

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 483 365**

51 Int. Cl.:

A01H 5/00 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.12.2003 E 03810076 (4)**

97 Fecha y número de publicación de la concesión europea: **14.05.2014 EP 1571900**

54 Título: **Generación de plantas con contenido de aceite alterado**

30 Prioridad:

18.12.2002 US 434601 P

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

06.08.2014

73 Titular/es:

**AGRIGENETICS, INC. (100.0%)
9330 ZIONSVILLE ROAD
INDIANAPOLIS, IN 46268-1053, US**

72 Inventor/es:

**LIGHTNER, JONATHAN;
COATE, JEREMY, E.;
CLENDENNEN, STEPHANIE, K. y
FEDERSPIEL, NANCY, ANNE**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 483 365 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

Generación de plantas con contenido de aceite alterado

Antecedentes de la invención

5 La capacidad para manipular la composición de las semillas de cultivos, concretamente el contenido y la composición de los aceites de semillas, tiene importantes aplicaciones en la industria agrícola, con respecto a los aceites alimentarios procesados como a aceites para la alimentación animal. Las semillas de los cultivos agrícolas contienen una variedad de componentes valiosos, incluyendo aceite, proteína y almidón. El procesamiento industrial puede separar algunos o todos estos componentes para su venta individual en aplicaciones específicas. Por ejemplo, casi 60% de la cosecha de soja de los Estados Unidos es triturada por la industria de procesamiento de soja. El procesamiento de la soja produce aceite purificado, que se comercializa a un alto valor, mientras que el resto se comercializa principalmente para la alimentación del ganado vacuno de menor valor (US Soybean Board, 2001 Soy Stats). La semilla de canola se tritura para producir aceite y co-producto de harina de canola (Canola Council of Canada). Casi 20% de la cosecha de maíz de los Estados Unidos 1999/2000 fue refinado industrialmente, principalmente para la producción de almidón, etanol y aceite (Corn Refiners Association). De este modo, a menudo es deseable maximizar el contenido de aceite de las semillas. Por ejemplo, para las semillas oleaginosas procesadas tales como la soja y la canola, el aumento del contenido de aceite absoluto de la semilla aumentará el valor de tales granos. Para el maíz procesado se puede desear aumentar o disminuir el contenido de aceite, dependiendo de la utilización de otros componentes principales. La disminución de aceite puede mejorar la calidad del almidón aislado mediante la reducción de sabores no deseados asociados con la oxidación del aceite. Por otra parte, en la producción de etanol, donde el sabor no es importante, el aumento del contenido de aceite puede aumentar el valor global. En muchos cereales forrajeros, tales como el maíz y el trigo, es deseable aumentar el contenido de aceite de la semilla, debido a que el aceite tiene un mayor contenido de energía que otros componentes de semillas tales como los hidratos de carbono. El procesamiento de semillas oleaginosas, como la mayoría de los negocios de procesamiento de grano, es un negocio intensivo en capital: por lo tanto los pequeños desplazamientos en la distribución de los productos de los componentes de escaso valor al componente oleoso de gran valor pueden tener consecuencias económicas considerables para los procesadores de grano.

La manipulación biotecnológica de los aceites puede proporcionar la alteración de la composición y la mejora del rendimiento de aceite. Las alteraciones en la composición incluyen soja y aceite de maíz de alto contenido oleico (Patentes de los Estados Unidos Núms. 6.229.033 y 6.248.939), y semillas que contienen laurato (Patente de los Estados Unidos Núm. 5.639.790), entre otras. El trabajo en la alteración de la composición se ha centrado principalmente en las semillas oleaginosas procesadas pero ha sido fácilmente ampliable a cultivos distintos de semillas oleaginosas, incluyendo el maíz. Si bien existe un interés considerable en el aumento del contenido de aceite, la única biotecnología llevada a la práctica en la actualidad en esta área es la tecnología de Maíz con Alto Contenido de Aceite (HOC por "High-Oil Corn") (DuPont, Patente de los Estados Unidos Núm. 5.704.160). En la HOC se emplean polinizadores de alto contenido de aceite desarrollados por cría selectiva clásica, junto con hembras híbridas (con esterilidad masculina) selectas en un sistema de producción al que se hace referencia como TopCross. El sistema TopCross High Oil eleva el contenido de aceite de los granos recolectados de maíz de ~3,5% a ~7%, mejorando el contenido de energía del grano.

Si bien ha sido fructífero, el sistema de producción HOC presenta limitaciones inherentes. En primer lugar, el sistema de tener un bajo porcentaje de polinizadores responsables del conjunto de semillas de un campo completo contiene riesgos inherentes, concretamente en años de sequía. En segundo lugar, los contenidos de aceite en los actuales campos de HOC han alcanzado una meseta de aproximadamente 9% de aceite. Por último, el maíz con alto contenido de aceite no es esencialmente un cambio bioquímico, sino más bien un mutante anatómico (aumento del tamaño del embrión) que presenta el resultado indirecto de un aumento del contenido de aceite. Por estas razones, sería especialmente valiosa una estrategia alternativa para alto contenido de aceite, concretamente una que derivara de una producción bioquímica alterada.

Los cultivos diana más evidentes para el mercado de aceite procesado son la soja y la colza, y una gran recopilación de trabajo comercial (p. ej., Patente de los Estados Unidos Núm. 5.952.544; Solicitud PCT WO 9411516) demuestra que *Arabidopsis* es un excelente modelo para el metabolismo de aceite en estos cultivos. Escrutinios bioquímicos de la composición del aceite de semillas han permitido identificar genes de *Arabidopsis* para muchas enzimas biosintéticas críticas y han conducido a la identificación de genes ortólogos de importancia agronómica. Por ejemplo, escrutinios en los que se utilizan poblaciones sometidas a mutagénesis química han identificado mutantes de lípidos cuyas semillas presentan una composición de ácidos grasos alterada (B. Lemieux, et al. 1990, Theor. Appl. Genet. 80, 234-240; James DW y Dooner HK (1990) Theor. Appl. Genet. 80, 241-245). Escrutinios de mutagénesis por ADN-t (Feldmann et al., Science 243: 1351-1354, 1989) que detectaban una composición de ácidos grasos alterada identificaron los genes de omega 3 desaturasa (*FAD3*) y delta-12 desaturasa (*FAD2*) (Patente de los Estados Unidos Núm. 5952544; Yadav NS et al. (1993) Plant Physiol. 103, 467-476; Okuley et al., Plant Cell. Enero de 1994; 6(1):147-58). En un escrutinio que se centró en el contenido de aceite en lugar de en la calidad del aceite, se analizaron mutantes inducidos químicamente para determinar semillas rugosas o la densidad de semillas alterada, a

partir de lo cual se infería un contenido alterado de aceite de semillas (Focks N y Benning C, Plant Physiol. 118:91-101, 1998). Otro escrutinio, diseñado para identificar las enzimas implicadas en la producción de ácidos grasos de cadena muy larga, identificó una mutación en el gen que codifica una diacilglicerol aciltransferasa (DGAT) como responsable de la acumulación reducida de triacilglicerol en las semillas (Katavic V. et al, Plant Physiol. Mayor de 1995; 108(1):399-409). Se demostró adicionalmente que la expresión en exceso, específica de semillas, del ADNc de DGAT estaba asociada con el aumento de contenido de aceite en las semillas (Jako et al., Plant Physiol. Junio de 2001; 126(2):S61-74).

Compendio de la Invención

La invención proporciona una planta transgénica que tiene un fenotipo de alto contenido de aceite. La planta transgénica comprende un vector de transformación que comprende una secuencia de nucleótidos que codifica o es complementaria a una secuencia que codifica un polipéptido de citrato sintasa. En realizaciones preferidas, la planta transgénica se selecciona del grupo que consiste de semillas de colza, soja, maíz, girasol, algodón, cacao, cártamo, aceite de palma, palma de coco, lino, ricino y cacahuete. La invención proporciona adicionalmente un método para producir aceite, que comprende cultivar la planta transgénica y recuperar el aceite de dicha planta.

La planta transgénica de la invención se produce mediante un método que comprende introducir en células progenitoras de la planta un vector de transformación vegetal que comprende una secuencia de nucleótidos que codifica o es complementaria a una secuencia que codifica un polipéptido de citrato sintasa, y cultivar las células progenitoras transformadas para producir una planta transgénica, en donde la secuencia de polinucleótidos de citrato sintasa es expresada ocasionando el fenotipo de alto contenido de aceite.

