

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 386 554**

51 Int. Cl.:

**C07H 21/04** (2006.01)

**C12N 15/82** (2006.01)

**A01H 5/00** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06785959 .5**

96 Fecha de presentación: **29.06.2006**

97 Número de publicación de la solicitud: **1919514**

97 Fecha de publicación de la solicitud: **14.05.2008**

54 Título: **Polinucleótidos que codifican enzimas modificadoras de isoprenoides y procedimientos de uso de los mismos**

30 Prioridad:  
**05.07.2005 US 697067 P**

45 Fecha de publicación de la mención BOPI:  
**22.08.2012**

45 Fecha de la publicación del folleto de la patente:  
**22.08.2012**

73 Titular/es:  
**THE REGENTS OF THE UNIVERSITY OF  
CALIFORNIA  
1111 FRANKLIN STREET, 12TH FLOOR  
OAKLAND, CA 94607, US**

72 Inventor/es:  
**RO, Dae-Kyun;  
NEWMAN, Karyn;  
PARADISE, Eric M.;  
KEASLING, Jay D.;  
OUELLET, Mario;  
EACHUS, Rachel;  
HO, Kimberly y  
HAM, Timothy**

74 Agente/Representante:  
**Ponti Sales, Adelaida**

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).

ES 2 386 554 T3

## DESCRIPCIÓN

Polinucleótidos que codifican enzimas modificadoras de isoprenoides y procedimientos de uso de los mismos

## 5 REFERENCIA RECÍPROCA

[0001] Esta solicitud reivindica el beneficio de la solicitud de patente provisional de los EE. UU. n° 60/697.067, presentada el 5 de julio de 2005.

## 10 CAMPO DE LA INVENCION

[0002] La presente invención corresponde al campo de la producción de compuestos isoprenoides y, en particular, de las enzimas que modifican compuestos isoprenoides.

## 15 ANTECEDENTES DE LA INVENCION

[0003] Los isoprenoides constituyen un grupo extremadamente amplio y diverso de productos naturales que tienen un origen biosintético común, es decir, un único precursor metabólico, isopentenildifosfato (IPP). Se han descrito al menos 20.000 isoprenoides. Por definición, los isoprenoides están compuestos por, así denominadas, unidades de isopreno (C5). Típicamente, el número de átomos de C presentes en los isoprenoides es divisible por cinco (C5, C10, C15, C25, C30 y C40), aunque se han descrito isoprenoides y politerpenos irregulares. Los compuestos isoprenoides también se denominan "terpenos" o "terpenoides". Algunos miembros importantes de los isoprenoides incluyen los carotenoides, sesquiterpenoides, diterpenoides y hemiterpenos. Los carotenoides incluyen, p. ej., licopeno,  $\beta$ -caroteno y similares, muchos de los cuales funcionan como antioxidantes. Los sesquiterpenoides incluyen, p. ej., artemisinina, un compuesto con actividad contra la malaria. Los diterpenoides incluyen, p. ej., taxol, un agente quimioterapéutico contra el cáncer.

[0004] Los isoprenoides comprenden la familia más numerosa y estructuralmente diversa de productos naturales. En esta familia, los terpenoides aislados de plantas y de otras fuentes naturales se usan como compuestos saborizantes y aromáticos comerciales y también como compuestos farmacéuticos, como fármacos contra la malaria, los virus o el cáncer. La mayoría de los compuestos terpenoides en uso en la actualidad son productos naturales o derivados de estos. Los organismos de origen (p. ej., árboles, invertebrados marinos) de muchos de estos productos naturales no son aptos para el cultivo a gran escala necesario para producir cantidades comercialmente viables ni para la manipulación genética para aumentar la producción o para la derivatización de estos compuestos. Por lo tanto, los productos naturales deben producirse de manera semisintética a partir de análogos o sintéticamente por medio de síntesis químicas convencionales. Además, muchos productos naturales tienen estructuras complejas y, como resultado, su síntesis no resulta económica o es imposible en la actualidad. Tales productos naturales deben extraerse de sus fuentes nativas, como árboles, esponjas, corales y microbios marinos o producirse de manera sintética o semisintética a partir de precursores más abundantes. La extracción de un producto natural de una fuente nativa está limitada por la disponibilidad de la fuente nativa y la producción sintética o semisintética de productos naturales puede adolecer de bajo rendimiento y/o altos costes. Estos problemas de producción y la limitada disponibilidad de la fuente natural pueden restringir el desarrollo comercial y clínico de tales productos.

[0005] Un ejemplo de un compuesto sesquiterpénico importante es la artemisinina. La artemisinina es un fármaco muy efectivo contra la malaria que actualmente se extrae de plantas (*Artemisia annua*) y se usa en medicamentos para tratamientos combinados. La artemisinina de origen vegetal es costosa y su disponibilidad está sujeta a las condiciones climáticas y políticas en los países que cultivan las plantas. El ácido artemisinínico es un intermedio clave en la biosíntesis de la artemisinina. La conversión de amorfa-4,11-dieno en alcohol artemisinínico, una etapa importante en la producción de artemisinina, mediante química tradicional es un proceso difícil y costoso.

[0006] En la técnica existe la necesidad de procedimientos para la generación de compuestos isoprenoides que eviten algunos de los inconvenientes mencionados. La presente invención aborda esta necesidad al proporcionar polinucleótidos que codifican enzimas que modifican compuestos isoprenoides y células huésped modificadas genéticamente para la producción de tales enzimas.

Bibliografía

[0007] Berteau y col. (2005) *Planta Med.* 71: 40 – 47; deKraker y col. (2003) *Tetrahedron* 59: 409 – 418; Martin

y col. (2003) Nat. Biotechnol. 21: 796 – 802; documento WO 03/025193; publicación de patente de los EE. UU. n° 20050019882; publicación de patente de los EE. UU. n° 20030148479; publicación de patente de los EE. UU. n° 20040005678; publicación de patente de los EE. UU. n° 20030166255.

## 5 RESUMEN DE LA INVENCION

[0008] La presente invención proporciona ácidos nucleicos aislados que comprenden secuencias nucleotídicas que codifican enzimas modificadoras de isoprenoides con actividad de amorfa-4,11-dieno-oxidasa, según se definen en las reivindicaciones adjuntas, así como vectores recombinantes que comprenden los ácidos nucleicos según se definen en las reivindicaciones adjuntas. Además, la presente invención proporciona células huésped modificadas genéticamente con un ácido nucleico o un vector recombinante objeto según se definen en las reivindicaciones adjuntas. Además, la presente invención proporciona una planta transgénica modificada genéticamente con un ácido nucleico objeto según se define en las reivindicaciones adjuntas. Además, la presente invención proporciona procedimientos para la producción de un compuesto isoprenoide, en que el procedimiento implica generalmente el cultivo de una célula huésped modificada genéticamente objeto en condiciones que permiten la síntesis de una enzima con actividad de amorfa-4,11-dieno-oxidasa por un ácido nucleico objeto, en que la enzima modifica un compuesto isoprenoide, según se define en las reivindicaciones adjuntas.

## BREVE DESCRIPCION DE LOS DIBUJOS

- [0009] La figura 1 muestra la secuencia nucleotídica codificante de un ADNc de *CYP71D-A4* (SEQ ID NO: 1).
- [0010] La figura 2 muestra la secuencia aminoacídica de una amorfadieno-12-oxidasa. (SEQ ID NO: 2).
- [0011] La figura 3 muestra la secuencia nucleotídica de la región codificante de un ADNc de citocromo-P450-reductasa de *Artemisia annua* (SEQ ID NO: 3).
- [0012] La figura 4 muestra la secuencia aminoacídica de una citocromo-P450-reductasa de *Artemisia annua*. (SEQ ID NO: 4).
- [0013] Las figuras 5A – C muestran los resultados de un experimento de suministro de sustrato *in vivo*.
- [0014] Las figuras 6A y 6B muestran la confirmación del producto por CG-EM.
- [0015] Las figuras 7A – C muestran la producción *de novo* de ácido artemisinico en levaduras.
- [0016] Las figuras 8A – C muestran ensayos enzimáticos *in vitro* para la amorfadieno-oxidasa.
- [0017] La figura 9 muestra la secuencia nucleotídica de un ADNc (clon 71D – B1) que codifica una enzima modificadora de isoprenoides (SEQ ID NO: 5).
- [0018] La figura 10 muestra la secuencia aminoacídica de una enzima modificadora de isoprenoides (71D – B1; SEQ ID NO: 6).
- [0019] Las figuras 11A – C muestran la actividad de hidroxilación de la enzima 71D – B1.
- [0020] La figura 12 muestra la secuencia nucleotídica de un ADN genómico que codifica una enzima modificadora de isoprenoides (SEQ ID NO: 7).
- [0021] La figura 13 es una representación esquemática de las rutas metabólicas de los isoprenoides que resultan en la producción de los poliprenildifosfatos intermedios de la ruta biosintética de los isoprenoides, geranildifosfato (GPP), farnesildifosfato (FPP) y geranilgeranildifosfato (GGPPP), a partir de isopentenildifosfato (IPP) y dimetilalildifosfato (DMAPP).
- [0022] La figura 14 es una representación esquemática de la ruta del mevalonato (MEV) para la producción de IPP.
- [0023] La figura 15 es una representación esquemática de la ruta de la DXP para la producción de IPP y dimetilalilpirofosfato (DMAPP).

## DEFINICIONES

**[0024]** Los términos “isoprenoide”, “compuesto isoprenoide”, “terpeno”, “compuesto terpénico”, “terpenoide” y “compuesto terpenoide” se usan de manera intercambiable en este documento. Los compuestos isoprenoides están formados por un número variable de, así denominadas, unidades de isopreno (C5). Típicamente, el número de átomos de carbono presentes en los isoprenoides es divisible por cinco (p. ej., C5, C10, C15, C20, C25, C30 y C40). Se han descrito isoprenoides y politerpenos irregulares, que también se incluyen en la definición de “isoprenoide”. Los compuestos isoprenoides incluyen, pero no se limitan a monoterpenos, sesquiterpenos, triterpenos, politerpenos y diterpenos.

**[0025]** Según se usa en este documento, el término “prenildifosfato” es intercambiable con “prenilpirofosfato” e incluye monoprenildifosfatos con un único grupo prenilo (p. ej. IPP y DMAPP), así como poliprenildifosfatos que incluyen dos o más grupos prenilo. Los monoprenildifosfatos incluyen isopentenilpirofosfato (IPP) y su isómero dimetilalilpirofosfato (DMAPP).

**[0026]** Según se usa en este documento, el término “terpeno-sintasa” se refiere a cualquier enzima que modifica enzimáticamente IPP, DMAPP o un poliprenilpirofosfato de modo que se produce un compuesto terpenoide. El término “terpeno-sintasa” incluye enzimas que catalizan la conversión de un prenildifosfato en un isoprenoide.

**[0027]** La palabra “pirofosfato” se usa en este documento de manera intercambiable con “difosfato”. Así, p. ej., los términos “prenildifosfato” y “prenilpirofosfato” son intercambiables; los términos “isopentenilpirofosfato” e “isopentenildifosfato” son intercambiables; los términos “farnesildifosfato” y “farnesilpirofosfato” son intercambiables; etc.

**[0028]** El término “ruta del mevalonato” o “ruta del MEV” se usa en este documento para referirse a la ruta biosintética que convierte acetil-CoA en IPP. La ruta del mevalonato comprende enzimas que catalizan las etapas siguientes: (a) condensación de dos moléculas de acetil-CoA para dar acetoacetil-CoA; (b) condensación de acetoacetil-CoA con acetil-CoA para formar HMG-CoA; (c) conversión de HMG-CoA en mevalonato; (d) fosforilación de mevalonato para dar mevalonato-5-fosfato; (e) conversión de mevalonato-5-fosfato en mevalonato-5-pirofosfato; y (f) conversión de mevalonato-5-pirofosfato en isopentenilpirofosfato. La ruta del mevalonato se ilustra esquemáticamente en la figura 14. La “mitad superior” de la ruta del mevalonato se refiere a las enzimas responsables de la conversión de acetil-CoA en mevalonato a través de un intermedio de la ruta del MEV.

**[0029]** El término “ruta de la 1-desoxi-D-xilulosa-5-difosfato” o “ruta de la DXP” se usa en este documento para referirse a la ruta que convierte gliceraldehído-3-fosfato y piruvato en IPP y DMAPP a través de un intermedio de la ruta de la DXP, en que la ruta de la DXP comprende enzimas que catalizan las reacciones mostradas esquemáticamente en la figura 15.

**[0030]** Según se usa en este documento, el término “prenil-transferasa” se usa de manera intercambiable con los términos “isoprenildifosfato-sintasa” y “poliprenil-sintasa” (p. ej., “GPP-sintasa”, “FPP-sintasa”, “OPP-sintasa”, etc.) para referirse a una enzima que cataliza la condensación 1'-4 consecutiva de isopentenildifosfato con sustratos iniciadores alílicos, lo que resulta en la formación de prenildifosfatos de diversas longitudes de cadena.

**[0031]** Los términos “polinucleótido” y “ácido nucleico”, usados de manera intercambiable en este documento, se refieren a una forma polimérica de nucleótidos de cualquier longitud, tanto ribonucleótidos como desoxinucleótidos. Por lo tanto, este término incluye, pero no se limita a ADN o ARN monocatenario, bicatenario o multcatenario, ADN genómico, ADNc, híbridos ADN-ARN o a un polímero que comprende bases de purina y pirimidina u otras bases nucleotídicas naturales, modificadas química o bioquímicamente, no naturales o derivatizadas.

**[0032]** Los términos “péptido”, “polipéptido” y “proteína”, usados de manera intercambiable en este documento, se refieren a una forma polimérica de aminoácidos de cualquier longitud que puede incluir aminoácidos codificados y no codificados, aminoácidos modificados química o bioquímicamente o derivatizados y polipéptidos con esqueletos peptídicos modificados.

**[0033]** La expresión “de origen natural”, según se usa en este documento aplicada a un ácido nucleico, célula u organismo, se refiere a un ácido nucleico, célula u organismo que se encuentra en la naturaleza. Por ejemplo, una secuencia polipeptídica o polinucleotídica que está presente en un organismo (incluidos los virus) que puede aislarse

de una fuente en la naturaleza y que no ha sido modificada intencionadamente por un humano en el laboratorio es de origen natural.

**[0034]** Según se usa en este documento, el término “aislado” pretende describir un polinucleótido, un polipéptido o una célula que está en un entorno diferente de aquel en el que el polinucleótido, el polipéptido o la célula se presentan de manera natural. Una célula huésped modificada genéticamente aislada puede estar presente en una población mixta de células huésped genéticamente modificadas.

**[0035]** Según se usa en este documento, el término “ácido nucleico exógeno” se refiere a un ácido nucleico que no se encuentra normalmente o naturalmente en una bacteria, organismo o célula dados en la naturaleza ni/o es producido por estos. Según se usa en este documento, el término “ácido nucleico endógeno” se refiere a un ácido nucleico que se encuentra normalmente en una bacteria, organismo o célula dados en la naturaleza y/o es producido por estos. Un “ácido nucleico endógeno” se denomina también “ácido nucleico nativo” o un ácido nucleico que es “nativo” de una bacteria, organismo o célula dados. Por ejemplo, los ácidos nucleicos que codifican HMGS, mevalonato-cinasa y fosfomevalonato-cinasa representan ácidos nucleicos exógenos a *E. coli*. Estos ácidos nucleicos de la ruta del mevalonato pueden clonarse a partir de *Saccharomyces cerevisiae*. En *S. cerevisiae*, las secuencias génicas que codifican HMGS, MK y PMK en el cromosoma serían ácidos nucleicos “endógenos”.

**[0036]** El término “ácido nucleico heterólogo”, según se usa en este documento, se refiere a un ácido nucleico para el que se cumple al menos uno de los puntos siguientes: (a) el ácido nucleico es extraño (“exógeno”) a (es decir, no se encuentra naturalmente en) un microorganismo huésped o célula huésped dados; (b) el ácido nucleico comprende una secuencia nucleotídica que se encuentra naturalmente en (es decir, es “endógena de”) un microorganismo huésped o célula huésped dados (p. ej., el ácido nucleico comprende una secuencia nucleotídica que es endógena del microorganismo huésped o la célula huésped), pero se produce en la célula en una cantidad no natural (p. ej., superior a la esperada o superior a la encontrada de manera natural) o difiere en secuencia de la secuencia nucleotídica endógena, de modo que la misma proteína codificada (con la misma o sustancialmente la misma secuencia aminoacídica) que se encuentra endógenamente se produce en una cantidad no natural en la célula (p. ej., superior a la esperada o superior a la que se encuentra de manera natural); (c) el ácido nucleico comprende dos o más secuencias nucleotídicas o segmentos que no se encuentran en la misma relación entre sí en la naturaleza, p. ej., el ácido nucleico es recombinante.

**[0037]** El término “recombinante”, según se usa en este documento, significa que un ácido nucleico concreto (ADN o ARN) es el producto de diversas combinaciones de etapas de clonación, restricción y/o ligación que resultan en una construcción con una secuencia estructural codificante o no codificante distinguible de los ácidos nucleicos endógenos que se encuentran en sistemas naturales. Generalmente, las secuencias de ADN que codifican la secuencia codificante estructural pueden ensamblarse a partir de fragmentos de ADNc y enzimas de restricción y ligación para proporcionar un ácido nucleico sintético capaz de expresarse a partir de una unidad transcripcional recombinante contenida en una célula o en un sistema de transcripción y traducción acelular. Tales secuencias pueden proporcionarse en forma de una pauta abierta de lectura no interrumpida por secuencias internas no traducidas o intrones, típicamente presentes en genes eucarióticos. El ADN genómico que comprende las secuencias pertinentes puede usarse también en la formación de un gen o unidad transcripcional recombinantes. La presencia de secuencias de ADN no traducido es posible delante o detrás de los extremos 5' o 3' de la pauta abierta de lectura, donde tales secuencias no interfieren con la manipulación o la expresión de las regiones codificantes y, de hecho, pueden modular la producción de un producto deseado por diversos mecanismos (véase más adelante “secuencias reguladoras de ADN”).

**[0038]** Por lo tanto, p. ej., el término polinucleótido o ácido nucleico “recombinante” se refiere a uno que no es de origen natural, p.ej., está hecho mediante la combinación artificial por intervención humana de dos segmentos de secuencia de otro modo separados. Con frecuencia, esta combinación artificial se lleva a cabo por procedimientos de síntesis química o por manipulación artificial de segmentos aislados de ácidos nucleicos, p. ej., por técnicas de ingeniería genética. Normalmente, esto se realiza para sustituir un codón por un codón redundante que codifica el mismo aminoácido o un aminoácido conservador, mientras que típicamente se introduce o se elimina un sitio de reconocimiento de secuencia. Alternativamente, se realiza para unir entre sí segmentos de ácido nucleico con las funciones deseadas para generar una combinación de funciones deseada. Con frecuencia, esta combinación artificial se lleva a cabo por procedimientos de síntesis química o por manipulación artificial de segmentos aislados de ácidos nucleicos, p. ej., por técnicas de ingeniería genética.

**[0039]** Por “construcción” se indica un ácido nucleico recombinante, generalmente ADN recombinante, que ha sido generado con el propósito de la expresión de una (s) secuencia (s) nucleotídica (s) específica (s) o para usarlo

en la construcción de otras secuencias nucleotídicas recombinantes.

**[0040]** Según se usa en este documento, los términos “operón” y “unidad de transcripción simple” se usan de manera intercambiable para referirse a dos o más regiones codificantes (secuencias nucleotídicas que codifican un producto génico como un ARN o una proteína) contiguas que se regulan de manera coordinada por uno o más elementos de control (p. ej., un promotor). Según se usa en este documento, el término “producto génico” se refiere a un ARN codificado por un ADN (o viceversa) o a una proteína codificada por un ARN o ADN, en que un gen comprende típicamente una o más secuencias nucleotídicas que codifican una proteína y puede incluir también intrones y otras secuencias nucleotídicas no codificantes.

**[0041]** Los términos “secuencias reguladoras de ADN”, “elementos de control” y “elementos reguladores”, usados de manera intercambiable en este documento, se refieren a secuencias de control transcripcional y traduccional como promotores, potenciadores, señales de poliadenilación, terminadores, señales de degradación de proteínas y similares, que dan lugar a, y/o regulan la expresión de una secuencia codificante y/o la producción de un polipéptido codificado en una célula huésped.

**[0042]** El término “transformación” se usa en este documento de manera intercambiable con “modificación genética” y se refiere a un cambio genético permanente o transitorio inducido en una célula después de la introducción de un nuevo ácido nucleico (es decir, ADN exógeno a la célula). Un cambio genético (“modificación”) puede efectuarse por incorporación del nuevo ADN en el genoma de la célula huésped o mediante el mantenimiento transitorio o estable del nuevo ADN como un elemento episómico. Si la célula es una célula eucariota, generalmente un cambio genético permanente se consigue por la introducción del ADN en el genoma de la célula. En células procariotas, los cambios permanentes pueden introducirse en el cromosoma o conseguirse por medio de elementos extracromosómicos como plásmidos y vectores de expresión, que pueden contener uno o más marcadores seleccionables para facilitar su mantenimiento en la célula huésped recombinante. Los procedimientos de modificación genética adecuados incluyen la infección vírica, transfección, conjugación, fusión de protoplastos, electroporación, tecnología de bombardeo de partículas, precipitación con fosfato de calcio, microinyección directa y similares. Generalmente, la elección del procedimiento depende del tipo de célula que se transforme y de las circunstancias en las que se lleva a cabo la transformación (es decir, *in vitro*, *ex vivo* o *in vivo*). Una discusión general de estos procedimientos puede encontrarse en Ausubel y col., Short Protocols in Molecular Biology, 3ª edición, Wiley & Sons, 1995.

**[0043]** El término “ligado operativamente” se refiere a una yuxtaposición en la que los componentes así descritos guardan una relación que les permite funcionar de la manera deseada. Por ejemplo, un promotor está ligado operativamente a una secuencia codificante si el promotor afecta a su transcripción o expresión. Según se usa en este documento, los términos “promotor heterólogo” y “regiones de control heterólogas” se refieren a promotores y a otras regiones de control que no están normalmente asociadas con un ácido nucleico concreto en la naturaleza. Por ejemplo, una “región de control transcripcional heteróloga con respecto a una región codificante” es una región de control transcripcional que normalmente no está asociada con la región codificante en la naturaleza.

**[0044]** Una “célula huésped”, según se usa en este documento, denota una célula eucariota, una célula procariota o una célula de un organismo pluricelular (p. ej., una línea celular) *in vivo* o *in vitro*, cultivada como una entidad unicelular, en que las células eucariotas o procariotas pueden ser o haber sido usadas como receptoras de un ácido nucleico (p. ej., un vector de expresión que comprende una secuencia nucleotídica que codifica uno o más productos génicos de una ruta biosintética como los productos génicos de la ruta del mevalonato) e incluyen la descendencia de la célula original que ha sido modificada genéticamente por el ácido nucleico. Se entiende que la descendencia de una célula individual no tiene por qué ser completamente idéntica en morfología o en el complemento de ADN genómico o total a la célula parental original, debido a mutaciones naturales, accidentales o deliberadas. Una “célula huésped recombinante” (denominada también “célula huésped modificada genéticamente”) es una célula huésped en la que se ha introducido un ácido nucleico heterólogo, p. ej., un vector de expresión. Por ejemplo, una célula huésped procariota objeto es una célula huésped procariota modificada genéticamente (p. ej., una bacteria) en virtud de la introducción en una célula huésped procariota adecuada de un ácido nucleico heterólogo, p. ej., un ácido nucleico exógeno que es extraño a (no se encuentra normalmente en la naturaleza en) la célula huésped procariota o un ácido nucleico recombinante que no se encuentra normalmente en la célula huésped procariota; y una célula huésped eucariota objeto es una célula huésped eucariota modificada genéticamente en virtud de la introducción en una célula huésped eucariota adecuada de un ácido nucleico heterólogo, p. ej., un ácido nucleico exógeno que es extraño a la célula huésped eucariota o un ácido nucleico recombinante que no se encuentra normalmente en la célula huésped eucariota.

**[0045]** Un ácido nucleico es “hibridable” con otro ácido nucleico, como un ADNc, ADN genómico o ARN, cuando una forma monocatenaria del ácido nucleico puede aparearse con el otro ácido nucleico en las condiciones apropiadas de temperatura y fuerza iónica de la disolución. Las condiciones de hibridación y lavado son bien conocidas y se ejemplifican en Sambrook, J., Fritsch, E. F. y Maniatis, T., *Molecular Cloning: A Laboratory Manual*, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), en particular el capítulo 11 y la tabla 11.1; Sambrook, J. y Russell, W., *Molecular Cloning: A Laboratory Manual*, tercera edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). Las condiciones de temperatura y fuerza iónica determinan la “restricción” de la hibridación. Las condiciones de restricción pueden ajustarse para realizar un análisis de cribado en busca de fragmentos moderadamente similares, como los genes que duplican enzimas funcionales de organismos muy próximos entre sí. Las condiciones de hibridación y los lavados después de la hibridación son útiles para obtener las condiciones de restricción deseadas para la hibridación. Un conjunto ilustrativo de lavados después de la hibridación es una serie de lavados que comienza con un lavado con 6 x SSC (en que SSC es NaCl 0,15 M y tampón de citrato 15 mM), SDS al 0,5 % a temperatura ambiente durante 15 minutos, que se repite después con 2 x SSC, SDS al 0,5 % a 45 °C durante 30 minutos y se repite después dos veces con 0,2 x SSC, SDS al 0,5 % a 50 °C durante 30 minutos. Otras condiciones restrictivas se obtienen mediante el uso de temperaturas más altas, en que los lavados son idénticos a los anteriores excepto porque la temperatura de los dos lavados finales de 30 minutos en 0,2 x SSC, SDS al 0,5 % se eleva a 60 °C. Otro conjunto de condiciones muy restrictivas usa dos lavados finales en 0,1 x SSC, SDS al 0,1 % a 65 °C. Otro ejemplo de condiciones de hibridación restrictivas es la hibridación a 50 °C o más y 0,1 x SSC (cloruro de sodio 15 mM, citrato de sodio 1,5 mM). Otro ejemplo de condiciones de hibridación restrictivas es la incubación durante la noche a 42 °C en una disolución de: formamida al 50 %, 5 x SSC (NaCl 150 mM, citrato de trisodio 15 mM), fosfato de sodio 50 mM (pH 7,6), 5 x disolución de Denhardt, sulfato de dextrano al 10 % y 20 µg/ml de ADN de esperma de salmón fragmentado y desnaturalizado, seguida del lavado de los filtros en 0,1 x SSC a aproximadamente 65 °C. Unas condiciones restrictivas de hibridación y de lavado después de la hibridación son condiciones de hibridación y de lavado después de la hibridación que son al menos tan restrictivas como las condiciones representativas anteriores.

**[0046]** La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque dependiendo de la restricción de la hibridación es posible el apareamiento incorrecto entre bases. La restricción apropiada para la hibridación de los ácidos nucleicos depende de la longitud de dichos ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor es el grado de similitud o de homología entre dos secuencias nucleotídicas, mayor es el valor de la temperatura de fusión ( $T_m$ ) para los híbridos de los ácidos nucleicos con estas secuencias. La estabilidad relativa (correspondiente a una mayor  $T_m$ ) de las hibridaciones de ácidos nucleicos disminuye en el orden siguiente: ARN: ARN, ADN: ARN, ADN: ADN. Para híbridos de más de 100 nucleótidos de longitud se han deducido ecuaciones para calcular la  $T_m$  (véase Sambrook y col., cita anterior, 9.50 – 9.51). Para hibridaciones de ácidos nucleicos de menor longitud, es decir, oligonucleótidos, la posición de los apareamientos incorrectos adquiere mayor importancia y la longitud del oligonucleótido determina su especificidad (véase Sambrook y col., cita anterior, 11.7 – 11.8). Típicamente, la longitud de un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos. Las longitudes mínimas ilustrativas para un ácido nucleico hibridable son: al menos aproximadamente 15 nucleótidos; al menos aproximadamente 20 nucleótidos; y al menos aproximadamente 30 nucleótidos. Además, el experto en la técnica reconocerá que la temperatura y la concentración de sales de las disoluciones de lavado pueden ajustarse según sea necesario de acuerdo con factores como la longitud de la sonda.

**[0047]** El término “sustitución conservadora de aminoácidos” se refiere a la intercambiabilidad en las proteínas de los restos aminoacídicos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos con cadenas laterales alifáticas consta de glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos con cadenas laterales alifáticas hidroxiladas consta de serina y treonina; un grupo de aminoácidos con cadenas laterales que contienen amidas consta de asparragina y glutamina; un grupo de aminoácidos con cadenas laterales aromáticas consta de fenilalanina, tirosina, y triptófano; un grupo de aminoácidos con cadenas laterales básicas consta de lisina, arginina e histidina; y un grupo de aminoácidos con cadenas laterales que contienen azufre consta de cisteína y metionina. Algunos grupos ejemplares de sustitución conservadora de aminoácidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparragina-glutamina.

**[0048]** Es posible ensamblar “ácidos nucleicos sintéticos” a partir de elementos oligonucleotídicos básicos que se sintetizan químicamente mediante procedimientos conocidos por los expertos en la técnica. Estos elementos básicos se ligan y aparean para formar segmentos génicos que después se ensamblan enzimáticamente para construir el gen entero. “Sintetizada químicamente”, en relación con una secuencia de ADN, significa que los nucleótidos componentes se han ensamblado *in vitro*. Puede llevarse a cabo una síntesis química manual de ADN mediante procedimientos bien establecidos o puede realizarse una síntesis química automática mediante una de las

máquinas disponibles comercialmente. La secuencia nucleotídica de los ácidos nucleicos puede modificarse para una expresión óptima, a base de la optimización de la secuencia nucleotídica para reflejar la preferencia de codones de la célula huésped. El experto en la técnica aprecia la probabilidad de éxito de la expresión si el uso de codones se altera hacia los codones favorecidos por el huésped. La determinación de los codones preferidos puede basarse en un estudio de los genes derivados de la célula huésped de los que se disponga de información de secuencia.

