

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 626 456**

51 Int. Cl.:

**C12P 5/00** (2006.01)

**C12N 9/88** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.12.2009 E 14152034 (6)**

97 Fecha y número de publicación de la concesión europea: **08.03.2017 EP 2772543**

54 Título: **Procedimiento enzimático para producir beta-santaleno y la enzima correspondiente**

30 Prioridad:

**11.12.2008 EP 08171298**  
**17.08.2009 WO PCT/IB2009/053623**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**25.07.2017**

73 Titular/es:

**FIRMENICH SA (100.0%)**  
**1, route des Jeunes, Case Postale 239**  
**1211 Genève 8, CH**

72 Inventor/es:

**SCHALK, MICHEL**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 626 456 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Procedimiento enzimático para producir beta-santaleno y la enzima correspondiente

**Campo técnico**

5 La presente invención proporciona un procedimiento de producción de  $\beta$ -santaleno, comprendiendo dicho procedimiento poner en contacto al menos un polipéptido con pirofosfato de farnesilo (FPP). En particular, dicho procedimiento se puede llevar a cabo in vitro o in vivo para producir  $\beta$ -santaleno, un compuesto muy útil en los campos de la perfumería y la aromatización. La presente invención también proporciona la secuencia de aminoácidos de un polipéptido útil en el procedimiento de la invención. Un ácido nucleico que codifica el polipéptido de la invención y un vector de expresión que contiene dicho ácido nucleico son también parte de la presente invención. Un organismo huésped no humano o una célula huésped no humana transformados para usarse en el procedimiento de producción de  $\beta$ -santaleno son también un objeto de la presente invención.

Otro objeto de la invención es un procedimiento para preparar polipéptidos variantes que tiene actividad  $\beta$ -santaleno sintasa.

**Técnica anterior**

15 Se encuentran terpenos en la mayoría de los organismos (microorganismos, animales y plantas). Estos compuestos están compuestos de cinco unidades de carbono llamadas unidades de isopreno y se clasifican por el número de estas unidades presentes en su estructura. Así, los monoterpenos, los sesquiterpenos y los diterpenos son terpenos que contienen 10, 15 y 20 átomos de carbono respectivamente. Los sesquiterpenos, por ejemplo, se encuentran ampliamente en el reino vegetal. Muchas moléculas de sesquiterpenos se conocen por sus propiedades de aroma y fragancia y sus efectos cosméticos, medicinales y antimicrobianos. Se han identificado por encima de 300 hidrocarburos sesquiterpenos y por encima de 3000 sesquiterpenoides (Joulain, D. y König, W.A., The Atlas of Spectral Data of Sesquiterpene Hydrocarbons, EB Verlag, Hamburgo, 1998; Connolly, J.D., Hill R.A., Dictionary of Terpenoids, vol. 1, Chapman and Hall (editor), 1991) y se identifican cada año muchas estructuras nuevas. Los extractos vegetales obtenidos por medios diferentes tales como destilación de vapor o extracción de disolventes se usan como fuente de terpenos. Las moléculas de terpenos se usan a menudo como tales, pero en algunos casos las reacciones químicas se usan para transformar los terpenos en otras moléculas de valor alto.

La producción biosintética de terpenos implica enzimas llamadas terpeno sintasas. Hay virtualmente una infinidad de sesquiterpeno sintasas presentes en el reino vegetal, usando todas el mismo sustrato (pirofosfato de farnesilo, FPP) pero que tienen perfiles de producto diferentes. Los genes y ADNc que codifican sintasas de sesquiterpenos se han clonado y sus las enzimas recombinantes correspondientes se han caracterizado. La biosíntesis de terpenos en plantas y otros organismos se ha estudiado exhaustivamente y no se detalla adicionalmente aquí, pero se hace referencia a Dewick, Nat. Prod. Rep., 2002, 19, 181-222, que revisa el estado de la técnica de las rutas de biosíntesis de terpenos.

35  $\beta$ -santaleno es una molécula de sesquiterpeno que se da en la naturaleza que se puede usar como material de partida para la síntesis química o para la biosíntesis de  $\beta$ -santalol (como se representa en la figura 2B), que es un constituyente principal de aceite de sándalo. El aceite de sándalo es un ingrediente de perfumería importante que se obtiene por destilación del duramen de las especies de Santalum. El sándalo se usa también grandemente para inciensos y medicina tradicional. El aceite contiene 90 % de alcoholes de sesquiterpenos. Entre los diferentes isómeros de santalol,  $\beta$ -santalol es el principal contribuyente para el típico olor dulce-leñoso y balsámico de aceite de sándalo. Otros constituyentes tales como epi- $\beta$ -santalol y  $\alpha$ -santalol pueden contribuir también a la nota de sándalo.

En general, el precio y la disponibilidad de los extractos naturales de plantas son dependientes de la abundancia, producción de aceite y origen geográfico de las plantas. Además, la disponibilidad y calidad de los extractos naturales es muy dependiente del clima y de otras condiciones locales que conducen a variabilidad de año en año, volviendo el uso de tales ingredientes en perfumería de alta calidad muy difícil o incluso imposible algunos años. Debido a la sobreexplotación de los recursos naturales, las dificultades de cultivo, el crecimiento lento de las plantas de *Santalum*, las disponibilidades de materia prima de sándalo han disminuido radicalmente durante las pasadas décadas. Por lo tanto, sería una ventaja proporcionar una fuente de  $\beta$ -santalol, que esté menos sujeta a fluctuaciones en disponibilidad y calidad. Una síntesis química de los constituyentes sesquiterpenos de sándalo hasta el momento no está disponible. Una ruta bioquímica que conduce a la síntesis de  $\beta$ -santaleno, que podría usarse después para producir  $\beta$ -santalol, sería por lo tanto de gran interés.

El santaleno tipo sesquiterpeno y particularmente sesquiterpenos con el esqueleto de  $\beta$ -santaleno, se han identificado en varias especies de plantas. Empero, no se ha descrito aún ninguna sesquiterpeno sintasa capaz de producir  $\beta$ -santaleno.

Una sesquiterpeno sintasa capaz de sintetizar al menos un sesquiterpeno bicíclico y/o tricíclico que tiene un esqueleto de carbono de santaleno, el ácido nucleico correspondiente y un procedimiento de producción de tal compuesto que tiene un esqueleto de carbono de santaleno se divulgan en la solicitud de patente internacional WO 2006/134523. No obstante, no se detectó ninguna traza de  $\beta$ -santaleno como producto de las sesquiterpeno sintasas divulgadas en los

ejemplos. El único producto con un esqueleto de santaleno fue epi-beta-santaleno. Las propiedades de epi-beta-santaleno son muy diferentes de aquellas de  $\beta$ -santaleno. En particular, no es de interés la síntesis de  $\beta$ -santalol. Además, la sesquiterpeno sintasa divulgada en el documento WO 2006/134523 comparte solo el 27 % de identidad con la secuencia de la invención.

- 5 El porcentaje de identidad entre sesquiterpeno sintasas conocido a partir de las bases de datos y los polipéptidos de la invención es muy bajo. La secuencia de proteínas más cercana a la  $\beta$ -santaleno sintasa de la invención es una monoterpeno sintasa de *Santalum album* (n.º de acceso ACF24767; Jones, C.G., Keeling, C.I., Ghisalberti, E.L., Barbour, E.L., Plummer, J.A. y Bohlmann, J. Arch. Biochem. Biophys., 2008, 477(1), 121-130) que comparte el 58 % de identidad de secuencia de aminoácidos con la  $\beta$ -santaleno sintasa de la invención. Cuando se pone en contacto con FPP, esta enzima produce por encima del 90 % de  $\beta$ -bisaboleno y no se forma ningún isómero de santaleno.

10 Además de las diferencias entre las propias secuencias, también ha de indicarse que la estructura y las propiedades de los productos sintetizados por la enzima mencionada anteriormente son muy diferentes de las del sesquiterpeno  $\beta$ -santaleno. En particular, los monoterpenos producidos por esta enzima, es decir alfa-terpineol, limoneno, geraniol, mirceno, linalool y algunos otros productos secundarios no son adecuados como un material de partida para la síntesis de  $\beta$ -santalol, que es un ingrediente muy útil en el campo de la perfumería.

15 A pesar de estudios exhaustivos de ciclación de terpenos, el aislamiento y la caracterización de las terpeno sintasas aún es difícil, en particular en plantas, debido a su baja abundancia, sus patrones de expresión a menudo transitorios y la complejidad de purificarlos a partir de las mezclas de resinas y de los compuestos fenólicos en tejidos donde se expresan.

- 20 Es un objetivo de la presente invención proporcionar procedimientos para fabricar  $\beta$ -santaleno de una forma económica, como se ha indicado anteriormente. De acuerdo con ello, la presente invención tiene el objetivo de producir  $\beta$ -santaleno produciendo a la vez pocos desechos, un procedimiento más eficiente en lo que respecta a energía y recursos y que reduzca a la vez la dependencia de los combustibles fósiles. Es un objetivo adicional proporcionar enzimas capaces de sintetizar  $\beta$ -santaleno, que es útil como ingrediente de perfumería y/o de aromas.

25 Abreviaturas usadas

ACC	ácido 1-aminociclopropanocarboxílico
pb	par de bases
kb	kilobase
BSA	seroalbúmina bovina
30 2,4D	ácido 2,4-diclorofenoxiacético
ADN	ácido desoxirribonucleico
ADNc	ADN complementario
dNTP	desoxinucleotido trifosfato
DTT	ditiotreitól
35 EDTA	ácido etilendiaminatetraacético
FPP	pirofosfato de farnesilo
CG	cromatografía de gases
IPTG	isopropil-D-tiogalacto-piranósido
Kin	cinetina
40 LB	caldo de lisogenia
EM	espectrómetro de masas
PCR	reacción en cadena de la polimerasa
RMCE	intercambio de casetes mediado por recombinasa
3'-/5'-RACE	amplificación rápida de ADNc en 3' y en 5'
45 ARN	ácido ribonucleico

ARNm ácido ribonucleico mensajero

RNAsa Ribonucleasa

### Descripción de la invención

5 La presente invención proporciona un procedimiento para producir biosintéticamente  $\beta$ -santaleno de un modo económico, fiable y reproducible, tal como se establece en cualquiera de las reivindicaciones 5 a 10, usando un polipéptido que tiene una actividad  $\beta$ -santaleno sintasa tal como se establece en la reivindicación 1 o la reivindicación 2. La presente invención es particularmente útil debido a que no se conoce en la técnica anterior tal polipéptido y debido a que no se ha descrito tal biosíntesis de  $\beta$ -santaleno. Esto resuelve el muy importante problema del suministro de  $\beta$ -santaleno, un compuesto que es muy útil para la industria de la perfumería. La presente invención también  
10 proporciona una secuencia de ácidos nucleicos tal como se establece en la reivindicación 3 o la reivindicación 4.

El polipéptido y el ácido nucleico son herramientas muy importantes, que son ambas necesarias para llevar a cabo el procedimiento de la invención tal como se establece en las reivindicaciones 5 a 10. Lo mismo es cierto para vectores tal como se establece en la reivindicación 11 y para organismos huésped no humanos o células huésped no humanas tal como se establece en las reivindicaciones 12 a 15.

15 Un procedimiento para preparar un polipéptido variante se proporciona también tal como se establece en la reivindicación 18. Debería indicarse que una variación de secuencia más amplia se aplica a las reivindicaciones del procedimiento de la invención, mientras que las reivindicaciones del producto están de algún modo más limitadas con respecto a las secuencias

20 Una "sesquiterpeno sintasa" o un "polipéptido que tiene una actividad sesquiterpeno sintasa" se desea para el propósito de la presente aplicación como un polipéptido capaz de catalizar la síntesis de una molécula de sesquiterpeno o de una mezcla de moléculas de sesquiterpenos a partir del precursor de terpeno acíclico FPP.

25 Como una " $\beta$ -santaleno sintasa" o como un "polipéptido que tiene una actividad  $\beta$ -santaleno sintasa", los autores de la presente invención quieren decir aquí un polipéptido capaz de catalizar la síntesis de  $\beta$ -santaleno, en forma de cualquiera de sus estereoisómeros o de una mezcla de ellos, comenzando a partir de FPP. El  $\beta$ -santaleno puede ser el único producto o puede ser parte de una mezcla de sesquiterpenos. El  $\beta$ -santaleno se define por su estructura, como se representa en la Figura 2A.

La capacidad de un polipéptido para catalizar la síntesis de un sesquiterpeno en particular (por ejemplo  $\beta$ -santaleno) se puede confirmar simplemente llevando a cabo el ensayo enzimático como se detalla en el Ejemplo 3.

30 De acuerdo con la presente invención, se pretende que los polipéptidos incluyan polipéptidos truncados siempre que conserven su actividad sesquiterpeno sintasa tal como se ha definido.

35 El porcentaje de identidad entre dos secuencias peptídicas o nucleotídicas es una función del número de residuos de aminoácidos o nucleótidos que son idénticos en las dos secuencias cuando un alineamiento de estas dos secuencias se ha generado. Los residuos idénticos se definen como residuos que son el mismo en las dos secuencias en una posición dada del alineamiento. El porcentaje de identidad de secuencia, como se usa en el presente documento, se calcula a partir del alineamiento óptimo tomando el número de residuos idénticos entre dos secuencias dividiéndolo por el número total de residuos en la secuencia más corta y multiplicando por 100. El alineamiento óptimo es el alineamiento en el que el porcentaje de identidad es el más alto posible. Se pueden introducir huecos dentro de una o dentro de ambas secuencias en una o más posiciones del alineamiento para obtener el alineamiento óptimo. Estos huecos se toman después en cuenta como residuos no idénticos para el cálculo del porcentaje de identidad de  
40 secuencia.