Descripción detallada de la invención

Definiciones

A menos que se indique lo contrario, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado que tendrían para un experto en la técnica de la presente invención. Para las definiciones y los términos de la técnica, los profesionales son particularmente remitidos a Sambrook y col. Molecular Cloning: A Laboratory Manual (Segunda Edición), Cold Spring Harbor Press, Plainview, Nueva York, 1989, y Ausubel FM et al. Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York, NY, 1993. Se ha de entender que esta invención no se limita a la metodología, los protocolos, y los reactivos concretos descritos, ya que estos pueden variar.

Según se utiliza en la presente memoria, el término "vector" se refiere a un constructo de ácido nucleico diseñado para la transferencia entre células anfitrionas diferentes. Un "vector de expresión" se refiere a un vector que tiene la capacidad de incorporar y expresar fragmentos de ADN heterólogos en una célula foránea. Se encuentran disponibles en el mercado muchos vectores de expresión procarióticos y eucarióticos. La selección de vectores de expresión apropiados se encuentra dentro del conocimiento de los expertos en la técnica.

Un constructo o secuencia de ácido nucleico "heterólogo" tiene una porción de la secuencia que no es nativa para la célula vegetal en la que se expresa. Heterólogo, con respecto a una secuencia de control se refiere a una secuencia de control (es decir, promotor o potenciador) que no funciona en la naturaleza regulando el mismo gen cuya expresión está regulando actualmente. Generalmente, las secuencias de ácidos nucleicos heterólogas no son endógenas para la célula o parte del genoma en el cual se encuentran presentes, y se han añadido a la célula, mediante infección, transfección, microinyección, electroporación, o similar. Un constructo de ácido nucleico "heterólogo" puede contener una combinación de secuencia de control/secuencia codificante de ADN que es la misma, o diferente de una combinación de secuencia de control/secuencia codificante de ADN encontrada en la planta nativa.

Según se utiliza en la presente memoria, el término "gen" significa el segmento de ADN implicado en la producción de una cadena polipeptídica, que puede incluir o no regiones que preceden y siguen a la región codificante, p. ej. secuencias no traducidas 5' (5'UTR) o "líder" y secuencias 3'UTR o "tráiler", así como secuencias intermedias (intrones) entre los segmentos codificantes individuales (exones) y la secuencia reguladora no transcrita.

Según se utiliza en la presente memoria, "recombinante" incluye la referencia a una célula o vector, que han sido modificados por medio de la introducción de una secuencia de ácido nucleico heteróloga o a la célula que deriva de una célula así modificada. De este modo, por ejemplo, las células recombinantes expresan genes que no se encuentran en forma idéntica dentro de la forma nativa (no recombinante) de la célula o expresan genes nativos que en cualquier caso se expresan anormalmente, se expresan poco, o no se expresan en absoluto como resultado de la deliberada intervención humana.

Según se utiliza en la presente memoria, el término "expresión génica" se refiere al proceso por el cual se produce un polipéptido basado en la secuencia de ácido nucleico de un gen. El proceso incluye tanto la transcripción como la traducción; en consecuencia, "expresión" puede referirse a una secuencia de polinucleótidos o polipéptidos, o ambas. A veces, la expresión de una secuencia de polinucleótidos no dará lugar a la traducción de proteínas. La "Expresión en exceso" se refiere al aumento de la expresión de una secuencia de polinucleótidos y/o polipéptidos

- con respecto a su expresión en una planta de tipo salvaje (u otra de referencia [p. ej., no transgénica]) y puede relacionarse con una secuencia de origen natural o de origen no natural. La "expresión ectópica" se refiere a la expresión en un momento, en un lugar y/o con un aumento de nivel que no se produce de forma natural en la planta no alterada o de tipo salvaje. La "subexpresión" se refiere a la disminución de la expresión de una secuencia de polinucleótidos y/o polipéptidos, generalmente de un gen endógeno, con respecto a su expresión en una planta de tipo salvaje. Los términos "expresión incorrecta" y "expresión alterada" abarcan la expresión en exceso, la subexpresión y la expresión ectópica.
- El término "introducido" en el contexto de la inserción de una secuencia de ácido nucleico en una célula, significa "transfección" o "transformación" o "transducción" e incluye la referencia a la incorporación de una secuencia de ácido nucleico en una célula eucariota o procariota donde la secuencia de ácido nucleico se puede incorporar al genoma de la célula (por ejemplo, cromosoma, plásmido, plástido, o ADN mitocondrial), convertir en un replicón autónomo, o expresar transitoriamente (por ejemplo, ARNm transfectado).
- Según se utiliza en la presente memoria, una "célula vegetal" se refiere a cualquier célula derivada de una planta, incluyendo células de tejido indiferenciado (p. ej., un callo) así como semillas, polen, propágulos y embriones de plantas.
- Según se utiliza en la presente memoria, los términos "nativo" y "de tipo salvaje" con respecto a un rasgo o fenotipo dado de una planta se refieren a la forma en la que ese rasgo o fenotipo se encuentran en la misma variedad de planta en la naturaleza.
- Según se utiliza en la presente memoria, el término "modificado" con respecto a un rasgo de la planta, se refiere a un cambio en el fenotipo de una planta transgénica con respecto a la planta no transgénica similar. Un "fenotipo de contenido de aceite alterado" se refiere a fenotipo medible de una planta modificada genéticamente, donde la planta presenta un aumento o disminución estadísticamente significativos en el contenido total de aceite (es decir, el porcentaje de masa de la semilla que es aceite), en comparación con una planta similar, pero no modificada. Un alto fenotipo de aceite se refiere a un aumento en el contenido global de aceite.
- Según se utiliza en la presente memoria, una secuencia de polinucleótidos o gen "mutantes" difieren de la correspondiente secuencia de polinucleótidos o gen de tipo salvaje, en términos de secuencia o de expresión, donde la diferencia contribuye a un fenotipo o rasgo vegetal modificado. Con respecto a una planta o línea vegetal, el término "mutante" se refiere a una planta o línea vegetal que tiene un fenotipo o rasgo vegetal modificado, donde el fenotipo o rasgo modificado está asociado con la expresión modificada de una secuencia de polinucleótidos o gen de tipo salvaje.
- Según se utiliza en la presente memoria, el término "T1" se refiere a la generación de plantas a partir de la semilla de plantas T0. La generación T1 es el primer conjunto de plantas transformadas que se puede seleccionar mediante la aplicación de un agente de selección, p. ej., un antibiótico o un herbicida, para el que la planta transgénica contiene el gen de resistencia correspondiente. El término "T2" se refiere a la generación de plantas mediante autofertilización de las flores de plantas T1, seleccionadas previamente como transgénicas. Las plantas T3 se generan a partir de plantas T2, etc. Según se utiliza en la presente memoria, la "progenie directa" de una planta dada procede de la semilla (o, a veces, de otro tejido) de esa planta y está en la generación inmediatamente subsiguiente; por ejemplo, para un linaje dado, una planta T2 es la progenie directa de una planta T1. La "progenie indirecta" de una planta dada procede de la semilla (o de otro tejido) de la progenie directa de esa planta, o de la semilla (o de otro tejido) de generaciones subsiguientes de ese linaje; por ejemplo, una planta T3 es la progenie indirecta de una planta T1.
- Según se utiliza en la presente memoria, el término "parte vegetal" incluye cualquier órgano o tejido vegetal, incluyendo, sin limitación, semillas, embriones, regiones meristemáticas, tejido de callo, hojas, raíces, brotes, gametofitos, esporofitos, polen y microsporas. Las células vegetales pueden obtenerse a partir de cualquier órgano o tejido vegetal y de cultivos preparados a partir de los mismos. La clase de plantas que se pueden utilizar en los métodos de la presente invención es, en general, tan amplia como la clase de plantas superiores susceptibles de técnicas de transformación, incluyendo plantas tanto monocotiledóneas como dicotiledóneas.
- Según se utiliza en la presente memoria, "planta transgénica" incluye una planta que comprende dentro de su genoma un polinucleótido heterólogo. El polinucleótido heterólogo puede ser integrado de forma estable en el genoma, o puede ser extra-cromosómico. Preferiblemente, el polinucleótido de la presente invención se integra de forma estable en el genoma de manera que el polinucleótido se haga pasar a las generaciones sucesivas. Una célula, tejido u órgano vegetal, o una planta en los que se han introducido los polinucleótidos heterólogos se consideran "transformados", "transfectados" o "transgénicos". La progenie directa e indirecta de las plantas o células vegetales transformadas que también contienen el polinucleótido heterólogo también se consideran transgénicas.
- Identificación de plantas con un fenotipo de contenido de aceite alterado
- Se produjeron plantas transgénicas para expresar en exceso diversos genes que codificaban enzimas de la ruta del glioxilato y se sometieron a ensayo las semillas de plantas transgénicas T1 para determinar el fenotipo de contenido de aceite. Se descubrió que la expresión en exceso de la citrato sintasa (At3g58750; Gl# 15231130) confiere un