[0049] Un polinucleótido o polipéptido tiene un cierto porcentaje de "identidad de secuencia" con otro polinucleótido o polipéptido, lo que significa que, al alinearse, este porcentaje de bases o de aminoácidos es idéntico y estos están en la misma posición relativa cuando se comparan las dos secuencias. La similitud de secuencias puede determinarse de varias maneras diferentes. Para determinar la identidad de secuencia, las secuencias pueden alinearse mediante los procedimientos y programas de ordenador, incluido BLAST, disponibles a través de internet en [ncbi.nlm.nih.gov/BLAST](http://ncbi.nlm.nih.gov/BLAST). Véase, p. ej., Altschul y col. (1990), J. Mol. Biol. 215: 403 – 10. Otro algoritmo de alineamiento es FASTA, disponible en el paquete del Genetics Computing Group (GCG) de Madison, Wisconsin, EE. UU., una filial de plena propiedad del Oxford Molecular Group, Inc. Otras técnicas de alineamiento se describen en Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., una división de Harcourt Brace & Co., San Diego, California, EE. UU. Un interés especial tienen los programas de alineamiento que permiten huecos en la secuencia. El algoritmo de Smith-Waterman es de un tipo que permite huecos en los alineamientos de secuencias. Véase Meth. Mol. Biol. 70: 173 – 187 (1997). El programa GAP que usa el método de alineamiento de Needleman y Wunsch también puede usarse para el alineamiento de secuencias. Véase J. Mol. Biol. 48: 443 – 453 (1970).

[0050] Antes de describir la presente invención más detalladamente, ha de entenderse que esta invención no se limita a las realizaciones concretas descritas, ya que, por supuesto, estas pueden variar. También ha de entenderse que la terminología usada en este documento solamente tiene la finalidad de describir realizaciones concretas y no pretende ser limitante, dado que el alcance de la presente invención solo quedará limitado por las reivindicaciones adjuntas.

[0051] Cuando se proporciona un intervalo de valores, se entiende que la invención comprende todo valor intermedio, hasta la décima parte de la unidad del límite inferior a no ser que el contexto claramente indique lo contrario, entre los límites superior e inferior de este intervalo y cualquier otro valor indicado o intermedio en el intervalo indicado. Los límites superior e inferior de estos intervalos menores pueden incluirse independientemente en los intervalos menores y también están comprendidos en la invención, sujetos a cualquier límite excluido específicamente en el intervalo indicado. Cuando el intervalo indicado incluye uno o los dos límites, los intervalos que excluyen uno o los dos de estos límites incluidos también están incluidos en la invención.

[0052] A menos que se definan de otro modo, todos los términos técnicos y científicos usados en este documento tienen el mismo significado entendido normalmente por un experto en la técnica a la que pertenece esta invención. Aunque cualquier procedimiento y material similar o equivalente a los descritos en este documento puede usarse también en la práctica o el ensayo de la presente invención, a continuación se describen los procedimientos y materiales preferidos. Todas las publicaciones mencionadas en este documento desvelan y describen los procedimientos y/o materiales en conexión con los que se citan las publicaciones.

[0053] Ha de señalarse que, según se usa en este documento y en las reivindicaciones adjuntas, las formas singulares "un", "una", "el" y "la" incluyen los referentes en plural a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "una enzima modificadora de isoprenoides" incluye una pluralidad de dichas enzimas y la referencia a "la citocromo-P450-reductasa" incluye la referencia a una o más citocromo-P450-reductasas y a equivalentes de estas conocidas por los expertos en la técnica y así sucesivamente. Ha de señalarse además que las reivindicaciones pueden redactarse para excluir cualquier elemento opcional. Como tal, esta afirmación pretende servir como base antecedente para el uso de la terminología exclusiva como "únicamente" "solo" y similar en conexión con la relación de los elementos reivindicados o el uso de una limitación "negativa".

[0054] Las publicaciones discutidas en este documento se proporcionan únicamente para su divulgación anterior a la fecha de presentación de la presente solicitud. Nada en este documento debe interpretarse como una admisión de que la presente invención no tiene derecho a preceder a tal publicación en virtud de una invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas reales de publicación, que pueden requerir una confirmación independiente.

**DESCRIPCIÓN DETALLADA DE LA INVENCION**

[0055] La presente invención proporciona ácidos nucleicos aislados que comprenden secuencias nucleotídicas que codifican enzimas modificadoras de isoprenoides, así como vectores recombinantes que comprenden los ácidos nucleicos. La presente invención proporciona además células huésped modificadas genéticamente que comprenden un ácido nucleico o un vector recombinante objeto. Además, la presente invención proporciona una planta transgénica que comprende un ácido nucleico objeto. Además, la presente invención proporciona procedimientos para la producción de un compuesto isoprenoide, en que el procedimiento implica generalmente el cultivo de una célula huésped modificada genéticamente objeto en condiciones que permiten la síntesis de una enzima modificadora de compuestos isoprenoides codificada por un ácido nucleico objeto.

**ÁCIDOS NUCLEICOS, VECTORES Y CÉLULAS HUÉSPED**

[0056] La presente invención proporciona un ácido nucleico aislado que comprende una secuencia nucleotídica que codifica una enzima que modifica un compuesto isoprenoide, en que una enzima que modifica un compuesto isoprenoide se denomina en este documento "enzima modificadora de isoprenoides". Un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides se denomina "ácido nucleico de enzima modificadora de isoprenoides". En realizaciones concretas, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una citocromo-P450-monooxigenasa. En realizaciones concretas, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una isoprenoide-oxidasa. En algunas realizaciones, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una terpeno-hidroxilasa. En algunas realizaciones, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una terpeno-oxidasa. En algunas realizaciones, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una sesquiterpeno-oxidasa. En algunas realizaciones, un ácido nucleico de enzima modificadora de isoprenoides aislado objeto comprende una secuencia nucleotídica que codifica una sesquiterpeno-hidroxilasa.

[0057] La NADPH-citocromo-P450-oxidoreductasa (CPR, EC 1.6.2.4) es la pareja redox de muchas P450-monooxigenasas. La presente invención proporciona además un ácido nucleico aislado que comprende una secuencia nucleotídica que codifica una citocromo-P450-reductasa. (CPR). Un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una CPR se denomina "ácido nucleico de CPR". Una CPR codificada por un ácido nucleico de CPR objeto transfiere electrones desde NADPH al citocromo P450. En general, una CPR codificada por un ácido nucleico de CPR objeto transfiere electrones desde NADPH a una enzima modificadora de isoprenoides, p. ej., una sesquiterpeno-oxidasa, codificada por un ácido nucleico de enzima modificadora de isoprenoides objeto.

Ácidos nucleicos que codifican enzimas modificadoras de isoprenoides

[0058] En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica un polipéptido que muestra actividad de isoprenoide-hidroxilasa y/o actividad de isoprenoide-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una citocromo-P450-monooxigenasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una isoprenoide-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una isoprenoide-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica un polipéptido que lleva a cabo reacciones sucesivas de hidroxilación y oxidación, p. ej., el polipéptido hidroxila un compuesto terpénico para generar un alcohol terpénico, oxida el alcohol terpénico para generar un aldehído terpénico y oxida el aldehído terpénico para generar un ácido carboxílico terpénico. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica un polipéptido que cataliza la hidroxilación y/u oxidación de un grupo isoprenilo de un terpeno, p. ej., cataliza la hidroxilación de un grupo isoprenilo de un monoterpeno, un diterpeno, un triterpeno, un sesquiterpeno o un politerpeno. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una monoterpeno-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una monoterpeno-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una politerpeno-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una politerpeno-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una diterpeno hidroxilasa. En algunas realizaciones, un ácido

nucleico aislado objeto comprende una secuencia nucleotídica que codifica una diterpeno-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una triterpeno-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una triterpeno-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una  
 5 secuencia nucleotídica que codifica una sesquiterpeno-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una sesquiterpeno-oxidasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una sesquiterpeno-C12-hidroxilasa. En algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica un polipéptido que lleva a cabo la oxidación del C12 de un sesquiterpeno. En  
 10 algunas realizaciones, un ácido nucleico aislado objeto comprende una secuencia nucleotídica que codifica una amorfadieno-12-oxidasa.

**[0059]** El producto de la acción de una terpeno-ciclasa (también denominada “terpeno-sintasa”) es el denominado “esqueleto terpénico”. En algunas realizaciones, un ácido nucleico aislado objeto comprende una  
 15 secuencia nucleotídica que codifica una enzima modificadora de isoprenoides que cataliza la hidroxilación y/u oxidación de un esqueleto terpénico o un producto posterior de este. En general, un sustrato de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto comprende un esqueleto terpénico o un esqueleto terpénico modificado. En muchas realizaciones, un sustrato de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto comprende un grupo isopropenilo.

**[0060]** Los sustratos monoterpénicos de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto incluyen, pero no se limitan a cualquier sustrato monoterpénico que da lugar a un producto de oxidación que es un compuesto monoterpénico o es un intermedio en una ruta biosintética que da lugar a un compuesto monoterpénico. Algunos sustratos monoterpénicos ejemplares incluyen, pero no se limitan a sustratos  
 25 monoterpénicos que corresponden a cualquiera de las familias siguientes: monoterpenos acíclicos, dimetiloctanos, mentanos, monoterpenoides irregulares, cineoles, canfanos, isocanfanos, monoterpenos monocíclicos, pinanos, fenchanos, tuyanos, caranos, iononas, iridanos y canabinoides. Algunos sustratos, intermedios y productos monoterpénicos ejemplares incluyen, pero no se limitan a limoneno, citranelol, geraniol, mentol, alcohol perfílico, linalol y tuyona.

**[0061]** Los sustratos diterpénicos de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto incluyen, pero no se limitan a cualquier sustrato diterpénico que da lugar a un producto de oxidación que es un compuesto diterpénico o es un intermedio en una ruta biosintética que da lugar a un compuesto diterpénico. Algunos sustratos diterpénicos ejemplares incluyen, pero no se limitan a sustratos diterpénicos que  
 35 corresponden a cualquiera de las familias siguientes: diterpenoides acíclicos, diterpenoides bicíclicos, diterpenoides monocíclicos, labdanos, clerodanos, taxanos, diterpenoides tricíclicos, diterpenoides tetracíclicos, caurenos, beyerenos, atisenos, afidicolinas, grayanotoxinas, giberelinas, diterpenos macrocíclicos y elizabetatrianos. Algunos sustratos, intermedios y productos diterpénicos ejemplares incluyen, pero no se limitan a casbeno, eleuterobina, paclitaxel, prostatina y psuedopterosina.

**[0062]** Los sustratos triterpénicos de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto incluyen, pero no se limitan a cualquier sustrato triterpénico que da lugar a un producto de oxidación que es un compuesto triterpénico o es un intermedio en una ruta biosintética que da lugar a un compuesto triterpénico. Algunos sustratos, intermedios y productos triterpénicos ejemplares incluyen, pero no se limitan a  
 45 arbrusido E, bruceantina, testosterona, progesterona, cortisona y digitoxina.

**[0063]** Los sustratos sesquiterpénicos de una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto incluyen, pero no se limitan a cualquier sustrato sesquiterpénico que da lugar a un producto de oxidación que es un compuesto sesquiterpénico o es un intermedio en una ruta biosintética que da lugar a un compuesto sesquiterpénico. Algunos sustratos sesquiterpénicos ejemplares incluyen, pero no se limitan a sustratos  
 50 sesquiterpénicos que corresponden a cualquiera de las familias siguientes: farnesanos, monociclofarnesanos, sesquiterpenos monocíclicos, sesquiterpenos bicíclicos, biciclofarnesanos, bisabolanos, santalanos, cupranos, herbertanos, gimnomitranos, tricotecanos, chamigranos, carotanos, acoranos, antisatines, cadinanos, oplopananos, copaanos, picrotoxanos, himachalanos, longipinanos, longiciclanos, cariofilanos, modhefanos, sifiperfolanos, humulanos, intergrifolios, lipifolios, protoiludanos, iludanos, hirsutanos, lactaranos, esterpuranos, fomanos, marasmanos, germacranos, elemanos, eudesmanos, bacanos, chilosifanos, guayanos, pseudoguayanos, sesquiterpenos tricíclicos, pachulianos, trixanos, aromadendranos, gorgonianos, nardosinanos, brasilanos, pinguisanos, sesquipinanos, sesquicanfanos, tuyopsanos, biciclohulmulanos, aliacanos, esterpuranos, lactaranos, africanos, integrifolios, protoiludanos, aristolanos y neolemanos. Algunos sustratos sesquiterpénicos ejemplares

incluyen, pero no se limitan a amorfadieno, aloisolongifoleno, (-)- $\alpha$ -*trans*-bergamoteno, (-)- $\beta$ -elemeno, ( + )-germacreno A, germacreno B, ( + )- $\gamma$ -gurjuneno, ( + )-ledeno, neointermediol, ( + )- $\beta$ -selineno y ( + )-valenceno.

**[0064]** Si un ácido nucleico objeto codifica una terpeno-oxidasa o una terpeno-hidroxilasa puede determinarse fácilmente mediante ensayos estándar para estas actividades enzimáticas con el sustrato adecuado. Los productos de la modificación enzimática se analizan generalmente por cromatografía de gases y espectrometría de masas. Si un ácido nucleico objeto codifica una sesquiterpeno-oxidasa o una sesquiterpeno-hidroxilasa puede determinarse fácilmente mediante ensayos estándar para estas actividades enzimáticas. Véase, p. ej., la publicación de patente de los EE. UU. n° 20050019882.

**[0065]** En algunas realizaciones, un ácido nucleico objeto comprende la secuencia nucleotídica mostrada en la figura 1 y expuesta en SEQ ID NO: 1. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 1. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, de aproximadamente diez a aproximadamente 15, de aproximadamente 15 a aproximadamente 20, de aproximadamente 20 a aproximadamente 25 o de aproximadamente 25 a aproximadamente 50 sustituciones de nucleótidos en comparación con la secuencia expuesta en SEQ ID NO: 1.

**[0066]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 1, en que el ácido nucleico codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa (p. ej., actividad de sesquiterpeno-oxidasa, actividad de sesquiterpeno-hidroxilasa, etc.).

**[0067]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con un tramo de al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 1.

**[0068]** En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 1. En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 1 y codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa, p. ej., actividad de sesquiterpeno-hidroxilasa y/u oxidasa.

**[0069]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que hibrida en condiciones de hibridación restrictivas con un ácido nucleico que comprende la secuencia nucleotídica expuesta

en SEQ ID NO: 1 o un complemento de esta.

**[0070]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica según se muestra en la figura 2 y se expone en SEQ ID NO: 2. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 2. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia aminoacídica con un tramo de al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450 o al menos aproximadamente 490 aminoácidos contiguos de la secuencia aminoacídica expuesta en SEQ ID NO: 2. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, de aproximadamente diez a aproximadamente 15, de aproximadamente 15 a aproximadamente 20 o de aproximadamente 20 a aproximadamente 25 sustituciones conservadoras de aminoácidos en comparación con la secuencia expuesta en SEQ ID NO: 2. En algunas realizaciones, el polipéptido codificado muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa. En algunas realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-oxidasa. En algunas realizaciones, el polipéptido codificado cataliza la oxidación del C12 de un sustrato sesquiterpénico. En otras realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-hidroxilasa.

**[0071]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450 o al menos aproximadamente 490 aminoácidos contiguos de una secuencia aminoacídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o el 100 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 2. En algunas realizaciones, el polipéptido codificado muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa. En algunas realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-oxidasa. En algunas realizaciones, el polipéptido codificado cataliza la oxidación del C12 de un sustrato sesquiterpénico. En otras realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-hidroxilasa.

**[0072]** En algunas realizaciones, un ácido nucleico objeto comprende la secuencia nucleotídica mostrada en la figura 9 y expuesta en SEQ ID NO: 5. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 5. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, de aproximadamente diez a aproximadamente 15, de aproximadamente 15 a aproximadamente 20, de aproximadamente 20 a aproximadamente 25 o de aproximadamente 25 a aproximadamente 50 sustituciones de nucleótidos en comparación con la secuencia expuesta en SEQ ID NO: 5.

**[0073]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 5, en que el ácido nucleico codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa (p. ej., actividad de sesquiterpeno-oxidasa, actividad de sesquiterpeno-hidroxilasa, etc.).

**[0074]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 57 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con un tramo de al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 5.

**[0075]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450 o al menos aproximadamente 480 aminoácidos contiguos de una secuencia aminoacídica con al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o el 100 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 6. En muchas realizaciones, el polipéptido codificado muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa. En muchas realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-oxidasa o sesquiterpeno-hidroxilasa. En muchas realizaciones, el polipéptido codificado cataliza la oxidación de un sustrato sesquiterpénico.

**[0076]** En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 5. En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400 o al menos aproximadamente 1.450 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 5 y codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/u oxidasa, p. ej., actividad de sesquiterpeno-oxidasa, actividad de sesquiterpeno-hidroxilasa, etc.

**[0077]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que hibrida en condiciones de hibridación restrictivas con un ácido nucleico que comprende la secuencia nucleotídica expuesta en SEQ ID NO: 5 o un complemento de esta.

**[0078]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica según se muestra en la figura 9 y se expone en SEQ ID NO: 6. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con al menos aproximadamente el 45 %, al menos

aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 6. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con al menos aproximadamente el 45 %, al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia aminoacídica con un tramo de al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450 o al menos aproximadamente 480 aminoácidos contiguos de la secuencia aminoacídica según se expone en SEQ ID NO: 6. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, de aproximadamente diez a aproximadamente 15, de aproximadamente 15 a aproximadamente 20 o de aproximadamente 20 a aproximadamente 25 sustituciones conservadoras de aminoácidos en comparación con la secuencia expuesta en SEQ ID NO: 6. En algunas realizaciones, el polipéptido codificado muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa. En algunas realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-oxidasa. En algunas realizaciones, el polipéptido codificado cataliza la hidroxilación de un sustrato sesquiterpénico. En otras realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-hidroxilasa.

**[0079]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450 o al menos aproximadamente 480 aminoácidos contiguos de una secuencia aminoacídica con al menos aproximadamente el 50 %, al menos aproximadamente el 55 %, al menos aproximadamente el 60 %, al menos aproximadamente el 65 %, al menos aproximadamente el 70 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o el 100 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 6. En algunas realizaciones, el polipéptido codificado muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa. En algunas realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-oxidasa. En algunas realizaciones, el polipéptido codificado cataliza la hidroxilación de un sustrato sesquiterpénico. En otras realizaciones, el polipéptido codificado muestra actividad de sesquiterpeno-hidroxilasa.

**[0080]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una variante de un polipéptido que comprende la secuencia aminoacídica expuesta en SEQ ID NO: 2 o SEQ ID NO: 6. Por ejemplo, en algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una enzima que muestra una o más de las propiedades siguientes en comparación con una enzima que comprende la secuencia aminoacídica expuesta en SEQ ID NO: 2 o SEQ ID NO: 6: 1) mayor actividad enzimática; 2) mayor estabilidad *in vitro* y/o *in vivo*; 3) mayor rendimiento de producto; 4) tasa de recambio de la proteína alterada; 5) especificidad de sustrato alterada (p. ej., de modo que la enzima variante modifica un (os) sustrato (s) seleccionado (s)); 6) mayor eficiencia enzimática (p. ej., mayor eficiencia de la conversión del sustrato para generar el producto); y 7) mayor solubilidad (p. ej., solubilidad en el citoplasma o citosol).

#### Ácidos nucleicos que codifican citocromo-P450-reductasas

**[0081]** La presente invención proporciona un ácido nucleico aislado que comprende una secuencia nucleotídica que codifica una citocromo-P450-reductasa (CPR). En algunas realizaciones, un ácido nucleico de CPR objeto comprende una secuencia nucleotídica que codifica una CPR que transfiere electrones desde NADPH a una citocromo-P450-oxidasa codificada por un ácido nucleico de enzima modificadora de isoprenoides objeto.

**[0082]** En algunas realizaciones, un ácido nucleico objeto comprende la secuencia nucleotídica mostrada en la figura 3 y expuesta en SEQ ID NO: 3. En algunas realizaciones, un ácido nucleico objeto comprende una

secuencia nucleotídica con al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 3.

- 5 **[0083]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que hibrida en condiciones de hibridación restrictivas con un ácido nucleico que comprende la secuencia nucleotídica expuesta en SEQ ID NO: 3 o un complemento de esta.

- [0084]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica según se muestra en la figura 4 y se expone en SEQ ID NO: 4. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 % o al menos aproximadamente el 99 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 4. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende una secuencia aminoacídica con una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez, de aproximadamente diez a aproximadamente 15, de aproximadamente 15 a aproximadamente 20 o de aproximadamente 20 a aproximadamente 25 sustituciones conservadoras de aminoácidos en comparación con la secuencia expuesta en SEQ ID NO: 4.

- 20 **[0085]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica un polipéptido que comprende al menos aproximadamente 50, al menos aproximadamente 75, al menos aproximadamente 100, al menos aproximadamente 150, al menos aproximadamente 200, al menos aproximadamente 250, al menos aproximadamente 300, al menos aproximadamente 350, al menos aproximadamente 400, al menos aproximadamente 450, al menos aproximadamente 500, al menos aproximadamente 550, al menos aproximadamente 600, al menos aproximadamente 650 o al menos aproximadamente 700 aminoácidos contiguos de una secuencia aminoacídica con al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o el 100 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 4. En algunas realizaciones, el polipéptido codificado transfiere electrones desde NADPH a un polipéptido (p. ej., una enzima modificadora de isoprenoides) codificado por un ácido nucleico de enzima modificadora de isoprenoides objeto.

- [0086]** En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400, al menos aproximadamente 1.500, al menos aproximadamente 1.600, al menos aproximadamente 1.700, al menos aproximadamente 1.800, al menos aproximadamente 1.900, al menos aproximadamente 2.000 o al menos aproximadamente 2.100 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 3. En algunas realizaciones, un ácido nucleico objeto comprende al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1.000, al menos aproximadamente 1.100, al menos aproximadamente 1.200, al menos aproximadamente 1.300, al menos aproximadamente 1.400, al menos aproximadamente 1.500, al menos aproximadamente 1.600, al menos aproximadamente 1.700, al menos aproximadamente 1.800, al menos aproximadamente 1.900, al menos aproximadamente 2.000 o al menos aproximadamente 2.100 nucleótidos contiguos de la secuencia nucleotídica expuesta en SEQ ID NO: 3 y codifica un polipéptido que transfiere electrones desde NADPH a una citocromo-P450-oxidasa codificada por un ácido nucleico de enzima modificadora de isoprenoides objeto, p. ej. el polipéptido codificado transfiere electrones desde NADPH a un polipéptido (p. ej., una enzima modificadora de isoprenoides) codificado por un ácido nucleico de enzima modificadora de isoprenoides objeto.

- [0087]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una variante de un polipéptido que comprende la secuencia aminoacídica expuesta en SEQ ID NO: 4. Por ejemplo, en algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una enzima que muestra una o más de las propiedades siguientes en comparación con una enzima que comprende la secuencia aminoacídica expuesta en SEQ ID NO: 4: 1) mayor actividad enzimática; 2) mayor estabilidad *in vitro* y/o *in vivo*; 3) mayor rendimiento de producto; 4) tasa de recambio de la proteína alterada; 5) especificidad de sustrato alterada (p. ej., de modo que la enzima variante modifica un (os) sustrato (s) seleccionado (s); 6) mayor eficiencia enzimática (p.

ej., mayor eficiencia de la conversión del sustrato para generar el producto); y 7) mayor solubilidad (p. ej., solubilidad en el citoplasma o citosol).

**[0088]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una proteína de fusión que comprende una secuencia aminoacídica de una enzima modificadora de isoprenoides que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa, según se describe anteriormente, fusionada con un polipéptido heterólogo (una "pareja de fusión"), p. ej., un polipéptido distinto de una enzima modificadora de isoprenoides según se describe anteriormente. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una proteína de fusión que comprende una secuencia aminoacídica de una CPR, según se describe anteriormente, y un polipéptido heterólogo, p. ej., un polipéptido distinto de una CPR. Las parejas de fusión adecuadas incluyen, pero no se limitan a polipéptidos que aumentan la solubilidad de la enzima modificadora de isoprenoides o la CPR; polipéptidos que proporcionan una señal detectable (p. ej., una proteína fluorescente; una enzima que da lugar a un producto detectable, p. ej.,  $\beta$ -galactosidasa, luciferasa, peroxidasa de rábano y similares); polipéptidos que dan lugar a la inclusión de la enzima modificadora de isoprenoides o la CPR en un compartimento celular concreto (p. ej., citosol, citoplasma); y similares.

**[0089]** En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides (p. ej., un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa) y una CPR. En algunas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que comprende una proteína de fusión que comprende una secuencia aminoacídica de una enzima modificadora de isoprenoides que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa, según se describe anteriormente, fusionada con un polipéptido CPR. En algunas realizaciones, la proteína de fusión codificada presenta la fórmula  $\text{NH}_2\text{-A-X-B-COOH}$ , en la que A es la enzima modificadora de isoprenoides que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa, X es un enlazante opcional y B es el polipéptido CPR. En algunas realizaciones, la proteína de fusión codificada presenta la fórmula  $\text{NH}_2\text{-A-X-B-COOH}$ , en la que A es el polipéptido CPR, X es un enlazante opcional y B es el polipéptido modificador de isoprenoides que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa.

**[0090]** El péptido enlazante puede tener diversas secuencias aminoacídicas. Las proteínas pueden unirse entre sí mediante un péptido espaciador, generalmente de naturaleza flexible, aunque no se excluyen otros enlaces químicos. El enlazante puede ser un enlazante escindible. Generalmente, las secuencias enlazantes adecuadas serán péptidos de entre aproximadamente 5 y aproximadamente 50 aminoácidos de longitud o de entre aproximadamente 6 y aproximadamente 25 aminoácidos de longitud. Generalmente se usarán péptidos enlazantes con un grado de flexibilidad. Los péptidos de enlace pueden tener prácticamente cualquier secuencia aminoacídica, teniendo en cuenta que los enlazantes preferidos tendrán una secuencia que resulte en un péptido generalmente flexible. Los aminoácidos pequeños como glicina y alanina son útiles para la creación de un péptido flexible. La creación de tales secuencias es una cuestión rutinaria para los expertos en la técnica. Existen diversos enlazantes disponibles comercialmente que se consideran adecuados para el uso de acuerdo con esta invención.

**[0091]** Los péptidos enlazantes adecuados incluyen frecuentemente secuencias aminoacídicas ricas en restos de alanina y prolina, que son conocidos por otorgar flexibilidad a la estructura de una proteína. Algunos enlazantes ejemplares tienen una combinación de restos de glicina, alanina, prolina y metionina, como AAAGGM (SEQ ID NO: 8), AAAGGMPPAAAGGM (SEQ ID NO: 9), AAAGGM (SEQ ID NO: 10) y PPAAAGGM (SEQ ID NO: 11). Otros péptidos enlazantes ejemplares incluyen IEGR (SEQ ID NO: 12) y GGKGGK (SEQ ID NO: 13). Sin embargo, puede usarse cualquier enlazante flexible, generalmente de entre aproximadamente 5 y aproximadamente 50 aminoácidos de longitud. Los enlazantes pueden tener prácticamente cualquier secuencia que resulte en un péptido generalmente flexible, incluidas secuencias ricas en alanina y prolina del tipo ejemplificado anteriormente.

#### Construcciones

**[0092]** La presente invención proporciona además vectores recombinantes ("construcciones") que comprenden un ácido nucleico objeto. En algunas realizaciones, un vector recombinante objeto da lugar a la amplificación de un ácido nucleico objeto. En algunas realizaciones, un vector recombinante objeto da lugar a la producción de una enzima modificadora de isoprenoides codificada o de una CPR codificada en una célula eucariota, en una célula procariota o en un sistema de transcripción y traducción acelular. Los vectores de expresión adecuados incluyen, pero no se limitan a vectores de baculovirus, vectores de bacteriófagos, plásmidos, fagémidos, cósmidos, fósmodos, cromosomas bacterianos artificiales, vectores víricos (p. ej., vectores víricos basados en el virus de la vacuna, poliovirus, adenovirus, virus adenoasociado, SV40, virus del herpes simple y similares), cromosomas artificiales basados en P1, plásmidos de levaduras, cromosomas artificiales de levaduras y cualquier otro vector para

huéspedes específicos de interés (como *E. coli*, levaduras y células vegetales).

**[0093]** En algunas realizaciones, un vector recombinante objeto comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides y un ácido nucleico objeto que codifica una CPR. En algunas de estas realizaciones, un vector recombinante objeto es un vector de expresión que da lugar a la producción de la enzima modificadora de isoprenoides codificada y de la CPR codificada en una célula eucariota, en una célula procariota o en un sistema de transcripción y traducción acelular.

**[0094]** Ciertos tipos de vectores permiten la amplificación de las casetes de expresión de la presente invención. Otros tipos de vectores son necesarios para la introducción eficiente del ácido nucleico objeto en las células y su expresión estable una vez introducido. Para los fines de la presente invención se contempla como vector recombinante adecuado cualquier vector capaz de aceptar un ácido nucleico objeto. El vector puede ser cualquier longitud circular o lineal de ADN que se integre en el genoma del huésped o se mantenga en forma episómica. Los vectores pueden requerir una manipulación adicional o condiciones especiales para incorporarse eficientemente en una célula huésped (p. ej., muchos plásmidos de expresión); o pueden ser parte de un sistema de autointegración específico de la célula (p. ej., un virus recombinante). En algunas realizaciones, el vector es funcional en una célula procariota, en la que tales vectores funcionan para propagar el vector recombinante y/o dar lugar a la expresión de un ácido nucleico objeto. En algunas realizaciones, el vector es funcional en una célula eucariota, en la que el vector será en muchas realizaciones un vector de expresión.