45 El alineamiento para el propósito de determinar el porcentaje de identidad de secuencia de aminoácidos o de ácidos nucleicos se puede lograr en distintas maneras usando programas de ordenador y por ejemplo programas de ordenador disponibles públicamente en la World Wide Web. Preferentemente, el programa BLAST (Tatiana y cols., FEMS Microbio/Lett., 1999, 174: 247-250, 1999) ajustado a los parámetros por defecto, disponible a partir del National Center for Biotechnology Information (NCBI) en <http://www.ncbi.nlm.nih.gov/BLAST/b12seq/wblast2.cgi>, se puede usar para obtener un alineamiento óptimo de las secuencias de nucleótidos y para calcular el porcentaje de identidad de secuencia.

De acuerdo con una realización preferente,  $\beta$ -santaleno representa al menos el 20 %, preferentemente al menos el 30 %, preferentemente al menos el 35 % de los productos producidos por el procedimiento de la invención.

50 El procedimiento puede llevarse a cabo in vitro así como in vivo, dado que los procedimientos que implican solo el metabolismo natural de la planta, sin ninguna transformación, no están comprendidos por los procedimientos de la presente invención, como se explicará en detalle adicionalmente.

El polipéptido que se va a poner en contacto con FPP in vitro se puede obtener por extracción a partir de cualquier organismo que lo exprese, usando proteína estándar o tecnología de extracción de enzimas. Si el organismo huésped

es un organismo unicelular o célula que libera el polipéptido de la invención en el medio de cultivo, el polipéptido puede recogerse simplemente del medio de cultivo, por ejemplo por centrifugación, opcionalmente seguido por etapas de lavado y de resuspensión en soluciones de tampón adecuadas. Si el organismo o célula acumula el polipéptido dentro de sus células, el polipéptido puede obtenerse por disrupción o lisis de las células y por extracción adicional del polipéptido a partir del lisado de la célula.

El polipéptido que tiene una actividad  $\beta$ -santaleno sintasa, bien en una forma aislada o bien conjuntamente con otras proteínas, por ejemplo en un extracto de proteína en bruto obtenida de células cultivadas o de microorganismos cultivados, puede estar suspendido en una solución tampón a pH óptimo. Si es adecuado, se pueden añadir sales, DTT, BSA y otras clases de cofactores enzimáticos con el fin de optimizar la actividad enzimática. La concentración de estos co-factores se puede ajustar con el fin de lograr un rendimiento optimizado. Por ejemplo, bajar la concentración de iones de  $Mg^{2+}$  en la suspensión polipeptídica es particularmente ventajosa para incrementar el rendimiento de  $\beta$ -santaleno. La concentración óptima de iones  $Mg^{2+}$  está comprendida entre 2 y 0,75 mM. Las condiciones apropiadas se describen en más detalle en los ejemplos adicionalmente.

El precursor FPP se añade a la suspensión polipeptídica, que se incuba a temperatura óptima, por ejemplo entre 15 y 40 °C, preferentemente entre 25 y 35 °C, más preferentemente a 30 °C. Después de incubación, el  $\beta$ -santaleno producido se puede aislar de la solución incubada por procedimientos de aislamiento estándar, tales como extracción y destilación de disolvente, opcionalmente después de eliminar los polipéptidos de la solución.

La invención también proporciona un procedimiento para producir un polipéptido tal como se establece en la reivindicación 16 o la reivindicación 17.

Estas realizaciones de la invención son particularmente ventajosas dado que es posible que lleven a cabo el procedimiento in vivo aislando anteriormente el polipéptido. La reacción tiene lugar directamente dentro del organismo o célula transformada para expresar dicho polipéptido.

De acuerdo con una realización particular de la invención, el, al menos un, ácido nucleico que codifica la  $\beta$ -santaleno sintasa comprende una secuencia de nucleótidos al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 2 o al complemento de la misma. De acuerdo con otra realización más preferente, dicho ácido nucleico comprende la secuencia nucleotídica SEC ID N.º: 2 o el complemento de la misma.

En otra realización preferente, el ácido nucleico consiste en una secuencia de nucleótidos al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 2 o al complemento de la misma. En una realización incluso más preferente, dicho ácido nucleico consiste en la secuencia nucleotídica SEC ID N.º: 2 o el complemento de la misma.

De acuerdo con otra realización, el, al menos un, ácido nucleico está aislado de una planta de la especie *Santalum*, preferentemente de *Santalum album*.

El organismo o célula se desea para "expresar" un polipéptido, siempre que el organismo o célula se transforme para albergar un ácido nucleico que codifica dicho polipéptido, este ácido nucleico se transcribe a ARNm y el polipéptido se encuentra en el organismo huésped o célula huésped. El término "expresa" comprende "expresa heterológicamente" y "sobreexpresa", refiriéndose el último a niveles de ARNm, polipéptido y/o actividad enzimática sobre y por encima de lo que se mide en un organismo o célula no transformado. Una descripción más detallada de procedimientos adecuados para transformar un organismo huésped no humano o una célula huésped no humana se describirá más adelante en la parte de la memoria descriptiva que está dedicada a tales organismos o células huéspedes no humanos transformados como objetos específicos de la presente invención y en los ejemplos.

Un organismo o célula particular se considera que es "capaz de producir FPP" cuando produce FPP de forma natural o cuando no produce FPP de forma natural pero se transforma para producir FPP, bien antes de la transformación con un ácido nucleico como se describe en el presente documento o bien conjuntamente con dicho ácido nucleico. Los organismos o células transformados para producir una cantidad más alta de FPP que el organismo o célula que se da en la naturaleza están también comprendidos por los "organismos o células capaces de producir FPP". Procedimientos para transformar organismos, por ejemplo microorganismos, de tal forma que produzcan FPP ya se conocen en la técnica. Tales procedimientos pueden encontrarse por ejemplo en la literatura, por ejemplo en las siguientes publicaciones: Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D. y Keasling, J.D. Nat Biotechnol., 2003, 21(7), 796-802 (transformación de E. coli); Wu, S., Schalk, M., Clark, A., Miles, R.B., Coates, R. y Chappell, J., Nat Biotechnol., 2006, 24 (11), 1441-1447 (transformación de plantas); Takahashi, S., Yeo, Y., Greenhagen, B. T., McMullin, T., Song, L., Maurina-Bruner, J., Rosson, R., Noel, J., Chappell, J. Biotechnology and Bioengineering, 2007, 97(1), 170-181 (transformación de levadura).

Para llevar a cabo la invención in vivo, el organismo huésped o la célula huésped se cultivan en condiciones que

conducen a la producción de  $\beta$ -santaleno. De acuerdo con ello, si el huésped es una planta transgénica, se proporcionan condiciones de crecimiento óptimas, tales como condiciones de nutrientes, de luz y de agua óptimas, por ejemplo. Si el huésped es un organismo unicelular, las condiciones que conducen a la producción de  $\beta$ -santaleno pueden comprender adición de cofactores adecuados al medio de cultivo del huésped. Además, se puede seleccionar un medio de cultivo, tal como para maximizar síntesis de  $\beta$ -santaleno. Las condiciones de cultivos óptimos se describen en una manera más detallada en los siguientes ejemplos.

Organismos huéspedes no humanos adecuados para llevar a cabo el procedimiento de la invención in vivo pueden ser cualesquiera organismos pluricelulares o unicelulares no humanos. En una realización preferida, el organismo huésped no humano usado para llevar a cabo la invención in vivo es una planta, un procarionta o un hongo. Se puede usar cualquier planta, procarionta u hongo. Plantas particularmente útiles son aquellas que producen de forma natural cantidades altas de terpenos. En una realización más preferida, la planta se selecciona a partir de la familia de las solanáceas, poáceas, brasicáceas, fabáceas, malváceas, asteráceas o lamiáceas. Por ejemplo, la planta está seleccionada de los géneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (colza), *Medicago* (alfalfa), *Gossypium* (algodón), *Artemisia*, *Salvia* y *Mentha*. Preferentemente, la planta pertenece a la especie *Nicotiana tabacum*.

En una realización más preferida el organismo huésped no humano usado para llevar a cabo el procedimiento de la invención in vivo es un microorganismo. Se puede usar cualquier microorganismo pero de acuerdo con una realización incluso más preferida dicho microorganismo es una bacteria o levadura. Lo más preferentemente, dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae*.

Algunos de estos organismos no producen FPP de forma natural. Para ser adecuados para llevar a cabo el procedimiento de la invención, estos organismos se han transformado para producir dicho precursor. Se pueden transformar así bien antes de la modificación con el ácido nucleico descrito de acuerdo con cualesquiera de las realizaciones anteriores o bien simultáneamente, como se explica anteriormente.

Se pueden usar también células eucariotas superiores aisladas, en vez de organismos completos, como huéspedes para llevar a cabo el procedimiento de la invención in vivo. Las células eucariotas adecuadas pueden ser cualesquiera células no humanas, pero son preferentemente células vegetales o fúngicas.

En el presente documento se describe al menos un polipéptido que tiene actividad  $\beta$ -santaleno sintasa usado en cualquiera de las realizaciones descritas anteriormente o codificado por el ácido nucleico usado en cualquiera de las realizaciones descritas anteriormente que comprende una secuencia de aminoácidos al menos el 55 %, preferentemente al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 1 o 3; incluyendo un polipéptido que comprende la secuencia de aminoácidos SEC ID N.º: 1 o 3.

También se describe en el presente documento un polipéptido que consiste en una secuencia de aminoácidos al menos el 50 %, preferentemente al menos el 55 %, preferentemente al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 1 o 3; incluyendo un polipéptido que consiste en la secuencia de aminoácidos SEC ID N.º: 1 o 3.

También se describe en el presente documento al menos un polipéptido que tiene una actividad  $\beta$ -santaleno sintasa usado en cualquiera de las realizaciones descritas anteriormente o codificado por el ácido nucleico usado en cualquiera de las realizaciones descritas anteriormente que comprende una secuencia de aminoácidos que es una variante de la SEC ID N.º: 1 o 3 obtenida mediante ingeniería genética, siempre que dicha variante conserve su actividad  $\beta$ -santaleno sintasa, tal como se ha definido anteriormente y que tenga el porcentaje de identidad requerido con la SEC ID N.º: 1 o 3. En otras palabras, dicho polipéptido comprende preferentemente una secuencia de aminoácidos codificada por una secuencia de nucleótidos que se ha obtenido modificando la SEC ID N.º: 2, 4 o el complemento de las mismas. En consecuencia, se describe también al menos un polipéptido que tiene una actividad  $\beta$ -santaleno sintasa usado en cualquiera de las realizaciones descritas anteriormente o codificado por el ácido nucleico usado en cualquiera de las realizaciones descritas anteriormente que consiste en una secuencia de aminoácidos que es una variante de la SEC ID N.º: 1 o 3 obtenida mediante ingeniería genética, es decir, una secuencia de aminoácidos codificada por una secuencia de nucleótidos que se ha obtenido modificando la SEC ID N.º: 2, 4 o el complemento de las mismas.

De acuerdo con otra realización preferida, el, al menos un, polipéptido que tiene una actividad  $\beta$ -santaleno sintasa usado en cualquiera de las realizaciones descritas anteriormente o codificado por el ácido nucleico usado en cualquiera de las realizaciones descritas anteriormente es una variante de la SEC ID N.º: 1 o 3 que puede encontrarse de forma natural en otros organismos, tales como otras especies de plantas, siempre que conserve su actividad  $\beta$ -santaleno sintasa tal como se ha definido anteriormente y tenga el porcentaje requerido de identidad con la SEC ID N.º: 1 o 3.

Tal como se describe en el presente documento, el polipéptido se pretende que sea un fragmento de polipéptido o de péptido que abarque las secuencias de aminoácidos identificadas en el presente documento, así como polipéptidos truncados o variantes, siempre que conserven su actividad β-santaleno sintasa tal como se ha definido anteriormente y compartan al menos el porcentaje de identidad definido con el fragmento correspondiente de la SEC ID N.º: 1 o 3.

5 Los ejemplos de polipéptidos variantes son proteínas de origen natural que se obtienen a partir de eventos de ajuste de ARNm alternos o a partir de escisión proteolítica de los polipéptidos descritos en el presente documento. Las variaciones atribuibles a proteólisis incluyen, por ejemplo, diferencias en los extremos N-terminal o C-terminal después de la expresión en diferentes tipos de células huésped, debido a la eliminación proteolítica de uno o más aminoácidos terminales de los polipéptidos de la invención. Los polipéptidos codificados por un ácido nucleico obtenido mediante mutación natural o artificial de un ácido nucleico de la invención, tal como se describen más adelante, también están abarcados por la invención.

15 Las variantes polipeptídicas que resultan de una fusión de secuencias peptídicas adicionales en los extremos amino-terminales y carboxi-terminales se pueden usar también en los procedimientos de la invención. En particular una fusión tal puede potenciar la expresión de los polipéptidos, ser útil en la purificación de la proteína o mejorar la actividad enzimática del polipéptido en un ambiente deseado o en un sistema de expresión deseado. Tales secuencias peptídicas adicionales pueden ser péptidos señal, por ejemplo. De acuerdo con ello, la presente invención comprende procedimientos que usan polipéptidos variantes, tales como aquellos obtenidos por fusión con otros oligopéptidos o polipéptidos y/o aquellos que están unidos a péptidos señal. Los polipéptidos que resultan de una fusión con otra proteína funcional, tales como otra proteína de la ruta de la biosíntesis de terpenos, se pueden usar también ventajosamente en los procedimientos de la invención.

20 De acuerdo con otra realización, el, al menos un, polipéptido que tiene una actividad β-santaleno sintasa usada en cualquiera de las realizaciones anteriormente descritas o codificada por el ácido nucleico usado en cualquiera de las realizaciones anteriormente descritas está aislado de una planta de las especies de *Santalum*, preferentemente de *Santalum album*.