fenotipo de contenido de aceite alterado (específicamente un fenotipo de alto contenido en aceite en la semilla). Por consiguiente, los genes y/o polipéptidos de citrato sintasa se pueden emplear en el desarrollo de plantas modificadas genéticamente que tienen un fenotipo de contenido en aceite modificado. Los genes de la citrato sintasa se pueden utilizar en la generación de cultivos de semillas oleaginosas que proporcionan un rendimiento mejorado de aceite a partir del procesamiento de semillas oleaginosas y en la generación de cultivos de cereales forrajeros que proporcionan aumento de energía para la alimentación animal. Los genes de citrato sintasa se pueden utilizar adicionalmente para incrementar el contenido de aceite de los cultivos de aceites especiales, con el fin de aumentar el rendimiento de ácidos grasos inusuales deseados. Las plantas transgénicas que han sido modificadas genéticamente para expresar citrato sintasa se pueden utilizar en la producción de aceite, donde las plantas transgénicas se cultivan, y el aceite se obtiene a partir de partes de la planta (p. ej., semillas) utilizando métodos convencionales.

Ácidos nucleicos y polipéptidos de citrato sintasa

Se proporciona el ácido nucleico de citrato sintasa de *Arabidopsis* en el SEC ID NO: 1 y en la entrada GI Núm. 30694870 de Genbank. La secuencia de proteínas correspondiente se proporciona en el SEC ID NO: 2 y en GI Núm. 15231130. Los ácidos nucleicos y/o proteínas que son ortólogos o parálogos de la citrato sintasa de *Arabidopsis*, se describen en el Ejemplo 2 de más abajo.

Según se utiliza en la presente memoria, el término "polipéptido de citrato sintasa" se refiere a una proteína de citrato sintasa completa o un fragmento, derivado (variante), u ortólogo de la misma que son "funcionalmente activos", lo que significa que el fragmento de proteína, derivado, u ortólogo muestran una o más de las actividades funcionales asociadas con el polipéptido de SEQ ID NO: 2. En una realización preferida, un polipéptido de citrato sintasa funcionalmente activo ocasiona un fenotipo alterado de contenido de aceite cuando se expresa incorrectamente en una planta. En una realización preferida adicional, la expresión incorrecta del polipéptido de citrato sintasa ocasiona un fenotipo de alto contenido de aceite en una planta. En otra realización, un polipéptido de citrato sintasa funcionalmente activo es capaz de rescatar actividad citrato sintasa endógena incompleta (incluyendo deficiente) cuando se expresa en una planta o en células vegetales; el polipéptido de rescate puede ser de la misma especie o de una especie diferente a aquella con la actividad incompleta. En otra realización, un fragmento funcionalmente activo de un polipéptido de citrato sintasa completo (es decir, un polipéptido nativo que tiene la secuencia de SEQ ID NO: 2 o un ortólogo de origen natural del mismo) conserva una o más de las propiedades biológicas asociadas con el polipéptido de citrato sintasa completo, tal como la actividad catalítica. Un fragmento de citrato sintasa comprende preferiblemente un dominio de citrato sintasa, tal como un dominio C- o N-terminal o catalítico, entre otros, y preferiblemente comprende al menos 10, preferiblemente al menos 20, más preferiblemente al menos 25, y lo más preferiblemente al menos 50 aminoácidos contiguos de una proteína citrato sintasa. Los dominios funcionales se pueden identificar utilizando el programa PFAM (Bateman A. et al, 1999 Nucleic Acids Res. 27:260-262; el sitio web en pfam.wustl.edu). Las variantes funcionalmente activas de los polipéptidos de citrato sintasa completos o los fragmentos de los mismos incluyen polipéptidos con inserciones, deleciones o sustituciones de aminoácidos que conservan una o más de las propiedades biológicas asociadas con el polipéptido de citrato sintasa completo. En algunos casos, se generan variantes que cambian el procesamiento post-traduccional de un polipéptido de citrato sintasa. Por ejemplo, las variantes pueden tener unas características de transporte proteico o localización proteica alteradas o una vida media proteica alterada en comparación con el polipéptido nativo.

Según se utiliza en la presente memoria, el término "ácido nucleico de citrato sintasa" incluye ácidos nucleicos con la secuencia proporcionada en o complementaria a la secuencia proporcionada en el SEQ ID NO: 1, así como fragmentos, derivados, u ortólogos de los mismos funcionalmente activos. Un ácido nucleico de citrato sintasa de esta invención puede ser ADN, un derivado de ADN genómico o ADNc, o ADN.

En una realización, un ácido nucleico de citrato sintasa funcionalmente activo codifica o es complementario a un ácido nucleico que codifica un polipéptido de citrato sintasa funcionalmente activo. Se encuentra incluido dentro de esta definición el ADN genómico que sirve como molde para un transcrito de ARN primario (es decir, un precursor de ARNm) que requiere procesamiento, tales como corte y empalme, antes de codificar el polipéptido de citrato sintasa funcionalmente activo. Un ácido nucleico de citrato sintasa puede incluir otras secuencias no codificantes, que pueden ser transcritas o no; tales secuencias incluyen 5' y 3' UTR, señales de poliadenilación y secuencias reguladoras que controlan la expresión génica, entre otras, como se conoce en la técnica. Algunos polipéptidos requieren eventos de procesamiento, tales como escisión proteolítica, modificación covalente, etc., con el objeto de que lleguen a ser totalmente activos. Por consiguiente, los ácidos nucleicos funcionalmente activos pueden codificar el polipéptido de citrato sintasa maduro o el pre-procesado, o una forma intermedia. Un polinucleótido de citrato sintasa también puede incluir secuencias codificantes heterólogas, por ejemplo, secuencias que codifican un marcador incluido para facilitar la purificación del polipéptido fusionado o un marcador de transformación.

Un ácido nucleico de citrato sintasa utilizado en los métodos de esta invención comprende una secuencia de ácido nucleico que codifica o es complementaria a una secuencia que codifica un de polipéptido citrato sintasa que tiene una identidad de secuencia de al menos 70%, 75%, 80 %, 85%, 90%, 95% o más con la secuencia de polipéptidos presentada en el SEC ID NO: 2.

En otra realización, un polipéptido de citrato sintasa de la invención comprende una secuencia de polipéptidos que

tiene una identidad de secuencia de al menos 70%, 80%, 85%, 90% o 95% o más con la secuencia del polipéptido de citrato sintasa del SEQ ID NO: 2. En otra realización, un polipéptido de citrato sintasa comprende una secuencia de polipéptidos con una identidad de secuencia de al menos 50%, 60%, 70%, 80%, 85%, 90% o 95% o más con un fragmento funcionalmente activo del polipéptido presentado en el SEC ID NO: 2. En otra realización más, un polipéptido de citrato sintasa comprende una secuencia de polipéptidos con una identidad de al menos 70%, 80%, o 90% con la secuencia del polipéptido del SEQ ID NO: 2 en toda su longitud.

En otro aspecto, una secuencia de polinucleótidos de citrato sintasa es idéntica al menos en 50% a 60% en toda su longitud a la secuencia de ácido nucleico de citrato sintasa presentada como SEQ ID NO: 1, o secuencias de ácido nucleico que son complementarias a tal secuencia de citrato sintasa, y pueden comprender una identidad de secuencia de al menos 70%, 80%, 85%, 90% o 95% o más con la secuencia de la citrato sintasa presentada como SEQ ID NO: 1 o un fragmento funcionalmente activo de la misma, o secuencias complementarias.