**[0095]** Los expertos en la técnica conocen numerosos vectores de expresión adecuados, muchos de los cuales están disponibles comercialmente. Los vectores siguientes se proporcionan como ejemplo; para células huésped bacterianas: pBluescript (Stratagene, San Diego, California, EE. UU.), vectores pQE (Qiagen), plásmidos pBluescript, vectores pNH, vectores  $\lambda$ -ZAP (Stratagene), pTrc (Amann y col., Gene, 69: 301 – 315 (1988)), pTrc99a, pKK223-3, pDR540 y pRIT2T (Pharmacia); para células huésped eucariotas: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG y pSVLSV40 (Pharmacia). Sin embargo, puede usarse cualquier otro plásmido u otro vector siempre que sea compatible con la célula huésped.

**[0096]** En muchas realizaciones, un vector recombinante objeto contendrá uno o más genes de marcadores seleccionables para proporcionar un carácter fenotípico para la selección de las células transformadas. Los marcadores seleccionables adecuados incluyen, pero no se limitan resistencia a dihidrofolato-reductasa o neomicina para un cultivo de células eucariotas y resistencia a tetraciclina o ampicilina en células huésped procariotas como *E. coli*.

**[0097]** En muchas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides, en que la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides está ligada operativamente a uno o más elementos de control transcripcional y/o traduccional. En muchas realizaciones, un ácido nucleico objeto comprende una secuencia nucleotídica que codifica una CPR, en que la secuencia nucleotídica que codifica la CPR está ligada operativamente a uno o más elementos de control transcripcional y/o traduccional.

**[0098]** En algunas realizaciones, según se señala anteriormente, un vector recombinante objeto comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides y un ácido nucleico objeto que codifica una CPR. En algunas de estas realizaciones, la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides y la secuencia nucleotídica que codifica la CPR están ligadas operativamente a diferentes elementos de control transcripcional. En otras realizaciones, la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides y la secuencia nucleotídica que codifica la CPR están ligadas operativamente al (os) mismo (s) elemento (s) de control transcripcional. En algunas realizaciones, la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides y la secuencia nucleotídica que codifica la CPR están las dos ligadas operativamente al mismo promotor inducible. En algunas realizaciones, la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides y la secuencia nucleotídica que codifica la CPR están las dos ligadas operativamente al mismo promotor constitutivo.

**[0099]** Los promotores adecuados para uso en células huésped procariotas incluyen, pero no se limitan a un promotor de la polimerasa de ARN del bacteriófago T7; un promotor *trp*; un promotor del operón *lac*; un promotor híbrido, p. ej., un promotor híbrido *lac/tac*, un promotor híbrido *tac/trc*, un promotor *trp/lac*, un promotor *T7/lac*, un promotor *trc*, un promotor *tac* y similares; un promotor *araBAD*; promotores regulados in vivo, como un promotor *ssaG* o un promotor relacionado (véase, p. ej., la publicación de patente de los EE. UU. n° 20040131637), un promotor *pagC* (Pulkkinen y Miller, J. Bacteriol., 1991: 173 (1): 86 – 93; Alpuche-Aranda y col., PNAS, 1992: 89 (21):

10079 – 83), un promotor *nirB* (Harborne y col. (1992) Mol. Micro. 6: 2805 – 2813) y similares (véanse, p. ej., Dunstan y col. (1999) Infect. Immun. 67: 5133 – 5141; McKelvie y col. (2004) Vaccine 22: 3243 – 3255; Chatfield y col. (1992) Biotechnol. 10: 888 – 892); un promotor  $\sigma 70$ , p. ej., un promotor  $\sigma 70$  consenso (véanse, p. ej., los n<sup>os</sup> de acceso de GenBank AX798980, AX798961 y AX798183); un promotor de fase estacionaria, p. ej., un promotor *dps*,  
 5 un promotor *spv* y similares; un promotor derivado de la isla de patogenicidad SPI-2 (véase, p. ej., el documento WO96/17951); un promotor *actA* (véase, p. ej., Shetron-Rama y col. (2002) Infect. Immun. 70: 1087 – 1096); un promotor *rpsM* (véase, p. ej., Valdivia y Falkow (1996) Mol. Microbiol. 22: 367 – 378); un promotor *tet* (véase p. ej., Hillen, W y Wissmann, A. (1989) en Saenger, W. y Heinemann U. (eds.), Topics in Molecular and Structural Biology, Protein-Nucleic Acid Interaction. Macmillan, Londres, Reino Unido, vol. 10, págs. 143 – 162); un promotor SP6  
 10 (véase, p. ej., Melton y col. (1984) Nucl. Acids. Res. 12: 7035 – 7056); y similares.

**[0100]** Algunos ejemplos no limitantes de promotores eucarióticos adecuados incluyen el promotor temprano inmediato del CMV, el promotor de la timidina-cinasa del VHS, los promotores temprano y tardío de SV40, el promotor LTR de retrovirus y el promotor de la metalotioneína I de ratón. En algunas realizaciones, p. ej., para  
 15 expresión en una célula de levadura, un promotor adecuado es un promotor constitutivo como un promotor *ADH1*, un promotor *PGK1*, un promotor *ENO*, un promotor *PYK1* y similares; o un promotor regulable como un promotor *GAL1*, un promotor *GAL10*, un promotor *ADH2*, un promotor *PHO5*, un promotor *CUP1*, un promotor *GAL7*, un promotor *MET25*, un promotor *MET3* y similares. La selección del vector y el promotor adecuados se encuentra dentro de las capacidades de los expertos en la técnica. El vector de expresión puede contener también un sitio de  
 20 unión a ribosomas para el inicio de la traducción y un terminador de transcripción. El vector de expresión puede incluir también secuencias apropiadas para amplificar la expresión.

**[0101]** En muchas realizaciones, una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides está ligada operativamente a un promotor inducible. En muchas realizaciones, una secuencia  
 25 nucleotídica que codifica una CPR está ligada operativamente a un promotor inducible. Los promotores inducibles son bien conocidos en la técnica. Los promotores inducibles adecuados incluyen, pero no se limitan a pL del bacteriófago  $\lambda$ ; *Plac*; *Ptp*; *Ptac* (promotor híbrido *Ptp-lac*); un promotor inducible por isopropil- $\beta$ -D-tiogalactopiranosido (IPTG), p. ej., un promotor *lacZ*; un promotor inducible por tetraciclina; un promotor inducible por arabinosa, p. ej., *P<sub>BAD</sub>* (véase, p. ej., Guzman y col., J. Bacteriol. 177: 4121 – 4130); un promotor inducible por xilosa,  
 30 p. ej., *P<sub>xyl</sub>* (véase, p. ej., Kim y col. (1996) Gene 181: 71 – 76); un promotor *GAL1*; un promotor de triptófano; un promotor *lac*; un promotor inducible por alcohol, p. ej., un promotor inducible por metanol, un promotor inducible por etanol; un promotor inducible por rafinosa; un promotor termoinducibile, p. ej., un promotor *P<sub>L</sub>* termoinducibile de  $\lambda$ , un promotor controlado por un represor termosensible (p. ej., los vectores de expresión basados en  $\lambda$ , reprimidos por C1857; véase, p. ej., Hoffmann y col. (1999) FEMS Microbiol. Lett. 177 (2): 327 – 34); y similares.

**[0102]** En levaduras, puede usarse una serie de vectores que contienen promotores constitutivos o inducibles. Como revisión, véanse Current Protocols in Molecular Biology, vol. 2, 1988, ed. Ausubel y col., Greene Publish. Assoc. y Wiley Interscience, capítulo 13; Grant, y col., 1987, Expression and Secretion Vectors for Yeast, en  
 40 Methods in Enzymology, eds. Wu y Grossman, 31987, Acad. Press, N.Y., EE. UU., vol. 153, págs. 516 – 544; Glover, 1986, DNA Cloning, vol. II, IRL Press, Washington, D.C., EE. UU., capítulo 3; y Bitter, 1987, Heterologous Gene Expression in Yeast, en Methods in Enzymology, eds. Berger y Kimmel, Acad. Press, N.Y., EE. UU., vol. 152, págs. 673 – 684; The Molecular Biology of the Yeast *Saccharomyces*, 1982, eds. Strathern y col., Cold Spring Harbor Press, vols. I y II. Puede usarse un promotor constitutivo de levaduras como *ADH* o *LEU2* o un promotor inducible como *GAL* (Cloning in Yeast, capítulo 3, R. Rothstein en: DNA Cloning vol. 11, A Practical Approach, ed.  
 45 DM. Glover, 1986, IRL Press, Washington., D.C., EE. UU.). Alternativamente pueden usarse vectores que facilitan la integración de secuencias de ADN extraño en el cromosoma de la levadura.

**[0103]** En algunas realizaciones, un ácido nucleico objeto o un vector objeto comprenden un promotor u otro (s) elemento (s) regulador (es) para expresión en una célula vegetal. Algunos ejemplos no limitantes de promotores  
 50 constitutivos adecuados que son funcionales en una célula vegetal son el promotor 35S del virus del mosaico de la coliflor, un promotor 35S en tándem (Kay y col., Science 236: 1299 (1987)), un promotor 19S del virus del mosaico de la coliflor, un promotor del gen de la nopalina-sintasa (Singer y col., Plant Mol. Biol. 14: 433 (1990); An, Plant Physiol. 81: 86 (1986)), un promotor del gen de la octopina-sintasa y un promotor de la ubiquitina. Los promotores inducibles adecuados que son funcionales en una célula vegetal incluyen, pero no se limitan a un promotor del gen  
 55 de la fenilalanina-amonioliasa, un promotor del gen de la chalcona-sintasa, un promotor del gen de la proteína relacionada con la patogénesis, un elemento regulador inducible por cobre (Mett y col., Proc. Natl. Acad. Sci. USA 90: 4567 – 4571 (1993); Furst y col., Cell 55: 705 – 717 (1988)); elementos reguladores inducibles por tetraciclina y clorotetraciclina (Gatz y col., Plant J. 2: 397 – 404 (1992); Röder y col., Mol. Gen. Genet. 243: 32 – 38 (1994); Gatz, Meth. Cell Biol. 50: 411 – 424 (1995)); elementos reguladores inducibles por ecdisona (Christopherson y col., Proc.

Natl. Acad. Sci. USA 89: 6314 – 6318 (1992); Kreutzweiser y col., Ecotoxicol. Environ. Safety 28: 14 – 24 (1994)); elementos reguladores inducibles por choque térmico (Takahashi y col., Plant Physiol. 99: 383 – 390 (1992); Yabe y col., Plant Cell Physiol. 35: 1207 – 1219 (1994); Ueda y col., Mol. Gen. Genet. 250: 533 – 539 (1996)); y elementos del operón *lac*, que se usan en combinación con un represor *lac* expresado constitutivamente para conferir, por ejemplo, expresión inducible por IPTG (Wilde y col., EMBO J. 11: 1251 – 1259 (1992)); un promotor inducible por nitrato derivado del gen de la nitrato-reductasa de espinaca (Back y col., Plant Mol. Biol. 17: 9 (1991)); un promotor fotoinducible, como el asociado con las familias génicas de la subunidad menor de la RuBP-carboxilasa o la LHCP (Feinbaum y col., Mol. Gen. Genet. 226: 449 (1991); Lam y Chua, Science 248: 471 (1990)); un elemento regulador de respuesta a la luz según se describe en la publicación de patente de los EE. UU. n° 20040038400; elementos reguladores inducibles por ácido salicílico (Uknes y col., Plant Cell 5: 159 – 169 (1993); Bi y col., Plant J. 8: 235 – 245 (1995)); elementos reguladores inducibles por hormonas vegetales (Yamaguchi-Shinozaki y col., Plant Mol. Biol. 15: 905 (1990); Kares y col., Plant Mol. Biol. 15: 225 (1990)); y elementos reguladores inducibles por hormonas humanas como el elemento de respuesta a glucocorticoides humanos (Schena y col., Proc. Natl. Acad. Sci. USA 88: 10421 (1991)).

15 **[0104]** En un ácido nucleico objeto o en un vector objeto pueden incluirse también elementos reguladores selectivos de tejidos vegetales. Los elementos reguladores selectivos de tejidos adecuados que pueden usarse para expresar ectópicamente un ácido nucleico en un solo tejido o en un número limitado de tejidos incluyen, pero no se limitan a un elemento regulador selectivo del xilema, un elemento regulador selectivo de traqueidas, un elemento regulador selectivo de fibras, un elemento regulador selectivo de tricomas (véase, p. ej., Wang y col. (2002) J. Exp. Botany 53: 1891 – 1897), un elemento regulador selectivo de tricomas glandulares y similares.

**[0105]** Los vectores adecuados para uso en células vegetales son conocidos en la técnica y cualquiera de estos vectores puede usarse para introducir un ácido nucleico objeto en una célula huésped vegetal. Los vectores adecuados incluyen, p. ej., un plásmido Ti de *Agrobacterium tumefaciens* o un plásmido Ri de *A. rhizogenes*. El plásmido Ti o Ri<sub>1</sub> se transmite a las células vegetales mediante la infección por *Agrobacterium* y se integra de manera estable en el genoma de la planta. J. Schell, Science 237: 1176 – 83 (1987). También es adecuado para uso un cromosoma vegetal artificial, según se describe, p. ej., en la patente de los EE. UU. n° 6.900.012.

### 30 Composiciones

**[0106]** La presente invención proporciona además composiciones que comprenden un ácido nucleico objeto. La presente invención proporciona además composiciones que comprenden un vector recombinante objeto. En muchas realizaciones, las composiciones que comprenden un ácido nucleico objeto o un vector de expresión objeto incluirán uno o más de los componentes siguientes: una sal, p. ej., NaCl, MgCl<sub>2</sub>, KCl, MgSO<sub>4</sub>, etc.; un agente tamponante, p. ej., un tampón de tris, ácido *N*-(2-hidroxietil)piperazin-*N'*-(2-etanosulfónico) (HEPES), ácido 2-(*N*-morfolino)etanosulfónico (MES), sal de sodio del ácido 2-(*N*-morfolino)etanosulfónico (MES), ácido 3-(*N*-morfolino)propanosulfónico (MOPS), ácido *N*-tris(hidroximetil)metil-3-aminopropanosulfónico (TAPS), etc.; un agente solubilizante; un detergente, p. ej., un detergente no iónico como Tween-20, etc.; un inhibidor de nucleasas; y similares. En algunas realizaciones, un ácido nucleico objeto o un vector recombinante objeto se liofiliza.

### Células huésped

**[0107]** La presente invención proporciona células huésped modificadas genéticamente, p. ej., células huésped que han sido modificadas genéticamente con un ácido nucleico objeto o un vector recombinante objeto. En muchas realizaciones una célula huésped modificada genéticamente objeto es una célula huésped *in vitro*. En otras realizaciones, una célula huésped modificada genéticamente objeto es una célula huésped *in vivo*. En otras realizaciones, una célula huésped modificada genéticamente es parte de un organismo pluricelular.

50 **[0108]** En muchas realizaciones, las células huésped son organismos unicelulares o crecen en cultivo como células individuales. En algunas realizaciones, la célula huésped es una célula eucariota. Las células huésped eucariotas adecuadas incluyen, pero no se limitan a células de levaduras, células de insectos, células vegetales, células de hongos y células de algas. Las células huésped eucariotas adecuadas incluyen, pero no se limitan a *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia tiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Neurospora crassa*, *Chlamydomonas reinhardtii* y similares. En algunas realizaciones, la célula huésped es una célula eucariota distinta de una célula

vegetal.

**[0109]** En otras realizaciones, la célula huésped es una célula vegetal. Las células vegetales incluyen células de monocotiledóneas y de dicotiledóneas.

5 **[0110]** En otras realizaciones, la célula huésped es una célula procariota. Las células procariotas adecuadas incluyen, pero no se limitan a cualquiera de una diversidad de cepas de laboratorio de *Escherichia coli*, *Lactobacillus* sp., *Salmonella* sp., *Shigella* sp., y similares. Véanse, p. ej., Carrier y col. (1992) J. Immunol. 148: 1176 – 1181; patente de los EE. UU. n° 6.447.784; y Sizemore y col. (1995) Science 270: 299 – 302. Algunos ejemplos de  
10 especies de *Salmonella* que pueden emplearse en la presente invención incluyen, pero no se limitan a *Salmonella typhi* y *S. typhimurium*. Las cepas adecuadas de *Shigella* incluyen, pero no se limitan a *Shigella flexneri*, *Shigella sonnei* y *Shigella dysenteriae*. Típicamente, la cepa de laboratorio es una cepa que no es patógena. Algunos ejemplos no limitantes de otras bacterias adecuadas incluyen, pero no se limitan a *Bacillus subtilis*, *Pseudomonas putida*, *Pseudomonas aeruginosa*, *Pseudomonas mavalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*,  
15 *Rhodospirillum rubrum*, *Rhodococcus* sp. y similares. En algunas realizaciones, la célula huésped es *Escherichia coli*.

**[0111]** Para generar una célula huésped modificada genéticamente objeto, un ácido nucleico objeto que comprende secuencias nucleotídicas que codifican una enzima modificadora de isoprenoides se introduce de  
20 manera estable o transitoria en una célula huésped parental mediante técnicas establecidas que incluyen, pero no se limitan a electroporación, precipitación con fosfato de calcio, transfección por medio de DEAE-dextrano, transfección por medio de liposomas y similares. Para una transformación estable, un ácido nucleico incluirá además generalmente un marcador seleccionable, p. ej., cualquiera de varios marcadores seleccionables bien conocidos como resistencia a neomicina, resistencia a ampicilina, resistencia a tetraciclina, resistencia a cloranfenicol,  
25 resistencia a kanamicina y similares.

**[0112]** En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula vegetal. Una célula vegetal modificada genéticamente objeto es útil para la producción de un compuesto isoprenoide seleccionado en un cultivo de células vegetales *in vitro*. En: Plant Cell and Tissue Culture, 1994, Vasil y Thorpe eds.,  
30 Kluwer Academic Publishers y en: Plant Cell Culture Protocols (Methods in Molecular Biology 111), 1999, Hall eds., Humana Press, por ejemplo, pueden encontrarse instrucciones concernientes al cultivo de tejidos vegetales.

#### Células huésped modificadas genéticamente

35 **[0113]** En algunas realizaciones, una célula huésped modificada genéticamente objeto comprende un vector de expresión objeto, en que el vector de expresión objeto comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides. En algunas realizaciones, una célula huésped modificada genéticamente objeto comprende un vector de expresión objeto, en que el vector de expresión objeto comprende una secuencia nucleotídica que codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa.

40 **[0114]** En algunas realizaciones, una célula huésped modificada genéticamente objeto comprende un primer vector de expresión objeto, en que el primer vector de expresión objeto comprende un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa; y además comprende un segundo vector de expresión objeto, en que el segundo vector de  
45 expresión objeto comprende un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una CPR. En otras realizaciones, una célula huésped modificada genéticamente objeto comprende un vector de expresión objeto, en que el vector de expresión objeto comprende un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides y un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una CPR. En otras realizaciones, una célula huésped  
50 modificada genéticamente objeto comprende un vector de expresión objeto, en que el vector de expresión objeto comprende un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica un polipéptido de fusión (p. ej., un polipéptido que incluye una enzima modificadora de isoprenoides y una CPR).

**[0115]** Los ácidos nucleicos que codifican CPR adecuados incluyen ácidos nucleicos que codifican CPR encontrados en plantas. Los ácidos nucleicos que codifican CPR adecuados incluyen ácidos nucleicos que codifican CPR encontrados en hongos. Algunos ejemplos de ácidos nucleicos que codifican CPR adecuados incluyen: n° de acceso de GenBank AJ303373 (CPR de *Triticum aestivum*); n° de acceso de GenBank AY959320 (CPR de *Taxus chinensis*); n° de acceso de GenBank AY532374 (CPR de *Ammi majus*); n° de acceso de GenBank AG211221 (CPR de *Oryza sativa*); n° de acceso de GenBank AF024635 (CPR de *Petroselinum crispum*).

**[0116]** En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula huésped que normalmente no sintetiza isopentenilpirofosfato (IPP) ni mevalonato por la ruta del mevalonato. La ruta del mevalonato comprende: (a) la condensación de dos moléculas de acetil-CoA para dar acetoacetil-CoA; (b) la condensación de acetoacetil-CoA con acetil-CoA para formar HMG-CoA; (c) la conversión de HMG-CoA en mevalonato; (d) la fosforilación de mevalonato para dar mevalonato-5-fosfato; (e) la conversión de mevalonato-5-fosfato en mevalonato-5-pirofosfato; y (f) la conversión de mevalonato-5-pirofosfato en isopentenilpirofosfato. Las enzimas de la ruta del mevalonato requeridas para la producción de IPP pueden variar dependiendo de las condiciones de cultivo.

10

**[0117]** Según se señala anteriormente, en algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula huésped que normalmente no sintetiza isopentenilpirofosfato (IPP) ni mevalonato por la ruta del mevalonato. En algunas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión objeto que comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides; y la célula huésped se modifica genéticamente con uno o más ácidos nucleicos heterólogos que comprenden secuencias nucleotídicas que codifican acetoacetil-CoA-tiolasa, hidroximetilglutaril-CoA-sintasa (HMGS), hidroximetilglutaril-CoA-reductasa (HMGR), mevalonato-cinasa (MK), fosfomevalonato-cinasa (PMK) y mevalonato-pirofosfato-descarboxilasa (MPD) (y opcionalmente también IPP-isomerasa). En muchas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión que comprende una secuencia nucleotídica que codifica una CPR. En algunas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión que comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides; y la célula huésped se modifica genéticamente con uno o más ácidos nucleicos heterólogos que comprenden secuencias nucleotídicas que codifican MK, PMK, MPD (y opcionalmente también IPP-isomerasa). En muchas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión que comprende una secuencia nucleotídica que codifica una CPR.

**[0118]** En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula huésped que normalmente no sintetiza IPP ni mevalonato por la ruta del mevalonato; la célula huésped se modifica genéticamente con un vector de expresión objeto que comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides; y la célula huésped se modifica genéticamente con uno o más ácidos nucleicos heterólogos que comprenden secuencias nucleotídicas que codifican acetoacetil-CoA-tiolasa, HMGS, HMGR, MK, PMK, MPD, IPP-isomerasa y una prenil-transferasa. En muchas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión que comprende una secuencia nucleotídica que codifica una CPR. En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula huésped que normalmente no sintetiza IPP ni mevalonato por la ruta del mevalonato; la célula huésped se modifica genéticamente con un vector de expresión objeto que comprende un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides; y la célula huésped se modifica genéticamente con uno o más ácidos nucleicos heterólogos que comprenden secuencias nucleotídicas que codifican MK, PMK, MPD, IPP-isomerasa y una prenil-transferasa. En muchas de estas realizaciones, la célula huésped se modifica genéticamente con un vector de expresión que comprende una secuencia nucleotídica que codifica una CPR.

**[0119]** En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula que normalmente sintetiza IPP o mevalonato por la ruta del mevalonato, p. ej., la célula huésped es una célula que comprende una ruta endógena del mevalonato. En algunas de estas realizaciones, la célula huésped es una célula de levadura. En algunas de estas realizaciones, la célula huésped es *Saccharomyces cerevisiae*.

**[0120]** En algunas realizaciones, una célula huésped modificada genéticamente se modifica genéticamente además con uno o más ácidos nucleicos que comprenden secuencias nucleotídicas que codifican una deshidrogenasa o deshidrogenasas, en que la deshidrogenasa modifica adicionalmente un compuesto isoprenoide. La deshidrogenasa codificada puede ser una que se encuentre naturalmente en una célula procariota o una célula eucariota o puede ser una variante de una deshidrogenasa semejante. En algunas realizaciones, la presente invención proporciona ácidos nucleicos aislados que comprenden secuencias nucleotídicas que codifican tales deshidrogenasas.

## 55 Ácidos nucleicos de la ruta del mevalonato

**[0121]** Las secuencias nucleotídicas que codifican los productos génicos de la ruta del MEV son conocidas en la técnica y cualquier secuencia nucleotídica conocida que codifique un producto génico de la ruta del MEV puede usarse para generar una célula huésped modificada genéticamente objeto. Por ejemplo, las secuencias

nucleotídicas que codifican acetoacetil-CoA-tiolasa, HMGS, HMGR, MK, PMK, MPD e IDI son conocidas en la técnica. Los siguientes son ejemplos no limitantes de secuencias nucleotídicas conocidas que codifican productos génicos de la ruta del MEV, con los números de acceso de GenBank y los organismos en paréntesis después de cada enzima de la ruta del MEV: acetoacetil-CoA-tiolasa: (NC\_000913, región: 2324131..2325315; *E. coli*), (D49362; *Paracoccus denitrificans*) y (L20428; *Saccharomyces cerevisiae*); HMGS: (NC\_001145, complemento 19061..20536; *Saccharomyces cerevisiae*), (X96617; *Saccharomyces cerevisiae*), (X83882; *Arabidopsis thaliana*), (AB037907; *Kitasatospora griseola*) y (BT007302; *Homo sapiens*); HMGR: (NM\_206548; *Drosophila melanogaster*), (NM\_204485; *Gallus gallus*), (AB015627; *Streptomyces* sp. KO-3988), (AF542543; *Nicotiana attenuata*), (AB037907; *Kitasatospora griseola*), (AX128213, que proporciona la secuencia que codifica una HMGR truncada; *Saccharomyces cerevisiae*) y (NC\_001145, complemento 115734..118898; *Saccharomyces cerevisiae*); MK: (L77688; *Arabidopsis thaliana*) y (X55875; *Saccharomyces cerevisiae*); PMK: (AF429385; *Hevea brasiliensis*), (NM\_006556; *Homo sapiens*), (NC\_001145, complemento 712315..713670; *Saccharomyces cerevisiae*); MPD: (X97557; *Saccharomyces cerevisiae*), (AF290095; *Enterococcus faecium*) y (U49260; *Homo sapiens*); e IDI: (NC\_000913, 3031087..3031635; *E. coli*) y (AF082326; *Haematococcus pluvialis*).

**[0122]** En algunas realizaciones, la región codificante de HMGR codifica una forma truncada de HMGR ("tHMGR") que carece del dominio transmembrana de la HMGR natural. El dominio transmembrana de la HMGR contiene las porciones reguladoras de la enzima y no tiene actividad catalítica.

**[0123]** La secuencia codificante de cualquier enzima conocida de la ruta del MEV puede alterarse de diversas maneras conocidas en la técnica para generar cambios específicos en la secuencia aminoacídica de la enzima codificada. Normalmente, la secuencia aminoacídica de una enzima variante de la ruta del MEV será sustancialmente similar a la secuencia aminoacídica de cualquier enzima conocida de la ruta del MEV, es decir, diferirá en al menos un aminoácido y puede diferir en al menos dos, al menos cinco, al menos diez o al menos 20 aminoácidos, pero, típicamente, en no más de aproximadamente 50 aminoácidos. Los cambios en la secuencia pueden ser sustituciones, inserciones o deleciones. Por ejemplo, según se describe más adelante, la secuencia nucleotídica puede alterarse por la preferencia de codones de una célula huésped concreta. Además, pueden introducirse una o más diferencias en la secuencia nucleotídica que resulten en cambios conservadores de aminoácidos en la proteína codificada.

#### Prenil-transferasas

**[0124]** En algunas realizaciones, una célula huésped modificada genéticamente objeto se modifica genéticamente para incluir un ácido nucleico que comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides; y en algunas realizaciones también se modifica genéticamente para incluir uno o más ácidos nucleicos que comprenden una (s) secuencia (s) nucleotídica (s) que codifican una o más enzimas de la ruta del mevalonato, según se describe anteriormente; y un ácido nucleico que comprende una secuencia nucleotídica que codifica una prenil-transferasa.

**[0125]** Las prenil-transferasas constituyen un amplio grupo de enzimas que catalizan la condensación consecutiva de IPP, lo que resulta en la formación de prenildifosfatos de diversas longitudes de cadena. Las prenil-transferasas adecuadas incluyen enzimas que catalizan la condensación de IPP con sustratos iniciadores alílicos para formar compuestos isoprenoides con aproximadamente dos unidades de isopreno hasta aproximadamente 6.000 unidades de isopreno o más, p. ej., dos unidades de isopreno (geranilpirofosfato-sintasa), tres unidades de isopreno (farnesilpirofosfato-sintasa), cuatro unidades de isopreno (geranilgeranilpirofosfato-sintasa), cinco unidades de isopreno, seis unidades de isopreno (hexadecilpirofosfato-sintasa), siete unidades de isopreno, ocho unidades de isopreno (fitoeno-sintasa, octaprenilpirofosfato-sintasa), nueve unidades de isopreno (nonaprenilpirofosfato-sintasa) diez unidades de isopreno (decaprenilpirofosfato-sintasa), de aproximadamente diez a aproximadamente 15 unidades de isopreno, de aproximadamente 15 unidades de isopreno a aproximadamente 20 unidades de isopreno, de aproximadamente 20 unidades de isopreno a aproximadamente 25 unidades de isopreno, de aproximadamente 25 unidades de isopreno a aproximadamente 30 unidades de isopreno, de aproximadamente 30 unidades de isopreno a aproximadamente 40 unidades de isopreno, de aproximadamente 40 unidades de isopreno a aproximadamente 50 unidades de isopreno, de aproximadamente 50 unidades de isopreno a aproximadamente 100 unidades de isopreno, de aproximadamente 100 unidades de isopreno a aproximadamente 250 unidades de isopreno, de aproximadamente 250 unidades de isopreno a aproximadamente 500 unidades de isopreno, de aproximadamente 500 unidades de isopreno a aproximadamente 1.000 unidades de isopreno, de aproximadamente 1.000 unidades de isopreno a aproximadamente 2.000 unidades de isopreno, de aproximadamente 2.000 unidades de isopreno a aproximadamente 3.000 unidades de isopreno, de aproximadamente 3.000 unidades de isopreno a aproximadamente 4.000 unidades de isopreno, de aproximadamente 4.000 unidades de isopreno a

aproximadamente 5.000 unidades de isopreno o de aproximadamente 5.000 unidades de isopreno a aproximadamente 6.000 unidades de isopreno o más.