25 Una herramienta importante para llevar a cabo el procedimiento de la invención es el polipéptido en sí mismo tal como se establece en la reivindicación 1 o en la reivindicación 2.

De acuerdo con una realización preferida, el polipéptido es capaz de producir una mezcla de sesquiterpenos en la que β-santaleno representa al menos el 20 %, preferentemente al menos el 30 %, preferentemente al menos el 35 %, de los sesquiterpenos producidos.

30 También se describe en el presente documento un polipéptido que comprende una secuencia de aminoácidos al menos el 50 %, preferentemente al menos el 55 %, preferentemente al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 1 o 3. También se incluye un polipéptido que comprende la secuencia de aminoácidos SEC ID N.º: 1 o 3.

35 También se describe en el presente documento adicionalmente un polipéptido que consiste en una secuencia de aminoácidos al menos el 50 %, preferentemente al menos el 55 %, preferentemente al menos el 60 %, preferentemente al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 1 o 3. También se incluye un polipéptido que consiste en la secuencia de aminoácidos SEC ID N.º: 1 o 3.

40 El, al menos un, polipéptido que comprende una secuencia de aminoácidos que es una variante de la SEC ID N.º: 1 o 3, puede obtenerse mediante ingeniería genética o encontrarse de forma natural en plantas de *Santalum* o en otras plantas o subespecies. En otras palabras, cuando el polipéptido variante se obtiene mediante ingeniería genética, dicho polipéptido comprende una secuencia de aminoácidos codificada por una secuencia de nucleótidos que se ha obtenido modificando la SEC ID N.º: 2, 4 o el complemento de las mismas. El, al menos un, polipéptido que tiene actividad β-santaleno sintasa puede consistir en una secuencia de aminoácidos que es una variante de la SEC ID N.º: 1 o 3 obtenida mediante ingeniería genética, es decir, una secuencia de aminoácidos codificada por una secuencia de nucleótidos que se ha obtenido modificando la SEC ID N.º: 2, 4 o el complemento de la misma.

45 De acuerdo con otra realización, el polipéptido se aísla de una planta de la especie *Santalum*, preferentemente de *Santalum album*.

50 Tal como se describe en el presente documento, el polipéptido se pretende que sea un fragmento de polipéptido o de péptido que abarque las secuencias de aminoácidos identificadas en el presente documento, así como polipéptidos truncados o variantes, siempre que conserven su actividad tal como se ha definido anteriormente y compartan al menos el porcentaje de identidad definido con el fragmento correspondiente de la SEC ID N.º: 1 o 3.

55 Los ejemplos de polipéptidos variantes son proteínas de origen natural que se obtienen a partir de eventos de ajuste de ARNm o a partir de escisión proteolítica de los polipéptidos descritos en el presente documento. Las variaciones

atribuibles a proteólisis incluyen, por ejemplo, diferencias en los extremos N-terminal o C-terminal después de la expresión en diferentes tipos de células huésped, debido a la eliminación proteolítica de uno o más aminoácidos terminales de los polipéptidos de la invención. Los polipéptidos codificados por un ácido nucleico obtenido mediante mutación natural o artificial de un ácido nucleico de la invención, tal como se describen más adelante, también están abarcados por la invención.

Las variantes polipeptídicas que resultan de una fusión de secuencias de péptidos adicionales en los extremos amino-terminal y carboxi-terminal están también comprendidas por los polipéptidos de la presente invención. En particular una fusión tal puede potenciar la expresión de los polipéptidos, ser útil en la purificación de la proteína o mejorar la actividad enzimática del polipéptido en un ambiente deseado o en un sistema de expresión deseado. Tales secuencias peptídicas adicionales pueden ser péptidos señal, por ejemplo. De acuerdo con ello, la presente invención comprende variantes de los polipéptidos de la invención, tales como aquellos obtenidos por fusión con otros oligopéptidos o polipéptidos y/o aquellos que están unidos a péptidos señal. Los polipéptidos que resultan de una fusión con otra proteína funcional, tal como otra proteína de la ruta de la biosíntesis de terpenos, están también comprendidos por los polipéptidos de la invención.

Como se ha mencionado anteriormente, el ácido nucleico que codifica el polipéptido de la invención es una herramienta útil para modificar organismos huéspedes no humanos o células deseadas para usarse cuando el procedimiento se lleva a cabo in vivo.

Un ácido nucleico que codifica un polipéptido de acuerdo con cualesquiera de las realizaciones descritas anteriormente es por lo tanto también un objeto de la presente invención.

El ácido nucleico de la invención se establece en la reivindicación 3 y el ácido nucleico para su uso en un procedimiento de la invención se define en la reivindicación 5.

De acuerdo con una realización preferida, el ácido nucleico comprende una secuencia de nucleótidos al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 2 o al complemento de la misma. De acuerdo con una realización más preferida, el ácido nucleico comprende la secuencia de nucleótidos SEC ID N.º: 2, o el complemento de la misma.

De acuerdo con otra realización preferida, el ácido nucleico consiste en una secuencia de nucleótidos al menos el 65 %, preferentemente al menos el 70 %, preferentemente al menos el 75 %, preferentemente al menos el 80 %, preferentemente al menos el 85 %, preferentemente al menos el 90 %, más preferentemente al menos el 95 % e incluso más preferentemente al menos el 98 % idéntica a la SEC ID N.º: 2 o al complemento de la misma. De acuerdo con una realización incluso más preferida, el ácido nucleico consiste en la secuencia de nucleótidos SEC ID N.º: 2, o el complemento de la misma.

De acuerdo con otra realización, el ácido nucleico está aislado de una planta de las especies de *Santalum*, preferentemente de *Santalum album*.

El ácido nucleico de la invención se puede definir como que incluye polímeros de desoxirribonucleótidos o de ribonucleótidos bien en forma de cadena simple o bien en forma de cadena doble (ADN y/o ARN). Los términos "secuencia de nucleótidos" deben entenderse también como que comprenden una molécula polinucleotídica o una molécula oligonucleotídica en forma de un fragmento separado o como un componente de un ácido nucleico más grande. Los ácidos nucleicos de la invención también abarcan ciertas secuencias de nucleótidos aisladas que incluyen aquellas que están sustancialmente libres de material endógeno contaminante. El ácido nucleico de la invención puede truncarse, siempre que codifique un polipéptido comprendido por la presente invención, como se ha descrito anteriormente.

El ácido nucleico de la invención puede o bien estar presente de forma natural en plantas de la especie *Santalum* u otras especies, o bien puede obtenerse modificando la SEC ID N.º: 2, o el complemento de la misma. Preferentemente dicho ácido nucleico consiste en una secuencia de nucleótidos que se ha obtenido modificando la SEC ID N.º: 2, o el complemento de la misma.

Los ácidos nucleicos que comprenden una secuencia obtenida mediante mutación de la SEC ID N.º: 2, o el complemento de la misma están abarcados por la invención, siempre que las secuencias que comprenden compartan al menos el porcentaje de identidad definido con los fragmentos correspondientes de SEC ID N.º: 2, o el complemento de la misma y siempre que codifiquen un polipéptido que tenga actividad  $\beta$ -santaleno sintasa, tal como se ha definido en cualquiera de las reivindicaciones anteriores. Las mutaciones pueden ser cualquier clase de mutaciones de estos ácidos nucleicos, tales como mutaciones puntuales, mutaciones de delección, mutaciones de inserción y/o mutaciones de cambio de fase. Se puede preparar un ácido nucleico variante con el fin de adaptar su secuencia de nucleótidos a un sistema de expresión específico. Por ejemplo, se sabe que los sistemas de expresión bacterianos expresan más eficientemente polipéptidos si los aminoácidos se codifican por un codón preferido. Debido a la degeneración del código genético, en el que más de un codón puede codificar el mismo aminoácido, múltiples secuencias de ADN pueden codificar para el mismo polipéptido, estando todas estas secuencias de ADN comprendidas por la invención.

Otra herramienta importante para transformar organismos o células huéspedes adecuados para llevar a cabo el procedimiento de la invención in vivo es un vector de expresión que comprende un ácido nucleico de acuerdo con cualquier realización de la invención. Un vector tal es por lo tanto también un objeto de la presente invención.

5 Un "vector de expresión" como se usa en el presente documento incluye cualquier vector lineal o recombinante incluyendo pero no limitado a vectores virales, bacteriófagos y plásmidos. La persona experta es capaz de seleccionar un vector adecuado de acuerdo con el sistema de expresión. En una realización, el vector de expresión incluye el ácido nucleico de la invención operativamente unido a al menos una secuencia reguladora, que controla transcripción, traducción, iniciación y terminación, tal como un promotor regional, operador o potenciador, o un sitio de unión ribosómico de ARNm y opcionalmente, que incluye al menos un marcador de selección. Las secuencias de nucleótidos están "enlazadas operativamente" cuando la secuencia reguladora está relacionada funcionalmente con el ácido nucleico de la invención.

15 Los vectores de expresión de la presente invención se pueden usar en los procedimientos para preparar un organismo huésped genéticamente transformado y/o una célula huésped genéticamente transformada, en organismos y/o células huéspedes que albergan los ácidos nucleicos de la invención y en los procedimientos de fabricación de polipéptidos que tienen una actividad  $\beta$ -santaleno sintasa, como se describe adicionalmente más adelante.

Los organismos y células huéspedes no humanos recombinantes transformados para albergar al menos un ácido nucleico de la invención de forma que ello expresa o sobreexpresa heterológicamente al menos un polipéptido de la invención son también herramientas muy útiles para llevar a cabo el procedimiento de la invención. Tales organismos huéspedes no humanos y tales células huéspedes no humanas son por lo tanto otro objeto de la presente invención.

20 Un ácido nucleico de acuerdo con cualquiera de las realizaciones anteriormente descritas se puede usar para transformar los organismos y células huéspedes no humanos y el polipéptido expresado puede ser cualquiera de los polipéptidos anteriormente descritos.

25 Los organismos huéspedes no humanos de la invención pueden ser cualesquiera organismos pluricelulares o unicelulares no humanos. En una realización preferida, el organismo huésped no humano es una planta, un procarionta o un hongo. Cualquier planta, procarionta u hongo es adecuado para transformarse de acuerdo con la presente invención. Plantas particularmente útiles son aquellas que producen de forma natural cantidades altas de terpenos. En una realización más preferida, la planta se selecciona a partir de la familia de las solanáceas, poáceas, brasicáceas, fabáceas, malváceas, asteráceas o lamiáceas. Por ejemplo, la planta está seleccionada de los géneros *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (colza), *Medicago* (alfalfa), *Gossypium* (algodón), *Artemisia*, *Salvia* y *Mentha*. Preferentemente, la planta pertenece a la especie *Nicotiana tabacum*.

30 En una realización más preferida el organismo huésped no humano es un microorganismo. Cualquier microorganismo es adecuado para la presente invención, pero de acuerdo con una realización más preferida dicho microorganismo es una bacteria o levadura. Lo más preferentemente, dicha bacteria es *E. coli* y dicha levadura es *Saccharomyces cerevisiae*.

35 Células eucariotas superiores aisladas se pueden transformar también, en lugar de organismos completos. Como células eucariotas superiores, los autores de la presente invención quieren decir aquí cualquier célula eucariota no humana salvo células de levaduras. Células eucariotas superiores preferidas son células vegetales o células fúngicas.

40 El término "transformado" se refiere al hecho de que el huésped está sometida a ingeniería genética para que comprenda una, dos o más copias de cada uno de los ácidos nucleicos requeridos en cualquier realización de las anteriormente descritas. Preferentemente el término "transformado" se refiere a huéspedes que expresan heterológicamente los polipéptidos codificados por el ácido nucleico con el que se transforman, así como que sobre-expresan dichos polipéptidos. De acuerdo con ello, en una realización, la presente invención proporciona un organismo transformado, en el que los polipéptidos se expresan en cantidad más alta que en el mismo organismo no transformado así.

45 Hay varios procedimientos conocidos en la técnica para la creación de organismos huésped transgénicos o células tales como plantas, hongos, procariontas o cultivos de células eucariotas superiores. Los vectores de clonación y de expresión adecuados para usar con huéspedes bacterianos, fúngicos, de levaduras, de plantas y de mamíferos se describen en Pouwels y cols., Cloning Vectors: A Laboratory Manual, 1985, Elsevier, Nueva York y Sambrook y cols., Molecular Cloning: A Laboratory Manual, 2ª edición, 1989, Cold Spring Harbor Laboratory Press. Vectores de clonación y expresión para plantas superiores y/o células vegetales en particular están disponibles para la persona experta. Véase por ejemplo Schardl y cols. Gen 61: 1-11, 1987.

50 Procedimientos para transformar organismos huéspedes o células huéspedes para albergar ácidos nucleicos transgénicos son familiares para la persona experta. Para la creación de plantas transgénicas, por ejemplo, los procedimientos actuales incluyen: electroporación de protoplastos vegetales, transformación mediada por liposomas, transformación mediada por *Agrobacterium*, transformación mediada por polietilenglicol, bombardeo de partículas, microinyección de células vegetales y transformación usando virus.

En una realización, el ADN transformado está integrado dentro de un cromosoma de un organismo huésped no

humano y/o célula huésped no humana de tal forma que dé como resultado un sistema recombinante estable. Cualquier procedimiento de integración cromosómica conocido en la técnica se puede usar en la práctica de la invención, incluyendo pero no limitado a intercambio de casetes mediado por recombinasas (RMCE), inserción cromosómica específica de sitio, adenovirus e inyección pronuclear.