Según se utiliza en la presente memoria, "porcentaje (%) de identidad de secuencia" con respecto a una secuencia sujeto especificada, o una porción especificada de la misma, se define como el porcentaje de nucleótidos o aminoácidos en la secuencia derivada candidata idéntica a los nucleótidos o aminoácidos de la secuencia sujeto (o porción especificada de la misma), después de alinear las secuencias e introducir huecos, si fuera necesario para lograr el máximo porcentaje de identidad de secuencia, como se genera por medio del programa WU-BLAST-2.0a19 (Altschul et al., J. Mol. Biol. (1997) 215:403-410; sitio web en blast.wustl.edu/blast/README.html) con los parámetros de búsqueda ajustados a los valores por defecto. Los valores HSP S y HSP S2 son valores dinámicos y son establecidos por el propio programa dependiendo de la composición de la secuencia concreta y la composición de la base de datos concreta contra la cual se está buscando la secuencia de interés. Un "% del valor de identidad" se determina por medio del número de nucleótidos o aminoácidos idénticos coincidentes dividido por la longitud de la secuencia para la que se está refiriendo el porcentaje de identidad. Se determina el "porcentaje (%) de similitud de secuencia de aminoácidos" realizando el mismo cálculo que para determinar el % de identidad de secuencia de aminoácidos, pero incluyendo sustituciones de aminoácidos conservativas además de los aminoácidos idénticos en el cálculo. Una sustitución de aminoácidos conservativa es aquella en la que un aminoácido es sustituido por otro aminoácido que tiene propiedades similares, de manera que el plegamiento o la actividad de la proteína no se vean significativamente afectados. Los aminoácidos aromáticos que pueden ser intercambiados el uno por el otro son fenilalanina, triptófano, y tirosina; los aminoácidos hidrófobos intercambiables son leucina, isoleucina, metionina, y valina; los aminoácidos polares intercambiables son glutamina y asparragina; los aminoácidos alcalinos intercambiables son arginina, lisina e histidina; los aminoácidos ácidos intercambiables son ácido aspártico y ácido glutámico; y los aminoácidos pequeños intercambiables son alanina, serina, treonina, cisteína y glicina.

Las moléculas de ácidos nucleicos derivadas de las moléculas de ácido nucleico objetivo incluyen secuencias que hibridan selectivamente con la secuencia de ácido nucleico del SEQ ID NO: 1. El rigor de la hibridación se puede controlar mediante la temperatura, la fuerza iónica, el pH, y la presencia de agentes desnaturizantes tales como formamida durante la hibridación y el lavado. Las condiciones utilizadas habitualmente son bien conocidas (véanse, p. ej., Current Protocol in Molecular Biology, vol.1, Cap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook y col., Molecular Cloning, Cold Spring Harbor (1989)). En algunas realizaciones, una molécula de ácido nucleico de la invención es capaz de hibridar con una molécula de ácido nucleico que contiene la secuencia de nucleótidos del SEQ ID NO: 1 en condiciones de hibridación rigurosas que son: prehibridación de los filtros que contienen el ácido nucleico durante un período 8 horas a toda la noche a 65°C en una disolución que comprende citrato de fuerza única (SSC del inglés Single Strength Citrate) 6X (1X SSC es NaCl 0,15 M, citrato de sodio 0,015 M, pH 7,0), disolución de Denhardt 5X, de pirofosfato de sodio al 0,05% y 100 µg/ml de ADN de esperma de arenque; hibridación durante 18-20 horas a 65°C en una disolución que contiene 6X SSC, 1X solución de Denhardt, 100 µg/ml de ARNt de levadura y pirofosfato de sodio al 0,05%; y lavado de los filtros a 65°C durante 1 h en una disolución que contiene 0,1 X SSC y SDS (dodecil sulfato de sodio) al 0,1%. En otras realizaciones, se utilizan condiciones de hibridación moderadamente rigurosas que son: pretratamiento de los filtros que contienen el ácido nucleico durante 6 horas a 40°C en una solución que contiene formamida al 35%, 5X SSC, Tris-HCl 50 mM (pH 7,5), EDTA 5 mM, PVP al 0,1%, Ficoll al 0,1%, BSA al 1%, y 500 µg/ml de ADN de esperma de salmón desnaturizado; hibridación durante 18-20 horas a 40°C en una disolución que contiene formamida al 35%, 5X SSC, Tris-HCl 50 mM (pH 7,5), EDTA 5 mM, PVP al 0,02%, Ficoll al 0,02%, BSA al 0,2%, 100 µg/ml de ADN de esperma de salmón, y sulfato de dextrano al 10% (peso/volumen); seguido de lavado dos veces durante 1 hora a 55°C en una disolución que contiene 2X SSC y SDS al 0,1%. Alternativamente, se pueden utilizar condiciones poco rigurosas que comprenden: incubación durante un período de 8 horas a toda la noche a 37°C en una disolución que comprende formamida al 20%, 5 x SSC, fosfato de sodio 50 mM (pH 7,6), 5X disolución de Denhardt, sulfato de dextrano al 10%, y 20 µg/ml de ADN de esperma de salmón sometido a cizalla desnaturizado; hibridación en el mismo tampón durante 18 a 20 horas; y lavado de los filtros en 1 x SSC a aproximadamente 37°C durante 1 hora.

Como resultado de la degeneración del código genético, se pueden producir diversas secuencias polinucleotídicas que codifiquen un polipéptido de citrato sintasa. Por ejemplo, se pueden seleccionar codones para aumentar la velocidad a la que tiene lugar la expresión del polipéptido en una especie anfitriona concreta, de acuerdo con la óptima utilización de codones dictada por el organismo anfitrión concreto (véase, p. ej., Nakamura et al, 1999, Nucleic Acids Res 27:292). Tales variantes de secuencia se pueden utilizar en los métodos de esta invención.

En los métodos de la invención se pueden utilizar ortólogos de la citrato sintasa de *Arabidopsis*. Los métodos para

identificar los ortólogos en otra especie vegetal son conocidos en la técnica. Normalmente, los ortólogos de diferentes especies conservan la misma función a causa de la presencia de uno o más motivos proteicos y/o estructuras tridimensionales. En la evolución, cuando un evento de duplicación génica sigue a la especiación, un único gen en una especie, tal como *Arabidopsis*, puede corresponder a múltiples genes (parálogos) en otra. Según se utiliza en la presente memoria, el término "ortólogos" incluye los parálogos. Cuando se dispone de datos de secuencias para una especie vegetal concreta los ortólogos son generalmente identificados mediante un análisis de homología de secuencias, tal como un análisis por BLAST, utilizando normalmente secuencias proteicas señuelo. Se asignan secuencias a un posible ortólogo si la secuencia más satisfactoria del resultado del BLAST directo recupera la secuencia original en cuestión en el BLAST inverso (Huynen MA y Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA et al., Genome Research (2000) 10:1204-1210). Se pueden utilizar programas para el alineamiento múltiple de secuencias, tal como CLUSTAL (Thompson JD et al, 1994, Nucleic Acids Res 22:4673-4680) para resaltar regiones y/o restos conservados de proteínas ortólogas y generar árboles filogenéticos. En un árbol filogenético que representa múltiples secuencias homólogas de diversas especies (p. ej., recuperadas a través del análisis por BLAST), las secuencias ortólogas de dos especies aparecen generalmente más cerca en el árbol con respecto a todas las demás secuencias de estas dos especies. El enhebramiento (threading) estructural u otros análisis del plegamiento de proteínas (p. ej. utilizando soporte lógico de ProCeryon, Biosciences, Salzburgo, Austria) también pueden identificar ortólogos potenciales. Los métodos de hibridación de ácido nucleico también se pueden utilizar para hallar genes ortólogos y se prefieren cuando no se dispone de datos sobre las secuencias. La PCR degenerada y el escrutinio de bibliotecas de ADNc o ADN genómico son métodos habituales para hallar secuencias génicas relacionadas y son bien conocidos en la técnica (véanse, p. ej., Sambrook, más arriba; Dieffenbach C y Dveksler G (Eds.) PCR Primer: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, 1989). Por ejemplo, en Sambrook *et al.*, más arriba, se describen métodos para generar una biblioteca de ADNc a partir de la especie vegetal de interés y sondear la biblioteca con sondas génicas parcialmente homólogas. Como sonda, se puede utilizar una porción muy conservada de la secuencia codificante de la citrato sintasa de *Arabidopsis*. Los ácidos nucleicos ortólogos de citrato sintasa pueden hibridar con el ácido nucleico del SEQ ID NO: 1 en condiciones de rigor elevado, moderado, o bajo. Después de la amplificación o el aislamiento de un segmento de un supuesto ortólogo, ese segmento puede ser clonado y secuenciado mediante técnicas convencionales y ser utilizado como sonda para aislar un clon de ADNc o genómico completo. Alternativamente, es posible iniciar un proyecto EST para generar una base de datos sobre información de secuencias para la especie vegetal de interés. En otro enfoque, para el aislamiento de ortólogos se utilizan anticuerpos que unen específicamente a polipéptidos de citrato sintasa conocidos (véase, p. ej., Harlow E y Lane D, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999, New York). El análisis mediante transferencia Western permite determinar que un ortólogo de citrato sintasa (esto es, una proteína ortóloga) está presente en un extracto bruto de una especie vegetal concreta. Cuando se observa reactividad, la secuencia que codifica el ortólogo candidato se puede aislar escrutando bibliotecas de expresión que representan la especie vegetal concreta. Se pueden construir bibliotecas de expresión en una diversidad de vectores asequibles comercialmente, incluyen lambda gt11, como se describe en Sambrook, *et al.*, más arriba. Una vez que se han identificado el ortólogo o los ortólogos candidatos por medio de cualquiera de estos métodos, se utilizan secuencias ortólogas candidatas como señuelo ("problema") para el BLAST inverso frente a secuencias de *Arabidopsis* u otra especie en la que se han identificado secuencias de ácido nucleico y/o polipéptido de la citrato sintasa.

