATTCTGTTC TTCGTCTGTG TTTTTTAAGC TAGCTTCCTC GAAGCAGCGT
101 ATAACTTTAA TTTGAATTTG GTTTTGGCGC GTTAGTGAAA TTGCGGCTGT
151 AAACGTGTCA AGTTGTGAGT GGCTGAAATA AGATAATAGA TATATTATTA
201 TTGTTTTAAT TTAATCCGC GAAGCGATAT GTTAAGTGAT AAATGAAACG
251 AAGCGTTTTG ATGACGTCAT ATGTCTCCGT GCCTACGTCA GCACGGGGCT
301 TAGTATTACC CCCGTGCCG GATCAGAGAC ATTTGACCAA TAGTGTACTA
351 GTATAATAGC CCTTGGATTA AATGACACGT GGACGCTCAG GATCTGTGAT
401 GCTAGTGAAG CGCTTAAGCT GAACGAATCT GACGGAAGAG CGTTCACACT
451 TAGATCTAGT TAGCGTACTT AGTACGCGTT GTCTTGGGTC TATAAATAGA
501 GTGCTTCTGA ACAGATTGTT CAGAATTTCA TAGCGCCGAT AACAATGATT
551 GAACAAGATG GATTGCACGC AGTTTCTCCG GCCGCTTGGG TGGAGAGGCT
601 ATTCGGCTAT GACTGGGCAC

```

Figura 13 continuación

Fragmento G, gen *NPTII*:

```

1 AACAGACAAT CGGCTGCTCT GATGCCGCCG TGTTCCGGCT GTCAGCGCAG