5 Con el fin de llevar a cabo el procedimiento de producción de  $\beta$ -santaleno in vitro, como se expone anteriormente en el presente documento, es muy ventajoso proporcionar un procedimiento de fabricación de al menos un polipéptido que tenga una actividad  $\beta$ -santaleno sintasa según se describe en cualquier realización de la invención. Por lo tanto, la invención proporciona un procedimiento de producción de al menos un polipéptido de acuerdo con cualquier realización de la invención tal como se establece en la reivindicación 16.

10 De acuerdo con una realización preferida, dicho procedimiento comprende adicionalmente, antes de la etapa a), transformar un organismo huésped no humano o una célula huésped no humana con el vector de expresión de la invención, de forma que albergue un ácido nucleico de acuerdo con la invención y exprese o sobreexpresa el polipéptido de la invención.

Se puede usar un ácido nucleico de acuerdo con cualquiera de las realizaciones descritas anteriormente.

15 La transformación y el cultivo del organismo huésped no humano o de la célula huésped no humana se puede llevar a cabo según se describe anteriormente por el procedimiento de producción de  $\beta$ -santaleno in vivo. La etapa b) se puede llevar a cabo usando cualquier técnica bien conocida en la técnica para aislar un polipéptido particular a partir de un organismo o célula.

20 Una "variante polipeptídica", tal como se denomina en el presente documento, significa un polipéptido que tiene una actividad  $\beta$ -santaleno sintasa y que es sustancialmente homóloga al polipéptido de acuerdo con cualquiera de las realizaciones anteriores, pero que tiene una secuencia de aminoácidos diferente de la codificada por cualquiera de las secuencias de ácidos nucleicos de la invención debido a una o más deleciones, inserciones o sustituciones.

25 Las variantes de polipéptido abarcadas por la invención son tal como se definen en la reivindicación 1. Las variantes pueden comprender secuencias sustituidas conservativamente, que significa que un residuo de aminoácido dado está reemplazado por un residuo que tiene características fisicoquímicas similares. Los ejemplos de sustituciones conservativas incluyen la sustitución de un residuo alifático por otro, tal como Ile, Val, Leu o Ala por otro, o sustituciones de un residuo polar por otro, tal como entre Lys y Arg; Glu y Asp; o Gln y Asn. Véase Zubay, Biochemistry, 1983, Addison-Wesley Pub. Co. Los efectos de dichas sustituciones pueden calcularse usando matrices de clasificación de sustituciones tales como PAM-120, PAM-200 y PAM-250 tal como se aborda en Altschul, J. Mol. Biol., 1991, 219, 555-565. Otras sustituciones conservativas de este tipo, por ejemplo sustituciones de regiones enteras que tienen características de hidrofobicidad similares, son bien conocidas.

35 Las variantes de péptidos de origen natural también están abarcadas por la invención. Los ejemplos de dichas variantes son proteínas que se obtienen a partir de eventos de ajuste de ARNm alternos o a partir de escisión proteolítica de los polipéptidos descritos en el presente documento. Las variaciones atribuibles a proteólisis incluyen, por ejemplo, diferencias en los extremos N-terminal o C-terminal después de la expresión en diferentes tipos de células huésped, debido a la eliminación proteolítica de uno o más aminoácidos terminales de los polipéptidos codificados por las secuencias de la invención.

40 Las variantes de los polipéptidos de la invención pueden usarse para conseguir, por ejemplo, la actividad enzimática potenciada o reducida deseada, regioquímica o estereoquímica modificada, o utilización de sustrato alterada o distribución de producto alterada, afinidad aumentada por el sustrato, especificidad mejorada para la producción de uno o más compuestos deseados, velocidad aumentada de la reacción enzimática, actividad o estabilidad superior en un ambiente específico (pH, temperatura, disolvente, etc.) o nivel de expresión mejorado en un sistema de expresión deseado. Una variante o un mutante dirigido a sitio pueden producirse mediante cualquier procedimiento conocido en la técnica. Las variantes y los derivados de polipéptidos nativos pueden obtenerse aislando variantes de origen natural, o la secuencia de nucleótidos de variantes, de otras o las mismas líneas o especies vegetales, por ejemplo plantas de la especie *Santalum*, o programando artificialmente mutaciones de secuencias de nucleótidos que codifican los polipéptidos de la invención. Las alteraciones de la secuencia de aminoácidos nativa pueden realizarse mediante cualquiera de una serie de procedimientos convencionales.

45 Las variantes polipeptídicas que resultan de una fusión de secuencias peptídicas adicionales en los extremos amino-terminales y carboxi-terminales de los polipéptidos de la invención se pueden usar para potenciar la expresión de los polipéptidos, ser útiles en la purificación de la proteína o mejorar la actividad enzimática del polipéptido en un ambiente deseado o en un sistema de expresión deseado. Tales secuencias peptídicas adicionales pueden ser péptidos señal, por ejemplo. De acuerdo con ello, la presente invención comprende variantes de los polipéptidos de la invención, tales como aquellos obtenidos por fusión con otros oligopéptidos o polipéptidos y/o aquellos que están unidos a péptidos señal. El polipéptido de fusión comprendido por la invención también comprende polipéptidos de fusión que resultan de una fusión de otras proteínas funcionales, tales como otras proteínas de la ruta de biosíntesis de terpenos.

Por lo tanto, en una realización, la presente invención proporciona un procedimiento para preparar un polipéptido

variante que tiene una actividad  $\beta$ -santaleno sintasa, tal como se establece en la reivindicación 18.

De acuerdo con una realización preferida, el polipéptido variante preparado es capaz de producir una mezcla de sesquiterpenos en la que  $\beta$ -santaleno representa al menos el 20 %, preferentemente al menos el 30 %, preferentemente al menos el 35 %, de los sesquiterpenos producidos.

5 En la etapa (b), se pueden crear un gran número de secuencias de ácidos nucleicos mutantes, por ejemplo por mutagénesis aleatorias, mutagénesis específicas de sitio, o reorganización del ADN. Los procedimientos detallados de reorganización de genes se encuentran en Stemmer, DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci EE.UU., 1994, 91(22). 10747-1075. Resumiendo, reorganización de ADN se refiere a un procedimiento de recombinación aleatoria de secuencias conocidas in vitro, que implican al menos dos ácidos nucleicos seleccionados para la recombinación. Por ejemplo se pueden introducir mutaciones en loci particulares sintetizando oligonucleótidos que contienen una secuencia mutante, flanqueada por sitios de restricción que permiten ligación a fragmentos de la secuencia nativa. Tras la ligación, la secuencia reconstruida resultante codifica un análogo que tiene la inserción, sustitución o delección aminoacídica deseada. Alternativamente, se pueden emplear procedimientos de mutagénesis específica de sitio dirigida a oligonucleótidos para proporcionar un gen alterado en los que se pueden alterar codones predeterminados por sustitución, delección o inserción.

De acuerdo con ello, el polipéptido se puede recombinar con cualesquiera otros ácidos nucleicos que codifiquen sesquiterpeno sintasa, por ejemplo aislados a partir de un microorganismo distinto de *Santalum album*. Así, se pueden obtener y separar ácidos nucleicos mutantes que se pueden usar para transformar una célula huésped de acuerdo con procedimientos estándar, por ejemplo tal como se divulga en los presentes ejemplos.

En la etapa (d), el polipéptido obtenido en la etapa (c) se criba para que tenga al menos una propiedad modificada, por ejemplo una actividad enzimática modificada deseada, pero que incluya al menos la comprobación de que posee actividad  $\beta$ -santaleno sintasa. Ejemplos de actividades enzimáticas deseadas, por las que se puede cribar un polipéptido deseado, incluyen actividad enzimática potenciada o reducida, según se mide por valor de  $K_M$  o  $V_{m\acute{a}x}$ , regioquímica o estereoquímica modificada y utilización de sustrato alterado o de distribución de producto alterada. El cribado de la actividad enzimática se puede llevar a cabo de acuerdo con procedimientos familiares para el experto y de acuerdo con aquellos procedimientos divulgados en los presentes ejemplos. La etapa (e) proporciona repetición de las etapas (a)-(d) del procedimiento, lo que se lleva a cabo preferentemente en paralelo. De acuerdo con ello, creando un número significativo de ácidos nucleicos mutantes, se pueden transformar muchas células huéspedes con diferentes ácidos nucleicos al mismo tiempo, permitiendo el cribado subsiguiente de un número elevado de polipéptidos. Las probabilidades de obtener un polipéptido variante deseado pueden incrementarse así a la discreción de la persona experta.

### Descripción de los dibujos

Figura 1: análisis de CG-EM del sesquiterpeno producido por la santaleno sintasa, recombinante de *Santalum album* (SaSantS).

Parte 1: Cromatograma iónico total. 1,  $\alpha$ -santaleno; 2, trans- $\alpha$ -bergamoteno; 3, *epi*- $\beta$ -santaleno; 4,  $\beta$ -santaleno; 5,  $\beta$ -farneseno

Partes 2 a 4: Espectros de masas de los picos identificados como sesquiterpenos.

Figura 2: Estructura molecular de  $\beta$ -santaleno y  $\beta$ -santalol.

### Ejemplos:

#### Ejemplo 1

##### Construcción de biblioteca de ADN, secuenciación y extracción de las secuencias relacionadas de terpeno sintasa

Se usaron segmentos de hipocotilos jóvenes obtenidos de semillas germinadas asépticamente de *Santalum album* L. (de 5 semanas de edad) induciendo formación de callos. Las semillas de *S. album* se obtuvieron de B&T World Seeds (Aigues-Vives, Francia) y de Sandeman Seeds (Lalongue, Francia). Las semillas se esterilizaron en superficie primero en HClO al 2,5 % durante 120 minutos y se aclararon tres veces en agua estéril ultrapura. Las semillas se descascararon después y se situaron en un medio basal de MS (Murashige & Skoog, 1962, Physiologia Plantarum 15, 473-497) suplementado con 15 g/l de sacarosa y 7,8 g/l de agar, pH 5,7. La germinación se observó típicamente después de 9 a 18 días con un rendimiento de aproximadamente el 40 %. Las plántulas se dejaron crecer in vitro durante 2 a 3 meses en una habitación de cultivo a una temperatura de 27 °C, con frío, luz fluorescente blanca y con 16 horas de fotoperiodo. Induciendo la formación de callo verde, los fragmentos de hipocotilos se cortaron en 3-4 mm segmentos transversos que se situaron en medio basal Gbg (Gamborg y cols., 1968, Exp Cell Res. 50(1), 151-158) suplementado con 2,4D (ácido 2,4-diclorofenoxiacético, Sigma-Aldrich Co.) 0,5  $\mu$ M y Kin (cinetina, Sigma-Aldrich Co.) 10  $\mu$ M en placas de Petri. El crecimiento del callo se perpetuó transfiriendo el tejido cada cuatro semanas a medio recién preparado en placas de Petri. Todos los cultivos de callos se llevaron a cabo en una cámara de crecimiento en

las mismas condiciones que anteriormente.

Los callos obtenidos después de un mes de cultivo en medio Gbg conteniendo Kin 5  $\mu\text{M}$  y ACC 2 mM se usaron para la extracción de ARN y para la construcción de bibliotecas de ADNc. El ARN total se extrajo siguiendo el protocolo descrito por Lefort y Douglas (Ann. For. Sci. 56 (1999), 259-263) salvo que se omitió el tratamiento de RNasa. El sedimento se resuspendió en 200  $\mu\text{l}$  de agua libre de RNasa y se centrifugó dos veces durante 10 minutos a 20000 g eliminando los polisacáridos. Se obtuvieron aproximadamente 125  $\mu\text{g}$  de ARN total a partir de 2,2 gramos de células. Los ARNm se purificaron usando el kit de aislamiento de ARNm FastTrack® 2.0 (Invitrogen) y se fabricó una biblioteca de ADNc usando el kit de síntesis de ADNc por PCR SMART® (Clontech Laboratories, Inc.) siguiendo las instrucciones del fabricante.

La tecnología de secuenciación en paralelo masiva de fragmentos de ADN pequeños desarrollada por Illumina (San Diego, California) se usó secuenciando la biblioteca de ADNc completa. La preparación de ADN para secuenciación, la secuenciación y el ensamblaje de las lecturas se llevaron a cabo en Fasteris SA (Plan-les-Ouates, Suiza). La biblioteca de ADNc se trató siguiendo el kit Genomic Sample Prep (Illumina) y se secuenció en el sistema de Genome Analyzer (Illumina). Se obtuvieron un total de 4,03 millones de secuencias de 35 pb (lecturas). Estas lecturas se ensamblaron usando EDENA 2.1.1, un software que halla solapamientos entre las lecturas y que ensambla contigios de novo (Hernández y cols., De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. Genome Res. 2008; 18: 802-809). El ensamblaje se hizo funcionar con apareamientos mínimos de 26 a 20 bases. Después de eliminar los contigios más cortos de 100 bases, se obtuvieron 1983 a 3473 contigios únicos con una longitud máxima de 1331 a 1914 dependiendo de los parámetros seleccionados para el ensamblaje. Se llevó a cabo otro ensamblaje usando el programa Velvet 1.0 (Zerbino y Birney (2008), Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18(5), 821-829), que proporciona 5905 contigios únicos de longitud entre 100 y 1616 bases.