Se pueden obtener ácidos nucleicos y polipéptidos de citrato sintasa utilizando cualquier método disponible por ejemplo, son bien conocidos en la técnica los mecanismos para aislar secuencias de ADNc o ADBN genómico de interés escrutando bibliotecas de ADN utilizando la reacción en cadena de la polimerasa (PCR) como se ha descrito anteriormente. Alternativamente, se puede sintetizar una secuencia de ácido nucleico. Se puede utilizar cualquier método conocido, tal como la mutagénesis dirigida al sitio (Kunkel TA et al., Methods Enzymol. 204:125-39, 1991), para introducir los cambios deseados en un ácido nucleico clonado.

En general, los métodos de la invención implican la incorporación de la forma deseada del ácido nucleico de la citrato sintasa en un vector de expresión vegetal para la transformación de células vegetales, y que el polipéptido de citrato sintasa se exprese en la planta anfitriona.

Una molécula de ácido nucleico de citrato sintasa aislada no está en la forma ni en la forma ni en el ambiente en el que se expresa en la naturaleza y es identificada y separada de al menos una molécula de ácido nucleico contaminante con la que se asocia normalmente en la fuente natural del ácido nucleico de la citrato sintasa. Sin embargo, una molécula de ácido nucleico de citrato sintasa aislada incluye moléculas de ácido nucleico de citrato sintasa contenidas en células que expresan normalmente la citrato sintasa donde, por ejemplo, la molécula de ácido nucleico está en una localización cromosómica distinta a la de las células naturales.

Generación de plantas modificadas genéticamente con un fenotipo alterado de contenido de aceite

Se pueden utilizar ácidos nucleicos y polipéptidos de citrato sintasa en la generación de plantas modificadas genéticamente que tengan un fenotipo de contenido de aceite modificado. Según se utiliza en la presente memoria, un "fenotipo de contenido de aceite modificado" puede hacer referencia a un contenido de aceite modificado en cualquier parte de la planta; el contenido de aceite modificado se observa a menudo en las semillas. En una realización preferida, la expresión alterada del gen de la citrato sintasa en una planta se utiliza para generar plantas con un fenotipo de alto contenido en aceite.

Los métodos descritos en la presente memoria son generalmente aplicables a todas las plantas. En una realización preferida, la invención se refiere a plantas productoras de aceite, que producen y almacenan triacilglicerol en órganos específicos, principalmente en semillas. Tales especies incluyen soja (*Glycine max*), colza y canola (incluyendo *Brassica napus*, *B. campestris*), girasol (*Helianthus annuus*), algodón (*Gossypium hirsutum*), maíz (*Zea mays*), cacao (*Theobroma cacao*), cártamo (*Carthamus tinctorius*), palma de aceite (*Elaeis guineensis*), palma de coco (*Cocos nucifera*), lino (*Linum usitatissimum*), ricino (*Ricinus communis*) y cacahuete (*Arachis hypogaea*). La invención también se puede referir a plantas portadoras de frutos y hortalizas, plantas productoras de grano, plantas productoras de nueces, especies de *Brassica* de ciclo rápido, alfalfa (*Medicago sativa*), tabaco (*Nicotiana*), gramíneas (familia Poaceae), otros cultivos de plantas forrajeras, y especies salvajes que pueden ser una fuente de ácidos grasos únicos.

El experto en la técnica reconocerá que en este campo existe una gran variedad de técnicas de transformación y que se están apareciendo continuamente técnicas nuevas. Cualquier técnica que sea adecuada para la planta anfitriona diana puede ser empleada dentro del alcance de la presente invención. Por ejemplo, los constructos se pueden introducir en una variedad de formas incluyendo, pero no limitadas a una hebra de ADN, en un plásmido, o en un cromosoma artificial. La introducción de los constructos en las células vegetales diana se puede completar por medio de una variedad de técnicas, incluyendo, pero no limitadas a, transformación mediada por *Agrobacterium*, electroporación, microinyección, bombardeo de microproyectiles, co-precipitación de fosfato cálcico-ADN o transformación mediada por liposomas de un ácido nucleico heterólogo. La transformación de la planta es preferiblemente permanente, es decir, mediante integración de los constructos de expresión introducidos en el genoma de la planta anfitriona, de manera que los constructos introducidos se hagan pasar a las generaciones sucesivas de la planta. Dependiendo del uso pretendido, un constructo de ácido nucleico heterólogo que comprende un polinucleótido de citrato sintasa puede codificar la proteína completa o una porción biológicamente activa de la misma.

En una realización, se pueden usar sistemas vectores binarios basados en Ti para transferir polinucleótidos. Los vectores binarios de *Agrobacterium* convencionales son conocidos por los expertos en la técnica, y muchos son asequibles comercialmente (p. ej., pBI121 Clontech Laboratories, Palo Alto, CA).

El procedimiento óptimo para la transformación de plantas con vectores de *Agrobacterium* variará con el tipo de planta que se esté transformando. Los métodos ilustrativos para la transformación mediada por *Agrobacterium* incluyen la transformación de explantes de tejido de hipocotilo, punta de brote, tallo u hoja, procedente de plántulas y/o plántulas estériles. Tales plantas transformadas se pueden reproducir sexualmente, o mediante cultivo de células o tejidos. La transformación con *Agrobacterium* para un gran número de tipos diferentes de plantas se ha descrito previamente, y en la bibliografía científica se pueden encontrar métodos para dicha transformación. Son particularmente relevantes los métodos para transformar cultivos comercialmente importantes, tales como los de semillas de colza (De Block et al., *Plant Physiol.* (1989) 91:694-701), girasol (Everett et al., *Bio/Technology* (1987) 5:1201), y soja (Christou et al., *Proc. Natl. Acad. Sci USA* (1989) 86:7500-7504; Kline et al., *Nature* (1987) 327:70).

Se puede regular la expresión (incluyendo transcripción y traducción) de citrato sintasa con respecto al nivel de expresión, el tipo o los tipos de tejido en los que tiene lugar la expresión y/o la fase de desarrollo de la expresión. Se encuentran disponibles numerosas secuencias reguladoras heterólogas (p. ej. promotores y potenciadores) para controlar la expresión de un ácido nucleico de citrato sintasa. Estos incluyen promotores constitutivos, inducibles y regulables, así como promotores y potenciadores que controlan la expresión de un modo específico tisular o temporal. Los promotores constitutivos ilustrativos incluyen el promotor E4 de la frambuesa (Patente de los Estados Unidos Núms. 5.783.393 y 5.783.394), el promotor 35S de CaMV (Jones JD et al, *Transgenic Res* 1:285-297 1992), el promotor de CsVMV (Verdaguer B et al., *Plant Mol Biol* 37:1055-1067, 1998) y el promotor de la actina de melón (solicitud PCT publicada WO 0056863). Los promotores específicos de tejido ilustrativos incluyen los promotores E4 y E8 de tomate (Patente de los Estados Unidos Núm. 5.859.330) y el promotor del gen 2All de tomate (Van Haaren MJJ et al., *Plant Mol Bio* 21:625-640, 1993).