51 GGGCGCCCGG TTCTTTTTGT CAAGACCGAC CTGTCCGGTG CCCTGAATGA
101 ACTGCAGGTA AATTTCTAGT TTTTCTCCTT CATTTTCTTG GTTAGGACCC
151 TTTTCTCTTT TTATTTTTTT GAGCTTTGAT CTTTCTTTAA ACTGATCTAT
201 TTTTAAATTG ATTGGTTATG GTGTAAATAT TACATAGCTT TAACTGATAA
251 TCTGATTACT TTATTTCTGT TGTCTATGAT GATGATGATA ACTGCAGGAC
301 GAGGCAGCGC GGCTATCGTG GCTGGCCACG ACGGGCGFTC CTTGCGCAGC
351 TGTGCTCGAC GTTGTCACTG AAGCGGGAAG GGACTGGCTG CTATTGGGCG
401 AAGTGCCGGG GCAGGATCTC CTGTCATCTC ACCTTGCTCC TGCCGAGAAA
451 GTATCCATCA TGGCTGATGC AATGCGGCGG CTGCATACGC TTGATCCGGC
501 TACCTGCCCA TTCGACCACC AAGCGAAACA TCGCATCGAG CGAGCACGTA
551 CTCGGATGGA AGCCGGTCTT GTCGATCAGG ATGATCTGGA CGAAGAGCAT
601 CAGGGGCTCG CGCCAGCCGA ACTGTTCCGC AGGCTCAAGG CGCGCATGCC
651 CGACGGCGAG GATCTCGTCG TGACCCATGG CGATGCCTGC TTGCCGAATA
701 TCATGGTGG

```

Fragmento H, terminador s3:

```