Todos los contigios generados se compararon con una base de datos de secuencias proteicas (secuencias de proteínas no redundantes, NCBI, <http://www.ncbi.nlm.nih.gov>) usando el algoritmo Blastx (Altschul y cols., J. Mol. Biol. 215, 403-410, 1990; <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). Los contigios que muestran homología de secuencia significativa con sesquiterpeno sintasas de plantas se retuvieron. Se seleccionaron así un total de 46 contigios con una longitud de 100 a 621 bases. Estos contigios se procesaron después usando el programa CAP (Huang, Genomics 14(1), 18-25, 1992) ensamblándolos y generando secuencias más largas. Se ensamblaron así cinco contigios únicos de longitud de 445 a 1064. Las secuencias de aminoácidos deducidas mostraron homología significativa con sintasas de terpenos vegetales y especialmente con secuencias descritas o anotadas como monoterpeno sintasas. El alineamiento de estas secuencias de aminoácidos mostraron que al menos dos ADNc estaban presentes (se encontraron dos secuencias en la mayoría de las posiciones a través del alineamiento). Este alineamiento mostró también que al menos una secuencia N-terminal y una secuencia C-terminal estuvieron presentes. Para obtener las secuencias completas y para asignar las secuencias de extremo 5' y de extremo 3' exactas a cada ADNc, se empleó un experimento de amplificación rápida de extremos de ADNc (RACE).

## Ejemplo 2

### Amplificación de las secuencias de longitud total de un ADNc de terpeno sintasa

Para los experimentos RACE, se diseñó un conjunto de cebadores desde uno de los cinco contigios obtenidos como se describe anteriormente. Así los cebadores directos SCH5-Ct58-R1 (SEC ID N.º: 6) y SCH5-Ct58-R2 (SEC ID N.º: 7) y los cebadores inversos SCH5-Ct58-F3 (SEC ID N.º: 8) y SCH5-Ct58-F4 (SEC ID N.º: 9) se dedujeron a partir de SCH5-cóntigo-5 (SEC ID N.º: 5).

La PCR se llevó a cabo con la Mezcla de Cebador Universal A (UPM) (kit de amplificación de ADNc de RACE SMART™, Clontech Laboratories, Inc.) en volumen final de 50  $\mu\text{l}$  conteniendo mezcla de dNTP 200  $\mu\text{M}$ , 5  $\mu\text{l}$  de biblioteca de ADNc (ejemplo 1), cebador específico de gen 0,2  $\mu\text{M}$ , mezcla de cebadores UPM 0,2  $\mu\text{M}$  (Clontech Laboratories, Inc.), 1  $\mu\text{l}$  de mezcla de polimerasas Advantage 2 (Clontech Laboratories, Inc.) y 5  $\mu\text{l}$  de tampón de reacción de PCR de ADNc 10x (Clontech Laboratories, Inc.). Las condiciones de ciclación térmica fueron como sigue: 3 minutos a 94 °C; 5 ciclos de 30 segundos a 94 °C y 3 minutos a 72 °C; 5 ciclos de 30 segundos a 94 °C y 3 minutos a 70 °C; 5 ciclos de 30 segundos a 94 °C y 3 minutos a 68 °C; 3 minutos a 72 °C. Con el RACE de 5', se obtuvo un fragmento de ADN de 610 pb (SCH5-Ct58\_RR1, SEC ID N.º: 10) incluyendo el extremo 5' del ADNc. Con el RACE de 3' se obtuvo un fragmento de 1049 pb (SCH5-Ct58-RF4, SEC ID N.º: 11) y la combinación de los dos productos de RACE con la secuencia de SCH5-cóntigo-5 (SEC ID N.º: 5) permitió la reconstitución de un ADNc de longitud completa nuevo (SCH5-Ct58, SEC ID N.º: 12). El ADNc SCH5-Ct58 de 2157 pb codificaba para una proteína de 569 aminoácidos (SEC ID N.º: 13) mostrando homología con secuencias de terpeno sintasas vegetales y conteniendo restos característicos de terpeno sintasas tales como el resto DDxxD presente en todas las monoterpeno sintasas y en todas las sesquiterpeno sintasas. De forma interesante la secuencia de aminoácidos mostró similitud superior a monoterpeno sintasas que a sesquiterpeno sintasas. Sin embargo la presencia de péptido señal de cloroplasto, una característica común en monoterpeno sintasas vegetales, no se predijo a partir del análisis de la secuencia N-terminal (Emanuelsson, O., Nielsen, N. y von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8, 978-984).

**Ejemplo 3**Expresión heteróloga y actividad enzimática in vitro de SCH5-Ct58

Los presentes inventores decidieron modificar la secuencia de ADN de SCH5-Ct58 (SEC ID N.º: 12) y rediseñar la secuencia para expresión heteróloga óptima en células de *E. coli*. Para empezar con la secuencia de aminoácidos verdadera, la secuencia nucleotídica exacta de SCH5-Ct58 en la biblioteca de ADNc hubo de estabilizarse primero. El Software Eland (Illumina) se usó recuperando todas las lecturas que coinciden con la secuencia SCH5-Ct58 (SEC ID N.º: 12) con un máximo de 2 apareamientos inadecuados. Un total de 5224 lecturas se recuperaron y se alinearon usando el programa CAP (Huang, Genomics 14(1), 18-25, 1992) con la secuencia de ADN SCH5-Ct58 (SEC ID N.º: 12) como una referencia. La cobertura promedio sobre la secuencia completa estaba por encima de 100x permitiendo la deducción no ambigua de la secuencia SCH5-Ct94 de ADNc (SEC ID N.º: 14). En esta secuencia nueva se corrigieron 5 bases comparadas con la secuencia SCH5-Ct58 (SEC ID N.º: 12) deducida de los resultados de RACE y esas correcciones resultaron en una nueva secuencia de aminoácidos (SCH5-Ct94, SEC ID N.º: 15) con una diferencia de dos residuos. Para expresión heteróloga, la secuencia de ADN de SCH5-Ct94 (SEC ID N.º: 14) se modificó retirando los 23 primeros codones y reemplazando por la secuencia de ATGGCT y el uso del codón se cambió optimizando la secuencia para expresión en *E. coli* (ADN 2.0, Menlo Park, CA, EE.UU.). El ADNc así diseñado (SCH5-Ct94-opt, SEC ID N.º: 2) se sintetizó (ADN 2.0, Menlo Park, CA, EE.UU.) y se sub-clonó en los sitios *NdeI-KpnI* del plásmido pETDuet-1 proporcionando el plásmido Ct94-pETDuet. Esta secuencia de ADNc optimizada codificó para el polipéptido SCH5-Ct94-opt (SEC ID N.º: 1).

La expresión heteróloga de Ct94 se llevó a cabo en células de *E. coli* BL21(DE3) usando el plásmido Ct94-pETDuet. Los ensayos enzimáticos in vitro se llevaron a cabo con FPP como sustrato en las condiciones descritas anteriormente y se observó actividad sesquiterpeno sintasa con formación de una mezcla de cinco sesquiterpenos. La identidad de estos sesquiterpenos se confirmó por CG-EM como que son los sesquiterpenos característicos de *Santalum album*:  $\alpha$ -santaleno, trans- $\alpha$ -bergamoteno, epi- $\beta$ -santaleno,  $\beta$ -santaleno y  $\beta$ -farneseno (Figura 1). A pH 7,0 y en la presencia de  $MgCl_2$  15 mM, la proporción relativa de los productos de sesquiterpeno recombinante fue 38,0 % de  $\alpha$ -santaleno, 18,2 % de trans- $\alpha$ -bergamoteno, 5,7 % de epi- $\beta$ -santaleno, 36,7 % de  $\beta$ -santaleno y 1,3 % de  $\beta$ -farneseno. Así el ADNc SCH-Ct98-opt codifica para una  $\beta$ -santaleno sintasa. La proporción de los productos fue muy similar a la proporción observada en aceite de *Santalum album* para los productos hidroxilados de estos sesquiterpenos. No se detectó ninguna actividad cuando  $MgCl_2$  se omitió y el medio se suplementó con EDTA 2,5 mM (quelando cationes residuales) que muestran el requerimiento estricto para cationes divalentes. La naturaleza y concentración del catión divalente presente en el ensayo tuvo un efecto en el perfil del producto (tabla 1). Por ejemplo, bajar la concentración de  $Mg^{2+}$  tuvo un efecto benéfico para  $\beta$ -santaleno, el último llegando a ser el producto principal del enzima. Además, la adición de  $Mn^{2+}$  tuvo un efecto negativo sobre la formación de  $\beta$ -santaleno dado que la proporción de los productos de santaleno sesquiterpeno decreció y la proporción de trans- $\alpha$ -bergamoteno y  $\beta$ -farneseno se incrementó, siendo trans- $\alpha$ -bergamoteno el producto principal de la enzima en presencia de  $MgCl_2$  1 mM.

Tabla 1: Efecto de la concentración de iones de  $Mg^{2+}$  y  $Mn^{2+}$  en la composición de la mezcla de sesquiterpenos obtenida poniendo en contacto SEC ID N.º: 1 con FPP

	Porcentaje, relativo a la mezcla de productos completa			
	$MgCl_2$ 15 mM	$MgCl_2$ 2 mM	$MgCl_2$ 0,75 mM	$MgCl_2$ 0,75 mM + $MnCl_2$ 1 mM
$\alpha$ -santaleno	38,0	33,0	36,5	24,5
trans- $\alpha$ -bergamoteno	18,2	11,8	12,6	35,4
epi- $\beta$ -santaleno	5,7	6,4	5,6	4,1
$\beta$ -santaleno	36,7	47,5	44,1	33,3
$\beta$ -farneseno	1,3	1,3	1,1	2,75

**Ejemplo 4**Producción in vivo de sesquiterpenos en *E. coli* usando el ADNc Ct94

El uso de la santaleno sintasa de *S. album* para la producción *in vivo* de sesquiterpenos en células de *E. coli* se evaluó coexpresando las enzimas de una ruta biosintética de cinco etapas que convierte ácido mevalónico en FPP.

El gen de la FPP sintasa de levaduras se amplificó a partir del ADN genómico de *S. cerevisiae* usando los cebadores FPPy\_NcoI (SEC ID N.º: 16) y fppY-Eco (SEC ID N.º: 17). El ADN amplificado se ligó al fragmento *NdeI-EcoRI* en el

primer sitio de multiclonación (MCS1) del plásmido pACYCDuet-1 proporcionando el plásmido FPPs-pACYCDuet que alberga el gen FPPs sometido al control del promotor T7. Un operón que incluye los genes que codifican para una mevalonato cinasa (mvaK1), una fosfomevalonato cinasa (mvaK2), una mevalonato difosfato decarboxilasa (MvaD) y una isopentenil difosfato isomerasa (idi) se amplificó a partir de ADN genómico de *Streptococcus pneumoniae* (ATCC BAA-334) con los cebadores MVA-up1-start (SEC ID N.º: 18) y MVA-up2-stop (SEC ID N.º: 19). La PCR se llevó a cabo usando la ADN polimerasa PfuUltra™ II Fusion HS (Stratagene). La composición de la mezcla de PCR estaba de acuerdo con las instrucciones del fabricante. La condición de ciclación térmica fueron 2 minutos a 95 °C; 30 ciclos de 20 segundos a 95 °C, 20 segundos a 58 °C y 90 segundos a 72 °C; y 3 minutos a 72 °C. El fragmento de 3,8 Kb se purificó en un gel de agarosa y se ligó usando el kit de clonación de PRC después de secar In-Fusion™ (Clontech) dentro del segundo MCS del plásmido FPPs-pACYCDuet digerido con *NdeI* y *XhoI* proporcionando el plásmido pACYCDuet-4506. Las secuencias de los dos insertos se secuenciaron totalmente excluyendo cualquier mutación.

Se cotransformaron células de *E. coli* BL21 Star™(DE3) (Invitrogen) con los plásmidos pACYCDuet-4506 y Ct94-pETDuet y las células transformadas se seleccionaron de placas de LB-agarosa con carbenicilina (50 µg/ml) cloranfenicol (34 µg/ml). Se usaron colonias individuales inoculando 5 ml de medio LB con 50 µg/ml de carbenicilina y 34 µg/ml de cloranfenicol. El cultivo se incubó durante toda una noche a 37 °C. El día siguiente se inocularon 2 ml de medio TB suplementado con los mismos antibióticos con 0,2 ml del cultivo durante toda una noche. Después de 6 horas de incubación a 37 °C, el cultivo se enfrió hasta 28 °C y se añadieron IPTG 1 mM, 2 mg/ml de mevalonato (preparados disolviendo mevalonolactona (Sigma) en NaOH 0,5 N a una concentración de 1 g/ml e incubando la solución durante 30 minutos a 37 °C) y 0,2 ml de decano a cada tubo. Los cultivos se incubaron durante 48 horas a 28 °C. Los cultivos se extrajeron después dos veces con 2 volúmenes de acetato de etilo, la fase orgánica se concentró a 500 µl y se analizó por CG-EM como se describe anteriormente en el ejemplo 3. En estas condiciones se logró rutinariamente producción de sesquiterpenos por encima de 200 mg/l. Se produjo beta-santaleno.

### Ejemplo 5

#### Producción in vivo de sesquiterpenos en *S. cerevisiae* usando el ADNc Ct94

Para producción in vivo de sesquiterpenos en células de levadura, una cepa de *Saccharomyces cerevisiae* YNP5 en la que el gen ERG9 (que codifica para escualeno sintasa, la enzima que convierte FPP a escualeno) se ha regulado a la baja reemplazando el promotor ERG9 nativo con el promotor regulable MET3. En trabajo anterior con sintasas de sesquiterpeno en plantas, esta estrategia condujo a una biosíntesis de ergosterol reducida en las células y a una acumulación de FPP disponible por sesquiterpeno sintasas (Asadollahi, Biotechnology and Bioengineering, 99(3), 666-677, 2008).