En una realización preferida, la expresión de la citrato sintasa se encuentra bajo el control de secuencias reguladoras de genes cuya expresión está asociada con el desarrollo precoz de semillas y/o embriones. Los genes de legumbres cuyos promotores están asociados con el desarrollo precoz de semillas y embriones incluyen *legumina* de *V. faba* (Baumlein et al., 1991, *Mol Gen Genet* 225:121-8; Baumlein et al., 1992, *Plant J* 2:233-9), *usp* de *V. faba* (Fiedler et al., 1993, *Plant Mol Biol* 22:669-79), *convicilina* de guisante (Bown et al., 1988, *Biochem J* 251:717-26), *lectina* de guisante (dePater et al., 1993, *Plant Cell* 5:877-86), *beta faseolina* de *P. vulgaris* (Bustos et al., 1991, *EMBO J* 10:1469-79), *DLEC2* y *PHS [beta]* de *P. vulgaris* (Bobb et al, 1997, *Nucleic Acids Res* 25:641-7), y *beta-Conglicinina* de soja, proteína de almacenamiento 7S (Chamberland et al., 1992, *Plant Mol Biol* 19:937-49). Los genes de cereales cuyos promotores están asociados con el desarrollo precoz de semillas y embriones incluyen *glutelina* de arroz ("GluA-3", Yoshihara y Takaiwa, 1996, *Plant Cell Physiol* 37:107-11; "GluB-1", Takaiwa et al., 1996, *Plant Mol Biol* 30:1207-21; Washida et al., 1999, *Plant Mol Biol* 40:1-12; "Gt3", Leisy et al., 1990, *Plant Mol Biol* 14:41-50), *prolamina* de arroz (Zhou y Fan, 1993, *Transgenic Res* 2:141-6), *prolamina* de trigo (Hammond-Kosack et al., 1993, *EMBO J* 12:545-54), *zeína* de maíz (Z4, Matzke et al., 1990, *Plant Mol Biol* 14:323-32), y *B-hordeínas* de cebada (Entwistle et al., 1991, *Plant Mol Biol* 17:1217-31). Otros genes cuyos promotores está asociados con el desarrollo precoz de semillas y embriones incluyen GL07A de palma de aceite (globulina 7S, Morcillo et al., 2001, *Physiol Plant* 112:233-243), *napina* de *Brassica napus*, proteína de almacenamiento 2S, y gen napA (Josefsson et

al., 1987, J Biol Chem 262:12196-201; Stalberg et al., 1993, Plant Mol Biol 1993 23:671-83; Ellerstrom et al., 1996, Plant Mol Biol 32:1019-27), *oleosina* de *Brassica napus* (Keddie et al., 1994, Plant Mol Biol 24:327-40), *oleosina* de *Arabidopsis* (Plant et al., 1994, Plant Mol Biol 25:193-205), FAE1 de *Arabidopsis* (Rossak et al., 2001, Plant Mol Biol 46:717-25), conA de *Canavalia gladiata* (Yamamoto et al., 1995, Plant Mol Biol 27:729-41), y estrictosidina sintasa de *Catharanthus roseus* (Str, Ouwerkerk y Memelink, 1999, Mol Gen Genet 261:635-43). En otra realización preferida, se utilizan secuencias reguladoras de genes expresados durante la biosíntesis de aceite (véase, p. ej., la Patente de los Estados Unidos Núm. 5.952.544). Son promotores alternativos los de los genes de proteínas de almacenamiento de plantas (Bevan et al., 1993, Philos Trans R Soc Lond B Biol Sci 342:209-15).

En otro aspecto más, en algunos casos puede ser deseable inhibir la expresión de la citrato sintasa endógena en una célula anfitriona. Los métodos ilustrativos para poner en práctica este aspecto de la invención incluyen, pero no se limitan a la supresión antisentido (Smith, et al., Nature 334:724-726, 1988; van derKrol et al., Biotechniques (1988) 6:958-976); la co-supresión (Napoli, et al., Plant Cell 2:279-289, 1990); ribozimas (Publicación PCT WO 97/10328); y combinaciones de efectores y antisentido (Waterhouse, et al., Proc. Natl. Acad. Sci. USA 95:13959-13964, 1998). En los métodos para la supresión de secuencias endógenas en una célula anfitriona se emplean típicamente la transcripción o la transcripción y traducción de al menos una porción de la secuencia que se va a suprimir. Tales secuencias pueden ser homólogas para las regiones codificantes así como no codificantes de la secuencia endógena. En la inhibición antisentido se puede utilizar la secuencia de ADNc completa (Sheehy et al., Proc. Natl. Acad. Sci. USA (1988) 85:8805-8809), una secuencia de ADNc parcial incluyendo fragmentos de la secuencia codificante 5' (Cannon et al., Plant Molec. Biol. (1990) 15:39-47), o secuencias no codificantes 3' (Ch'ng et al., Proc. Natl. Acad. Sci. USA (1989) 86:10006-10010). En las técnicas de co-supresión se puede utilizar la secuencia completa de ADNc (Napoli et al., más arriba; van der Krol et al., The Plant Cell (1990) 2:291-299) o una secuencia parcial de ADNc (Smith et al., Mol. Gen. Genetics (1990) 224:477-481).

Se pueden llevar a cabo ensayos moleculares y genéticos convencionales para analizar adicionalmente la asociación entre un gen y un fenotipo observado. Las técnicas ilustrativas se describen a continuación.

25 1. Análisis de ADN/ARN

Los patrones de expresión génica específicos de la fase y el tejido en líneas mutantes frente a líneas de tipo salvaje se pueden determinar, por ejemplo, mediante hibridación in situ. Se puede realizar el análisis del estado de metilación del gen, especialmente en regiones reguladoras colindantes. Otras técnicas adecuadas incluyen la expresión en exceso, la expresión ectópica, la expresión en otra especie vegetal y la inactivación (knock-out) génica (genética inversa, inactivación dirigida, silenciamiento génico inducido víricamente [VIGS, véase Baulcombe D, Arch Virol Supl 15:189-201, 1999]).

En una aplicación preferida, se utiliza el perfilado de la expresión, generalmente mediante análisis con micromatrices, para medir simultáneamente diferencias o cambios inducidos en la expresión de muchos genes diferentes. Los mecanismos para el análisis con micromatrices son bien conocidos en la técnica (Scheda M et al., Science (1995) 270:467-470; Baldwin D et al., Cor Opin Plant Biol. 2(2):96-103, 1999; Dangond F, Physiol Genomics (2000) 2:53-58; van Hal NL et al., J Biotechnol (2000) 78:271-280; Richmond T y Somerville S, Curr Opin Plant Biol (2000) 3:108-116). Se puede llevar a cabo el perfilado de la expresión de líneas etiquetadas individuales. Tales análisis pueden identificar otros genes que son regulados coordinadamente como consecuencia de la expresión en exceso del gen de interés, lo que puede ayudar a colocar un gen desconocido en una ruta concreta.

40 2. Análisis de productos génicos

El análisis de productos génicos puede incluir la expresión de proteínas recombinantes, la producción de antisueros, la inmunolocalización, análisis bioquímicos para determinar la actividad catalítica u otra actividad, el análisis del estado de fosforilación, y el análisis de la interacción con otras proteínas por medio de análisis de dos híbridos en levadura.

45 3. Análisis de rutas

El análisis de rutas puede incluir la colocación de un gen o producto génico en una ruta bioquímica, metabólica o de señalización concreta basándose en su fenotipo de expresión incorrecta o mediante su homología secuencial con genes relacionados. Alternativamente, el análisis puede comprender cruces genéticos con líneas de tipo salvaje y otras líneas mutantes (creándose mutantes dobles) con el fin de ordenar el gen en una ruta, o determinar el efecto de una mutación sobre la expresión de genes "informadores" aguas abajo en una ruta.

Generación de plantas mutadas con un fenotipo de contenido de aceite alterado

La invención proporciona adicionalmente un método para identificar plantas que tienen mutaciones en la citrato sintasa endógena que confieren un contenido alterado de aceite, y generar una progenie de estas plantas con el contenido de aceite alterado que no están modificadas genéticamente. En un método, denominado "TILLING" (detección de lesiones locales inducidas, en genomas; del inglés Targeting Induced Local Lesions In Genomes), se inducen mutaciones en la semilla de una planta de interés, por ejemplo, utilizando tratamiento con EMS. Las plantas resultantes se cultivan y autofertilizan, y se utiliza la progenie para preparar muestras de ADN. Se utiliza una PCR

específica de citrato sintasa para identificar si una planta mutada presenta una mutación de citrato sintasa. Las plantas que presentan mutaciones de citrato sintasa se pueden someter a ensayo a continuación para determinar el contenido de aceite alterado, o alternativamente, las plantas se pueden someter a ensayo para determinar el contenido de aceite alterado, y a continuación utilizar la PCR específica de citrato sintasa para determinar si una planta que tiene un contenido alterado tiene un gen mutado de citrato sintasa. El TILLING permite identificar mutaciones que pueden alterar la expresión de genes específicos o la actividad de las proteínas codificadas por estos genes (véanse Colbert et al (2001) *Plant Physiol* 126:480-484; McCallum et al (2000) *Nature Biotechnology* 18:455-457).