1 AAAATGGCCG CTTTTCTGGA TTCATCGACT GTGGCCGGCT GGGTGTGGCG
51 GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG CTGAAGAGCT
101 TGGCGGCGAA TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC
151 CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA
201 GCGGGACTCT GGGGTTGAA ATGACCGACC AAGCGACGCC CAACCTGCCA
251 TCACGAGATT TCGATTCCAC CGCCGCCTTC TATGAAAGGT TGGGCTTCGG
301 AATCGTTTTC CGGGACGCCG GCTGGATGAT CCTCCAGCGC GGGGATCTCA
351 TGCTGGAGTT CTTCCGCCAC CCCGGGTCGA CGGTATTGAG TTTATGTCTA
401 TTGTAATTGA TAGAGGTTCT ATTAAGATAG AATTATGAGA TGTAATTGTG
451 ATTAATGAAT AAAGAGTTGT TATTATTCTT TGAATTACTC CGCGAAGCGG
501 TGTGTTATGT TTTTGTGGA AGCTTTCTAG ACTCGACTAG AGCGGCCG

```

Figura 13 continuación

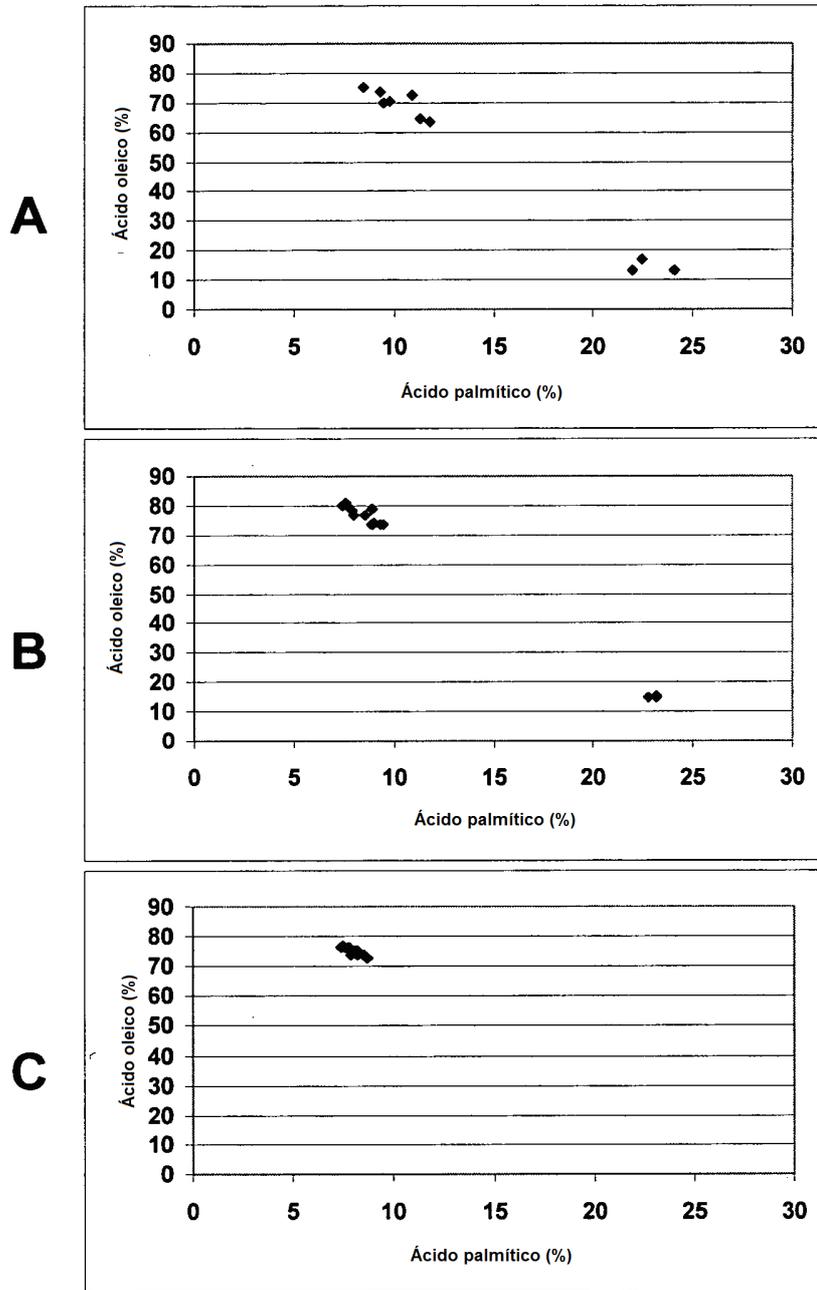


Figura 14

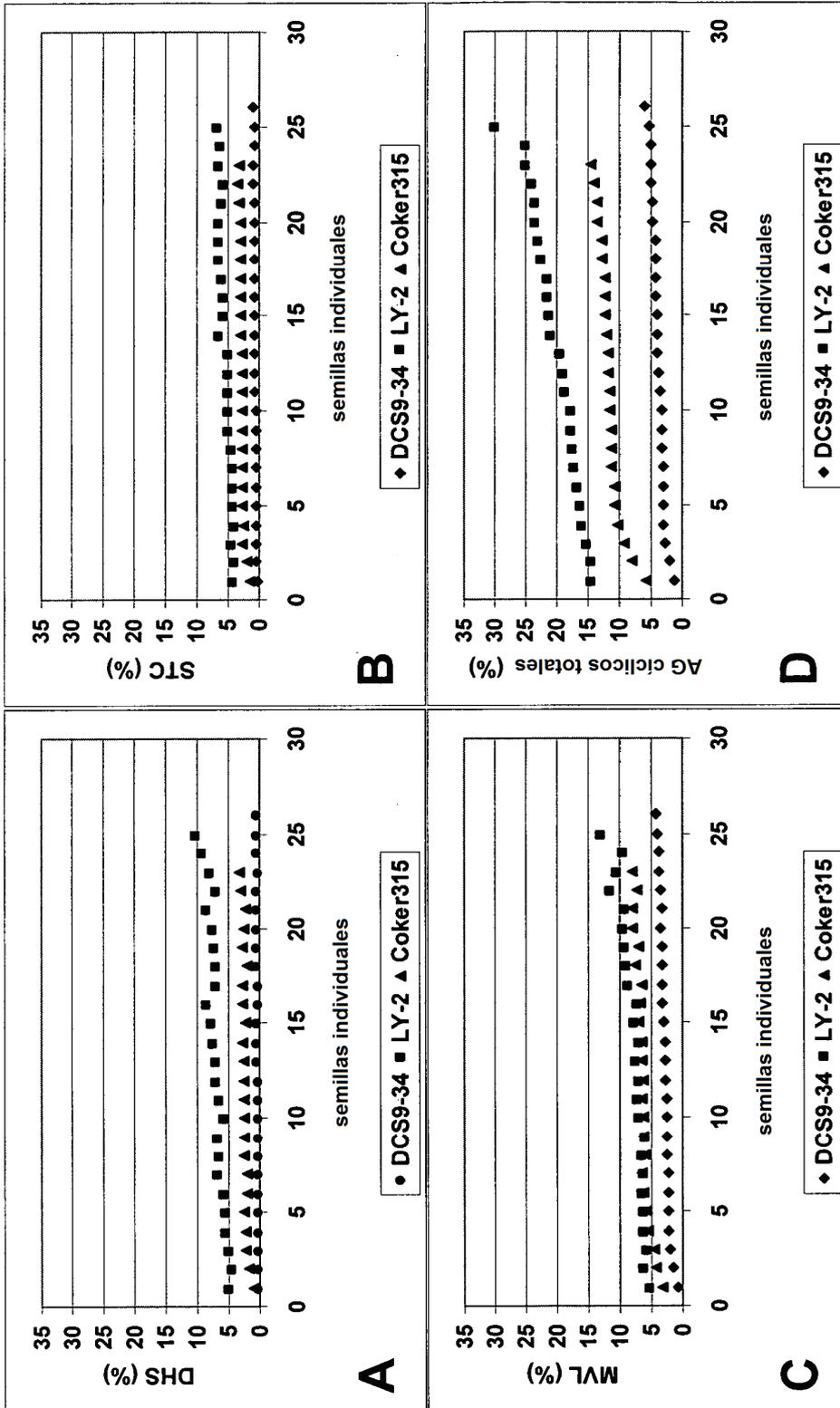


Figura 15