El ADNc SCH5-Ct94-opt (SEC ID N.º: 2) se amplificó a partir del Ct94-pETDuet con los cebadores Ct94\_BamHI (SEC ID N.º: 20) y T7term (SEC ID N.º: 21). La PCR se llevó a cabo con la ADN polimerasa Pfu (Promega) usando las siguientes condiciones de ciclación térmica: 90 segundos a 94 °C; 35 ciclos de 45 segundos a 94 °C, 45 segundos a 55 °C, 4 minutos a 72 °C; y 10 minutos a 72 °C. El ADNc amplificado se digirió con los sitios de restricción de *BamHI* y de *XhoI* y se ligó en los sitios correspondientes del plásmido pESC-URA (Stratagene) proporcionando el plásmido Ct94-pESC-ura. S.c. Las células YNP5 se transformaron usando el kit de transformación EasyComp™ S.c. (Invitrogen).

Una sola colonia de cepas de levaduras transformadas se usó inoculando 20 ml de medio YNB (5 g/l de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 3 g/l de KH<sub>2</sub>PO<sub>4</sub>; 0,5 g/l de MgSO<sub>4</sub>·7 H<sub>2</sub>O; 1 ml/l de solución de metal traza) suplementado con glucosa al 2 %. El cultivo se incubó durante 24 horas a 28 °C. Las células se recuperaron por centrifugación y se resuspendieron en 20 ml de medio YNB suplementado con galactosa al 2 %. Después de cultivo de 1 hora, se añadieron al cultivo metionina a concentración final de 0,5 mM y 2 ml de decano. Después de 24 horas de incubación a 28 °C, los cultivos se extrajeron con acetato de etilo y se analizaron por CG-EM como se describe en el ejemplo 4. La cantidad total de sesquiterpenos producidos por las células de levaduras en estas condiciones se estimó en 50 mg/l.

### Ejemplo 6

#### Aislamiento de una santaleno sintasa a partir de raíces de *Santalum album*

Plantones de *Santalum album* obtenidos de semillas germinadas asépticamente se transfirieron al suelo 5 a 10 semanas después de la germinación. Dado que las especies de *Santalum* son hemiparásitos de raíces, la adaptación al suelo se hizo en estrecho contacto con plantas de *Citrus* de 6 meses a 1 año (*Citrus sinensis*). Las raíces de plantas de *Santalum* se cosecharon, 2-3 años después se transfirieron a los suelos y se separaron de las raíces de la planta huésped. El análisis de CG-EM de un extracto de estas raíces mostró la presencia de los sesquiterpenos característicos del aceite de sándalo. El ARN total se extrajo a partir de las raíces usando el reactivo de ARN de plantas Concert™ (Invitrogen). A partir de 12 g de tejido, se aislaron 640 µg de ARN total. El ARNm se purificó usando el kit de aislamiento de ARNm FastTrack® 2.0 (Invitrogen) y una biblioteca de ADNc se fabricó usando el kit de amplificación de ADNc Marathon™ (Clontech Laboratories, Inc.) siguiendo las instrucciones del fabricante.

Una cantidad de 1 µg de ADNc se usó para secuenciación usando el Sistema Analizador de Genoma (Illumina). Se obtuvieron un total de 10,3 millones de lecturas de 35 pares de bases de longitud. Estas lecturas se ensamblaron usando en paralelo los softwares ensambladores Edena (Hernández y cols., 2008, Genome Res. 18, 802-809) y Velvet

(Zerbino y Birney, 2008, Genome Res. 18: 821-829) que dan como resultado contigios únicos 18'937 y 22'414 con un promedio de 242 y 211 pb. Las lecturas se registraron usando el programa tBlastn (Altschul y cols., 1990, J. Mol. Biol. 215, 403-410) con la secuencia de aminoácidos SCH5-CT94 (SEC ID N.º: 15) como secuencia de consulta. Se seleccionaron quince contigios mostrando homología significativa de sus secuencias de aminoácidos deducidas con sesquiterpeno sintasas de plantas. Estos contigios seleccionados se reensamblaron en dos secuencias distintas, de las que SCH10-Ct8201 (SEC ID N.º: 22) fue de 383 pb en longitud y mostró la homología más alta con la secuencia de ADN SCH5-CT94 (SEC ID N.º: 14). El cebador directo SCH10-Ctg8201-F2 (SEC ID N.º: 23) se diseñó a partir de la secuencia de SCH10-Ct8201 y se usó sucesivamente para RACE de 3' usando el kit de amplificación de Marathon™ (Clontech Laboratories, Inc.). A partir de la secuencia del producto de 3'RACE así obtenida, se diseñaron dos cebadores inversos (SCH10-Ct19779-R3 (SEC ID N.º: 24) y SCH10-Ct19779-R4 (SEC ID N.º: 25)) y se usaron sucesivamente para la amplificación por RACE de 5' del extremo 5' del ADNc correspondiente. A partir de las secuencias del RACE de 5' y del RACE de 3' se reconstituyó así una secuencia de longitud total de una nueva terpeno sintasa. Con el fin de verificar la secuencia, se usó el programa MAQ (Li y cols., 2008, Genome Res. 18(11), 1851-1858) buscando y mapeando las lecturas con un máximo de 2 apareamientos inadecuados. Esta aproximación proporciona una secuencia de ADN de 1725 pares de bases de longitud (SEC ID N.º: 26) que codifica una proteína de 570 aminoácidos de longitud (SEC ID N.º: 27) que tiene una identidad del 91,9 % con la secuencia de aminoácidos de SCH5-Ct94 (SEC ID N.º: 15).

Para expresión heteróloga en *E. coli*, una ADNc optimizada se diseñó delecionando los 21 primeros codones, añadiendo la secuencia ATGGCTACC como los 3 primeros codones y optimizando el uso del codón por *E. coli*. Esta secuencia optimizada (SCH10-Ct8201-opt, SEC ID N.º: 4) que codifica para la proteína modificada en su parte N-terminal SCH10-Tps8201-opt (SEC ID N.º: 3) se sintetizó (ADN 2.0; Menlo Park, CA, EE.UU.) y se subclonó en los sitios NdeI-KpnI del plásmido de expresión pETDuet-1 (Novagen). La expresión heteróloga u la caracterización enzimática de SCH10-Tps8201-opt (SEC ID N.º: 3) se llevó a cabo como se describe en el ejemplo 3. La proteína recombinante mostró actividad sesquiterpeno sintasa y produjo a partir de FPP la misma mezcla de sesquiterpenos como la proteína recombinante SCH5-CT94-opt (SEC ID N.º: 1, ejemplo 3) con las mismas proporciones relativas.

#### Listado de secuencias

<110> Firmenich SA  
 <120> Procedimiento de producción de beta-santaleno  
 <130> P210451EP1  
 30 <141> 09/12/2009  
 <150> EP 08171298.6  
 <151> 11/12/2008  
 <150> PCT/IB2009/053623  
 <151> 17/08/2009  
 35 <160> 27  
 <170> PatentIn versión 3.5  
 <210> 1  
 <211> 548  
 <212> PRT  
 40 <213> Santalum album  
 <400> 1

ES 2 626 456 T3

Met Ala Thr Asp Asn Asp Ser Ser Glu Asn Arg Arg Met Gly Asn Tyr  
 1 5 10 15

Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln Ser Leu Ala Thr Arg  
 20 25 30

His Asn Ile Met Glu Glu Arg His Leu Lys Leu Ala Glu Lys Leu Lys  
 35 40 45

Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met Glu Pro Leu Ala Lys  
 50 55 60

Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly Leu Asn His Arg Phe  
 65 70 75 80

Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile Tyr Lys Asp Glu Ser  
 85 90 95

Asn Gly Trp Trp Phe Gly His Leu His Ala Thr Ser Leu Arg Phe Arg  
 100 105 110

Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln Asp Val Phe Lys Thr  
 115 120 125

Phe Gln Ser Lys Thr Gly Glu Phe Asp Met Lys Leu Cys Asp Asn Val  
 130 135 140

Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Phe Leu Gly Trp Arg Asp  
 145 150 155 160

Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Ala Thr Lys Tyr Leu Lys  
 165 170 175

Asn Ala Trp Glu Asn Ile Ser Gln Lys Trp Leu Ala Lys Arg Val Lys  
 180 185 190

His Ala Leu Ala Leu Pro Leu His Trp Arg Val Pro Arg Ile Glu Ala

ES 2 626 456 T3

	195					200						205			
Arg	Trp 210	Phe	Val	Glu	Ala	Tyr 215	Gly	Glu	Glu	Glu	Asn 220	Met	Asn	Pro	Thr
Leu	Leu	Lys	Leu	Ala	Lys 230	Leu	Asp	Phe	Asn	Met 235	Val	Gln	Ser	Ile	His 240
Gln	Lys	Glu	Ile	Gly 245	Glu	Leu	Ala	Arg	Trp 250	Trp	Val	Thr	Thr	Gly 255	Leu
Asp	Lys	Leu	Ala 260	Phe	Ala	Arg	Asn	Asn 265	Leu	Leu	Gln	Ser	Tyr 270	Met	Trp
Ser	Cys	Ala 275	Ile	Ala	Ser	Asp	Pro 280	Lys	Phe	Lys	Leu	Ala 285	Arg	Glu	Thr
Ile	Val 290	Glu	Ile	Gly	Ser	Val 295	Leu	Thr	Val	Val	Asp 300	Asp	Ala	Tyr	Asp
Val	Tyr	Gly	Ser	Met	Asp 310	Glu	Leu	Asp	Leu	Tyr 315	Thr	Asn	Ser	Val	Glu 320
Arg	Trp	Ser	Cys	Thr 325	Glu	Ile	Asp	Lys	Leu 330	Pro	Asn	Thr	Leu	Lys 335	Leu
Ile	Phe	Met	Ala 340	Met	Phe	Asn	Lys	Thr 345	Asn	Glu	Val	Gly	Leu	Arg	Val
Gln	His	Glu 355	Arg	Gly	Tyr	Ser	Gly 360	Ile	Thr	Thr	Phe	Ile 365	Lys	Ala	Trp
Val	Glu 370	Gln	Cys	Lys	Ser	Tyr 375	Gln	Lys	Glu	Ala	Arg 380	Trp	Tyr	His	Gly
Gly	His	Thr	Pro	Pro	Leu 390	Glu	Glu	Tyr	Ser	Leu 395	Asn	Gly	Leu	Val	Ser 400
Ile	Gly	Phe	Pro	Leu 405	Leu	Leu	Ile	Thr	Gly 410	Tyr	Val	Ala	Ile	Ala	Glu 415
Asn	Glu	Ala	Ala 420	Leu	Asp	Lys	Val	His 425	Pro	Leu	Pro	Asp	Leu	Leu	His 430
Tyr	Ser	Ser 435	Leu	Leu	Ser	Arg	Leu 440	Ile	Asn	Asp	Met	Gly 445	Thr	Ser	Ser
Asp	Glu 450	Leu	Glu	Arg	Gly	Asp 455	Asn	Leu	Lys	Ser	Ile 460	Gln	Cys	Tyr	Met
Asn	Gln	Thr	Gly	Ala	Ser	Glu	Lys	Val	Ala	Arg	Glu	His	Ile	Lys	Gly



ES 2 626 456 T3

atggctaccg ataatgacag ctctgaaaac cgctcgtatgg gtaattacaa gccgtccatc 60  
 tggaaactacg acttcctgca gtccctggct acccgccaca atatcatgga agagcgccac 120  
 ttgaaactgg cggagaaact gaaaggccag gtgaagtta tgtttggtgc cccgatggag 180  
 ccgctggcca aactggagct ggttgatggt gttcagcgcc tgggtctgaa tcatcgcttc 240  
 gagacggaga ttaaggaggc cctgttcagc atctacaagg atgagagcaa cggttggtgg 300  
 tttggccacc tgcattgccac cagcctgcgt tttcgctgc tgcgccagtg tggctgttc 360  
 attccgcaag acgttttcaa gacgttccaa agcaagaccg gcgagttcga catgaaactg 420  
 tgcgacaacg tcaagggttt gctgagcctg tacgaggctt cctttctggg ctggcgtgac 480  
 gaaaatatcc tggacgaagc gaaagctttt gccacgaagt acctgaagaa cgcattggaa 540  
 aacattagcc agaagtggct ggcgaaacgc gtgaagcatg cgttggcact gccgttgac 600  
 tggcgtgtgc ctcgtattga agcacgctgg tttgttgagg cgtacggcga ggaggaaaat 660  
 atgaatccga ccttgcctgaa gctggctaag ttggatttta acatggtgca atctattcac 720  
 caaaaggaaa tcggtgaatt ggcacgttgg tgggtcacca ccggtctgga caaactggca 780  
 ttcgcgcgca ataatttgct gcaaagctac atgtggagct gcgcatcgc atctgaccgc 840  
 aagtttaagc tggctcgcga aaccatcgtg gagatcgggt ccgctgctgac tgttggtgat 900  
 gacgcctacg atgtttacgg tagcatggac gaactggact tgtataccaa tagcgtggag 960  
 cgttggagct gtacggaaat cgataagctg ccgaatacgc tgaactgat ttttatggct 1020  
 atgtttaaca agaccaatga agttggtctg cgtgttcagc acgagcgtgg ttactccggc 1080  
 atcaccacct tcattaaggc atgggtcgaa cagtgtgaaga gctatcaaaa agaagcgcgc 1140  
 tggatcatg gtggtcacac gcctccgctg gaagagtact ccttgaatgg cttggtgagc 1200  
  
 attggtttcc cgctgctgct gattaccggc tacgtcgcca ttgccgaaaa cgaagcagcg 1260  
 ctggacaaag tgcattccgct gccggatctg ctgcactata gctctctgct gagccgcctg 1320  
 atcaacgaca tgggtacgag cagcgacgag ctggagcgcg gcgataatct gaaaagcattc 1380  
 caatgctata tgaatcagac cggcgcgagc gagaaggtgg cgcgcgagca catcaagggc 1440  
 atcattgagg agaattggaa gattctgaac gaatgttgc tgcaccaag ccaatttcaa 1500  
 gagccgttcg tgacgttcaa cctgaacagc gttcgtgggt cccatttctt ttacgagttt 1560  
 ggtgacgggt tcggtgtgac gaatagctgg accaagggtg acatgaagag cgtcctgatt 1620  
 gatccgattc cactggatga agaataatga 1650

<210> 3

<211> 551

<212> PRT

5 <213> Secuencia artificial

<220>

<223> Polipéptido codificado por la secuencia optimizada SCH10-ctg8201-opt para expresión en E. coli.