En otro método, se puede utilizar un enfoque de gen candidato/Locus de Rasgo Cuantitativo (QTL; del inglés, Quantitative Trait Locus) en un programa de cría asistido por marcadores para identificar, en el gen de citrato sintasa o en ortólogos de citrato sintasa, alelos o mutaciones que puedan conferir un contenido alterado de aceite (véanse Bert et al., *Theor Appl Genet.* Junio 2003; 107(1):181-9; y Lionneton et al., *Genome.* Diciembre de 2002; 45(6):1203-15). De este modo, en un aspecto adicional de la invención, se utiliza un ácido nucleico de citrato sintasa para identificar si una planta que tiene un contenido alterado de aceite tiene una mutación en la citrato sintasa endógena o tiene un alelo concreto que causa el contenido alterado de aceite.

Ejemplos

Ejemplo 1

Generación de plantas transgénicas que expresan en exceso citrato sintasa

Un subgrupo de genes que codificaban las enzimas de la ruta del glioxilato se expresaron en exceso en *Arabidopsis* de tipo salvaje, y las semillas de las plantas transgénicas T1 se sometieron a ensayo para determinar un fenotipo de elevado contenido de aceite. Para expresar en exceso los genes, el ADN genómico se amplificó mediante PCR con cebadores específicos para cada gen. El producto de la PCR se clonó detrás del promotor de CsVMV en un vector de ADN-T que contenía el gen NPT2 (que sirve como marcador seleccionable) y se transformó en plantas de *Arabidopsis* de tipo salvaje. Las plantas transgénicas se seleccionaron mediante germinación de las semillas sobre medio de agar que contenía kanamicina. Las plántulas resistentes a kanamicina se trasplantaron a tierra y se cultivaron hasta la madurez. Las plantas (Col-0) de tipo salvaje no transformadas se utilizaron como control. Las semillas se hicieron germinar sobre medio de agar (que carecía de kanamicina) y las plántulas se trasplantaron a tierra y se cultivaron hasta la madurez. El contenido de aceite de las semillas cosechadas de plantas tanto transgénicas como de control se midió utilizando la espectroscopia infrarroja cercana (NIR). El espectro de infrarrojos NIR se capturó utilizando un Bruker 22 N/F. Se utilizó el soporte lógico Bruker para estimar el contenido de aceite total de la semilla utilizando los datos del análisis NIR y los métodos de referencia de acuerdo con las instrucciones de los fabricantes.

Los resultados demostraron que la expresión en exceso del gen de citrato sintasa de *Arabidopsis* (At3g58750) confiere un fenotipo de elevado contenido en aceite a la semilla. Las semillas de 4 líneas transgénicas que expresaban en exceso At3g58750 (que codificaban citrato sintasa) tuvieron más aceite que todas las plantas de control sometidas a ensayo. Los valores oscilaron entre 41,6% y 41,2% de aceite para las plantas transgénicas, mientras los valores de las semillas de control oscilaron entre 40,9% y 38,7%. El contenido medio de aceite en las semillas de control fue de 39,9%. La expresión en exceso de la citrato sintasa puede conferir un incremento tan grande como 7% del aceite de la semilla.

Ejemplo 2

Análisis de citrato sintasa de *Arabidopsis*

Los análisis de secuencias se realizaron con BLAST (Altschul et al., 1997, *J. Mol. Biol.* 215:403-410), PFAM (Bateman et al., 1999, *Nucleic Acids Res.* 27:260-262), PSORT (Nakai K y Horton P, 1999, *Trends Biochem Sci.* 24:34-6), y/o CLUSTAL (Thompson JD et al, 1994, *Nucleic Acids Res.* 22:4673-4680). Se identificaron numerosos ortólogos que tenían una alta identidad de secuencia, y por lo tanto se esperaba que también confirieran un fenotipo de alto contenido de aceite cuando se expresaran en exceso en una planta. Las citrato sintasa ortólogas de diversas especies vegetales incluyen: GI Núm. 1345933, y GI Núm. 975633 (*Cucurbita* cv.); GI Núm. 15231128, y GI Núm. 11268305 (*Arabidopsis thaliana*); GI Núm. 8928010 (*Daucus carota*); GI Núm. 6647461 (*Solanum tuberosum*); GI Núm. 1352088 (*Citrus maxima*); GI Núm. 15982952 (*Prunus persica*); GI Núm. 11066954 (*Oryza sativa*); GI Núm. 2300712 (*Nicotiana tabacum*); y GI Núm. 2300710 (*Beta vulgaris*).

LISTA DE SECUENCIAS

- <110> Agrinomics LLC
- <120> Generación de plantas con contenido de aceite alterado
- <130> AG03-084C-PC
- 5 <150> 60/434.601
- <151> 18-12-2002
- <160> 2
- <170> PatentIn versión 3.2
- <210> 1
- 10 <211> 1545
- <212> ADN
- <213> Arabidopsis thaliana
- <400> 1

ES 2 483 365 T3

atggagattt cgagagagt gaaagctaga ttagccgttc tcaccgcgca tttggcggtg 60
 tctgataccg tccgattgga acaggtgttg cggcgatcg cgccatggtg tacatcggct 120
 cacattaccg ctgcacctca tggatcactc aaaggaaact tgacgatcgt cgatgagcgt 180
 acggggaaga aatatcaggt ccctgtctca gagcatggta ccgttaaagc cgttgatctc 240
 aagaagataa cgacggggaa ggatgataag gggctgaagt tgtacgatcc tggttacttg 300
 aacacggctc cggttcgatc ttcgatttgt tacatcgacg gagatgaagg aatcttacgt 360
 tatcggggat acccaattga agagttggct gagagcagta cttttattga ggttgcttat 420
 ctctcatgt atggaaatct gccttctcaa agtcagctag ctgattggga gttcactggt 480
 tctcagcatt cagctgtgcc acaaggagta ttggatatca tacagtccat gcctcatgat 540
 gcacacccaa tgggagttct tgtgagtgcc atgagtgac tttctatctt tcaccctgat 600
 gcaaatcctg ctcttagtgg ccaagacatt tacaagtcaa aacaagttcg tgataaacag 660
 attgttcgca ttctcggaaa ggcaccaaca attgcagcag ctgcttattt gaggacggca 720
 ggcaggcctc ctgttcttcc ttcggccaac ctttcttatt cagagaattt cctctatatg 780
 ctggattcaa tgggcaatag gtcttacaag cctaatactc gtttggtctcg agtgctggac 840
 atcctcttca tactgcatgc tgaacatgaa atgaactgct ctactgctgc tgctcggcat 900
 cttgcctcta gtgggtgttg tgtgtacacc gcatgtgctg gagctgttgg ggcgctttat 960
 ggtccacttc atggtggcgc gaacgaggcc gtgcttaaga tgtagcaga gattgggact 1020
 gctgaaaata ttccagattt cattgaaggc gtgaagaaca gaaagaggaa gatgtcaggt 1080
 tttggacatc gtgtttacaa aaactatgac ccccgagcaa aagttataaa aaaactggca 1140
 gatgaagtgt tctccattgt tggtagggat cctctcatcg aggtagcagt tgctctagag 1200
 aaggcggcac tgtctgatga atattttgtt aagagaaagc tgtacccaaa tgttgatttc 1260

 tactctggat taatctatag ggcaatggga ttcccaccag aattcttcac agtctgttc 1320
 gcagtcccgc gtatggctgg atacttgtca cactggcgtg agtcgtaga tgatcctgac 1380
 actaggatca tgagaccca acaggcctat actggagtgt ggatgaggca ttacgagcca 1440
 gtgagagaac gaacgttate aagtgattcg gataaggata agtttggta agtttcatt 1500
 tcgaatgcat caagaaggcg tttagctgga tcatctgccc tttag 1545