**[0126]** Las prenil-transferasas adecuadas incluyen, pero no se limitan a una E-isoprenildifosfato-sintasa que incluye, pero no se limita a geranildifosfato (GPP)-sintasa, farnesildifosfato (FPP)-sintasa, geranilgeranildifosfato (GGPP)-sintasa, hexaprenildifosfato (HexPP)-sintasa, heptaprenildifosfato (HepPP)-sintasa, octaprenildifosfato (OPP)-sintasa, solanesildifosfato (SPP)-sintasa, decaprenildifosfato (DPP)-sintasa, sintasa de chicle y sintasa de gutapercha; y una Z-isoprenildifosfato-sintasa que incluye, pero no se limita a nonaprenildifosfato (NPP)-sintasa, undecaprenildifosfato (UPP)-sintasa, deshidrodoliquildifosfato-sintasa, eicosaprenildifosfato-sintasa, sintasa de la goma natural y otras Z-isoprenildifosfato-sintasas.

**[0127]** Las secuencias nucleotídicas de numerosas prenil-transferasas de diversas especies son conocidas y pueden usarse o modificarse para su uso en la generación de una célula huésped modificada genéticamente objeto. Las secuencias nucleotídicas que codifican prenil-transferasas son conocidas en la técnica. Véanse, p. ej., el ARNm de la farnesilpirofosfato-sintetasa humana (n° de acceso de GenBank J05262; *Homo sapiens*); el gen de la farnesildifosfato-sintetasa (FPP) (n° de acceso de GenBank J05091; *Saccharomyces cerevisiae*); el gen de la isopentenildifosfato: dimetialildifosfato-isomerasa (J05090; *Saccharomyces cerevisiae*); Wang y Ohnuma (2000) Biochim. Biophys. Acta 1529: 33 – 48; patente de los EE. UU. n° 6.645.747; el ARNm de la farnesilpirofosfato-sintetasa 2 (FPS2) / FPP-sintetasa 2 / farnesildifosfato-sintetasa 2 (At4g17190) de *Arabidopsis thaliana* (n° de acceso de GenBank NM\_202836); el ARNm de la geranilgeranildifosfato-sintasa (ggpps) de *Ginkgo biloba* (n° de acceso de GenBank AY371321); el ARNm de la geranilgeranilpirofosfato-sintasa (GGPS1) / GGPP-sintetasa / farnesil-transtransferasa de *Arabidopsis thaliana* (At4g36810) (n° de acceso de GenBank NM\_119845); el gen de *Synechococcus elongatus* para farnesil, geranilgeranil, geranilfarnesil, hexaprenil, heptaprenildifosfato-sintasa (Self-HepPS) (n° de acceso de GenBank AB016095); etc.

#### Terpeno-sintasas

**[0128]** En algunas realizaciones, una célula huésped modificada genéticamente objeto se modifica genéticamente para incluir un ácido nucleico que comprende una secuencia nucleotídica que codifica una terpeno-sintasa. En algunas realizaciones, la terpeno-sintasa es una sintasa que modifica FPP para generar un sesquiterpeno. En otras realizaciones, la terpeno-sintasa es una sintasa que modifica GPP para generar un monoterpene. En otras realizaciones, la terpeno-sintasa es una sintasa que modifica GGPP para generar un diterpeno.

**[0129]** Las secuencias nucleotídicas que codifican terpeno-sintasas son conocidas en la técnica y cualquier secuencia nucleotídica conocida que codifique una terpeno-sintasa puede usarse para modificar genéticamente una célula huésped. Por ejemplo, son conocidas y pueden usarse las siguientes secuencias nucleotídicas que codifican terpeno-sintasas, indicadas seguidas de sus números de acceso de GenBank y de los organismos en que se identificaron: ARNm de (-)-germacreno-D-sintasa (AY438099; *Populus balsamifera* subsp. *trichocarpa* x *Populus deltoides*); ARNm de E,E- $\alpha$ -farneseno-sintasa (AY640154; *Cucumis sativus*); ARNm de 1,8-cineol-sintasa (AY691947; *Arabidopsis thaliana*); ARNm de terpeno-sintasa 5 (TPS5) (AY518314; *Zea mays*); ARNm de terpeno-sintasa 4 (TPS4) (AY518312; *Zea mays*); ARNm de mirceno/ocimeno-sintasa (TPS10) (At2g24210) (NM\_127982; *Arabidopsis thaliana*); ARNm de geraniol-sintasa (GES) (AY362553; *Ocimum basilicum*); ARNm de pineno-sintasa (AY237645; *Picea sitchensis*); ARNm de mirceno-sintasa 1e20 (AY195609; *Antirrhinum majus*); ARNm de (E)- $\beta$ -ocimeno-sintasa (0e23) (AY195607; *Antirrhinum majus*); ARNm de E- $\beta$ -ocimeno-sintasa (AY151086; *Antirrhinum majus*); ARNm de terpeno-sintasa (AF497492; *Arabidopsis thaliana*); ARNm de (-)-canfeno-sintasa (AG6.5) (U87910; *Abies grandis*); gen de (-)-4S-limoneno-sintasa (p. ej., secuencia genómica) (AF326518; *Abies grandis*); gen de  $\delta$ -selineno-sintasa (AF326513; *Abies grandis*); ARNm de amorfa-4,11-dieno-sintasa (AJ251751; *Artemisia annua*); ARNm de E- $\alpha$ -bisaboleno-sintasa (AF006195; *Abies grandis*); ARNm de  $\gamma$ -humuleno-sintasa (U92267; *Abies grandis*); ARNm de  $\delta$ -selineno-sintasa (U92266; *Abies grandis*); ARNm de pineno-sintasa (AG3.18) (U87909; *Abies grandis*); ARNm de mirceno-sintasa (AG2.2) (U87908; *Abies grandis*); etc.

#### Uso de codones

**[0130]** En algunas realizaciones, una secuencia nucleotídica usada para generar una célula huésped modificada genéticamente objeto se modifica de modo que la secuencia nucleotídica refleje la preferencia de codones de la célula huésped concreta. Por ejemplo, en algunas realizaciones, la secuencia nucleotídica se modificará para reflejar la preferencia de codones de levaduras. Véase p. ej., Bennetzen y Hall (1982) J. Biol. Chem. 257 (6): 3026 – 3031. Como otro ejemplo no limitante, en otras realizaciones, la secuencia nucleotídica se modificará

para reflejar la preferencia de codones de *E. coli*. Véanse, p. ej., Gouy y Gautier (1982) *Nucleic Acids Res.* 10 (22): 7055 – 7074; Eyre-Walker (1996) *Mol. Biol. Evol.* 13 (6): 864 – 872. Véase también Nakamura y col. (2000) *Nucleic Acids Res.* 28 (1): 292.

## 5 Modificaciones genéticas adicionales

**[0131]** En algunas realizaciones, una célula huésped modificada genéticamente objeto es una célula que se modifica genéticamente para incluir uno o más ácidos nucleicos que comprenden una secuencia (s) nucleotídica (s) que codifica (n) una enzima modificadora de isoprenoides; y que además se modifica genéticamente para conseguir un aumento de la producción de un intermedio de una ruta biosintética de terpenos y/o que se modifica genéticamente además de modo que se inactiva funcionalmente un gen de una ruta biosintética de terpenos endógena. El término “inactivado funcionalmente”, según se usa en este documento en el contexto de un gen de una ruta biosintética de terpenos endógena, se refiere a una modificación genética de un gen de una ruta biosintética de terpenos, en que dicha modificación resulta en la producción de un producto génico codificado por el gen que se produce a niveles inferiores a los normales y/o no es funcional.

**[0132]** Las modificaciones genéticas que aumentan la producción de un intermedio de una ruta biosintética de terpenos endógena incluyen, pero no se limitan a modificaciones genéticas que resultan en un nivel y/o actividad reducidos de una fosfotransacetilasa en la célula huésped. La concentración intracelular de un intermedio de una ruta biosintética de terpenos aumenta al aumentar la concentración intracelular de acetil-CoA. *E. coli* secreta una fracción significativa de acetil-CoA en forma de acetato al medio. La delección del gen que codifica la fosfotransacetilasa, *pta*, la primera enzima responsable de la transformación de acetil-CoA en acetato, reduce la secreción de acetato. Las modificaciones genéticas que reducen el nivel y/o la actividad de la fosfotransacetilasa en una célula huésped procariota son especialmente útiles en los casos en que la célula huésped modificada genéticamente es una célula que se modifica genéticamente con un ácido nucleico que comprende secuencias nucleotídicas que codifican uno o más productos génicos de la ruta del MEV.

**[0133]** En algunas realizaciones, una modificación genética que resulta en una reducción del nivel de fosfotransacetilasa en una célula huésped procariota es una mutación genética que inactiva funcionalmente el gen *pta* endógeno de la célula huésped procariota que codifica la fosfotransacetilasa. El gen *pta* puede inactivarse funcionalmente de cualquiera de diversas maneras, incluidas la inserción de un elemento genético móvil (p. ej., un transposón, etc.); la delección de todo o de parte del gen, de modo que el producto génico no se produce o está truncado y no es funcional para convertir acetil-CoA en acetato; la mutación del gen de modo que el producto génico no se produce o está truncado y no es funcional para convertir acetil-CoA en acetato; la delección o mutación de uno o más elementos de control que controlan la expresión del gen *pta*, de modo que el producto génico no se produce; y similares.

**[0134]** En algunas realizaciones, el gen *pta* endógeno de una célula huésped modificada genéticamente se deleciona. Puede usarse cualquier procedimiento para deleccionar un gen. Un ejemplo no limitante de un procedimiento para la delección de un gen *pta* es el uso del sistema de recombinación  $\lambda$ Red. Datsenko y Wanner (2000) *Proc. Natl. Acad. Sci. USA* 97 (12): p. 6640 – 5. En algunas realizaciones, el gen *pta* se deleccionará de una célula huésped (p. ej., *E. coli*) que se modifica genéticamente con un ácido nucleico que comprende secuencias nucleotídicas que codifican MK, PMK, MPD e IDI. En algunas realizaciones, el gen *pta* se deleccionará de una célula huésped (p. ej., *E. coli*) que se modifica genéticamente con un ácido nucleico que comprende secuencias nucleotídicas que codifican MK, PMK, MPD e IPP. En algunas realizaciones, el gen *pta* se deleccionará de una célula huésped (p. ej., *E. coli*) que se modifica genéticamente con un ácido nucleico que comprende secuencias nucleotídicas que codifican MK, PMK, MPD, IPP y una prenil-transferasa.

**[0135]** En algunas realizaciones, una célula modificada genéticamente objeto es una célula que se modifica genéticamente para incluir uno o más ácidos nucleicos que comprenden una secuencia (s) nucleotídica (s) que codifica (n) un producto (s) génico (s) de la ruta biosintética del MEV; y que además se modifica genéticamente de modo que se inactiva funcionalmente un gen de la ruta biosintética de la DXP endógena. En otras realizaciones, una célula modificada genéticamente objeto es una célula que se modifica genéticamente para incluir uno o más ácidos nucleicos que comprenden una secuencia (s) nucleotídica (s) que codifica (n) un producto (s) génico (s) de la ruta biosintética de la DXP; y que además se modifica genéticamente de modo que se inactiva funcionalmente un gen de la ruta biosintética del MEV endógena.

**[0136]** En algunas realizaciones en las que la célula huésped modificada genéticamente objeto es una célula procariota que se modifica genéticamente con un ácido (s) nucleico (s) que comprende (n) secuencias nucleotídicas

que codifican uno o más productos génicos de la ruta del MEV, la célula huésped se modificará genéticamente además de modo que se inactiven funcionalmente uno o más genes de la ruta de la DXP endógena. Los genes de la ruta de la DXP que pueden inactivarse funcionalmente incluyen uno o más de los genes que codifican cualquiera de los productos génicos de la ruta de la DXP siguientes: 1-desoxi-D-xilulosa-5-fosfato-sintasa, 1-desoxi-D-xilulosa-5-fosfato-reductoisomerasa, 4-difosfocitidil-2C-metil-D-eritritol-sintasa, 4-difosfocitidil-2C-metil-D-eritritol-cinasa, 2C-metil-D-eritritol-2,4-ciclodifosfato-sintasa y 1-hidroxi-2-metil-2- (E)-butenil-4-difosfato-sintasa.

**[0137]** Un gen de la ruta de la DXP endógena puede inactivarse funcionalmente de cualquiera de diversas maneras, incluidas la inserción de un elemento genético móvil (p. ej., un transposón, etc.); la delección de todo o de parte del gen, de modo que el producto génico no se produce o está truncado y es enzimáticamente inactivo; la mutación del gen de modo que el producto génico no se produce o está truncado y no es enzimáticamente funcional; la delección o mutación de uno o más elementos de control que controlan la expresión del gen, de modo que el producto génico no se produce; y similares.

**[0138]** En otras realizaciones, en las que la célula huésped modificada genéticamente objeto es una célula procariota que se modifica genéticamente con un ácido (s) nucleico (s) que comprende (n) secuencias nucleotídicas que codifican uno o más productos génicos de la ruta de la DXP, la célula huésped se modificará genéticamente además de modo que se inactiven funcionalmente uno o más genes de la ruta del MEV endógena. Los genes de la ruta del MEV endógena que pueden inactivarse funcionalmente incluyen uno o más de los genes que codifican cualquiera de los productos génicos de la ruta del MEV siguientes: HMGS, HMGR, MK, PMK, MPD e IDI. Un gen de la ruta del MEV endógena puede inactivarse funcionalmente de cualquiera de diversas maneras, incluidas la inserción de un elemento genético móvil (p. ej., un transposón, etc.); la delección de todo o de parte del gen, de modo que el producto génico no se produce o está truncado y es enzimáticamente inactivo; la mutación del gen de modo que el producto génico no se produce o está truncado y no es enzimáticamente funcional; la delección o mutación de uno o más elementos de control que controlan la expresión del gen, de modo que el producto génico no se produce; y similares.

#### Composiciones que comprenden una célula huésped modificada genéticamente objeto

**[0139]** La presente invención proporciona además composiciones que comprenden una célula huésped modificada genéticamente objeto. Una composición objeto comprende una célula huésped modificada genéticamente objeto y, en algunas realizaciones, comprenderá uno o más componentes adicionales, en que tales componentes se seleccionan basados en parte en el uso deseado para la célula huésped modificada genéticamente. Los componentes adecuados incluyen, pero no se limitan a sales; tampones; estabilizantes; agentes inhibidores de proteasas; agentes inhibidores de nucleasas; compuestos para protección de la membrana celular y/o la pared celular, p. ej., glicerol, dimetilsulfóxido, etc.; medios nutricionales apropiados para la célula; y similares. En algunas realizaciones las células se liofilizan.

#### Plantas transgénicas

**[0140]** En algunas realizaciones, un ácido nucleico objeto o un vector de expresión objeto (p. ej., un ácido nucleico de enzima modificadora de isoprenoides objeto o un vector de expresión objeto que comprende un ácido nucleico de enzima modificadora de isoprenoides) se usa como un transgén para generar una planta transgénica que produce la enzima modificadora de isoprenoides codificada. Por lo tanto, la presente invención proporciona además una planta transgénica, en que la planta comprende un transgén que comprende un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una enzima que muestra actividad de terpeno-hidroxilasa y/o terpeno-oxidasa, según se describe anteriormente. En algunas realizaciones, el genoma de la planta transgénica comprende un ácido nucleico objeto. En algunas realizaciones, la planta transgénica es homocigótica con respecto a la modificación genética. En algunas realizaciones, la planta transgénica es heterocigótica con respecto a la modificación genética.

**[0141]** En algunas realizaciones, una planta transgénica objeto produce un polipéptido codificado por un transgén que muestra actividad de terpeno-hidroxilasa y/u oxidasa en una cantidad que es al menos aproximadamente el 50 %, al menos aproximadamente dos veces, al menos aproximadamente cinco veces, al menos aproximadamente diez veces, al menos aproximadamente 25 veces, al menos aproximadamente 50 veces o al menos aproximadamente 100 veces o más la cantidad del polipéptido producido por la planta control, p. ej., una planta no transgénica (una planta que no contiene el transgén que codifica el polipéptido) de la misma especie.

**[0142]** En algunas realizaciones, una planta transgénica objeto es una versión transgénica de una planta de

control no transgénica que produce normalmente un compuesto isoprenoide que se genera por, o es un producto posterior de un polipéptido codificado por un transgén que muestra actividad de terpeno-hidroxilasa y/u oxidasa; en que la planta transgénica produce el compuesto isoprenoide en una cantidad que es al menos aproximadamente el 50 %, al menos aproximadamente dos veces, al menos aproximadamente cinco veces, al menos aproximadamente diez veces, al menos aproximadamente 25 veces, al menos aproximadamente 50 veces o al menos aproximadamente 100 veces o más la cantidad del isoprenoide producido por la planta control, p. ej., una planta no transgénica (una planta que no contiene el transgén que codifica el polipéptido) de la misma especie.

**[0143]** Los procedimientos de introducción de ácidos nucleicos exógenos en células vegetales son bien conocidos en la técnica. Tales células vegetales se consideran "transformadas", según se define anteriormente. Los procedimientos adecuados incluyen la infección vírica (como con los virus de ADN bicatenario), transfección, conjugación, fusión de protoplastos, electroporación, tecnología de bombardeo de partículas, precipitación con fosfato de calcio, microinyección directa, tecnología de fibras de carburo de silicio, transformación por medio de *Agrobacterium tumefaciens* y similares. La elección del procedimiento depende generalmente del tipo de célula que se transforma y de las circunstancias en que tiene lugar la transformación (es decir, *in vitro*, *ex vivo*, o *in vivo*).

**[0144]** Los procedimientos de transformación basados en la bacteria del suelo *Agrobacterium tumefaciens* son especialmente útiles para la introducción de una molécula de ácido nucleico exógena en una planta vascular. La forma natural de *Agrobacterium* contiene un plásmido Ti (inductor de tumores) que dirige la producción de un crecimiento de agallas tumorígenas en el cuello de las plantas huésped. La transferencia de la región inductora de tumores, ADN-T, del plásmido Ti al genoma de una planta requiere los genes de virulencia codificados por el plásmido Ti, así como los bordes del ADN-T, que son una serie de repeticiones directas que delimitan la región que va a transferirse. Un vector basado en *Agrobacterium* es una forma modificada del plásmido Ti, en el que las funciones inductoras de tumores se sustituyen por la secuencia de ácido nucleico de interés que ha de introducirse en la planta huésped.

**[0145]** Generalmente, la transformación por medio de *Agrobacterium* emplea vectores cointegrados o, preferentemente, sistemas de vectores binarios en los que los componentes del plásmido Ti se reparten entre un plásmido auxiliar, que reside permanentemente en el *Agrobacterium* huésped y porta los genes de virulencia, y un vector de transferencia que contiene el gen de interés flanqueado por secuencias del ADN-T. En la técnica se conocen diversos vectores binarios que pueden obtenerse comercialmente, p. ej., de Clontech (Palo Alto, California, EE. UU.). Los procedimientos para el cocultivo de *Agrobacterium* con células vegetales en cultivo o con tejidos lesionados como, p. ej., tejido foliar, explantes radiculares, hipocotiledones, piezas de tallo o tubérculos también son bien conocidos en la técnica. Véase, p. ej., Glick y Thompson, (eds.). *Methods in Plant Molecular Biology and Biotechnology*, Boca Raton, Florida, EE. UU.: CRC Press (1993).

**[0146]** La transformación por medio de *Agrobacterium* es útil para la producción de diversas plantas transgénicas vasculares (Wang y col., cita anterior, 1995), incluidas al menos una especie de *Eucalyptus* y leguminosas forrajeras como alfalfa (lucerna), loto de cuernecillo, trébol blanco, *Stylosanthes*, *Lotononis bainesii* y esparceta.

**[0147]** La transformación por medio de microproyectiles también puede usarse para producir una planta transgénica objeto. Este procedimiento, descrito por primera vez por Klein y col. (Nature 327: 70 – 73 (1987)) utiliza microproyectiles de oro o de wolframio que se recubren con la molécula de ácido nucleico deseado por precipitación con cloruro de calcio, espermidina o polietilenglicol. Las partículas de los microproyectiles se introducen a gran velocidad en el tejido de angiospermas mediante un dispositivo como el BIOLISTIC PD-1000 (Biorad; Hercules, California, EE. UU.).

**[0148]** Un ácido nucleico objeto puede introducirse en una planta de manera tal que el ácido nucleico puede penetrar en la (s) célula (s) vegetal (es), p. ej., por medio de un protocolo *in vivo* o *ex vivo*. Con "*in vivo*" quiere decirse que el ácido nucleico se administra a una planta viva, p. ej., por infiltración. Con "*ex vivo*" quiere decirse que se modifican fuera de la planta células o explantes y después, a partir de estas células o explantes, se regenera una planta. Se ha descrito una serie de vectores adecuados para la transformación estable de células vegetales o para el establecimiento de plantas transgénicas, incluidos los descritos en Weissbach y Weissbach (1989) *Methods for Plant Molecular Biology*, Academic Press y Gelvin y col. (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Algunos ejemplos específicos incluyen los derivados de un plásmido Ti de *Agrobacterium tumefaciens*, así como los descritos por Herrera-Estrella y col. (1983) *Nature* 303: 209, Bevan (1984) *Nucl. Acids Res.* 12: 8711 – 8721, Klee (1985) *Bio/Technol.* 3: 637 – 642. Alternativamente, es posible usar plásmidos distintos del Ti para transferir el ADN a plantas y células mediante el uso de técnicas de administración de ADN libre. Mediante el uso de

estos procedimientos pueden producirse plantas transgénicas de trigo, arroz (Christou (1991) *Bio/Technology* 9: 957 – 962) y maíz (Gordon-Kamm (1990) *Plant Cell* 2: 603 – 618). Un embrión inmaduro puede ser también un tejido diana adecuado en monocotiledóneas para las técnicas de administración directa de ADN mediante bombardeo de partículas (Weeks y col. (1993) *Plant Physiol.* 102: 1077 – 1084; Vasil (1993) *Bio/Technol.* 10: 667 – 674; Wan y Lemeaux (1994) *Plant Physiol.* 104: 37 – 48) y para la transferencia de ADN por medio de *Agrobacterium* (Ishida y col. (1996) *Nature Biotech.* 14: 745 – 750). Algunos procedimientos ejemplares para la introducción de ADN en cloroplastos son el bombardeo de partículas, la transformación de protoplastos con polietilenglicol y la microinyección (Danieli y col., *Nat. Biotechnol.* 16: 345 – 348, 1998; Staub y col., *Nat. Biotechnol.* 18: 333 – 338, 2000; O'Neill y col. *Plant J.* 3: 729 – 738, 1993; Knoblauch y col. *Nat. Biotechnol.* 17: 906 – 909; patentes de los EE. UU. n<sup>os</sup> 5.451.513, 5.545.817, 5.545.818 y 5.576.198; solicitud internacional n<sup>o</sup> WO 95/16783; y en Boyton y col., *Methods in Enzymology* 217: 510 – 536 (1993), Svab y col., *Proc. Natl. Acad. Sci. USA* 90: 913 – 917 (1993) y McBride y col., *Proc. Natl. Acad. Sci. USA* 91: 7301 – 7305 (1994)). Cualquier vector adecuado para los procedimientos de bombardeo de partículas, transformación de protoplastos con polietilenglicol y microinyección será adecuado como vector específico para la transformación de cloroplastos. Como vector de transformación puede usarse cualquier vector de ADN bicatenario, especialmente cuando el procedimiento de introducción no utiliza *Agrobacterium*.

**[0149]** Las plantas que pueden modificarse genéticamente incluyen cereales, cultivos forrajeros, frutas, verduras, cultivos de semillas oleaginosas, palmeras, plantas forestales y enredaderas. Algunos ejemplos específicos de plantas que pueden modificarse son: maíz, plátano, cacahuate, guisantes, girasol, tomate, colza, tabaco, trigo, cebada, avena, patata, soja, algodón, claveles, sorgo, altramuz y arroz. Otros ejemplos incluyen *Artemisia annua* u otras plantas de las que se sabe que producen compuestos isoprenoides de interés.

**[0150]** La invención objeto también proporciona células vegetales transformadas, tejidos, plantas y productos que contienen las células vegetales transformadas. Una característica de las células transformadas objeto y de los tejidos y productos que las incluyen es la presencia de un ácido nucleico objeto integrado en el genoma y la producción por las células vegetales de un polipéptido que muestra actividad de terpeno-hidroxilasa y/o terpeno oxidasa, p. ej., una sesquiterpeno-oxidasa. Las células vegetales recombinantes de la presente invención son útiles como poblaciones de células recombinantes o como tejido, semilla, planta completa, tallo, fruto, hoja, raíz, flor, tallo, tubérculo, grano, pienso, un campo de cultivo y similares.

**[0151]** La invención objeto proporciona también material reproductivo de una planta transgénica objeto, en que el material reproductivo incluye semillas, plantas descendientes y material clonal.

### 35 PROCEDIMIENTOS DE PRODUCCIÓN DE COMPUESTOS ISOPRENOIDES

**[0152]** La presente invención proporciona un procedimiento para la producción de un compuesto isoprenoide. En algunas realizaciones, los procedimientos implican generalmente el cultivo de una célula huésped modificada genéticamente en un medio adecuado, en que dicha célula huésped se modifica genéticamente con un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides. En otras realizaciones, los procedimientos implican generalmente el mantenimiento de una planta transgénica objeto en condiciones que favorecen la producción de la enzima modificadora de isoprenoides. La producción de la enzima modificadora de isoprenoides resulta en la producción del compuesto isoprenoide. Por ejemplo, en algunas realizaciones, los procedimientos implican generalmente el cultivo de una célula huésped modificada en un medio adecuado, en que dicha célula huésped se modifica genéticamente con un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una terpeno-oxidasa. La producción de la terpeno-oxidasa resulta en la producción del compuesto isoprenoide. Típicamente, el procedimiento se lleva a cabo *in vitro*, aunque también se contempla la producción de un compuesto isoprenoide *in vivo*. En algunas de estas realizaciones, la célula huésped es una célula eucariota, p. ej., una célula de levadura. En otras realizaciones, la célula huésped es una célula procariota. En algunas de estas realizaciones, la célula huésped es una célula vegetal. En algunas realizaciones, el procedimiento se lleva a cabo en una planta transgénica objeto.

**[0153]** Las células usan típicamente una de dos rutas para generar isoprenoides o precursores de isoprenoides (p. ej., IPP, poliprenildifosfatos, etc.). Las figuras 13 – 15 sirven para ilustrar las rutas usadas por las células para generar compuestos isoprenoides o precursores como poliprenildifosfatos.

**[0154]** La figura 13 representa las rutas de isoprenoides que implican la modificación de isopentenildifosfato (IPP) y/o su isómero dimetilalildifosfato (DMAPP) por prenil-transferasas para generar los poliprenildifosfatos geranildifosfato (GPP), farnesildifosfato (FPP) y geranilgeranildifosfato (GGPPP). GPP y FPP son modificados

posteriormente por terpeno-sintasas para generar monoterpenos y sesquiterpenos, respectivamente; y GGPP es modificado posteriormente por terpeno-sintasas para generar diterpenos y carotenoides. IPP y DMAPP se generan por una de las dos rutas siguientes: la ruta del mevalonato (MEV) y la ruta de la 1-desoxi-D-xilulosa-5-fosfato (DXP).

- 5 **[0155]** La figura 14 muestra esquemáticamente la ruta del MEV, en la que se convierte acetil-CoA en IPP a través de una serie de reacciones.

**[0156]** La figura 15 muestra esquemáticamente la ruta del DXP, en la que piruvato y D-gliceraldehído-3-fosfato se convierten en IPP y DMAPP a través de una serie de reacciones. Las células eucariotas distintas de las vegetales  
10 usan exclusivamente la ruta de isoprenoides del mevalonato para convertir acetil-coenzima A (acetil-CoA) en IPP, que se isomeriza subsiguientemente a DMAPP. Las plantas usan tanto la ruta del mevalonato como la ruta independiente de mevalonato, o de la DXP, para la síntesis de isoprenoides. Con algunas excepciones, los procariotas usan la ruta de la DXP para producir IPP y DMAPP separadamente a través de un punto de ramificación.

15 **[0157]** En algunas realizaciones, una célula huésped se modifica genéticamente con un ácido nucleico objeto que comprende una secuencia nucleotídica que codifica una sesquiterpeno oxidasa y la célula huésped se cultiva en un medio que contiene el sesquiterpeno. El sesquiterpeno penetra en la célula, donde es modificado por la sesquiterpeno-oxidasa. En muchas realizaciones, el sesquiterpeno se selecciona entre amorfadieno, aloisolongifoleno, (-)- $\alpha$ -*trans*-bergamoteno, (-)- $\beta$ -elemeno, (+)-germacreno A, germacreno B, (+)- $\gamma$ -gurjuneno, (+)-  
20 ledeno, neointermediol, (+)- $\beta$ -selineno y (+)-valenceno. En algunas realizaciones, la sesquiterpeno-oxidasa es una amorfadieno-oxidasa y la célula huésped se cultiva en un medio que contiene amorfa-4,11-dieno-oxidasa.

**[0158]** En otras realizaciones, la célula huésped se modifica genéticamente además con un ácido nucleico que comprende una secuencia nucleotídica que codifica una terpeno-sintasa. Así, p. ej., la célula huésped se  
25 modifica genéticamente con uno o más ácidos nucleicos que comprenden secuencias nucleotídicas que codifican una terpeno-sintasa y una enzima modificadora de isoprenoides (p. ej., una sesquiterpeno-oxidasa). El cultivo de un huésped tal en un medio de cultivo adecuado da lugar a la producción de la terpeno-sintasa y de la enzima modificadora de isoprenoides (p. ej., una sesquiterpeno-oxidasa). Por ejemplo, la terpeno-oxidasa modifica un farnesilpirofosfato para generar un sustrato sesquiterpénico para dicha sesquiterpeno-oxidasa.