<400> 3

ES 2 626 456 T3

Met Ala Thr Leu Lys Thr Asp Thr Asp Ala Ser Glu Asn Arg Arg Met  
 1 5 10 15

Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln Ser Leu  
 20 25 30

Ala Thr His His Asn Ile Val Glu Glu Arg His Leu Lys Leu Ala Glu  
 35 40 45

Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met Glu Pro  
 50 55 60

Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly Leu Asn  
 65 70 75 80

His Leu Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile Tyr Lys  
 85 90 95

Asp Gly Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr Ser Leu  
 100 105 110

Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln Asp Val  
 115 120 125

Phe Lys Thr Phe Gln Asn Lys Thr Gly Glu Phe Asp Met Lys Leu Trp  
 130 135 140

Asp Asn Val Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Tyr Leu Gly  
 145 150 155 160

ES 2 626 456 T3

Trp Lys Gly Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Thr Thr Lys  
 165 170 175

Cys Leu Lys Ser Ala Trp Glu Asn Ile Ser Glu Lys Trp Leu Ala Lys  
 180 185 190

Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val Pro Arg  
 195 200 205

Ile Glu Ala Arg Trp Phe Ile Glu Val Tyr Glu Gln Glu Ala Asn Met  
 210 215 220

Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met Val Gln  
 225 230 235 240

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

Thr Gly Leu Asp Lys Leu Asp Phe Ala Arg Asn Asn Leu Leu Gln Ser  
 260 265 270

Tyr Met Trp Ser Cys Ala Ile Ala Ser Asp Pro Lys Phe Lys Leu Ala  
 275 280 285

Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val Asp Asp  
 290 295 300

Gly Tyr Asp Val Tyr Gly Ser Met Asp Glu Leu Asp Leu Tyr Thr Ser  
 305 310 315 320

Ser Val Glu Arg Trp Ser Cys Val Lys Ile Asp Lys Leu Pro Asn Thr  
 325 330 335

Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu Val Gly  
 340 345 350

Leu Arg Val Gln His Glu Arg Gly Tyr Asn Ser Ile Pro Thr Phe Ile  
 355 360 365

Lys Ala Trp Val Glu Gln Cys Lys Ser Tyr Gln Lys Glu Ala Arg Trp  
 370 375 380

Phe His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu Asn Gly  
 385 390 395 400

Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr Val Ala  
 405 410 415

Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu Pro Asp  
 420 425 430

ES 2 626 456 T3

Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp Ile Gly  
435 440 445

Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser Ile His  
450 455 460

Cys Tyr Met Asn Glu Thr Gly Ala Ser Glu Glu Val Ala Arg Glu His  
465 470 475 480

Ile Lys Gly Val Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln Cys Cys  
485 490 495

Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn Leu Asn  
500 505 510

Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly Phe Gly  
515 520 525

Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu Ile Asp  
530 535 540

Pro Ile Pro Leu Gly Glu Glu  
545 550

<210> 4

<211> 1656

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Secuencia optimizada para expresión en E. coli.

<400> 4

ES 2 626 456 T3

atggcaacct tgaagactga caccgacgct agcgagaatc gtcgcatggg caactataaa 60  
 ccgagcattt ggaactacga tttcctgcaa agcctggcta cccaccacaa tatcgtggag 120  
 gagcgtcacc tgaaaactggc agaaaaattg aaaggccaag tgaaatcat gttcggcgca 180  
 ccgatggaac cgctggcgaa actggagctg gtcgacgtgg tccaacgcct gggctcgaat 240  
 cacctgtttg aaaccgaaat taaagaggca ctgttcagca tctataagga cggttcgaac 300  
 ggttgggtgg tgggtcacct gcatgcaacc agcctgcgct ttcgtctgct gcgtcagtgt 360  
 ggctgtttca ttccgcagga cgtctttaa accittcaga acaaaaccgg cgagtttgac 420  
 atgaagctgt gggacaatgt gaaaggcctg ttgagcctgt atgaggcgag ctacctgggt 480  
 tggagggtg aaaacatcct ggatgaagca aaggcattta ccaccaagtg tctgaagagc 540  
 gcgtgggaaa atatctctga gaaatggtg gcgaaacgtg tgaagcacgc gctggcgctg 600  
 ccgctgcact ggcgcgttcc gcgcatcgaa gcgcgctggt ttatcgaagt ttatgaacag 660  
 gaagctaata tgaacccgac cctgctgaag ctggcgaagc tggatttcaa catggttcaa 720  
 agcattcatc aaaaggagat cggcgagctg gcccgctggt gggtgaccac gggtttgac 780  
  
 aagctggact ttgcacgtaa taatctgttg caaagctaca tgtggagctg cgctatcgca 840  
 tccgacccga aatttaagtt ggcacgtgaa accatcgttg aaattggtag cgtgctgact 900  
 gtgggtgatg acggttacga tgtttacggt agcatggacg aactggacct gtacacgctg 960  
 agcgtcgagc gctggagctg tgtcaaaatt gataagctgc cgaacacgct gaaactgatc 1020  
 ttcatgagca tgttcaacaa aaccaacgaa gtgggcctgc gcgtgcagca cgaacgtggc 1080  
 tataatagca ttccgacggt tatcaaggca tgggtggagc aatgtaaaag ctaccaaaaa 1140  
 gaggcccggt ggtttcatgg cggccatacc ccgcctctgg aggaatatag cctgaacggc 1200  
 ctgggtgtcca ttggttttcc gctgctgctg atcaccggct acgtggcaat cgcggaaaat 1260  
 gaagccgcgc tggacaaggt ccatccactg ccggacctgt tgcattatag ctctctgctg 1320  
 agccgtctga tcaatgatat cggtagcagc ccggacgaga tggctcgtgg tgacaacctg 1380  
 aaaagcatcc attgttatat gaacgagacg ggtgcgtccg aagaggctgc ccgagcat 1440  
 atcaagggcg ttattgagga gaactggaaa atcctgaatc aatgttgctt cgatcaaagc 1500  
 cagttccaag agccgttcat cacgttcaat ctgaacagcg ttcgcggtag ccacttttcc 1560  
 tacgaatttg gcgacggttt tggcgttacg gacagctgga ccaaagttga tatgaaatcc 1620  
 gttctgatcg acccgatccc gttgggtgaa gagtag 1656

<210>5

<211> 1064

<212> ADN

5 <213> Santalum album

<400> 5

ES 2 626 456 T3

cccccgcat gagagctcca ttcattgatc atactgatca tgtgaatctc agaactgata 60  
 acgattcctc agagaatcga aggatgggga attataaacc cagtatttg aactatgatt 120  
 ttttgcaatc gcttgcgact cgccacaata ttatggaaga gaggcattca aagctagctg 180  
 agaagctgaa gggccaagtg aagtttatgt ttggggcacc aatggagccg ttagcaaagc 240  
 tggagcttgt ggatgtgggtt caaaggctcg ggctaaacca ccgatttgag acagagatca 300  
 aggaagcgtc atttagtatt tataaggatg agagcaatgg atggtgggtt ggccacctcc 360  
 atgcgacatc tctccgattt aggctgctac gacagtgtgg gctttttatc ccccaggatg 420  
 tgtttaaacc atttcagagc aaaactgggtg aatttgatat gaaactgtgt gacaatgtaa 480  
 aaggattgct gagcttgat gaagcttcat tcttgggggtg gagggatgaa aacatcttag 540  
 atgaagccaa agccttcgcc accaagtact tgaaaaatgc atgggaaaac atatcccaa 600  
 agtggcttgc caaaagagtg aagcatgcac tggctttgcc tttgactgg agagtcctta 660  
 gaatcgaagc tagatgggtc gttgaggcat atggggaaga agagaatatg aacccaacac 720  
 ttctcaaaact tgcaaaattg gactttaaca tgggtcaatc aattcatcag aaagagattg 780  
 gggaaattagc gaggtgggtg gtgactacgg ggttgataa gttagcgtt gctaggaata 840  
 atttactgca aagctatatg tggagctgcg cgattgcttc cgacccaag ttcaaacttg 900  
 ctagagaac tattgttgaa atcggaagtg tactcacagt tgttgacgat gcatatgacg 960

tctatggtc aatggatgaa ctgatctct acacgaactc cgttgaaagg tggagctgta 1020

cagaaatga caagttacca aacacattaa aattgattt tatg 1064

<210> 6

<211> 29

<212> ADN

5 <213> Artificial

<220>

<223> Cebador sintetizado en base a SCH5-Cóntigo5.

<400> 6

cttcactctt ttggcaagcc acttttggg 29

10 <210> 7

<211> 30

<212> ADN

<213> Artificial

<220>

15 <223> Cebador sintetizado en base a SCH5-cóntigo5.

<400> 7

gtggcgaagg ctttgcttc atctaagatg 30

<210> 8

<211> 31  
 <212> ADN  
 <213> Artificial  
 <220>  
 5 <223> Cebador sintetizado en base a SCH5-Cóntigo5  
 <400> 8  
 gcatatgacg tctatggtc aatggatgaa c 31  
 <210> 9  
 <211> 30  
 10 <212> ADN  
 <213> Artificial  
 <220>  
 <223> Cebador sintetizado en base a SCH5-Cóntigo5  
 <400> 9  
 15 gttgaaagggt ggagctgtac agaaattgac 30  
 <210> 10  
 <211> 610  
 <212> ADN  
 <213> Santalum album  
 20 <400> 10  
 acaaaataaa tctcttggtc tgttctttgg atctcgtttt cttcccctca gctctctcac 60  
 taatggattc ttccaccgcc accgccatga gagctccatt cattgatcat actgatcatg 120  
 tgaatctcag aactgataac gattcctcag agaatcgaag gatggggaat tataaaccca 180  
 gtatttgtaa ctatgatattt ttgcaatcgc ttgcgactcg ccacaatatt atggaagaga 240  
 ggcattctaaa gctagctgag aagctgaagg gccaaagtga gtttatgttt ggggcaccaa 300  
 tggagccggt agcaaagctg gagcttgtgg atgtggttca aaggctcggg ctaaaccacc 360  
 gatttgagac agagatcaag gaagcgctat ttagtattta taaggatgag agcaatggat 420  
 ggtggtttgg ccacctccat gcgacatctc tccgatttag gctgctacga cagtgtgggc 480  
 tttttatccc ccaggatgtg tttaaaacat ttcagagcaa aactggtgaa tttgatatga 540  
 aactgtgtga caatgtaaaa ggattgctga gcttgtatga agcttcattc ttgggggtgga 600  
 gggatgaaaa 610  
 <210> 11  
 <211> 1049  
 <212> ADN  
 25 <213> Santalum album

ES 2 626 456 T3

<400> 11

caagttacca aacacattaa aattgatttt tatgtctatg tttacaaga ccaatgaagt	60
tggccttcga gtccagcatg agcgaggcta cagtggcatc actactttta tcaaagcgtg	120
ggttgaacag tgtaaactgt accagaaaga agcaagatgg taccatgggg gacacacgcc	180
tccactggaa gaatatagct tgaatggact ggtttccata ggattccctc tcttgttgat	240
cacaggctac gtggcaatcg ctgagaacga ggctgcactg gataaagtgc acccccttcc	300
tgatcttctg cactactcct ccctccttag tcgcctcatc aatgatatgg gaacctctc	360
ggacgagttg gaaaggggag ataactctgaa gtcaattcaa tgttacatga accaaactgg	420
ggcttctgag aaagttgctc gtgagcacat aaaggaata atcgaggaaa actggaaaat	480
actgaatgag tgttgcttg atcaatctca gtttcaggag cttttgtaa cattcaattt	540
gaactctgtt cgagggctc atttcttcta cgaatttggg gatggctttg gggtgacgga	600
tagctggaca aaggttgata tgaagtctgt tttgatcgat cctattcctc tcgacgagga	660
gtagaaaact caaagcttgt gcttggttta cggtaatagt gattcagtat aaatataaaa	720
atcgacgaa cttgaggaat atgtgaggca taactatttt taatgatcat gagttaaata	780
attaagaaat atctattcgg ctcatgattc ttgagtatat attattcctt atgcgttata	840
tttccatcaa ataattagtc cgctcctgta agtcgactgt aacattactc taagggtcgc	900
tattggtttt atgttatatt aagtctacta gtttgaagtg atggaataaa tgtttgtttt	960
taagggggtt atgcactatg ttctcggttg ccttttacta ataaattttt tatgaaactc	1020
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa	1049

<210> 12

<211> 2157

5 <212> ADN

<213> Santalum album

<400> 12

acaaaataaa tctcttgctc tgttctttgg atctcgtttt ctccccctca gctctctcac	60
taatggattc ttccaccgcc accgcatga gagctccatt cattgatcat actgatcatg	120