<210> 2

<211> 514

<212> PRT

5 <213> Arabidopsis thaliana

<400> 2

ES 2 483 365 T3

Met Glu Ile Ser Gln Arg Val Lys Ala Arg Leu Ala Val Leu Thr Ala
 1 5 10 15

His Leu Ala Val Ser Asp Thr Val Gly Leu Glu Gln Val Leu Pro Ala
 20 25 30

Ile Ala Pro Trp Cys Thr Ser Ala His Ile Thr Ala Ala Pro His Gly
 35 40 45

Ser Leu Lys Gly Asn Leu Thr Ile Val Asp Glu Arg Thr Gly Lys Lys
 50 55 60

Tyr Gln Val Pro Val Ser Glu His Gly Thr Val Lys Ala Val Asp Leu
 65 70 75 80

Lys Lys Ile Thr Thr Gly Lys Asp Asp Lys Gly Leu Lys Leu Tyr Asp
 85 90 95

Pro Gly Tyr Leu Asn Thr Ala Pro Val Arg Ser Ser Ile Cys Tyr Ile
 100 105 110

Asp Gly Asp Glu Gly Ile Leu Arg Tyr Arg Gly Tyr Pro Ile Glu Glu
 115 120 125

Leu Ala Glu Ser Ser Thr Phe Ile Glu Val Ala Tyr Leu Leu Met Tyr
 130 135 140

Gly Asn Leu Pro Ser Gln Ser Gln Leu Ala Asp Trp Glu Phe Thr Val
 145 150 155 160

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

ES 2 483 365 T3

Met Pro His Asp Ala His Pro Met Gly Val Leu Val Ser Ala Met Ser
 180 185 190

Ala Leu Ser Ile Phe His Pro Asp Ala Asn Pro Ala Leu Ser Gly Gln
 195 200 205

Asp Ile Tyr Lys Ser Lys Gln Val Arg Asp Lys Gln Ile Val Arg Ile
 210 215 220

Leu Gly Lys Ala Pro Thr Ile Ala Ala Ala Tyr Leu Arg Thr Ala
 225 230 235 240

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

Phe Leu Tyr Met Leu Asp Ser Met Gly Asn Arg Ser Tyr Lys Pro Asn
 260 265 270

Pro Arg Leu Ala Arg Val Leu Asp Ile Leu Phe Ile Leu His Ala Glu
 275 280 285

His Glu Met Asn Cys Ser Thr Ala Ala Ala Arg His Leu Ala Ser Ser
 290 295 300

Gly Val Asp Val Tyr Thr Ala Cys Ala Gly Ala Val Gly Ala Leu Tyr
 305 310 315 320

Gly Pro Leu His Gly Gly Ala Asn Glu Ala Val Leu Lys Met Leu Ala
 325 330 335

Glu Ile Gly Thr Ala Glu Asn Ile Pro Asp Phe Ile Glu Gly Val Lys
 340 345 350

Asn Arg Lys Arg Lys Met Ser Gly Phe Gly His Arg Val Tyr Lys Asn
 355 360 365

Tyr Asp Pro Arg Ala Lys Val Ile Lys Lys Leu Ala Asp Glu Val Phe
 370 375 380

Ser Ile Val Gly Arg Asp Pro Leu Ile Glu Val Ala Val Ala Leu Glu
 385 390 395 400

Lys Ala Ala Leu Ser Asp Glu Tyr Phe Val Lys Arg Lys Leu Tyr Pro
 405 410 415

Asn Val Asp Phe Tyr Ser Gly Leu Ile Tyr Arg Ala Met Gly Phe Pro

ES 2 483 365 T3

	420		425		430														
Pro	Glu	Phe	Phe	Thr	Val	Leu	Phe	Ala	Val	Pro	Arg	Met	Ala	Gly	Tyr				
	435						440					445							
Leu	Ser	His	Trp	Arg	Glu	Ser	Leu	Asp	Asp	Pro	Asp	Thr	Arg	Ile	Met				
	450					455					460								
Arg	Pro	Gln	Gln	Ala	Tyr	Thr	Gly	Val	Trp	Met	Arg	His	Tyr	Glu	Pro				
465					470					475					480				
Val	Arg	Glu	Arg	Thr	Leu	Ser	Ser	Asp	Ser	Asp	Lys	Asp	Lys	Phe	Gly				
				485					490					495					
Gln	Val	Ser	Ile	Ser	Asn	Ala	Ser	Arg	Arg	Arg	Leu	Ala	Gly	Ser	Ser				
			500					505					510						
Ala	Leu																		

REIVINDICACIONES

- 5 1. Una planta transgénica que comprende un vector de transformación vegetal que comprende una secuencia de nucleótidos que codifica o es complementaria a una secuencia que codifica un polipéptido de citrato sintasa que tiene una identidad de secuencia de al menos 70% con la SEQ ID NO: 2, en donde la secuencia de nucleótidos es expresada en exceso con respecto a una la planta de control no transgénica, y por medio de la cual la planta transgénica tiene una un mayor porcentaje de masa de la semilla que es aceite con respecto a las plantas de control.
2. Una planta transgénica de la reivindicación 1, en donde dicho polipéptido tiene una identidad de secuencia de al menos 80% con el SEQ ID NO: 2.
- 10 3. La planta transgénica de la reivindicación 1 ó 2, en donde dicho polipéptido tiene una identidad de secuencia de al menos 90% con el SEQ ID NO: 2.
4. La planta transgénica de una cualquiera de las reivindicaciones 1 a 3, en donde dicho polipéptido comprende la secuencia de aminoácidos del SEQ ID NO: 2.
- 15 5. La planta transgénica de una cualquiera de las reivindicaciones 1 a 4, que se selecciona del grupo que consiste en semillas de colza, soja, maíz, girasol, algodón, cacao, cártamo, palma de aceite, palma de coco, lino, ricino y cacahuete.
6. Un tejido u órgano transgénico de la planta de acuerdo con una cualquiera de las reivindicaciones 1 a 4.
7. El tejido u órgano transgénico de la reivindicación 6, que es una semilla.
8. Un método de producción de aceite que comprende el cultivo de la planta transgénica de una cualquiera de las reivindicaciones 1 a 4 y la recuperación del aceite de dicha planta.
- 20 9. Un método para producir una planta que tiene un mayor porcentaje de masa de la semilla que es aceite, comprendiendo dicho método:
- a) introducir en células progenitoras de la planta un vector de transformación vegetal que comprende una secuencia de nucleótidos que codifica o es complementaria a una secuencia que codifica un polipéptido de citrato sintasa que tiene una identidad de secuencia de al menos 70% con el SEQ ID NO: 2, en donde la secuencia de nucleótidos es expresada en exceso con respecto a una planta de control no transgénica, y
- 25 b) cultivar las células progenitoras transformadas para producir una planta transgénica, en donde dicha secuencia de polinucleótidos es expresada, y dicha planta transgénica muestra un aumento del porcentaje de masa de la semilla que es aceite con respecto a las plantas de control.
- 30 10. El método de la reivindicación 9, en donde dicho polipéptido tiene una identidad de secuencia de al menos 80% con el SEQ ID NO: 2.
11. El método de la reivindicación 9 ó 10, en donde dicho polipéptido tiene una identidad de secuencia de al menos 90% con el SEQ ID NO: 2.
12. El método de una cualquiera de las reivindicaciones 9 a 11, en donde dicho polipéptido comprende la secuencia de aminoácidos del SEQ ID NO: 2.
- 35 13. El método de una cualquiera de las Reivindicaciones 9 a 12, en donde dicha planta se selecciona del grupo que consiste de semillas de colza, soja, maíz, girasol, algodón, cacao, cártamo, palma de aceite, palma de coco, lino, ricino y cacahuete.
- 40 14. El uso de una molécula de ácido nucleico de citrato sintasa que comprende una secuencia mostrada como SEQ ID NO: 1 para identificar si una planta que tiene un mayor porcentaje de masa de la semilla que es aceite comprende una mutación en el SEQ ID NO: 1 o un alelo del SEQ ID NO: 1 que causa el aumento del porcentaje de la masa de la semilla que es aceite.
15. El uso de acuerdo con la reivindicación 14, en donde el alelo o el SEQ ID NO: 1 mutado codifica un polipéptido que tiene una identidad de secuencia de al menos 80% con el SEQ ID NO: 2.
- 45 16. El uso de acuerdo con la reivindicación 14 o la reivindicación 15, en donde el alelo o el SEQ ID NO: 1 mutado codifica un polipéptido que tiene una identidad de secuencia de al menos 90% con el SEQ ID NO: 2.
17. El uso de una cualquiera de las reivindicaciones 14 a 16, que emplea la metodología del gen candidato/QTL.