30 **[0159]** Dependiendo del medio de cultivo en el que se cultiva la célula huésped y dependiendo de si la célula huésped sintetiza IPP por la ruta de la DXP o por la ruta del mevalonato, en algunas realizaciones, la célula huésped puede incluir otras modificaciones genéticas adicionales. Por ejemplo, en algunas realizaciones, la célula huésped es una célula que no tiene una ruta del mevalonato endógena, p. ej., la célula huésped es una célula que  
35 normalmente no sintetiza IPP o mevalonato por la ruta del mevalonato y la célula huésped se modifica genéticamente con uno o más ácidos nucleicos que comprenden secuencias nucleotídicas que codifican dos o más enzimas de la ruta del mevalonato, una IPP-isomerasa, una prenil-transferasa, una terpeno-sintasa y una enzima modificadora de isoprenoides (p. ej., una enzima modificadora de isoprenoides codificada por un ácido nucleico objeto). El cultivo de una célula huésped tal da lugar a la producción de las enzimas de la ruta del mevalonato, la  
40 IPP-isomerasa, la prenil-transferasa, la terpeno-sintasa y la enzima modificadora de isoprenoides (p. ej., una sesquiterpeno-oxidasa). La producción de las enzimas de la ruta del mevalonato, la IPP-isomerasa, la prenil-transferasa, la terpeno-sintasa y la enzima modificadora de isoprenoides (p. ej., una sesquiterpeno-oxidasa) resulta en la producción de un compuesto isoprenoide. En muchas realizaciones, la prenil-transferasa es una FPP-sintasa, que genera un sustrato sesquiterpénico para una sesquiterpeno-oxidasa codificada por un ácido nucleico objeto; y la  
45 producción de la sesquiterpeno-oxidasa resulta en la oxidación del sustrato sesquiterpénico en la célula huésped. Cualquier ácido nucleico que codifique las enzimas de la ruta del mevalonato, la IPP-isomerasa, la prenil-transferasa y la terpeno-sintasa es adecuado para su uso. Por ejemplo, algunos ácidos nucleicos adecuados se describen en Martin y col. (2003), cita anterior.

50 **[0160]** En algunas de las realizaciones descritas anteriormente, en las que la célula huésped se modifica genéticamente con uno o más ácidos nucleicos que comprenden secuencias nucleotídicas que codifican dos o más enzimas de la ruta del mevalonato, las dos o más enzimas de la ruta del mevalonato incluyen MK, PMK y MDP y la célula huésped se cultiva en un medio que contiene mevalonato. En otras realizaciones, las dos o más enzimas de la ruta del mevalonato incluyen acetoacetil-CoA-tiolasa, HMGS, HMGR, MK, PMK y MPD.

55 **[0161]** En algunas realizaciones, la célula huésped es una célula que normalmente no sintetiza IPP por la ruta del mevalonato, la célula huésped se modifica genéticamente según se describe anteriormente y la célula huésped comprende además una ruta de la DXP inactivada funcionalmente.

- [0162]** En algunas realizaciones, la célula huésped se modifica genéticamente con un ácido nucleico que comprende una secuencia nucleotídica que codifica una citocromo-P450-reductasa (CPR). Se conoce una gran diversidad de secuencias nucleotídicas de CPR y puede usarse cualquier ácido nucleico conocido que codifique una CPR, siempre que la CPR codificada muestre actividad de transferencia de electrones desde NADPH. En algunas realizaciones, el ácido nucleico que codifica la CPR codifica una CPR que transfiere electrones desde NADPH a una enzima modificadora de isoprenoides, p. ej., una sesquiterpeno-oxidasa, codificada por un ácido nucleico objeto que codifica una enzima modificadora de isoprenoides. En algunas realizaciones, el ácido nucleico que codifica la CPR es un ácido nucleico de CPR objeto.
- 10 **[0163]** Un procedimiento objeto es útil para la producción de diversos compuestos isoprenoides, incluidos, pero sin limitarse a ácido artemisinico (p. ej., en que el sustrato sesquiterpénico es amorfa-4,11-dieno), alcohol de aloisolongifoleno (p. ej., en que el sustrato es aloisolongifoleno), (E)-*trans*-bergamota-2,12-dien-14-ol (p. ej., en que el sustrato es (-)- $\alpha$ -*trans*-bergamoteno), (-)-elema-1,3,11 (13)-trien-12-ol (p. ej., en que el sustrato es (-)- $\beta$ -elemeno), germacra-1 (10),4,11 (13)-trien-12-ol (p. ej., en que el sustrato es (+)-germacreno A), alcohol de germacreno B (p. ej., en que el sustrato es germacreno B), 5,11 (13)-guayadieno-12-ol (p. ej., en que el sustrato es (+)- $\gamma$ -gurjuneno), alcohol de ledeno (p. ej., en que el sustrato es (+)-ledeno), 4 $\beta$ -H-eudesm-11 (13)-eno-4,12-diol (p. ej., en que el sustrato es neointermediol), (+)- $\beta$ -costol (p. ej., en que el sustrato es (+)- $\beta$ -selineno, y similares; y derivados posteriores de cualquiera de los anteriores.
- 15 **[0164]** En algunas realizaciones, una célula huésped modificada genéticamente objeto se cultiva en un medio adecuado (p. ej., caldo de Luria-Bertani, enriquecido opcionalmente con uno o más agentes adicionales, como un inductor (p. ej., cuando la enzima modificadora de isoprenoides está bajo el control de un promotor inducible), etc.); y el medio de cultivo se cubre con un disolvente orgánico, p. ej., dodecano, que forma una capa orgánica. El compuesto isoprenoide producido por la célula huésped modificada genéticamente pasa a la fase orgánica, a partir de la cual puede purificarse. En algunas realizaciones, en las que la secuencia que codifica la enzima modificadora de isoprenoides está ligada operativamente a un promotor inducible, se añade un inductor al medio de cultivo; y, después de un tiempo adecuado, el compuesto isoprenoide se aísla de la fase orgánica que cubre el medio de cultivo.
- 20 **[0165]** En algunas realizaciones, el compuesto isoprenoide se separará de otros productos que pueden estar presentes en la capa orgánica. La separación del compuesto isoprenoide de los otros productos que pueden estar presentes en la capa orgánica se consigue fácilmente, p. ej., mediante técnicas cromatográficas estándar.
- 25 **[0166]** En algunas realizaciones, un compuesto isoprenoide sintetizado por un procedimiento objeto se somete a una modificación química adicional en una reacción acelular. Por ejemplo, en algunas realizaciones se aísla ácido artemisinico del medio de cultivo y/o lisado celular y dicho ácido artemisinico se somete a una modificación química adicional en una reacción acelular para generar artemisinina.
- 30 **[0167]** En algunas realizaciones, el compuesto isoprenoide es puro, p. ej., puro al menos en aproximadamente el 40 %, puro al menos en aproximadamente el 50 %, puro al menos en aproximadamente el 60 %, puro al menos en aproximadamente el 70 %, puro al menos en aproximadamente el 80 %, puro al menos en aproximadamente el 90 %, puro al menos en aproximadamente el 95 %, puro al menos en aproximadamente el 98 % o puro en más del 98 %, en que "puro" en el contexto de un compuesto isoprenoide se refiere a un compuesto isoprenoide que no contiene otros compuestos isoprenoides, macromoléculas, contaminantes, etc.

## EJEMPLOS

- [0168]** Los siguientes ejemplos se exponen para proporcionar a los expertos en la técnica una divulgación y una descripción completas de cómo realizar y usar la presente invención y no pretenden limitar el alcance de lo que los inventores consideran como su invención ni con estos se pretende representar que los experimentos siguientes son todos o los únicos experimentos realizados. Se han hecho esfuerzos para asegurar la exactitud con respecto a los números usados (p. ej., cantidades, temperatura, etc.), pero debe contarse con algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular promedio, la temperatura es en grados Celsius y la presión es la atmosférica o próxima a esta. Pueden usarse las abreviaturas estándar, p. ej., pb, par (es) de bases; kb, kilobase (s); pl, picolitro (s); s, segundo (s); min, minuto (s); h, hora (s); aa, aminoácido (s); kb, kilobase (s); pb, par (es) de bases; nt, nucleótido (s); i.m., intramuscular; i.p., intraperitoneal; s.c., subcutáneo/a; y similares.

**Ejemplo 1:** Clonación y secuenciación de enzimas modificadoras de isoprenoides

**[0169]** La mayor parte de las enzimas de las que se sabe que hidroxilan un terpeno son citocromos P450. Todas las secuencias aminoacídicas disponibles de terpeno-hidroxilasas se alinearon con las secuencias aminoacídicas de los citocromos P450 de girasol y de lechuga. Estas dos especies vegetales pertenecen a la familia de las asteráceas, a la que también pertenece *Artemisia annua*. Las enzimas modificadoras de isoprenoides, p. ej., la familia CYP71D, formaron un grupo, lo que sugiere un antecesor común. Se diseñaron cebadores degenerados para la reacción en cadena de la polimerasa (PCR) para amplificar los genes de la familia CYP71D de las asteráceas.

**[0170]** **Clonación del ADNc de CYP71AV1 (también denominado CYP71D-A4 o AMO) y de CPR.** Se preparó un banco de ADNc mediante el kit de síntesis de ADNc Super SMART PCR (BD Bioscience), con 50 ng de ARN total purificado de células enriquecidas en tricomas de *A. annua*. Los cebadores degenerados para P450 se diseñaron a partir de un motivo de aminoácidos conservado de la subfamilia CYP71 de lechuga y girasol: el cebador 1 de [Y/Q]G[E/D][H/Y]WR (directo) y el cebador 2 de FIPERF (inverso) (la tabla I proporciona información sobre la secuencia de los cebadores).

**Tabla I. Cebadores usados para la construcción de plásmidos**

| Número de cebador | Secuencia (5' a 3')                                                      |
|-------------------|--------------------------------------------------------------------------|
| 1                 | TCCGACCA(C/T)ANGGNGAN(C/T)A(C/T)TGGAG; SEQ ID NO: 14                     |
| 2                 | TCCGACCAAANC(G/T)(C/T)TCNGG(A/G/T)AT(A/G)AA; SEQ ID NO: 15               |
| 3                 | CCAGCACA(A/G)TA(C/T)GA(A/G)CA(C/T)TT(C/T)AA(C/T)AA(A/G)AT; SEQ ID NO: 16 |
| 4                 | CCAGCAGCCATNCE(C/T)TINGC(A/G)TCNCC(A/G)CA; SEQ ID NO: 17                 |
| 5                 | ACGTCTAGAAATGAAGAGTATACTAAAAGCAATG; SEQ ID NO: 18                        |
| 6                 | ACGTCTAGAGCGAAACTTGGAACGAGTAACAACT; SEQ ID NO: 19                        |
| 7                 | ATGGATCCTATGCAATCAACAACTTCCGTAAAGTTAT; SEQ ID NO: 20                     |
| 8                 | TATGTGACCCATACATCACGGAGATATCTTCCT SEQ ID NO: 21                          |
| 9                 | GGACTAGTAAAACAATGGCCCTGACCGAAGAG; SEQ ID NO: 22                          |
| 10                | CCAAGCTTTCAGATGGACATCGGGTAAAC; SEQ ID NO: 23                             |
| 11                | CTGCCGCGGGGCCGCAAATTAAGCCTTC; SEQ ID NO: 24                              |
| 12                | CTGCCGCGGTAGTACGGATTAGAAGCCGC; SEQ ID NO: 25                             |
| 13                | CGGTATCCAAAACAATGGCTGCAGACCAATTGGTG; SEQ ID NO: 26                       |
| 14                | GCGTCGACTTAGGATTTAATGCAGGTGACG; SEQ ID NO: 27                            |
| 15                | CGGGATCCAAAACAATGAGCGAAGTCGGTATACAG; SEQ ID NO: 28                       |
| 16                | GCGTCGACTCATAACGAAAAATCAGAGAAATTTG; SEQ ID NO: 29                        |
| 17                | GGACTAGTAAAACAATGGCTTCAGAAAAAGAAATTAG; SEQ ID NO: 30                     |
| 18                | TCCCCCGGGCTATTTGCTTCTCTTGTAAC; SEQ ID NO: 31                             |

**[0171]** La reacción en cadena de la polimerasa (PCR) mediante el uso de estos cebadores y los ADNc de *A. annua* produjo un fragmento de ADN de 1 kb. El programa de PCR usado constó de siete ciclos con una temperatura de hibridación de 48 °C y 27 ciclos adicionales con una temperatura de hibridación de 55 °C. Los aminoácidos deducidos del fragmento génico amplificado mostraron una identidad aminoacídica del 85 % y el 88 %, respectivamente, con cóntigos de girasol (QH\_CA\_Contig1442) y lechuga (QG\_CA\_Contig7108). El banco de datos de EST de las compuestas puede encontrarse en cgpdb.ucdavis.edu. Se aisló un fragmento de *CPR* de *A. annua* mediante un cebador directo (cebador 3) y un cebador inverso (cebador 4), diseñados, respectivamente, a partir de los motivos conservados QYEHFNKI (SEQ ID NO: 32) y CGDAKGMA (SEQ ID NO: 33). El programa de PCR usado constó de 30 ciclos con una temperatura de hibridación de 50 °C. Las secuencias de los extremos 5' y 3' de *CYP71AV1* ("*CYP71D-A4*") y *CPR* se determinaron con un kit RLM-RACE (Ambion), a lo que siguió una recuperación de los ADNc de longitud completa a partir de los ADNc de hojas de *A. annua*. L. Las pautas abiertas de lectura de *CYP71AV1* y *CPR* se amplificaron por PCR y se ligaron en los sitios *SpeI* y *BamHI/SalI* de pESC-URA (Stratagene), en las etiquetas FLAG y cMyc, respectivamente. Para la amplificación por PCR de *CYP71AV1* se usaron los cebadores 5 y 6; para la amplificación por PCR de *CPR* se usaron los cebadores 7 y 8. El programa de PCR usado constó de 35 ciclos con una temperatura de hibridación de 55 °C. Todos los clones se secuenciaron para confirmar las secuencias.

**[0172]** **Análisis de extractos de plantas.** El material foliar de *A. annua* (100 a 200 mg de peso fresco) se agitó vigorosamente durante dos horas en 1 ml de hexano con octadecano 5,8 µM añadido como estándar interno.

Los extractos hexanólicos se concentraron hasta 200 µl y 1 µl de la muestra se usó para el análisis de CG-EM mediante una columna DB-XLB (0,25 mm d.i. x 0,25 µm x 30 m, J & W Scientific) para determinar el contenido de artemisinina de 14 muestras de plantas según se describe en Woerdenbag y col. (1991) *Phytochem. Anal.* 2: 215 – 219. El programa del horno de CG usado fue de 100 °C a 250 °C con un incremento de 5 °C/min. Los extractos hexanólicos de las plantas se derivatizaron con TMS-diazometano para determinar el contenido de ácido artemisinínico mediante CG-FID, equipada con una columna DB5 (n = 8). El programa del horno de CG usado fue de 80 °C (durante 2 minutos), un incremento de 20 °C/min hasta 140 °C y la separación del producto con un incremento de 5 °C/min hasta 220 °C. Los estándares de artemisinina auténtica se adquirieron de Sigma-Aldrich (San Luis, MO, EE. UU.).

10

**[0173] Síntesis de alcohol artemisinínico.** Se disolvió ácido artemisinínico (100,0 mg, 0,43 mmol) en THF (10,0 ml) y se añadió LiAlH<sub>4</sub> (17,0 mg, 0,45 mmol). La mezcla heterogénea se calentó a reflujo (70 °C) durante 15 h. Después del enfriamiento, la reacción se detuvo con agua (3,0 ml) y NaOH acuoso al 15 % (3,0 ml), se agitó durante 10 min y se filtró a través de celita. La fase orgánica se separó, se secó sobre MgSO<sub>4</sub> y se concentró mediante un rotavapor. El producto se purificó por cromatografía en columna (2: 1 hexanos/EtOAc) para dar 61,0 mg (rendimiento del 65 %) del alcohol como un aceite incoloro. Una pequeña cantidad de ácido artemisinínico contaminante se eliminó posteriormente por cromatografía en columna sobre alúmina neutra (actividad de Brockmann 1). Los datos de la caracterización fueron coherentes con los valores de la bibliografía.

15

**[0174] Síntesis de aldehído artemisinínico.** El alcohol artemisinínico se oxidó a aldehído artemisinínico siguiendo un procedimiento descrito en la bibliografía. Sharpless y col., *Tetrahedron Letters* 17: 2503 – 2506 (1976). A un matraz de 10 ml secado a la llama que contenía RuCl<sub>2</sub>(PPh<sub>3</sub>)<sub>3</sub> (17,0 mg, 0,018 mmol) y N-óxido de N-metilmorfolina (60,0 mg, 0,51 mmol) en atmósfera de argón se le añadió acetona (4,0 ml). A la disolución se le añadió alcohol artemisinínico (55,0 mg, 0,25 mmol) disuelto en acetona (1,0 ml) por medio de una jeringa. La mezcla se agitó a 23 °C durante 2 h y se concentró al vacío. El producto crudo se purificó por cromatografía en columna (4: 1 hexanos/EtOAc) para dar 32,0 mg (rendimiento del 59 %) de aldehído artemisinínico como un aceite incoloro. Los datos de la caracterización fueron coherentes con la descripción de la bibliografía.

25

### **Generación y caracterización de las cepas EPY**

30

**[0175]** Productos químicos. El dodecano y el cariofileno se adquirieron de Sigma-Aldrich (San Luis, MO, EE. UU.). El ácido 5-fluoroórótico (5-FOA) se adquirió de Zymo Research (Orange, CA, EE. UU.). Las mezclas completas de complementos para la formulación del medio sintético definido (SD) se adquirieron de Qbiogene (Irvine, CA, EE. UU.). Todos los demás componentes de los medios se adquirieron de Sigma-Aldrich o Becton, Dickinson (Franklin Lakes, NJ, EE. UU.).

35

**[0176] Cepas y medios.** Las cepas DH10B y DH5α de *Escherichia coli* se usaron para la transformación bacteriana y la amplificación plasmídica en la construcción de los plásmidos de expresión usados en este estudio. Las cepas se cultivaron a 37 °C en el medio de Luria-Bertani con 100 mg/l de ampicilina, con la excepción de los plásmidos basados en pδ-UB que se cultivaron con 50 mg/l de ampicilina en DH5α.

40

**[0177]** La cepa BY4742 de *Saccharomyces cerevisiae* (Brachmann y col., *Yeast* 14: 115 – 132 (1998)), derivada de S288C, se usó como la cepa parental para todas las cepas de levadura. La cepa se cultivó en medio rico YPD. Burke y col., *Methods in yeast genetics: a Cold Spring Harbor laboratory course manual* (Cold Spring Harbor Laboratory Press, Plainview, NY, EE. UU., 2000). Las cepas de levadura manipuladas genéticamente se cultivaron en medio SD (Burke y col., cita anterior) carente de leucina, uracilo, histidina y/o metionina en el caso apropiado. Para la inducción de los genes expresados a partir del promotor *GAL1*, las cepas de *S. cerevisiae* se cultivaron con galactosa al 2 % como única fuente de carbono.

45

**[0178] Construcción de plásmidos.** Para crear el plásmido pRS425ADS para la expresión de ADS con el promotor *GAL1*, ADS se amplificó por PCR a partir de pADS (Martin y col., *Nat. Biotechnol.* 21: 796 – 802 (2003)) mediante el par de cebadores 9 y 10. (Tabla I). Con estos cebadores se clonó la secuencia nucleotídica 5'-AAAACA-3' inmediatamente por delante del codón de inicio de ADS. Esta secuencia consenso se usó para la traducción eficiente de ADS y de otros genes inducibles por galactosa usados en este estudio. El producto amplificado se escindió con *SpeI* y *HindIII* y se clonó en pRS425GAL1 digerido con *SpeI* y *HindIII* (Mumberg y col., *Nucleic Acids Research* 22: 5767 – 5768 (1994)).

55

**[0179]** Para la integración de una casete de expresión para *tHMGR* se construyó el plásmido pδ-HMGR. Primeramente se introdujeron sitios de restricción para *SacII* en pRS426GAL1 (Mumberg y col., cita anterior) en el

extremo 5' del promotor *GAL1* y en el extremo 3' del terminador *CYC1*. Para conseguir esto se amplificó por PCR la casete con el promotor – sitio de clonación múltiple – terminador de pRS426GAL1 con el par de cebadores 11 y 12. El producto amplificado se clonó directamente en pRS426GAL1 digerido con *PvuII* para construir el vector pRS426-SacII. El dominio catalítico de *HMG1* se amplificó por PCR a partir del plásmido pRH127-3 (Donald y col., Appl. Environ. Microbiol. 63: 3341 – 3344 (1997)) con el par de cebadores 13 y 14. El producto amplificado se escindió con *BamHI* y *Sall* y se clonó en pRS426-SacII digerido con *BamHI* y *XhoI*. El plásmido pRS-HMGR se escindió con *SacII* y el fragmento con la casete de expresión se extrajo de un gel y se clonó en pδ-UB digerido con *SacII* (Lee y col. Biotechnol. Prog. 13: 368 – 373 (1997)).

10 **[0180]** El alelo *upc2-1* de *UPC2* se amplificó por PCR a partir del plásmido pBD33 con el par de cebadores 15 y 16. El producto amplificado se escindió con *BamHI* y *Sall* y se clonó en pRS426-SacII digerido con *BamHI* y *XhoI* para crear el plásmido pRS-UPC2. Para la integración de *upc2-1* se creó pB-UPC2 de manera idéntica, mediante la digestión de pRS-UPC2 con *SacII* y el traslado del fragmento apropiado a pS-UB.

15 **[0181]** Para sustituir el promotor *ERG9* por el promotor *MET3* se construyó el plásmido pRS-ERG9. El plásmido pRH973 (Gardner y col., J. Biol. Chem. 274: 31671 – 31678 (1999)) contenía un segmento de *ERG9* truncado en el extremo 5' y colocado detrás del promotor *MET3*. El plásmido pRH973 se escindió con *ApaI* y *ClaI* y se clonó en pRS403 digerido con *AlaI* y *ClaI*, que tiene un marcador de selección HIS3 (Sikorski y col., Genetics 122: 19 – 27 (1989)).

20 **[0182]** Para la expresión de *ERG20* se construyó el plásmido pB-ERG20. El plásmido pRS-SacII se digirió primeramente con *Sall* y *XhoI* lo que creó extremos cohesivos compatibles. El plásmido se religó, con lo que se eliminaron los sitios *Sall* y *XhoI*, para crear el plásmido pRS-SacII-DX. *ERG20* se amplificó por PCR a partir del ADN genómico de BY4742 mediante el par de cebadores 17 y 18. El producto amplificado se escindió con *SpeI* y *SmaI* y se clonó en pRS-SacII-DX digerido con *SpeI* y *SmaI*. Después se escindió pRS-ERG20 con *SacII* y la casete de expresión se extrajo de un gel y se clonó en pδ-UB digerido con *SacII*.

**[0183] Transformación de levaduras y construcción de cepas.** La cepa BY4742 de *Saccharomyces cerevisiae* (Brachmann y col., cita anterior), derivada de S288C se usó como cepa parental para todas las cepas de *S. cerevisiae*. La transformación de todas las cepas de *S. cerevisiae* se realizó por el procedimiento estándar del acetato de litio. Gietz, R D. y Woods, R A. en Guide to Yeast Genetics and Molecular and Cell Biology, parte B, 87 – 96 (Academic Press Inc, San Diego, EE. UU., 2002). Se analizaron de tres a diez colonias de cada transformación para seleccionar el transformante con la mayor producción de amorfadieno. La cepa EPY201 se construyó por transformación de la cepa BY4742 con el plásmido pRS425ADS y selección en placas de SD-LEU. El plásmido pδ-HMGR se digirió con *XhoI* antes de la transformación del ADN en la cepa EPY201. Después de una selección inicial en placas de SD-LEU-URA, los transformantes se cultivaron y se sembraron en placas de SD-LEU con 1 g/l de 5-FOA como selección por la pérdida del marcador *URA3*. El auxótrofo para uracilo resultante, EPY208, se transformó entonces con el ADN del plásmido pS-UPC2 digerido con *XhoI*. Después de una selección inicial en placas de SD-LEU-URA, los transformantes se cultivaron y se sembraron en placas de SD-LEU con 1 g/l de 5-FOA para la construcción de EPY210. El plásmido pRS-ERG9 se escindió con *HindIII* para la integración de la fusión *P<sub>MET3</sub>-ERG9* en los locus *ERG9* de EPY208 y EPY210 para la construcción de EPY213 y EPY225, respectivamente. Estas cepas se seleccionaron en placas de SD-LEU-HIS-MET. Después EPY213 se transformó con ADN del plásmido pδ-HMGR digerido con *XhoI*. Después de una selección inicial en placas de SD-LEU-URA-HIS-MET, los transformantes se cultivaron y se sembraron en placas de SD-LEU-HIS-MET con 1 g/l de 5-FOA para la construcción de EPY219. EPY219 se transformó con ADN del plásmido pδ-ERG20 digerido con *XhoI*. Después de una selección inicial en placas de SD-LEU-URA-HIS-MET, los transformantes se cultivaron y se sembraron en placas de SD-LEU-HIS-MET con 1 g/l de 5-FOA para la construcción de EPY224.

**[0184]** La integración de pRS-ERG9 se verificó por análisis de PCR con dos juegos de cebadores. Cada juego contenía un oligonucleótido para unirse al ADN insertado y un oligonucleótido para unirse al ADN genómico alrededor de la inserción. Todas las demás integraciones se verificaron en cuanto a la longitud completa de la inserción con cebadores de unión al extremo 5' del promotor *GAL1* y al extremo 3' del gen fusionado.

**[0185] Cultivo de levaduras.** Todas las medidas de densidad óptica a 600 nm ( $DO_{600}$ ) se llevaron a cabo con un espectrofotómetro Beckman DU-640. Para medir la producción de amorfadieno, se inocularon tubos de cultivo que contenían 5 ml del medio SD (con galactosa al 2 %) (con las omisiones de aminoácidos apropiadas, según se describe anteriormente) con las especies de interés. Estos inóculos se cultivaron a 30 °C hasta una  $DO_{600}$  de entre 1 y 2. Con estos cultivos de siembra se inocularon matraces de fondo liso (250 ml) que contenían 50 ml de medio SD hasta una  $DO_{600}$  de 0,05. La producción de amorfadieno se determinó después de seis días de cultivo. Todos los

cultivos contenían metionina 1mM para reprimir la fusión  $P_{MET3}$ -*ERG9* en los locus *ERG9*. Todos los matraces contenían también 5 ml de dodecano. Para la determinación de la producción de amorfadieno por CG-EM se tomaron muestras de esta capa de dodecano y se diluyeron en acetato de etilo.

## 5 RESULTADOS

[0186] La artemisinina se produce en los tricomas glandulares, células especializadas de la planta. Se aislaron células de los tricomas glandulares de *A. annua* y se extrajo el ARN de dichas células. Mediante el uso de los cebadores degenerados se aisló un ADNc parcial de un gen nuevo que se denominó *CYP71D-A4*. El gen de longitud completa se recuperó mediante una rápida amplificación de los extremos del ADNc (RACE). La secuencia nucleotídica de la región codificante del ADNc se presenta en la figura 1 (SEQ ID NO: 1); la secuencia aminoacídica traducida se muestra en la figura 2 (SEQ ID NO: 2).

[0187] El ADNc de longitud completa de *CYP71D-A4* se expresó en células de levadura. Para determinar la actividad de amorfadieno-oxidasa, *CYP71D-A4* se puso bajo el control transcripcional de un promotor *GAL10* en un soporte plasmídico pESC-URA (Stratagene) en el que el gen *CPR* de *A. annua* (*AACPR*; figura 3; la secuencia aminoacídica de la proteína codificada se muestra en la figura 4) se expresa a partir de un promotor *GAL1*. El gen *AACPR* se obtuvo del ARNm de tricomas glandulares de *A. annua* mediante un procedimiento de PCR con cebadores degenerados y RACE, según se describe anteriormente.

[0188] Para determinar la actividad de amorfadieno-oxidasa *in vivo*, este plásmido (p71D-A4/CPR: : pESC-URA) y un plásmido control que carecía del gen *CYP71D-A4* se transformaron en células de *S. cerevisiae* manipuladas genéticamente para producir amorfadieno. Brevemente, estas células son la cepa BY4742 con un gen integrado que codifica una HMG-CoA-reductasa truncada que es soluble en levaduras. Estas células contienen pRS425ADS que tiene un gen *ADS* con codones optimizados bajo el control del promotor *GAL1*. Las células transformadas se cultivaron en medio sintético carente de leucina y uracilo y se indujeron por galactosa al 2 % durante 29 horas, después de las cuales el medio se extrajo con éter. Los extractos se concentraron y 1 µl de estos se analizó por cromatografía de gases-espectrometría de masas (CG-EM) con una columna EXL y usando un programa de temperaturas con incrementos de 5 °C por minuto desde 50 °C a 250 °C. Para sintetizar alcohol artemisinínico y aldehído artemisinínico, usados como estándares, se utilizó ácido artemisinínico auténtico. Mediante este procedimiento se detectaron dos picos a partir de las células que expresaban *CPR* y *CYP71D-A4*, pero no de las células de control que expresaban solo *CPR*. Por comparación de los tiempos de retención y los espectros de masas con los estándares auténticos se determinó que estos picos correspondían a alcohol artemisinínico y a aldehído artemisinínico. No se detectó ácido artemisinínico; su aparición sin haberlo derivatizado no es de esperar al usar CG, debido a su volatilidad.

[0189] Se realizó un ensayo de suministro *in vivo* para determinar la actividad de amorfadieno-oxidasa, en el que los dos mismos plásmidos se transformaron individualmente en la cepa natural de *S. cerevisiae*, YPH499. Las células de levadura se cultivaron en 50 ml de medio carente de uracilo con dextrosa al 2 % y se indujeron con galactosa al 2 % durante 24 horas. Se recolectaron por centrifugación 5 ml de células de levadura inducidas y se usó medio fresco con amorfadieno-4,11-dieno 150 µM, alcohol artemisinínico o aldehído artemisinínico para resuspender las células de levadura. Después, las células de levadura se cultivaron a 30 °C durante 5 horas. El medio se extrajo con éter y después se llevó a cabo una derivatización con *N*-(*terc*-butildimetilsilil)-*N*-metiltrifluoroacetamida para permitir la detección de cualquier ácido artemisinínico mediante CG-EM. Los estándares de alcohol artemisinínico y aldehído artemisinínico auténticos también se derivatizaron de manera similar. Una cantidad de 1 µl de cada uno de los controles y muestras derivatizados se analizó por CG-EM. El programa de temperatura usado constó de incrementos de 5 °C por minuto desde 50 °C a 250 °C.

[0190] Al suministrar amorfadieno a las células solamente se detectó una acumulación significativa de ácido artemisinínico junto con una pequeña cantidad de los compuestos alcohol y aldehído artemisinínico en las células de levadura que expresaban *CPR* y *CYP71D-A4* (figura 5A). Al suministrar alcohol artemisinínico o aldehído artemisinínico a las células, la acumulación relativa de ácido artemisinínico fue mayor en el medio de cultivo de las células de levadura transformadas con *CPR/CYP71D-A4* que en el de la cepa de control transformada solo con *CPR* (figuras 5B y 5C).

[0191] Figuras 5A-C. Se añadieron amorfadieno (figura 5A) y los dos otros intermedios de la artemisinina, alcohol artemisinínico (figura 5B) y aldehído artemisinínico (figura 5C), a una concentración de 150 µM, al medio en el que se cultivaron e indujeron con galactosa al 2 % las células de levadura transformadas solo con *CPR* (cromatograma superior) o con ambos *CPR* y *CYP71D-A4* (cromatograma inferior). Amorfadieno (1), alcohol

artemisínico (2), aldehído artemisínico (3) y ácido artemisínico (4) se indican con flechas. El alcohol artemisínico (2) y el ácido artemisínico (4) se detectaron después de la derivatización con *N*- (terc-butildimetilsilil)-*N*-metiltrifluoroacetamida. Los asteriscos indican la adición de los sustratos al medio.

5 **[0192]** La autenticidad del ácido artemisínico derivatizado en las muestras se confirmó mediante el estándar de ácido artemisínico auténtico (figuras 6A y 6B). Estos datos indicaron que la primera hidroxilación es catalizada por la enzima citocromo P450 codificada en el clon *CYP71D-A4* y las subsiguientes conversiones oxidativas del alcohol artemisínico en aldehído artemisínico y del aldehído artemisínico en ácido artemisínico son probablemente catalizadas por la enzima recombinante de *CYP71D-A4* junto con las actividades de oxidación endógenas de la  
10 levadura.

**[0193]** Figuras 6A y 6B. El espectro de masas y el tiempo de retención del nuevo compuesto producido después del suministro de amorfadieno a las células de levadura transformadas con *CPR/71D-A4* se muestran en la figura 6A y los del estándar de ácido artemisínico auténtico se muestran en la figura 6B. Ambos, producto y estándar  
15 se detectaron por CG-EM después de la derivatización, lo que añadió 114 unidades de masa al peso molecular básico.

**[0194]** La síntesis *de novo* de ácido artemisínico en las levaduras manipuladas genéticamente a partir de un azúcar simple como galactosa se demostró mediante la modificación genética de EPY224 con pESC-URA que contenía *CPR* ("AACPR") y *AMO* ("CYP71D-A4") (pESC-URA: : AACPR/AMO). Una construcción que contenía una HMGCoA-reductasa de levadura troncada se integró dos veces en la cepa de levadura BY4742. El factor de transcripción *upc2-1* se expresó en exceso para aumentar el nivel de transcripción de varios genes de la ruta biosintética del ergosterol. El gen de la escualeno-sintasa (*ERG9*) se reguló por disminución mediante el promotor reprimible por metionina, *MET3*. La FPP-sintasa se expresó en exceso mediante el promotor *GAL1* en el plásmido  
20 pRS425. La cepa EPY224 de levadura que contenía pESC-URA: : AACPR/AMO se cultivó en medio sintético con galactosa al 1,8 % y glucosa al 0,2 % durante 5 días a 30 °C. Las células de levadura se sedimentaron y el sedimento se lavó con tampón alcalino (tampón de tris pH 9). El tampón se acidificó a pH 2 mediante la adición de HCl y el tampón acidificado se extrajo con acetato de etilo. Se añadieron TMS-diazometano y metanol a la fracción de acetato de etilo para derivatizar el ácido artemisínico. La forma de éster metílico del ácido artemisínico se detectó  
25 por CG-EM.

**[0195]** Las figuras 7A-7C muestran la producción *de novo* de ácido artemisínico en levaduras cuando se expresan *AACPR* y *AMO*. En contraste, no se detectó ácido artemisínico en una cepa de levadura de control que solo expresaba *AACPR*. El nuevo pico a 13,62 min (figura 7A, pico 1) mostró el mismo patrón de fragmentación de  
35 masas que el auténtico ácido artemisínico de la planta *Artemisia annua* (figuras 7B y C).

**[0196]** Las figuras 8A-8C muestran ensayos enzimáticos para *AMO in vitro*. Se aislaron microsomas de *S. cerevisiae* YPH499 que expresaba *AACPR* o *CPR/AMO*. Los picos cromatográficos para los sustratos usados se indican con asteriscos. Para cada ensayo enzimático se usaron amorfadieno 10 µM (a), alcohol artemisínico 25 µM  
40 (b) o aldehído artemisínico 25 µM (c). Las fracciones extraíbles con éter se derivatizaron y se analizaron por CG-EM en el modo de ión selectivo (m/z: 121,189, 204, 218, 220 y 248). Los productos enzimáticos son según se indica: 1, alcohol artemisínico [tiempo de retención (tr) = 13,20]; 2 aldehído artemisínico (tr = 11,79); ácido artemisínico (tr = 13,58, detectado como éster metílico).

45 **[0197]** La figura 9 muestra la secuencia nucleotídica de un clon de ADNc, designado *71D-B1* (también denominado "AMH", por amorfadieno-hidroxilasa), que codifica una terpeno-hidroxilasa.

**[0198]** La figura 10 muestra la secuencia aminoacídica de la proteína codificada por *71D-B1* (AMH).

50 **[0199]** Las figuras 11A-C muestran la actividad de hidroxilación de la enzima recombinante codificada en el clon *AMH* (*71D-B1*). El pico a 16,82 min en A es ácido artemisínico cuando *AMO* se expresa en levaduras que expresan HMGCoA en exceso y el pico a 18,50 min en B es amorfadieno hidroxilado cuando *AMH* y *AACPR* se expresan en exceso en levaduras que expresan HMGCoA en exceso. Los patrones de fragmentación de masas del amorfadieno hidroxilado se presentan en la figura 11C. Se muestra el pico para el ión parental (220) del amorfadieno  
55 hidroxilado y también se muestran otros patrones de fragmentación de iones típicos para sesquiterpenos y terpenos (p. ej., 93, 119, 132, 145, 159 y 177).

**[0200]** La figura 12 muestra la secuencia nucleotídica de un ADN genómico que codifica una terpeno-hidroxilasa/oxidasa.

LISTADO DE SECUENCIAS

[0201]

< 110 > RO, DAE-KYUN

5 NEWMAN, KARYN  
PARADISE, ERIC M.  
KEASLING, JAY D.  
OUELLET, MARIO  
EACHUS, RACHEL

10 HO, KIMBERLY  
HAM, TIMOTHY

< 120 > POLINUCLEÓTIDOS QUE CODIFICAN ENZIMAS MODIFICADORAS DE ISOPRENOIDES Y  
PROCEDIMIENTOS DE USO DE LOS MISMOS

< 130 > BERK-049WO

15 < 150 > 60/697.067

< 151 > 05-07-2005

< 160 > 33

< 170 > FastSEQ para Windows, versión 4.0

< 210 > 1

20 < 211 > 1.488

< 212 > ADN

< 213 > *Artemisia annua*

< 400 > 1

|            |            |             |             |             |             |      |
|------------|------------|-------------|-------------|-------------|-------------|------|
| atgaagagta | tactaaaagc | aatggcactc  | tcactgacca  | cttccattgc  | tcttgcaacg  | 60   |
| atccttttgt | tcgtttacaa | gttcgctact  | cgttccaaat  | ccaccaaaaa  | aagccttcoct | 120  |
| gagccatggc | ggcttcccat | tattgggtcac | atgcatcact  | tgattgggtac | aacgccacat  | 180  |
| cgtgggggta | gggatttagc | cagaaaagat  | ggatccttga  | tgcatttaca  | gcttgggtgaa | 240  |
| gttccaacaa | tcgtgggtgc | atctccgaaa  | tgggctaaaag | agattttgac  | aacgtacgac  | 300  |
| attacctttg | ctaacaggcc | cgagacttta  | actggtgaga  | ttgttttata  | tcacaatacg  | 360  |
| gatgttggtc | ttgcacctta | tgggtgaatac | tggaggcaat  | tacgtaaaat  | ttgcacattg  | 420  |
| gagcttttga | gtgttaagaa | agtaaagtca  | tttcagtcac  | ttcgtgaaga  | ggagtgttgg  | 480  |
| aatttggttc | aagagattaa | agcttcagggt | tcaggggagac | cggttaacct  | ttcagagaaat | 540  |
| gttttcaagt | tgattgcaac | gatacttagt  | agagccgcat  | ttgggaaagg  | gatcaaggac  | 600  |
| cagaaagagt | taacggagat | tgtgaaagag  | atactgaggc  | aaactggttg  | ttttgatgtg  | 660  |
| gcgatatctc | ttccttcaaa | gaaatttctt  | catcatcttt  | cgggcaagag  | agctcgggta  | 720  |
| actagccttc | gcaaaaagat | cgataattta  | atcgataacc  | ttgtagctga  | gcatactgtt  | 780  |
| aacacctcca | gtaaaactaa | cgagacactc  | ctcgatgttc  | ttttaaggct  | caaagacagt  | 840  |
| gctgaattcc | cattaacatc | tgataacatt  | aaagccatca  | ttttggatat  | gtttggagca  | 900  |
| ggcacagaca | cttcctcatc | cacaatcgaa  | tgggcgattt  | cggaaactcat | aaagtgtccg  | 960  |
| aaagcaatgg | agaaagtaca | agcgggaattg | aggaaagcat  | tgaacggaaa  | agaaaagatc  | 1020 |
| catgaggaag | acattcaaga | actaagctac  | ttgaacatgg  | taatcaaaga  | aacattgagg  | 1080 |
| ttgcaccctc | cactaccctt | ggttctgcca  | agagagtgcc  | gccaaccagt  | caatttggct  | 1140 |
| ggatacaaca | taccaataa  | gaccaaactt  | attgtcaacg  | tctttgcgat  | aaatagggac  | 1200 |
| cctgaatatt | ggaaagacgc | tgaagctttc  | atccctgaac  | gatttgaaaa  | tagttctgca  | 1260 |
| actgtcatgg | gtgcagaata | cgagtatctt  | ccgtttggag  | ctgggagaag  | gatgtgtcct  | 1320 |
| ggagccgcac | ttgggttagc | taacgtgcag  | ctcccgtcgc  | ctaataact   | atatcatttc  | 1380 |
| aactggaaac | tccccaatgg | tgtgagctat  | gaccagatcg  | acatgaccga  | gagctctgga  | 1440 |
| gccacgatgc | aaagaaagac | tgagttgtta  | ctcgttccaa  | gtttctag    |             | 1488 |

25

< 210 > 2

< 211 > 495

< 212 > Proteína

30 < 213 > *Artemisia annua*

< 400 > 2

```

Met Lys Ser Ile Leu Lys Ala Met Ala Leu Ser Leu Thr Thr Ser Ile
 1      5      10      15
Ala Leu Ala Thr Ile Leu Leu Phe Val Tyr Lys Phe Ala Thr Arg Ser
      20      25      30
Lys Ser Thr Lys Lys Ser Leu Pro Glu Pro Trp Arg Leu Pro Ile Ile
      35      40      45
Gly His Met His His Leu Ile Gly Thr Thr Pro His Arg Gly Val Arg
      50      55      60
Asp Leu Ala Arg Lys Tyr Gly Ser Leu Met His Leu Gln Leu Gly Glu
      65      70      75      80
Val Pro Thr Ile Val Val Ser Ser Pro Lys Trp Ala Lys Glu Ile Leu
      85      90      95
Thr Thr Tyr Asp Ile Thr Phe Ala Asn Arg Pro Glu Thr Leu Thr Gly
      100      105      110
Glu Ile Val Leu Tyr His Asn Thr Asp Val Val Leu Ala Pro Tyr Gly
      115      120      125
Glu Tyr Trp Arg Gln Leu Arg Lys Ile Cys Thr Leu Glu Leu Leu Ser
      130      135      140
Val Lys Lys Val Lys Ser Phe Gln Ser Leu Arg Glu Glu Glu Cys Trp
      145      150      155      160
Asn Leu Val Gln Glu Ile Lys Ala Ser Gly Ser Gly Arg Pro Val Asn
      165      170      175
Leu Ser Glu Asn Val Phe Lys Leu Ile Ala Thr Ile Leu Ser Arg Ala
      180      185      190
Ala Phe Gly Lys Gly Ile Lys Asp Gln Lys Glu Leu Thr Glu Ile Val
      195      200      205
Lys Glu Ile Leu Arg Gln Thr Gly Gly Phe Asp Val Ala Asp Ile Phe
      210      215      220
Pro Ser Lys Lys Phe Leu His His Leu Ser Gly Lys Arg Ala Arg Leu
      225      230      235      240
Thr Ser Leu Arg Lys Lys Ile Asp Asn Leu Ile Asp Asn Leu Val Ala
      245      250      255
Glu His Thr Val Asn Thr Ser Ser Lys Thr Asn Glu Thr Leu Leu Asp
      260      265      270
Val Leu Leu Arg Leu Lys Asp Ser Ala Glu Phe Pro Leu Thr Ser Asp
      275      280      285
Asn Ile Lys Ala Ile Ile Leu Asp Met Phe Gly Ala Gly Thr Asp Thr
      290      295      300
Ser Ser Ser Thr Ile Glu Trp Ala Ile Ser Glu Leu Ile Lys Cys Pro
      305      310      315      320
Lys Ala Met Glu Lys Val Gln Ala Glu Leu Arg Lys Ala Leu Asn Gly
      325      330      335
Lys Glu Lys Ile His Glu Glu Asp Ile Gln Glu Leu Ser Tyr Leu Asn
      340      345      350
Met Val Ile Lys Glu Thr Leu Arg Leu His Pro Pro Leu Pro Leu Val
      355      360      365
Leu Pro Arg Glu Cys Arg Gln Pro Val Asn Leu Ala Gly Tyr Asn Ile
      370      375      380
Pro Asn Lys Thr Lys Leu Ile Val Asn Val Phe Ala Ile Asn Arg Asp
      385      390      395      400
Pro Glu Tyr Trp Lys Asp Ala Glu Ala Phe Ile Pro Glu Arg Phe Glu
      405      410      415
Asn Ser Ser Ala Thr Val Met Gly Ala Glu Tyr Glu Tyr Leu Pro Phe
      420      425      430
Gly Ala Gly Arg Arg Met Cys Pro Gly Ala Ala Leu Gly Leu Ala Asn
      435      440      445
Val Gln Leu Pro Leu Ala Asn Ile Leu Tyr His Phe Asn Trp Lys Leu
      450      455      460
Pro Asn Gly Val Ser Tyr Asp Gln Ile Asp Met Thr Glu Ser Ser Gly
      465      470      475      480
Ala Thr Met Gln Arg Lys Thr Glu Leu Leu Leu Val Pro Ser Phe

```

285

490

495

< 210 > 3  
< 211 > 2.157  
5 < 212 > ADN  
< 213 > *Artemisia annua*  
< 400 > 3

|             |             |            |            |            |             |      |
|-------------|-------------|------------|------------|------------|-------------|------|
| atgcaatcaa  | caacttcggt  | taagttatct | cccttcgac  | taatgacggc | gttacttaac  | 60   |
| ggcaaggat   | cgttcgacac  | atcaaacaca | tcggatacga | atattccggt | agcgggtgtt  | 120  |
| atggagaatc  | gtgagctttt  | gatgatttta | actacttcgg | ttgcgggtgt | gatcggatgc  | 180  |
| gttggtggtg  | ttgtgtggag  | acggtcgctg | tcggcggcga | agaaagcggc | ggagtcgccg  | 240  |
| gtgattgttg  | tgccgaagaa  | agtacggag  | gatgaggttg | atgacggacg | gaagaaagt   | 300  |
| actgtgtttt  | ttggaactca  | gactggtact | gctgaagggt | ttgctaaggc | gcttgttgaa  | 360  |
| gaagctaaag  | cgcgatatga  | aaaggcgggt | tttaaagtga | ttgatttgga | tgattatgct  | 420  |
| gctgaagatg  | atgagtatga  | ggagaagtta | aagaaagaat | ctcttgcttt | tttcttttta  | 480  |
| gctacgtatg  | gagatggtga  | gccgacagat | aatgctgcta | gattctataa | atggtttacc  | 540  |
| gaggggtgaag | agaaagggtga | atggcttgac | aagcttcaat | acgcagtgtt | tggacttggg  | 600  |
| aacagacagt  | atgagcattt  | caacaagatt | gcgaagggtg | tcgatgaaaa | acttgtggag  | 660  |
| caggggtgcaa | agcgccttgt  | tcctgttggc | atgggagacg | atgatcaatg | tatcgaagac  | 720  |
| gacttcactg  | catggaaaga  | gttggtgtgg | cctgagttgg | atcaattact | tcgtgatgag  | 780  |
| gatgatacat  | ctgttgccac  | tccatacaca | gctgctgttg | gagaataccg | tgttgtgttc  | 840  |
| catgacaaac  | cagagacata  | tgatcaggat | caactgacaa | atggccatgc | tggtcatgat  | 900  |
| gctcaacatc  | catgcagatc  | caatgtcgct | gtcaaaaagg | agctccattc | ccctctatct  | 960  |
| gaccggtcct  | gcactcattt  | ggaatttgat | atctctaata | ctggattatc | gtatgaaact  | 1020 |
| ggggaccatg  | ttggagtcta  | cgttgagaat | ctaagtgaag | ttgtggacga | agctgaaaaa  | 1080 |
| ttaatagggt  | taccgcccga  | cacttatttc | tcagtacata | ctgataacga | agacgggaca  | 1140 |
| ccacttgggt  | gagcctcttt  | gccacctcct | ttccctccat | gcactttaag | aaaagcattg  | 1200 |
| gcttcctatg  | ccgatgtttt  | gagctctcct | aaaaagtcag | ctttgcttgc | tttagctgct  | 1260 |
| catgtactatg | attctactga  | agctgataga | ctgaaatttt | ttgcgtctcc | tgctggaaag  | 1320 |
| gatgaatatg  | ctcagtggat  | agttgcaagc | cacagaagtc | tccttgaggt | catggaggcc  | 1380 |
| ttcccatcag  | ctaagcctcc  | gcttggtgtt | ttttttgcat | ctgtcgcccc | acgtttgcag  | 1440 |
| ccgagatact  | attccatttc  | ttcttcccca | aagtttgccg | caaataggat | tcagttaact  | 1500 |
| tgtgcattag  | tgtatgagca  | aacaccatca | ggccgcgttc | acaagggagt | ctgttcaaca  | 1560 |
| tggatgaaga  | atgccgtgcc  | tatgacagaa | agccaggatt | gcagttgggc | cccaatttat  | 1620 |
| gttagaacat  | ccaatttcag  | acttccttct | gatcctaagg | tcccagttat | catgattggc  | 1680 |
| ccaggcactg  | gattggctcc  | atthagaggt | ttccttcagg | aaaggtagc  | tcagaaggaa  | 1740 |
| gctgggactg  | agctcggaac  | agccatctta | ttcttcggat | gcaggaatcg | caaagtggat  | 1800 |
| ttcatatatg  | aggacgagct  | taataatttc | gtggagacgg | gggctctttc | cgagcttggt  | 1860 |
| acggccttct  | ctcgtgaagg  | tgccactaag | gagtacgtgc | aacacaagat | gactcagaag  | 1920 |
| gcttcggata  | tctggaattt  | actctctgag | ggagcatatt | tgtatgtttg | cgggtgatgcc | 1980 |
| aaaggcatgg  | ccaaagatgt  | acatcggact | ctgcacacta | ttgtgcaaga | acagggatct  | 2040 |
| ctagactcct  | caaaggcggg  | gctctacgtg | aagaatctac | aaatggcagg | aagatatctc  | 2100 |
| cgtgatgtat  | gggtcgacat  | ggaacagaag | ttgatttccg | aagaagacct | cgagtaa     | 2157 |

10 < 210 > 4  
< 211 > 718  
< 212 > Proteína  
< 213 > *Artemisia annua*  
15 < 400 > 4

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

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Leu | Val | Val | Pro | Lys | Lys | Val | Thr | Glu | Asp | Glu | Val | Asp | Asp | Gly |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Arg | Lys | Lys | Val | Thr | Val | Phe | Phe | Gly | Thr | Gln | Thr | Gly | Thr | Ala | Glu |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Gly | Phe | Ala | Lys | Ala | Leu | Val | Glu | Glu | Ala | Lys | Ala | Arg | Tyr | Glu | Lys |
|     |     |     | 115 |     |     |     |     | 120 |     |     |     | 125 |     |     |     |
| Ala | Val | Phe | Lys | Val | Ile | Asp | Leu | Asp | Asp | Tyr | Ala | Ala | Glu | Asp | Asp |
|     |     |     | 130 |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Glu | Tyr | Glu | Glu | Lys | Leu | Lys | Lys | Glu | Ser | Leu | Ala | Phe | Phe | Phe | Leu |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Ala | Thr | Tyr | Gly | Asp | Gly | Glu | Pro | Thr | Asp | Asn | Ala | Ala | Arg | Phe | Tyr |
|     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Lys | Trp | Phe | Thr | Glu | Gly | Glu | Glu | Lys | Gly | Glu | Trp | Leu | Asp | Lys | Leu |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Gln | Tyr | Ala | Val | Phe | Gly | Leu | Gly | Asn | Arg | Gln | Tyr | Glu | His | Phe | Asn |
|     |     |     | 195 |     |     |     |     | 200 |     |     |     | 205 |     |     |     |
| Lys | Ile | Ala | Lys | Val | Val | Asp | Glu | Lys | Leu | Val | Glu | Gln | Gly | Ala | Lys |
|     |     |     | 210 |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Arg | Leu | Val | Pro | Val | Gly | Met | Gly | Asp | Asp | Asp | Gln | Cys | Ile | Glu | Asp |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Asp | Phe | Thr | Ala | Trp | Lys | Glu | Leu | Val | Trp | Pro | Glu | Leu | Asp | Gln | Leu |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Leu | Arg | Asp | Glu | Asp | Asp | Thr | Ser | Val | Ala | Thr | Pro | Tyr | Thr | Ala | Ala |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Val | Gly | Glu | Tyr | Arg | Val | Val | Phe | His | Asp | Lys | Pro | Glu | Thr | Tyr | Asp |
|     |     |     | 275 |     |     |     |     | 280 |     |     |     | 285 |     |     |     |
| Gln | Asp | Gln | Leu | Thr | Asn | Gly | His | Ala | Val | His | Asp | Ala | Gln | His | Pro |
|     |     |     | 290 |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Cys | Arg | Ser | Asn | Val | Ala | Val | Lys | Lys | Glu | Leu | His | Ser | Pro | Leu | Ser |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Asp | Arg | Ser | Cys | Thr | His | Leu | Glu | Phe | Asp | Ile | Ser | Asn | Thr | Gly | Leu |
|     |     |     | 325 |     |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Ser | Tyr | Glu | Thr | Gly | Asp | His | Val | Gly | Val | Tyr | Val | Glu | Asn | Leu | Ser |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Glu | Val | Val | Asp | Glu | Ala | Glu | Lys | Leu | Ile | Gly | Leu | Pro | Pro | His | Thr |
|     |     |     | 355 |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
| Tyr | Phe | Ser | Val | His | Thr | Asp | Asn | Glu | Asp | Gly | Thr | Pro | Leu | Gly | Gly |
|     |     |     | 370 |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
| Ala | Ser | Leu | Pro | Pro | Pro | Phe | Pro | Pro | Cys | Thr | Leu | Arg | Lys | Ala | Leu |
| 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| Ala | Ser | Tyr | Ala | Asp | Val | Leu | Ser | Ser | Pro | Lys | Lys | Ser | Ala | Leu | Leu |
|     |     |     | 405 |     |     |     |     |     | 410 |     |     |     |     | 415 |     |
| Ala | Leu | Ala | Ala | His | Ala | Thr | Asp | Ser | Thr | Glu | Ala | Asp | Arg | Leu | Lys |
|     |     |     | 420 |     |     |     | 425 |     |     |     |     | 430 |     |     |     |
| Phe | Phe | Ala | Ser | Pro | Ala | Gly | Lys | Asp | Glu | Tyr | Ala | Gln | Trp | Ile | Val |
|     |     |     | 435 |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| Ala | Ser | His | Arg | Ser | Leu | Leu | Glu | Val | Met | Glu | Ala | Phe | Pro | Ser | Ala |
|     |     |     | 450 |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
| Lys | Pro | Pro | Leu | Gly | Val | Phe | Phe | Ala | Ser | Val | Ala | Pro | Arg | Leu | Gln |
| 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| Pro | Arg | Tyr | Tyr | Ser | Ile | Ser | Ser | Ser | Pro | Lys | Phe | Ala | Pro | Asn | Arg |
|     |     |     | 485 |     |     |     |     |     | 490 |     |     |     |     | 495 |     |
| Ile | His | Val | Thr | Cys | Ala | Leu | Val | Tyr | Glu | Gln | Thr | Pro | Ser | Gly | Arg |
|     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| Val | His | Lys | Gly | Val | Cys | Ser | Thr | Trp | Met | Lys | Asn | Ala | Val | Pro | Met |
|     |     |     | 515 |     |     |     | 520 |     |     |     |     | 525 |     |     |     |
| Thr | Glu | Ser | Gln | Asp | Cys | Ser | Trp | Ala | Pro | Ile | Tyr | Val | Arg | Thr | Ser |
|     |     |     | 530 |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| Asn | Phe | Arg | Leu | Pro | Ser | Asp | Pro | Lys | Val | Pro | Val | Ile | Met | Ile | Gly |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Pro | Gly | Thr | Gly | Leu | Ala | Pro | Phe | Arg | Gly | Phe | Leu | Gln | Glu | Arg | Leu |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Gln | Lys | Glu | Ala | Gly | Thr | Glu | Leu | Gly | Thr | Ala | Ile | Leu | Phe | Phe |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Gly | Cys | Arg | Asn | Arg | Lys | Val | Asp | Phe | Ile | Tyr | Glu | Asp | Glu | Leu | Asn |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |
| Asn | Phe | Val | Glu | Thr | Gly | Ala | Leu | Ser | Glu | Leu | Val | Thr | Ala | Phe | Ser |
|     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Arg | Glu | Gly | Ala | Thr | Lys | Glu | Tyr | Val | Gln | His | Lys | Met | Thr | Gln | Lys |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Ala | Ser | Asp | Ile | Trp | Asn | Leu | Leu | Ser | Glu | Gly | Ala | Tyr | Leu | Tyr | Val |
|     |     |     | 645 |     |     |     |     |     | 650 |     |     |     |     | 655 |     |
| Cys | Gly | Asp | Ala | Lys | Gly | Met | Ala | Lys | Asp | Val | His | Arg | Thr | Leu | His |
|     |     | 660 |     |     |     |     |     | 665 |     |     |     |     | 670 |     |     |
| Thr | Ile | Val | Gln | Glu | Gln | Gly | Ser | Leu | Asp | Ser | Ser | Lys | Ala | Glu | Leu |
|     | 675 |     |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |
| Tyr | Val | Lys | Asn | Leu | Gln | Met | Ala | Gly | Arg | Tyr | Leu | Arg | Asp | Val | Trp |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Val | Asp | Met | Glu | Gln | Lys | Leu | Ile | Ser | Glu | Glu | Asp | Leu | Glu |     |     |
| 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     |     |

- < 210 > 5
- < 211 > 1.467
- 5 < 212 > ADN
- < 213 > *Artemisia annua*
- < 400 > 5

|            |            |             |            |             |             |      |
|------------|------------|-------------|------------|-------------|-------------|------|
| atggcacttt | cactgaccac | ctccattgct  | cttgccacga | tccttttctt  | cgtaatttac  | 60   |
| aagttcgcta | ctcgttccaa | atccacaaaa  | aacagccttc | ctgagccatg  | gcgacttccc  | 120  |
| attattggtc | acatgcatca | cttgattggt  | acaataccac | atcgtgggct  | tatggattta  | 180  |
| gccagaaagt | atggatcttt | aatgcattta  | cagcttggtg | aagtttcaac  | aatcgtgggtg | 240  |
| tcatctccga | aatgggctaa | agagattttg  | acaacgtacg | acattgcctt  | tgctaacagg  | 300  |
| ccctggactt | tggtggtga  | gattgttgta  | tatcgcaata | caaataattgc | tgctgcacct  | 360  |
| tatggtgaat | actggaggcg | attacgtaaa  | ctttgcacat | cggagcttat  | gagtgttaag  | 420  |
| aaagtaaagt | catatcagtc | gcttcgtgaa  | gaggagtgtt | ggaatttggt  | tcaagagatt  | 480  |
| aaagcttcag | gttcagggat | accgggttaac | ctttcagaga | acattttcaa  | gttgattgca  | 540  |
| acgatacttt | gtagagccgc | gtttggaata  | ggagtcaagg | accagaagga  | gtgtacggag  | 600  |
| attatgaaag | agatgttgag | ggaagttggt  | ggttttgatg | tggcagatat  | ctttccttcg  | 660  |
| aagaaatttc | ttcatcatct | ttcaggcaag  | agagccaggt | taactagcat  | tcataagaag  | 720  |
| ctcgataatt | ttatcaataa | ccttggtgct  | gagcatactt | tcaaaacttc  | aagtaaaact  | 780  |
| gaggagacac | ttcttgatgt | tcttctaagg  | ctcaaagata | gcgctgaatt  | cccatataca  | 840  |
| gctgacaatg | ttaaagccat | cattttggat  | atatttgcat | cagggacaga  | cacttcatca  | 900  |
| accacaatcg | aatgggcat  | ttcggaactc  | ataaagtgtc | cgagagcgat  | ggagaaagta  | 960  |
| caagcagaac | tgaggaaagc | acttaacgga  | aaagaaaaga | tccatgagga  | agatattcaa  | 1020 |
| ggactaagct | acttaaaact | ggtaatacaa  | gaaacattaa | ggttgacccc  | tccactaccc  | 1080 |
| ttgttgccaa | gagagtgcgc | tgaaccagtc  | aatttggtcg | gatacgacat  | acccaataag  | 1140 |
| acaagactta | ttgtcaacgt | ctttgcgata  | aatagggacc | cagaatactg  | gaaagacgct  | 1200 |
| gaaattttca | tcccgaacg  | atttgaaaat  | agttctacaa | ctctcatggg  | tgacagaatat | 1260 |
| gagtatcttc | cgtttgagc  | tggtgagaag  | atgtgtcctg | gagccgcact  | tggttttagcc | 1320 |
| aacgtgcagc | taccgctcgc | taatatacta  | tatcatttca | actggaaact  | ccccaacggt  | 1380 |
| gcgagctatg | atcagatcga | catgaccgag  | aggtttggaa | tctcggttga  | aagaaagact  | 1440 |
| cagttgttac | tcgtaccaag | tttctag     |            |             |             | 1467 |

- 10 < 210 > 6
- < 211 > 488
- < 212 > Proteína
- < 213 > *Artemisia annua*
- 15 < 400 > 6

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Leu | Ser | Leu | Thr | Thr | Ser | Ile | Ala | Leu | Ala | Thr | Ile | Leu | Phe |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     | 15  |     |     |
| Phe | Val | Ile | Tyr | Lys | Phe | Ala | Thr | Arg | Ser | Lys | Ser | Thr | Lys | Asn | Ser |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Leu | Pro | Glu | Pro | Trp | Arg | Leu | Pro | Ile | Ile | Gly | His | Met | His | His | Leu |
|     | 35  |     |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Ile | Gly | Thr | Ile | Pro | His | Arg | Gly | Leu | Met | Asp | Leu | Ala | Arg | Lys | Tyr |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Gly | Ser | Leu | Met | His | Leu | Gln | Leu | Gly | Glu | Val | Ser | Thr | Ile | Val | Val |
|     | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| Ser | Ser | Pro | Lys | Trp | Ala | Lys | Glu | Ile | Leu | Thr | Thr | Tyr | Asp | Ile | Ala |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Phe | Ala | Asn | Arg | Pro | Trp | Thr | Leu | Ala | Gly | Glu | Ile | Val | Val | Tyr | Arg |
|     |     |     |     | 100 |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Asn | Thr | Asn | Ile | Ala | Ala | Ala | Pro | Tyr | Gly | Glu | Tyr | Trp | Arg | Arg | Leu |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Arg | Lys | Leu | Cys | Thr | Ser | Glu | Leu | Met | Ser | Val | Lys | Lys | Val | Lys | Ser |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Tyr | Gln | Ser | Leu | Arg | Glu | Glu | Glu | Cys | Trp | Asn | Leu | Val | Gln | Glu | Ile |
|     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| Lys | Ala | Ser | Gly | Ser | Gly | Ile | Pro | Val | Asn | Leu | Ser | Glu | Asn | Ile | Phe |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Lys | Leu | Ile | Ala | Thr | Ile | Leu | Cys | Arg | Ala | Ala | Phe | Gly | Lys | Gly | Val |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Lys | Asp | Gln | Lys | Glu | Cys | Thr | Glu | Ile | Met | Lys | Glu | Met | Leu | Arg | Glu |
|     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Val | Gly | Gly | Phe | Asp | Val | Ala | Asp | Ile | Phe | Pro | Ser | Lys | Lys | Phe | Leu |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| His | His | Leu | Ser | Gly | Lys | Arg | Ala | Arg | Leu | Thr | Ser | Ile | His | Lys | Lys |
|     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Leu | Asp | Asn | Phe | Ile | Asn | Asn | Leu | Val | Ala | Glu | His | Thr | Phe | Lys | Thr |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Ser | Ser | Lys | Thr | Glu | Glu | Thr | Leu | Leu | Asp | Val | Leu | Leu | Arg | Leu | Lys |
|     |     |     | 260 |     |     |     | 265 |     |     |     |     |     | 270 |     |     |
| Asp | Ser | Ala | Glu | Phe | Pro | Leu | Thr | Ala | Asp | Asn | Val | Lys | Ala | Ile | Ile |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Leu | Asp | Ile | Phe | Ala | Ala | Gly | Thr | Asp | Thr | Ser | Ser | Thr | Thr | Ile | Glu |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Trp | Ala | Ile | Ser | Glu | Leu | Ile | Lys | Cys | Pro | Arg | Ala | Met | Glu | Lys | Val |
|     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gln | Ala | Glu | Leu | Arg | Lys | Ala | Leu | Asn | Gly | Lys | Glu | Lys | Ile | His | Glu |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Glu | Asp | Ile | Gln | Gly | Leu | Ser | Tyr | Leu | Asn | Leu | Val | Ile | Lys | Glu | Thr |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Leu | Arg | Leu | His | Pro | Pro | Leu | Pro | Leu | Leu | Pro | Arg | Glu | Cys | Arg | Glu |
|     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
| Pro | Val | Asn | Leu | Ala | Gly | Tyr | Asp | Ile | Pro | Asn | Lys | Thr | Arg | Leu | Ile |
|     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
| Val | Asn | Val | Phe | Ala | Ile | Asn | Arg | Asp | Pro | Glu | Tyr | Trp | Lys | Asp | Ala |
|     | 385 |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| Glu | Ile | Phe | Ile | Pro | Glu | Arg | Phe | Glu | Asn | Ser | Ser | Thr | Thr | Leu | Met |
|     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |
| Gly | Ala | Glu | Tyr | Glu | Tyr | Leu | Pro | Phe | Gly | Ala | Gly | Arg | Arg | Met | Cys |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| Pro | Gly | Ala | Ala | Leu | Gly | Leu | Ala | Asn | Val | Gln | Leu | Pro | Leu | Ala | Asn |
|     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| Ile | Leu | Tyr | His | Phe | Asn | Trp | Lys | Leu | Pro | Asn | Gly | Ala | Ser | Tyr | Asp |
|     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
| Gln | Ile | Asp | Met | Thr | Glu | Arg | Phe | Gly | Ile | Ser | Val | Glu | Arg | Lys | Thr |
|     | 465 |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| Gln | Leu | Leu | Leu | Val | Pro | Ser | Phe |     |     |     |     |     |     |     |     |
|     |     |     |     | 485 |     |     |     |     |     |     |     |     |     |     |     |

< 210 > 7  
< 211 > 1.038

< 212 > ADN  
 < 213 > *Artemisia annua*  
 < 400 > 7

```

gcccttcgag ccgatatggg attactggcg gcaattacgt aaactttgca cattggagct 60
tttgagtgct aagaaagtag agtcatatca gtcgcttcgt gaagaggagt gttggaattt 120
agttcaagag attaaagctt cagggttcagg gataccgggt aacctttcag agaataattt 180
caagttgggt gcaatgatac ttagtagagc tgcgtttggg aaaagaatca aggaccataa 240
ggagtttacg gagcttggtg aacagatggt gagggaaact ggtggttttg atgtggcaga 300
tatctttcct tcgcagaaat ttctacatca tatttcgggc aagagatcta ggtaactag 360
cattcacaaa aagotcgata atttaataca taaccttggt gctgagcata ttgttgaagc 420
ctcaagtaaa actaaggaga cgctccttga tgttcttcta aggcacaaag atagccttga 480
attcccattg acagctgata acgttaaagc catcattttg gtatgaatta atccaatata 540
tttttttttt caaaaggcca taatagtgtt aaacaagctt gaaatttttt ataactaagt 600
acatgcacta acttttagtac tcgtgaaaat ataatgagtc atcatagggg ttccatgaaa 660
tatacaggac atgttttacag caggcacaga cacttcgtca accacaatcg aatgggtgat 720
ttcggaactc ataaagtgtc cgagagctat ggagaaaata caagcggaac tgaggaaagc 780
acttaacgga aaagaaaaga tccacgagga agacatccaa gaactaagct acttaaactt 840
ggtaatcaaa gaaacattaa ggttgacccc tccactaccc ttggttttgc cacgagagtg 900
ccgtcaacca gtcaatttgg ctggatatga catacccaat aagaccaaac ttattgtcaa 960
cgtctttgcg ataaataggg accctgaata ctggaaagac gctgaatctt tcatcccaga 1020
gcgcttctta actctggt 1038
    
```

5  
 < 210 > 8  
 < 211 > 6  
 < 212 > Proteína  
 10 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > péptido enlazante  
 < 400 > 8

Ala Ala Ala Gly Gly Met  
 1 5

15  
 < 210 > 9  
 < 211 > 14  
 < 212 > Proteína  
 20 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > péptido enlazante  
 < 400 > 9

Ala Ala Ala Gly Gly Met Pro Pro Ala Ala Ala Gly Gly Met  
 1 5 10

25  
 < 210 > 10  
 < 211 > 6  
 < 212 > Proteína  
 30 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > péptido enlazante  
 < 400 > 10

Ala Ala Ala Gly Gly Met  
 1 5

35  
 < 210 > 11

- < 211 > 8
- < 212 > Proteína
- < 213 > Secuencia artificial
- < 220 >
- 5 < 223 > Péptido enlazante
- < 400 > 11

Pro Pro Ala Ala Ala Gly Gly Met  
1 5

- 10 < 210 > 12
- < 211 > 4
- < 212 > Proteína
- < 213 > Secuencia artificial
- < 220 >
- 15 < 223 > péptido enlazante
- < 400 > 12

Ile Glu Gly Arg  
1

- 20 < 210 > 13
- < 211 > 6
- < 212 > Proteína
- < 213 > Secuencia artificial
- < 220 >
- 25 < 223 > péptido enlazante
- < 400 > 13

Gly Gly Lys Gly Gly Lys  
1 5

- 30 < 210 > 14
- < 211 > 25
- < 212 > ADN
- < 213 > Secuencia artificial
- < 220 >
- 35 < 223 > cebador sintético
- < 220 >
- < 221 > elemento mixto
- < 222 > 9, 18, 20
- < 223 > N = C o T
- 40 < 220 >
- < 221 > elemento mixto
- < 222 > 11, 14, 17
- < 223 > n = A, T, C o G
- < 400 > 14

45 tccgaccana nggngannan tggag 25

- < 210 > 15
- < 211 > 25
- 50 < 212 > ADN
- < 213 > Secuencia artificial
- < 220 >
- < 223 > cebador sintético
- < 220 >

- < 221 > elemento mixto
- < 222 > (13) ... (13)
- < 223 > N = G o T
- < 220 >
- 5 < 221 > elemento mixto
- < 222 > (14) ... (14)
- < 223 > N = C o T
- < 220 >
- < 221 > elemento mixto
- 10 < 222 > (20) ... (20)
- < 223 > N = A, G o T
- < 220 >
- < 221 > elemento mixto
- < 222 > (23) ... (23)
- 15 < 223 > N = A o G
- < 220 >
- < 221 > elemento mixto
- < 222 > 11, 17
- < 223 > n = A, T, C o G
- 20 < 400 > 15

tccgacaaa ncnntcnggn atnaa

25

- < 210 > 16
- 25 < 211 > 29
- < 212 > ADN
- < 213 > Secuencia artificial
- < 220 >
- < 223 > cebador sintético
- 30 < 220 >
- < 221 > elemento mixto
- < 222 > (9) ... (9)
- < 223 > N = A o G
- < 220 >
- 35 < 221 > elemento mixto
- < 222 > (12) ... (12)
- < 223 > N = C o T
- < 220 >
- < 221 > elemento mixto
- 40 < 222 > (15) ... (15)
- < 223 > N = A o G
- < 220 >
- < 221 > elemento mixto
- < 222 > (18) ... (18)
- 45 < 223 > N = C o T
- < 220 >
- < 221 > elemento mixto
- < 222 > (21) ... (21)
- < 223 > N = C o T
- 50 < 220 >
- < 221 > elemento mixto
- < 222 > (24) ... (24)
- < 223 > N = C o T
- < 220 >
- 55 < 221 > elemento mixto
- < 222 > (27) ... (27)
- < 223 > N = A o G
- < 400 > 16

|    |                                         |    |
|----|-----------------------------------------|----|
|    | ccagcacant angancantt naanaanat         | 29 |
|    | < 210 > 17                              |    |
|    | < 211 > 29                              |    |
| 5  | < 212 > ADN                             |    |
|    | < 213 > Secuencia artificial            |    |
|    | < 220 >                                 |    |
|    | < 223 > cebador sintético               |    |
|    | < 220 >                                 |    |
| 10 | < 221 > elemento mixto                  |    |
|    | < 222 > (15) ... (15)                   |    |
|    | < 223 > N = C o T                       |    |
|    | < 220 >                                 |    |
|    | < 221 > elemento mixto                  |    |
| 15 | < 222 > (21) ... (21)                   |    |
|    | < 223 > N = A o G                       |    |
|    | < 220 >                                 |    |
|    | < 221 > elemento mixto                  |    |
|    | < 222 > (27) ... (27)                   |    |
| 20 | < 223 > N = A o G                       |    |
|    | < 220 >                                 |    |
|    | < 221 > elemento mixto                  |    |
|    | < 222 > 12, 18, 24                      |    |
|    | < 223 > n = A, T, C o G                 |    |
| 25 | < 400 > 17                              |    |
|    | ccagcagcca tncnttngc ntcncnca           | 29 |
|    | < 210 > 18                              |    |
| 30 | < 211 > 33                              |    |
|    | < 212 > ADN                             |    |
|    | < 213 > Secuencia artificial            |    |
|    | < 220 >                                 |    |
|    | < 223 > cebador sintético               |    |
| 35 | < 400 > 18                              |    |
|    | acgtctagaa tgaagagtat actaaaagca atg    | 33 |
|    | < 210 > 19                              |    |
|    | < 211 > 34                              |    |
| 40 | < 212 > ADN                             |    |
|    | < 213 > Secuencia artificial            |    |
|    | < 220 >                                 |    |
|    | < 223 > cebador sintético               |    |
|    | < 400 > 19                              |    |
| 45 | acgtctagag cgaaacttgg aacgagtaac aact   | 34 |
|    | < 210 > 20                              |    |
|    | < 211 > 37                              |    |
| 50 | < 212 > ADN                             |    |
|    | < 213 > Secuencia artificial            |    |
|    | < 220 >                                 |    |
|    | < 223 > cebador sintético               |    |
|    | < 400 > 20                              |    |
| 55 | atggatccta tgcaatcaac aactccgtt aagttat | 37 |
|    | < 210 > 21                              |    |
|    | < 211 > 34                              |    |

< 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 5 < 400 > 21  
  
 tatgtcgacc catacatcac ggagatatct tcct 34  
  
 < 210 > 22  
 10 < 211 > 32  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 15 < 400 > 22  
  
 ggactagtaa aacaatggcc ctgaccgaag ag 32  
  
 < 210 > 23  
 20 < 211 > 29  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 25 < 400 > 23  
  
 ccaagcttgc agatggacat cgggtaaac 29  
  
 < 210 > 24  
 30 < 211 > 29  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 35 < 400 > 24  
  
 ctgccgcggg gccgcaaatt aaagccttc 29  
  
 < 210 > 25  
 40 < 211 > 33  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 45 < 400 > 25  
  
 ctgccgcggt agtacggatt agaagccgc 29  
  
 < 210 > 26  
 50 < 211 > 35  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 < 223 > cebador sintético  
 55 < 400 > 26  
  
 cgggatccaa aacaatggct gcagaccaat tgggtg 35  
  
 < 210 > 27

< 211 > 30  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 5 < 223 > cebador sintético  
 < 400 > 27  
  
 gcgtcgactt aggatttaat gcaggtgacg 30  
  
 10 < 210 > 28  
 < 211 > 35  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 15 < 223 > cebador sintético  
 < 400 > 28  
  
 cgggatccaa aacaatgagc gaagtcggtat tacag 35  
  
 20 < 210 > 29  
 < 211 > 34  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 25 < 223 > cebador sintético  
 < 400 > 29  
  
 gcgtcgactc ataacgaaaa atcagagaaa ttg 34  
  
 30 < 210 > 30  
 < 211 > 37  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 35 < 223 > cebador sintético  
 < 400 > 30  
  
 ggactagtaa aacaatggct tcagaaaaag aaattag 37  
  
 40 < 210 > 31  
 < 211 > 30  
 < 212 > ADN  
 < 213 > Secuencia artificial  
 < 220 >  
 45 < 223 > cebador sintético  
 < 400 > 31  
  
 tcccccgggc tatttgcttc tctgtaaac 30  
  
 50 < 210 > 32  
 < 211 > 8  
 < 212 > Proteína  
 < 213 > Secuencia artificial  
 < 220 >  
 55 < 223 > motivo  
 < 400 > 32

Gln Tyr Glu His Phe Asn Lys Ile  
1 5

< 210 > 33  
< 211 > 8  
5 < 212 > Proteína  
< 213 > Secuencia artificial  
< 220 >  
< 223 > motivo  
< 400 > 33

10

Cys Gly Asp Ala Lys Gly Met Ala  
1 5

## REIVINDICACIONES

1. Un polinucleótido aislado que comprende una secuencia nucleotídica que codifica una enzima que modifica un compuesto isoprenoide, en que la secuencia nucleotídica codifica un polipéptido que tiene al menos el 85 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 2, en que el polipéptido muestra actividad de amorfa-4,11-dieno-oxidasa
2. El polinucleótido de la reivindicación 1, en que la secuencia nucleotídica tiene al menos el 90 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 1.
3. El polinucleótido de la reivindicación 1, en que el polinucleótido comprende una secuencia nucleotídica que codifica un polipéptido que tiene al menos el 90 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 2.
4. El polinucleótido de la reivindicación 3, en que la secuencia nucleotídica codifica un polipéptido que tiene al menos el 95 % de identidad de secuencia aminoacídica con la secuencia aminoacídica expuesta en SEQ ID NO: 2.
5. Un vector recombinante que comprende el polinucleótido de una cualquiera de las reivindicaciones 1 a 4, en que opcionalmente dicho polinucleótido está ligado operativamente a un promotor.
6. Una célula huésped modificada genéticamente con el polinucleótido de una cualquiera de las reivindicaciones 1 a 4 o el vector de la reivindicación 5.
7. La célula huésped de la reivindicación 6, que se selecciona del grupo que consta de una célula procariota, una célula de levadura y una célula vegetal.
8. Un procedimiento para la producción de un compuesto isoprenoide en una célula huésped, en que dicho procedimiento comprende: el cultivo de una célula huésped modificada genéticamente de las reivindicaciones 6 ó 7 en un medio adecuado para producir una enzima modificadora de isoprenoides, en que, en presencia de un sustrato terpénico, dicha producción de dicha enzima modificadora de isoprenoides resulta en la modificación enzimática del sustrato terpénico y la producción del compuesto isoprenoide; y en que dicha enzima modificadora de isoprenoides es una amorfadieno-oxidasa y en que el sustrato terpénico es amorfa-4,11-dieno.
9. El procedimiento de la reivindicación 8, en que dicha célula huésped se modifica genéticamente además con un ácido nucleico que comprende una secuencia nucleotídica que codifica una terpeno-sintasa, en que dicho cultivo da lugar a la producción de dicha terpeno-sintasa, en que dicha terpeno-sintasa se selecciona de una terpeno sintasa que modifica un farnesilpírofosfato para generar un sustrato sesquiterpénico para dicha enzima modificadora de isoprenoides.
10. El procedimiento de la reivindicación 8, en que dicha célula huésped es una célula que no sintetiza normalmente isopentenilpírofosfato (IPP) por la ruta del mevalonato y en que la célula huésped se modifica genéticamente con uno o más ácidos nucleicos que comprenden secuencias nucleotídicas que codifican dos o más enzimas de la ruta del mevalonato, una IPP-isomerasa, una prenil-transferasa y una terpeno-sintasa, en que dicho cultivo da lugar a la producción de las enzimas de la ruta del mevalonato, en que dicha producción de dichas dos o más enzimas de la ruta del mevalonato, dicha IPP-isomerasa, dicha prenil-transferasa, dicha terpeno-sintasa y dicha enzima modificadora de isoprenoides, resulta en la producción de un compuesto isoprenoide.
11. El procedimiento de la reivindicación 10, en que dichas dos o más enzimas de la ruta del mevalonato comprenden mevalonato-cinasa, fosfomevalonato-cinasa y mevalonato-pírofosfato-descarboxilasa y en que la célula huésped se cultiva en presencia de mevalonato.
12. El procedimiento de la reivindicación 10, en que dichas dos o más enzimas de la ruta del mevalonato comprenden acetoacetil-CoA-tiolasa, hidroximetilglutaril-CoA-sintasa, hidroximetilglutaril-CoA-reductasa, mevalonato-cinasa, fosfomevalonato-cinasa y mevalonato-pírofosfato-descarboxilasa.
13. El procedimiento de la reivindicación 10, en que dicha prenil-transferasa es una farnesilpírofosfato-sintasa.

14. El procedimiento de una cualquiera de las reivindicaciones 8 a 13, en que dicha célula huésped se modifica genéticamente con un ácido nucleico que comprende una secuencia nucleotídica que codifica una citocromo-P450-reductasa, en que la reductasa transfiere electrones desde NADPH a una enzima modificadora de isoprenoides.
- 5 15. El procedimiento de la reivindicación 14, en que dicho ácido nucleico tiene al menos el 90 % de identidad de secuencia nucleotídica con la secuencia nucleotídica expuesta en SEQ ID NO: 3.
16. El procedimiento de la reivindicación 8, en que la célula huésped es una célula procariota, en que  
10 dicha célula es una célula que normalmente sintetiza isopentenildifosfato a través de la ruta de la 1-desoxi-D-xilulosa-5-fosfato.
17. El procedimiento de la reivindicación 16, en que la ruta de la 1-desoxi-D-xilulosa-5-fosfato se inactiva.
- 15 18. El procedimiento de una cualquiera de las reivindicaciones 7 a 17, que comprende además el aislamiento del compuesto isoprenoide.
19. El procedimiento de la reivindicación 9, en que el compuesto isoprenoide es ácido artemisínico y dicho procedimiento comprende además la modificación del ácido artemisínico para generar artemisinina y el aislamiento  
20 de dicha artemisinina.
20. Una planta transgénica modificada genéticamente con un ácido nucleico que comprende una secuencia nucleotídica que codifica una enzima modificadora de isoprenoides, en que la secuencia nucleotídica codifica un polipéptido que tiene al menos el 85 % de identidad de secuencia aminoacídica con la secuencia  
25 aminoacídica expuesta en SEQ ID NO: 2, en que el polipéptido muestra actividad de amorfa-4,11-dieno-oxidasa y en que el ácido nucleico se expresa en una célula de la planta para producir la enzima modificadora de isoprenoides en la célula.
21. La planta transgénica de la reivindicación 20, en que la planta es tabaco o *Artemisia annua*.
- 30 22. La planta transgénica de las reivindicaciones 20 ó 21, en que la secuencia nucleotídica que codifica la enzima modificadora de isoprenoides está ligada operativamente a un promotor constitutivo, un promotor inducible o un promotor específico de un tejido.
- 35 23. La planta transgénica de la reivindicación 22, en que el promotor específico de un tejido es un promotor específico de tricomas.
24. Un procedimiento para la producción de un compuesto isoprenoide, en que el procedimiento comprende el mantenimiento de la planta transgénica de una cualquiera de las reivindicaciones 20 a 23, en que la  
40 enzima modificadora de isoprenoides se produce en dicha planta transgénica y en que la producción de la enzima modificadora de isoprenoides resulta en la modificación de un sustrato terpénico y la producción de un compuesto isoprenoide.

# FIG. 1

## Secuencia del ADNc de CYP71D-A4

Secuencia codificante del ADNc de CYP71D-A4

ATGAAGAGTATACTAAAAGCĀATGĠCACTCTCACTGACCACCTCCATT  
GCTCTTGCAACGATCCTTTTGTTCGTTTACAAGTTCGCTACTCGTTCC  
AAATCCACCAAAAAAGCCTTCCTGAGCCATGGCGGCTTCCCATTATT  
GGTCACATGCATCACTTGATTGGTACAACGCCACATCGTGGGGTTAGG  
GATTTAGCCAGAAAGTATGGATCTTTGATGCATTTACAGCTTGGTGAA  
GTTCCAACAATCGTGGTGTCTATCTCCGAAATGGGCTAAAGAGATTTTG  
ACAACGTACGACATTACCTTTGCTAACAGGCCCGAGACTTTAACTGGT  
GAGATTGTTTTATATCACAATACGGATGTTGTTCTTGACCTTATGGT  
GAATACTGGAGGCAATTACGTAAAATTTGCACATTGGAGCTTTTGAGT  
GTTAAGAAAGTAAAGTCATTTCAAGTCACCTTCGTGAAGAGGAGTGTGG  
AATTTGGTTCAAGAGATTAAAGCTTCAGGTTTCAGGGAGACCGGTTAAC  
CTTTCAGAGAATGTTTTCAAGTTGATTGCAACGATACTTAGTAGAGCC  
GCATTTGGGAAAGGGATCAAGGACCAGAAAGAGTTAACGGAGATTGTG  
AAAGAGATACTGAGGCAAACCTGGTGGTTTTGATGTGGCAGATATCTTT  
CCTTCAAAGAAATTTCTTCATCATCTTTTCGGGCAAGAGAGCTCGGTTA  
ACTAGCCTTCGCAAAAAGATCGATAATTTAATCGATAACCTTGTAGCT  
GAGCATACTGTTAACACCTCCAGTAAAACCTAACGAGACACTCCTCGAT  
GTTCTTTTAAAGGCTCAAAGACAGTGCTGAATTTCCATTAAACATCTGAT  
AACATTAAAGCCATCATTTTGGATATGTTTGGAGCAGGCACAGACACT  
TCCTCATCCACAATCGAATGGGCGATTTTCGGAACTCATAAAGTGTCG  
AAAGCAATGGAGAAAGTACAAGCGGAATTGAGGAAAGCATTGAACGGA  
AAAGAAAAGATCCATGAGGAAGACATTCAAGAACTAAGCTACTTGAAC  
ATGGTAATCAAAGAAACATTGAGGTTGCACCCTCCACTACCCTTGTT  
CTGCCAAGAGAGTGCCGCCAACCAGTCAATTTGGCTGGATACAACATA  
CCCAATAAGACCAAACCTTATTGTCAACGTCTTTGCGATAAATAGGGAC  
CCTGAATATTGGAAAGACGCTGAAGCTTTCATCCCTGAACGATTTGAA  
AATAGTTCTGCAACTGTCATGGGTGCAGAATACGAGTATCTTCCGTTT  
GGAGCTGGGAGAAGGATGTGTCTGGAGCCGCACCTGGTTTAGCTAAC  
GTGCAGCTCCCGCTCGCTAATATACTATATCATTTCAACTGGAAACTC  
CCCAATGGTGTGAGCTATGACCAGATCGACATGACCGAGAGCTCTGGA  
GCCACGATGCAAAGAAAGACTGAGTTGTTACTCGTTCCAAGTTTCTAG  
(SEQ ID NO:1)

**FIG. 2**

Secuencia aminoacídica de la amorfadieno-12-oxidasa  
MKSILKAMALSLTTSIALATILLFVYKFATRSKSTKKSLPEPWRLPII  
GHMHHLIGTTPHRGVRDLARKYGSLMHLQLGEVPTIVVSSPKWAKEIL  
TTYDITFANRPETLTGEIVLYHNTDVVLAPYGEYWRQLRKICTLELLS  
VKKVKSFQSLREEECWNLVQEIKASGSGRPVNLSENVFKLIATILSRA  
AFGKGIKDQKELTEIVKEILRQTGGFDVADIFPSKKFLHHLSGKRARL  
TSLRKKIDNLIIDNLVAEHTVNTSSKTNETLLDVLLRLKDSAEFPLTSD  
NIKAIILDMFGAGTDTSSSTIEWAISELIKCPKAMEKVQAE LRKALNG  
KEKIHEEDIQELSYLNMVIKETLRLHPPLPLVLPRECRQPVNLAGYNI  
PNKTKLIVNVFAINRDPEYWKDAEAFIPERFENS SATVMGA EYEYLPF  
GAGRRMCPGAALGLANVQLPLANILYHFNWKLPNGVSYDQIDMTESSG  
ATMQRKTELLLVPSF (SEQ ID NO:2)

## FIG. 3

ADN de la citocromo-P450-reductasa de *Artemisia annua*

Secuencia codificante del ADNc de la AACPR

ATGcAATCAACAACCTCCGTTAAGTTATCTCCCTTCGATCTAATGA  
CGGCGTTACTTAACGGCAAGGTATCGTTCGACACATCAAACACATC  
GGATACGAATATTCCGTTAGCGGTGTTTATGGAGAATCGTGAGCTT  
TTGATGATTTTAACTACTTCGGTTGCGGTGTTGATCGGATGCGTTG  
TGGTGCTTGTGTGGAGACGGTCGTCGTCGGCGGCGAAGAAAGCGGC  
GGAGTCGCCGGTGATTGTTGTGCCGAAGAAAGTGACGGAGGATGAG  
GTTGATGACGGACGGAAGAAAGTTACTGTGTTTTTTGGAACCTCAGA  
CTGGTACTGCTGAAGGTTTTGCTAAGGCGCTTGTGGAAGAAGCTAA  
AGCGCGATATGAAAAGGCGGTGTTTAAAGTGATTGATTTGGATGAT  
TATGCTGCTGAAGATGATGAGTATGAGGAGAAGTTAAAGAAAGAAT  
CTCTTGCTTTTTTCTTTTCTAGCTACGTATGGAGATGGTGAGCCGAC  
AGATAATGCTGCTAGATTCTATAAATGGTTTACCGAGGGTGAAGAG  
AAAGGTGAATGGCTTGACAAGCTTCAATACGCAGTGTTTGGACTTG  
GTAACAGACAGTATGAGCATTTCAACAAGATTGCGAAGGTGGTCTGA  
TGAAAAACTTGTGGAGCAGGGTGCAAAGCGCCTTGTTCCTGTTGGC  
ATGGGAGACGATGATCAATGTATCGAAGACGACTTCACTGCATGGA  
AAGAGTTGGTGTGGCCTGAGTTGGATCAATTACTTCGTGATGAGGA  
TGATACATCTGTTGCCACTCCATACACAGCTGCTGTTGGAGAATAC  
CGTGTTGTGTTCCATGACAAACCAGAGACATATGATCAGGATCAAC  
TGACAAATGGCCATGCTGTTTCATGATGCTCAACATCCATGCAGATC  
CAATGTCGCTGTCAAAAAGGAGCTCCATTCCTCTATCTGACCGG  
TCTTGCACTCATTGGAATTTGATATCTCTAATACTGGATTATCGT  
ATGAACTGGGGACCATGTTGGAGTCTACGTTGAGAATCTAAGTGA  
AGTTGTGGACGAAGCTGAAAAATTAATAGGTTTACCGCCGCACACT  
TATTTCTCAGTACATACTGATAACGAAGACGGGACACCACCTGGTG  
GAGCCTCTTTGCCACCTCCTTTCCCTCCATGCACCTTTAAGAAAAGC  
ATTGGCTTCCTATGCCGATGTTTTGAGCTCTCCTAAAAAGTCAGCT  
TTGCTTGCTTTAGCTGCTCATGCTACTGATTCTACTGAAGCTGATA  
GACTGAAATTTTTTGCGTCTCCTGCTGGAAAGGATGAATATGCTCA  
GTGGATAGTTGCAAGCCACAGAAGTCTCCTTGAGGTCATGGAGGCC  
TTCCCATCAGCTAAGCCTCCGCTTGGTGTTTTTTTTTGCATCTGTCTG  
CCCCACGTTTGCAGCCGAGATACTATTCCATTTCTTCTTCCCCAAA  
GTTTGCGCCAAATAGGATTCATGTAACCTTGTGCATTAGTGATGAG  
CAAACACCATCAGGCCGCGTTTCAAGGGAGTCTGTTCAACATGGA  
TGAAGAATGCCGTGCCTATGACAGAAAGCCAGGATTGCAGTTGGGC  
CCCAATTTATGTTAGAACATCCAATTTTCAGACTTCCTTCTGATCCT  
AAGGTCCCAGTTATCATGATTGGCCCAGGCACTGGATTGGCTCCAT  
TTAGAGGTTTTCTTTCAGGAAAGGTTAGCTCAGAAGGAAGCTGGGAC  
TGAGCTCGGAACAGCCATCTTATTCTTCGGATGCAGGAATCGCAAA  
GTGGATTTTCATATATGAGGACGAGCTTAATAATTTTCGTGGAGACGG  
GGGCTCTTTCCGAGCTTGTTACGGCCTTCTCTCGTGAAGGTGCCAC  
TAAGGAGTACGTGCAACACAAGATGACTCAGAAGGCTTCGGATATC  
TGGAATTTACTCTCTGAGGGAGCATATTTGTATGTTTGCGGTGATG  
CCAAAGGCATGGCCAAAGATGTACATCGGACTCTGCACACTATTGT  
GCAAGAACAGGGATCTCTAGACTCCTCAAAGGCGGAGCTCTACGTG  
AAGAATCTACAAATGGCAGGAAGATATCTCCGTGATGTATGGGTCTG  
ACATGGAACAGAAGTTGATTTCCGAAGAAGACCTCGAGTAA (SEQ  
ID NO:3)

## FIG. 4

Secuencia aminoacídica de la citocromo-P450-reductasa de *Artemisia annua*

MQSTTSVKLSPFDLMTALLNGKVSFDTSNTSDTNIPLAVFMENREL  
LMILTTSVAVLIGCVVVLVWRRSSSAAKKAAESPVIVVPKKVTEDE  
VDDGRKKVTVFFGTQTGTAEGRFAKALVEEAKARYEKAVFKVIDLDD  
YAAEDDEYEEKLKKESLAFFFLATYGDGEPTDNAARFYKWFTEGEE  
KGEWLDKLQYAVFGLGNROYEHFNKIAKVVDEKLVEQGAKRLVPVG  
MGDDDQCIEDDFTAWKELVWPELDQLLRDEDDTSVATPYTAAVGEY  
RVVFHDKPETYDQDQLTNGHAVHDAQHPCRSNVAVKKELHSPLSDR  
SCTHLEFDISNTGLSYETGDHVG VYVENLSEVVDEAEKLI GLPPHT  
YFSVHTDNEDGTPLGGASLPPFPFPPCTLRKALASYADVLSSPKKSA  
LLALAAHATDSTEADRLKFFASPAGKDEYAQWIVASHRSLLEVMEA  
FPSAKPPLGVFFASVAPRLQPRYYSISSSPKFAPNRIHVTCALVYE  
QTPSGRVHKGVCSTWMKNAVPMTESQDCSWAPIYVRTSNFRLPSDP  
KVPVIMIGPGTGLAPFRGFLQERLAQKEAGTELGTAILFFGCRNRK  
VDFIYEDELNNFVETGALSELVTAFSREGATKEYVQHKMTQKASDI  
WNLLSEGAYLYVCGDAKGMADVHRTLHTIVQEQSLDSSKAELYV  
KNLQMAGRYLRDVWVDMEQKLI SEEDLE (SEQ ID NO:4)

FIG. 5

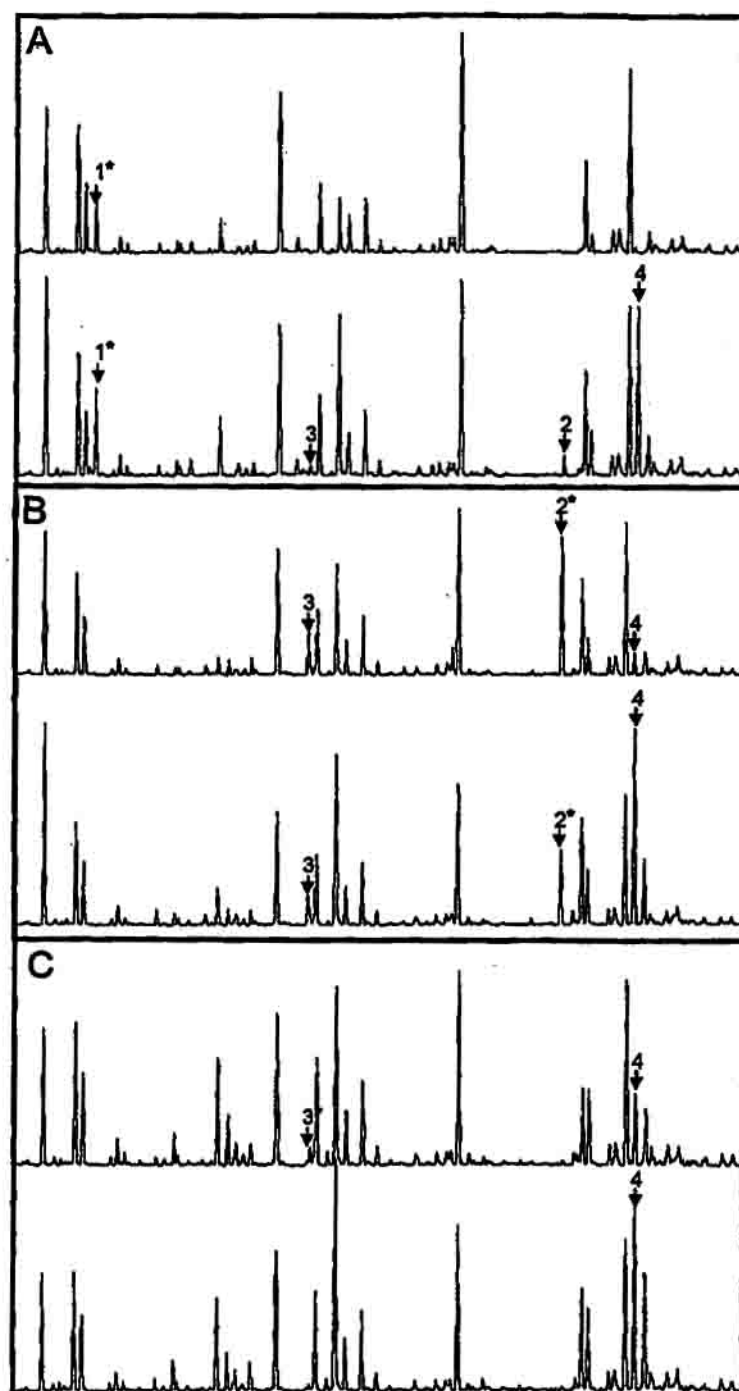


FIG. 6A

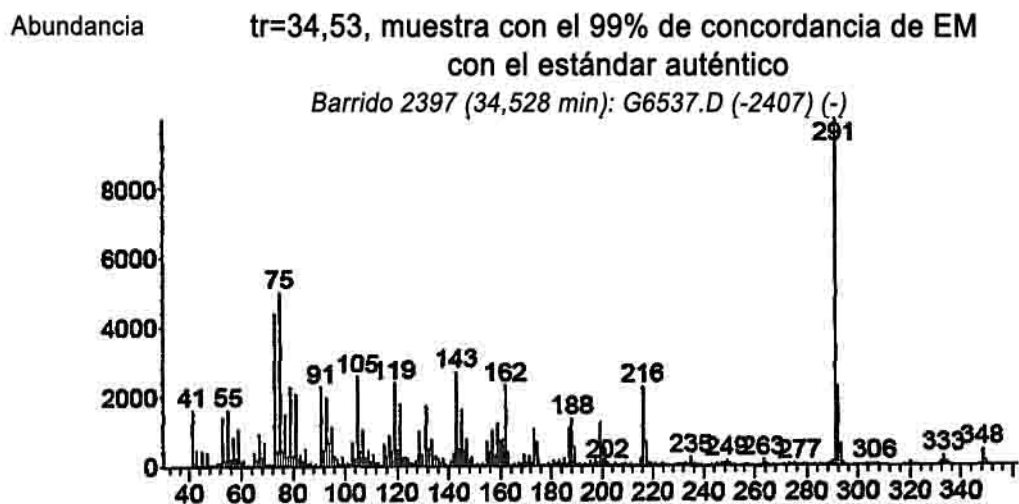


FIG. 6B

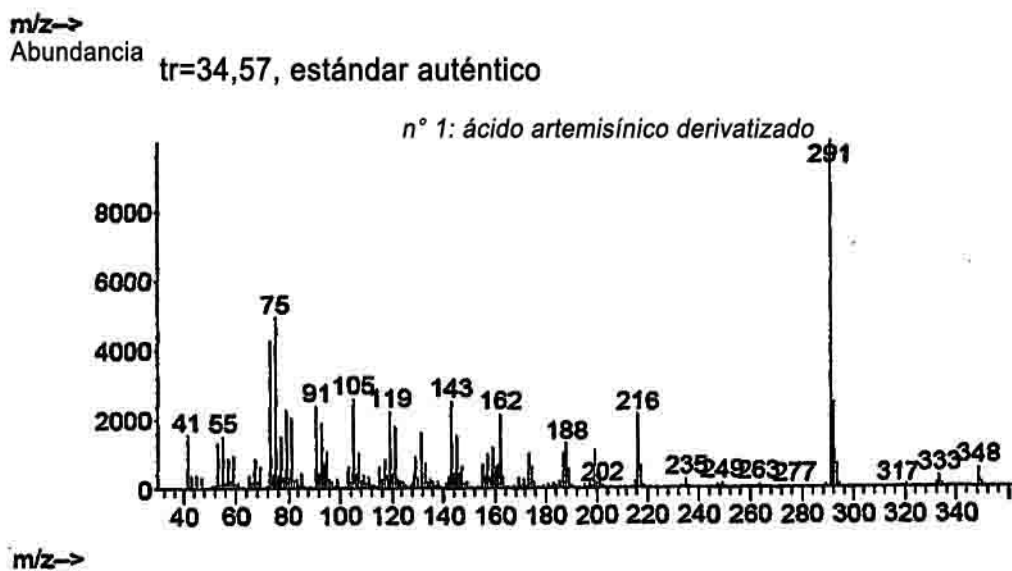


FIG. 7

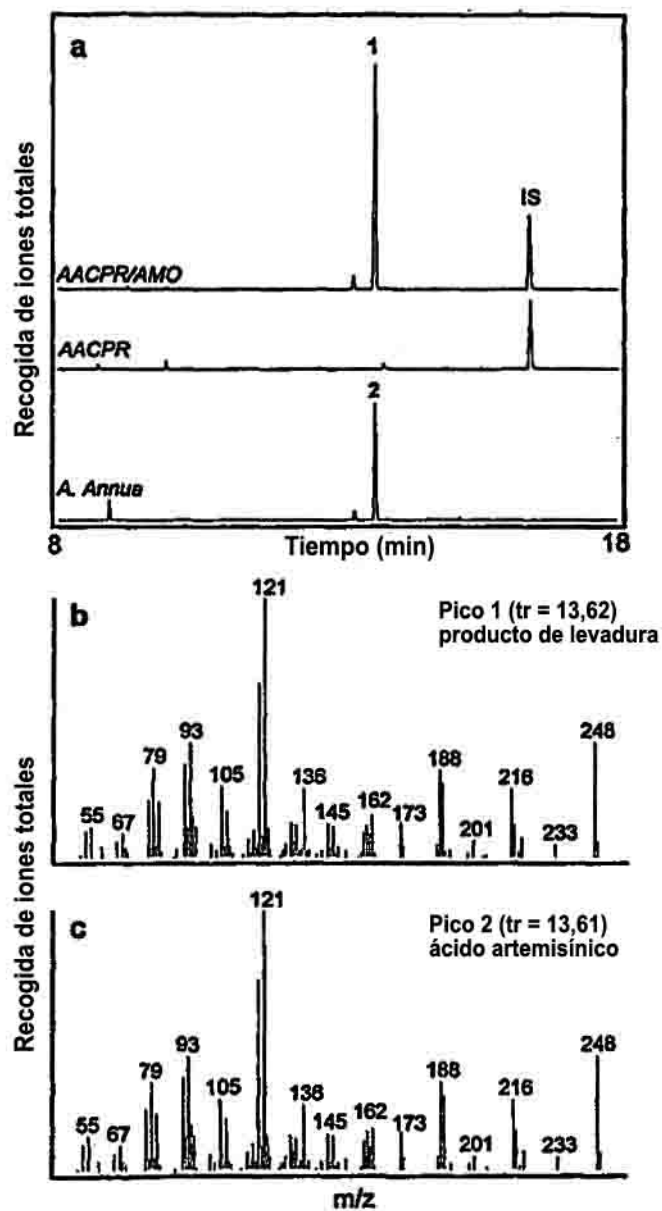
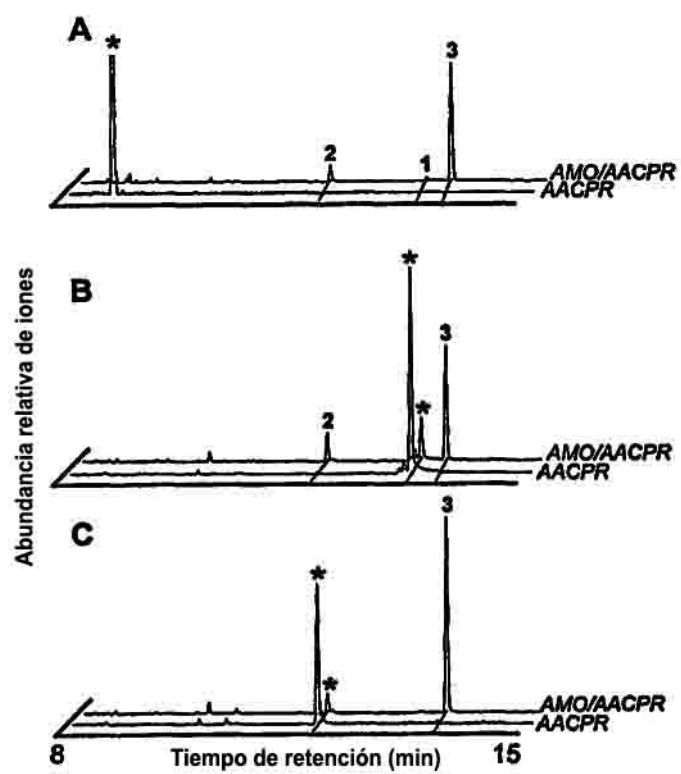


FIG. 8



## FIG. 9

&gt;71D-B1-nt

ATGGCACTTTCACTGACCACCTCCATTGCTCTTGCCACGATCCTTTTCTTCGTAATTTACAAGTTCGCTAC  
 TCGTTCCAAATCCACAAAAACAGCCTTCCTGAGCCATGGCGACTTCCCATTATTGGTCACATGCATCACT  
 TGATTGGTACAATACCACATCGTGGGCTTATGGATTTAGCCAGAAAGTATGGATCTTTAATGCATTTACAG  
 CTTGGTGAAGTTTCAACAATCGTGGTGTCTCCTCGAAATGGGCTAAAGAGATTTTGACAACGTACGACAT  
 TGCCTTTGCTAACAGGCCCTGGACTTTGGCTGGTGAGATTGTTGTATATCGCAATACAAATATTGCTGCTG  
 CACCTTATGGTGAATACTGGAGGCGATTACGTAAACTTTGCACATCGGAGCTTATGAGTGTTAAGAAAGTA  
 AAGTCATATCAGTCGCTTCGTGAAGAGGAGTGTTGGAATTTGGTTCAAGAGATTAAAGCTTCAGGTTCAAG  
 GATACCGGTAAACCTTTCAGAGAACATTTTCAAGTTGATTGCAACGATACTTTGTAGAGCCGCGTTTGGA  
 AAGGAGTCAAGGACCAGAAGGAGTGACGGAGATTATGAAAGAGATGTTGAGGGAAGTTGGTGGTTTTGAT  
 GTGGCAGATATCTTTCCTTCGAAGAAATTTCTTCATCATCTTTCAGGCAAGAGAGCCAGGTTAACTAGCAT  
 TCATAAGAAGCTCGATAATTTTATCAATAACCTTGTTGCTGAGCATACTTTCAAACTTCAAGTAAACTG  
 AGGAGACACTTCTTGATGTTCTTCTAAGGCTCAAAGATAGCGCTGAATTCCTTAAACAGCTGACAATGTT  
 AAAGCCATCATTTTGGATATATTTGCAGCAGGGACAGACACTTCATCAACCACAATCGAATGGGCGATTTT  
 GGAACATATAAGTGTCGAGAGCGATGGAGAAAGTACAAGCAGAACTGAGGAAAGCACTTAACGGAAAAG  
 AAAAGATCCATGAGGAAGATATTCAAGGACTAAGCTACTTAACTTGGTAATCAAAGAAACATTAAGGTTG  
 CACCTTCCACTACCTTGTTGCCAAGAGAGTGCCGTGAACCACTCAATTTGGCTGGATACGACATACCCAA  
 TAAGACAAGACTTATTGTCAACGCTCTTTCGATAAAATAGGGACCCAGAATACTGGAAAGACGCTGAAATTT  
 TCATCCCCGAACGATTTGAAAATAGTTCTACAACCTCTCATGGGTGCAGAATATGAGTATCTTCCGTTTGGA  
 GCTGGGAGAAGGATGTGTCCTGGAGCCGCACCTGGTTTAGCCAACGTGCAGCTACCGCTCGCTAATATAC  
 ATATCATTTCAACTGGAACTCCCAACGGTGCAGCTATGATCAGATCGACATGACCGAGAGGTTTGGA  
 TCTCGGTTGAAAGAAAGACTCAGTTGTTACTCGTACCAAGTTTCTAG (SEQ ID NO:5)

## FIG. 10

>71D-B1-aa

MALSLTTSIALATILFFVIYKFATRSKSTKNSLPEPWRLPIIGHMHHLIGTIPHRGLMDLARKYGSLMHLQ  
LGEVSTIVSSPKWAKEILTTYDIAPANRPWTLAGEIVVYRNTNIAAAPYGEYWRRLRKLCTSELM SVKKV  
KSYQSLREEECWNLVQEIKASGSGIPVNLSNIFKLIATILCRAAFGKGVKDQKECTEIMKEMLEVGGFD  
VADIFPSKKFLHHLSGKRRLTSIHKKLDNFINNVAEHTFKTSSKTEETLLDVLLRLKDSAEFPLTADNV  
KAILDIFAAGTDTSSTTIEWAISELIKCPRAMEKVQAE LRKALNGKEKIHEEDIQGLSYLNLVIKETLRL  
HPPLPLLPRECREPVNLAGYDIPNKTRLIVNVFAINRDPEYWKDAEIFIPERFENSSTLMGA EYEYLPFG  
AGR RMCPGAALGLANVQLPLANILYHFNWKL PNGASYDQIDMTERFGISVERKTQLLLVP SF (SEQ ID  
NO:6)

FIG. 11

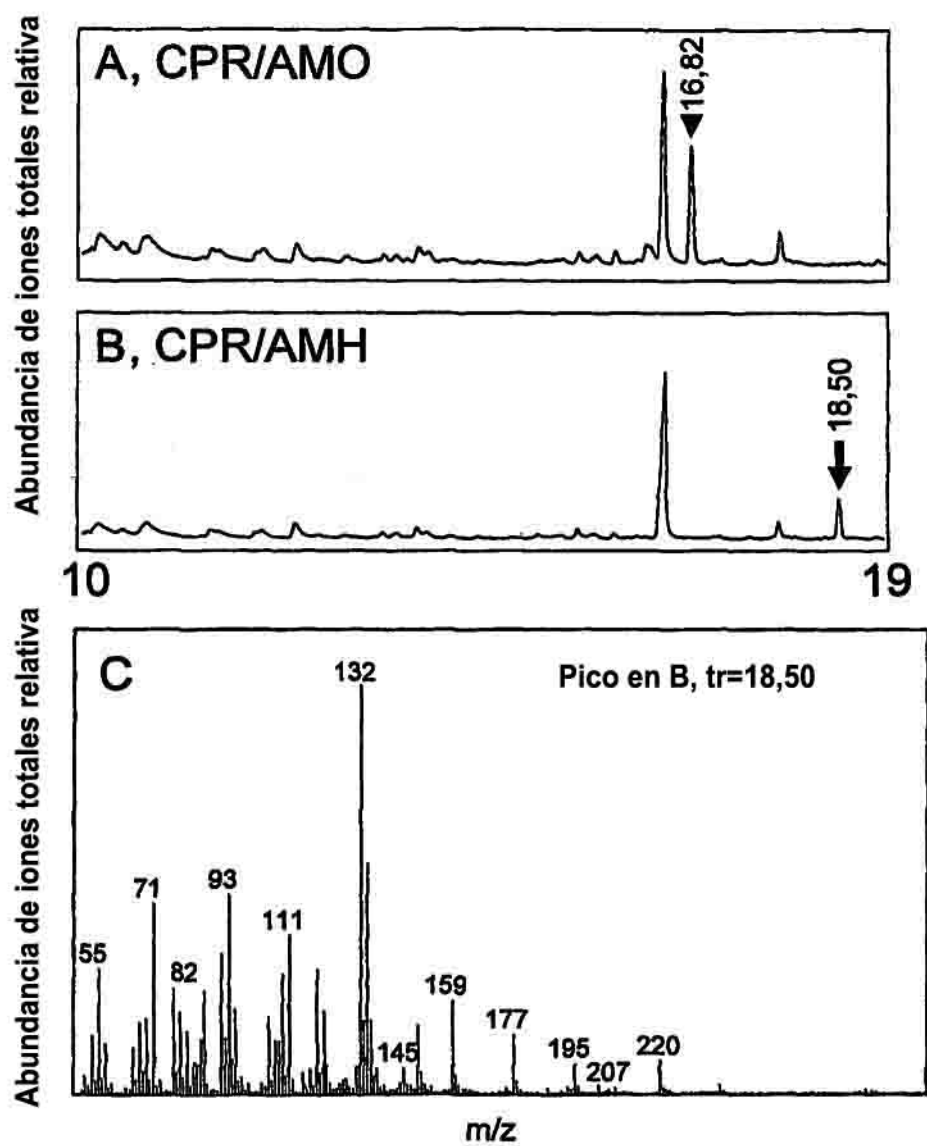
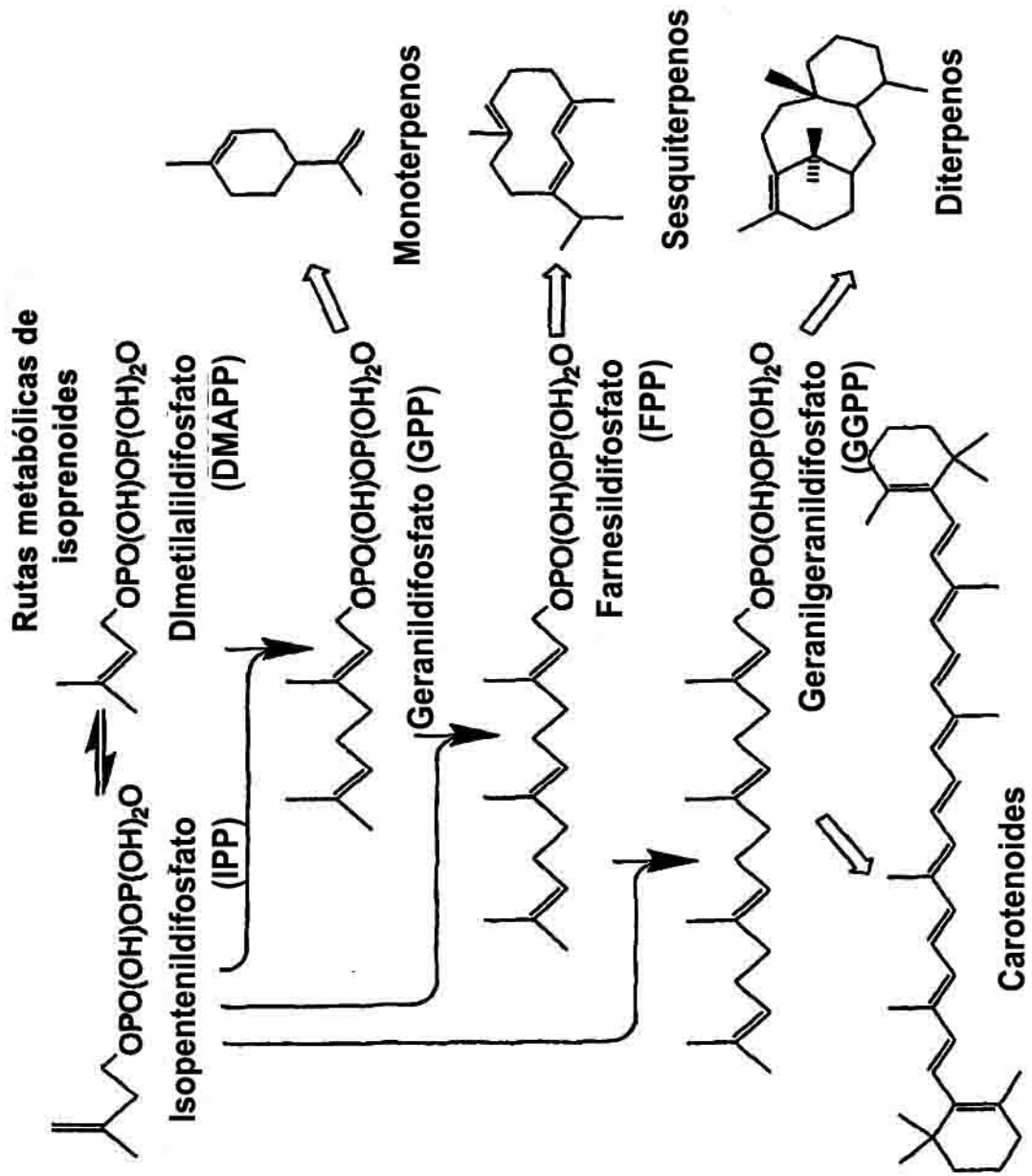


FIG. 12

&gt;71D-C1-nt

GCCCTTCGAGCCGTATGGGGATTACTGGCGGCAATTACGTAAACTTTGCACATTGGAGCTTTTGAGTGCTA  
 AGAAAGTAGAGTCATATCAGTCGCTTCGTGAAGAGGAGTGTTGGAATTTAGTTCAAGAGATTAAAGCTTCA  
 GGTCAGGGATACCGGTAAACCTTTCAGAGAATATTTACAAGTTGGTTGCAATGATACTTAGTAGAGCTGC  
 GTTTGGGAAAAGAATCAAGGACCATAAGGAGTTTACGGAGCTTGTGGAACAGATGTTGAGGGAACCTGGTG  
 GTTTGATGTGGCAGATATCTTTCCTTCGCAGAAAATTTCTACATCATATTTTCGGGCAAGAGATCTAGGTTA  
 ACTAGCATTACAAAAAGCTCGATAATTTAATCAATAACCTTGTTGCTGAGCATATTGTTGAAGCCTCAAG  
 TAAAACTAAGGAGACGCTCCTTGATGTTCTTCTAAGGCACAAAGATAGCCTTGAATCCCATTGACAGCTG  
 ATAACGTTAAAGCCATCATTTTGGTATGAATTAATCCAATATATTTTTTTTTTCAAAGGCCATAATAGTG  
 TTAAACAAGCTTGAAATTTTTTATAACTAAGTACATGCACCTAAGTTAGTACTCGTGAAAATATAATGAGT  
 CATCATAGGGGTTCCATGAAATATACAGGACATGTTTACAGCAGGCACAGACACTTCGTCAACCACAATCG  
 AATGGGTGATTTTCGGAACCTATAAAGTGTCGAGAGCTATGGAGAAAATACAAGCGGAACCTGAGGAAAGCA  
 CTTAACGGAAGAAAAGATCCACGAGGAAGACATCCAAGAACTAAGCTACTTAAACTTGGTAATCAAAGA  
 AACATTAAGGTTGCACCCCTCCACTACCCTTGGTTTTGCCACGAGAGTGCCGTCAACCAGTCAATTTGGCTG  
 GATATGACATACCCAATAAGACCAAACTTATTGTCAACGCTTTTTCGATAAATAGGGACCCTGAATACTGG  
 AAAGACGCTGAATCTTTCATCCCAGAGCGCTTCTTAACTCTGGT (SEQ ID NO:7)

FIG. 13



**FIG. 14**  
**Ruta del mevalonato**