ES 2 626 456 T3

tgaatctcag aactgataac gattcctcag agaatcgaag gatggggaat tataaaccca	180
gtatttgga ctatgatttt ttgcaatcgc ttgcgactcg ccacaatatt atggaagaga	240
ggcatctaaa gctagctgag aagctgaagg gccaaagtga gtttatgttt ggggcaccaa	300
tggagccggt agcaaagctg gagcttggtg atgtggttca aaggctcggg ctaaaccacc	360
gatttgagac agagatcaag gaagcgctat ttagtattta taaggatgag agcaatggat	420
ggtggtttgg ccacctccat gcgacatctc tccgatttag gctgctacga cagtgtgggc	480
tttttatccc ccaggatgtg tttaaaacat ttcagagcaa aactggtgaa tttgatatga	540
aactgtgtga caatgtaaaa ggattgctga gcttgatga agcttcattc ttgggggtgga	600
gggatgaaaa catcttagat gaagccaaag ccttcgccac caagtacttg aaaaatgcat	660
gggaaaacat atcccaaaag tggcttgcca aaagagtga gcatgcactg gctttgcctt	720
tgcactggag agtccctaga atcgaagcta gatggttcgt tgaggcatat ggggaagaag	780
agaatatgaa cccaacactt ctcaaacttg caaaattgga ctttaacatg gtgcaatcaa	840
ttcatcagaa agagattggg gaattagcga ggtgggtgggt gactacgggg ttggataagt	900
tagcgtttgc taggaataat ttactgcaaa gctatatgtg gagctgcgcg attgcttccg	960
acccaaagt caaacttgc agagaaacta ttgttgaaat cggaaagtga ctcacagtgt	1020
ttgacgatgc atatgacgtc tatggttcaa tggatgaact tgatctctac acgaactccg	1080
ttgaaagggt gagctgtaca gaaattgaca agttaccaa cacattaana ttgattttta	1140
tgtctatgtt taacaagacc aatgaagttg gccttcgagt ccagcatgag cgaggctaca	1200
gtggcatcac tacttttacc aaagcgtggg ttgaacagtg taaatcgtac cagaaagaag	1260
caagatggta ccatggggga cacacgcctc cactggaaga atatagcttg aatggactgg	1320
ttccatagg attccctctc ttgttgatca caggctacgt ggcaatcgct gagaacgagg	1380
ctgcactgga taaagtgcac ccccttctg atcttctgca ctactcctcc ctcttagtc	1440
gcctcatcaa tgatatggga acctcttcgg acgagttgga aaggggagat aatctgaagt	1500
caattcaatg ttacatgaac caaactgggg cttctgagaa agttgctcgt gagcacataa	1560
agggaaataat cgaggaaaac tggaaaatac tgaatgagtg ttgctttgat caatctcagt	1620
ttcaggagcc ttttgaaca ttcaatttga actctgttcg agggctctcat ttcttctacg	1680
aatttgagga tggctttggg gtgacggata gctggacaaa ggttgatatg aagtctgttt	1740
tgatcgatcc tattcctctc gacgaggagt agaaaactca aagcttgtgc ttggtttacg	1800
gtaatagtga ttcagtataa atataaaaat cggacgaact tgaggaatat gtgaggcata	1860
actattttta atgatcatga gttaaataat taagaaatat ctattcggct catgattctt	1920
gagtatatat tattccttat gcgttatatt tccatcaaat aattagtcg ctctgtgaag	1980
tcgactgtaa cattactcta agggctcgtc ttggttttat gttatattaa gtctactagt	2040
ttgaagtgat ggaataaatg tttgttttta agggggttat gcactatggt ctcggttggc	2100
ttttactaat aaatttttta tgaaactcaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	2157

ES 2 626 456 T3

<210> 13

<211> 569

<212> PRT

<213> Santalum album

5 <400> 13

Met Asp Ser Ser Thr Ala Thr Ala Met Arg Ala Pro Phe Ile Asp His  
 1 5 10 15

Thr Asp His Val Asn Leu Arg Thr Asp Asn Asp Ser Ser Glu Asn Arg  
 20 25 30

Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln  
 35 40 45

Ser Leu Ala Thr Arg His Asn Ile Met Glu Glu Arg His Leu Lys Leu  
 50 55 60

Ala Glu Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met  
 65 70 75 80

Glu Pro Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly  
 85 90 95

Leu Asn His Arg Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile  
 100 105 110

Tyr Lys Asp Glu Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr  
 115 120 125

Ser Leu Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln  
 130 135 140

Asp Val Phe Lys Thr Phe Gln Ser Lys Thr Gly Glu Phe Asp Met Lys  
 145 150 155 160

Leu Cys Asp Asn Val Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Phe  
 165 170 175

Leu Gly Trp Arg Asp Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Ala  
 180 185 190

Thr Lys Tyr Leu Lys Asn Ala Trp Glu Asn Ile Ser Gln Lys Trp Leu  
 195 200 205

Ala Lys Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val  
 210 215 220

Pro Arg Ile Glu Ala Arg Trp Phe Val Glu Ala Tyr Gly Glu Glu Glu  
 225 230 235 240



ES 2 626 456 T3

Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Val Thr Phe Asn  
 515 520  
 Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly  
 530 535 540  
 Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu  
 545 550 555 560  
 Ile Asp Pro Ile Pro Leu Asp Glu Glu  
 565

<210> 14

<211> 1710

<212> ADN

5 <213> Santalum album

<400> 14

```

atggattcct ccaccgccac cgccatgaga gctccattca ttgatcatac tgatcatgtg      60
aatctcagaa ctgataacga ttcctcagag aatcgaagga tggggaatta taaaccagtg      120
at ttggaact atgatttttt gcaatcgctt gcgactcgcc acaatattat ggaagagagg      180
catctaaagc tagctgagaa gctgaagggc caagtgaagt ttatgtttgg ggcaccaatg      240
gagccgtag caaagctgga gcttgtagat gtggttcaaa ggctcgggct aaaccaccga      300
tttgagacag agatcaagga agcgtatatt agtatttata aggatgagag caatggatgg      360
tggtttggcc acctccatgc gacatctctc cgatttaggc tgctacgaca gtgtgggctt      420
tttatcccc aggatgtggt taaaacattt cagagcaaaa ctggtgaatt tgatatgaaa      480
ctgtgtgaca atgtaaaagg attgctgagc ttgtatgaag cttcattctt ggggtggagg      540
gatgaaaaca tcttagatga agccaaagcc ttcgccacca agtacttgaa aaatgcatgg      600
gaaaacatat cccaaaagtg gcttgccaaa agagtgaagc atgcaactggc tttgcctttg      660
cactggagag tccctagaat cgaagctaga tggttcgttg aggcataatgg ggaagaagag      720
aatatgaacc caacacttct caaacttgca aaattggact ttaacatggt gcaatcaatt      780
catcagaaag agattgggga attagcgagg tggtaggtga ctacggggtt ggataagtta      840
gcgtttgcta ggaataattt actgcaaagc tatatgtgga gctgcgcgat tgcttccgac      900
ccaaagttca aacttgctag agaaactatt gttgaaatcg gaagtgtact cacagttggt      960
gacgatgcat atgacgtcta tggttcaatg gatgaacttg atctctacac gaactccggt      1020
gaaaggtgga gctgtacaga aattgacaag ttaccaaaaca cattaaaatt gatttttatg      1080
gctatgttta acaagaccaa tgaagttggc cttcgagtcc agcatgagcg aggctacagc      1140
ggcatcacta cttttatcaa agcatggggt gaacagtgta aatcgtacca gaaagaagca      1200
agatggtacc atgggggaca cacgcctcca ctggaagaat atagcttgaa tggacttggt      1260
tccataggat tccctctctt gttgatcaca ggctacgtgg caatcgctga gaacgaggct      1320
gcactggata aagtgcacc ccttctctgat cttctgcact actcctccct ccttagtcgc      1380

ctcatcaatg atatgggaac ctcttcggac gagttggaag ggggagataa tctgaagtca      1440
attcaatggt acatgaacca aactggggct tctgagaaag ttgctcgtga gcacataaag      1500
ggaataatcg aggaaaactg gaaaatactg aatgagtggt gctttgatca atctcagttt      1560
caggagcctt ttgtaacatt caatttgaac tctgttcgag ggtctcattt cttctacgaa      1620
tttgagatg gctttggggg gacgaatagc tggacaaagg ttgatatgaa gtctgttttg      1680
atcgatccta ttcctctcga cgaggagtag
1710
    
```

ES 2 626 456 T3

<210> 15

<211> 569

<212> PRT

<213> Santalum album

5 <400> 15

Met Asp Ser Ser Thr Ala Thr Ala Met Arg Ala Pro Phe Ile Asp His  
 1 5 10 15

Thr Asp His Val Asn Leu Arg Thr Asp Asn Asp Ser Ser Glu Asn Arg  
 20 25 30

Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln  
 35 40 45

Ser Leu Ala Thr Arg His Asn Ile Met Glu Glu Arg His Leu Lys Leu  
 50 55 60

Ala Glu Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met  
 65 70 75 80

Glu Pro Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly  
 85 90 95

Leu Asn His Arg Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile  
 100 105 110

Tyr Lys Asp Glu Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr  
 115 120 125

Ser Leu Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln  
 130 135 140

Asp Val Phe Lys Thr Phe Gln Ser Lys Thr Gly Glu Phe Asp Met Lys  
 145 150 155 160

Leu Cys Asp Asn Val Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Phe  
 165 170 175

Leu Gly Trp Arg Asp Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Ala  
 180 185 190

ES 2 626 456 T3

Thr Lys Tyr Leu Lys Asn Ala Trp Glu Asn Ile Ser Gln Lys Trp Leu  
 195 200 205  
 Ala Lys Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val  
 210 215 220  
 Pro Arg Ile Glu Ala Arg Trp Phe Val Glu Ala Tyr Gly Glu Glu Glu  
 225 230 235 240  
 Asn Met Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met  
 245 250 255  
 Val Gln Ser Ile His Gln Lys Glu Ile Gly Glu Leu Ala Arg Trp Trp  
 260 265  
 Val Thr Thr Gly Leu Asp Lys Leu Ala Phe Ala Arg Asn Asn Leu Leu  
 275 280 285  
 Gln Ser Tyr Met Trp Ser Cys Ala Ile Ala Ser Asp Pro Lys Phe Lys  
 290 295 300  
 Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val  
 305 310 315 320  
 Asp Asp Ala Tyr Asp Val Tyr Gly Ser Met Asp Glu Leu Asp Leu Tyr  
 325 330 335  
 Thr Asn Ser Val Glu Arg Trp Ser Cys Thr Glu Ile Asp Lys Leu Pro  
 340 345 350  
 Asn Thr Leu Lys Leu Ile Phe Met Ala Met Phe Asn Lys Thr Asn Glu  
 355 360 365  
 Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Ser Gly Ile Thr Thr  
 370 375 380  
 Phe Ile Lys Ala Trp Val Glu Gln Cys Lys Ser Tyr Gln Lys Glu Ala  
 385 390 395 400  
 Arg Trp Tyr His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu  
 405 410  
 Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr  
 420 425 430  
 Val Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu  
 435 440 445  
 Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp  
 450 455 460

ES 2 626 456 T3

Met Gly Thr Ser Ser Asp Glu Leu Glu Arg Gly Asp Asn Leu Lys Ser  
465 470 475 480

Ile Gln Cys Tyr Met Asn Gln Thr Gly Ala Ser Glu Lys Val Ala Arg  
485 490

Glu His Ile Lys Gly Ile Ile Glu Glu Asn Trp Lys Ile Leu Asn Glu  
500 505 510

Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Val Thr Phe Asn  
515 520 525

Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly  
530 535 540

Phe Gly Val Thr Asn Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu  
545 550 555 560

Ile Asp Pro Ile Pro Leu Asp Glu Glu  
565

<210> 16

<211> 34

<212> ADN

5 <213> Artificial

<220>

<223> Cebador

<400> 16

ctagccatgg cttcagaaaa agaaattagg agag 34

10 <210> 17

<211> 40

<212> ADN

<213> Artificial

<220>

15 <223> Cebador

<400> 17

ccggaattcc tatttgcttc tcttgtaaac tttgtcaag 40

<210> 18

<211> 42

20 <212> ADN

<213> Artificial

<220>

## ES 2 626 456 T3

<223> Cebador  
<400> 18  
aaggagatat acatatgaca aaaaaagttg gtgtcggta gg 42  
<210> 19

5 <211> 43  
<212> ADN  
<213> Artificial  
<220>  
<223> Cebador

10 <400> 19  
cttaccaga ctcgagttac gccttttca tctgatcctt tgc 43  
<210> 20  
<211> 35  
<212> ADN

15 <213> Cebador  
<400> 20  
cccgggggat ccatggctac cgataatgac agctc 35  
<210> 21  
<211> 23

20 <212> ADN  
<213> Artificial  
<220>  
<223> Cebador  
<400> 21

25 caccgctgag caataactag cat 23  
<210> 22  
<211> 383  
<212> ADN  
<213> Santalum album

30 <400> 22

ES 2 626 456 T3

gatcaaggaa gcgctgttta gtatttaciaa ggatgggagc aatggatggt ggtttgcca 60  
 ccttcatgcg acatctctcc gatttaggct gctacgacag tgtgggcttt ttattcccca 120  
 agatgtgttt aaaacgttcc aaaacaaaac tggggaatth gatatgaaac tgtgggaciaa 180  
 cgtaaaaggc ctgctgagct tatatgaagc ttcatacttg ggatggaagg gtgaaaacat 240  
 cctagatgaa gccaaggcct tcaccaccaa gtgcttgaaa agtgcacggg aaaatatac 300  
 cgaaaagtgg ttagcaciaa gagtgaagca tgcattggct ttgcctttgc attggagagt 360  
 ccctcgaatc gaagctagat ggt 383

<210> 23

<211> 26

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Cebador

<400> 23

ccgaaaagtg gtagcaciaa agagtg 26

10 <210> 24

<211> 23

<212> ADN

<213> Secuencia artificial

<220>

15 <223> Cebador

<400> 24

cggtcggaa gcaatcgcgc agc 23

<210> 25

<211> 26

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador

<400> 25

25 ccgatttcga caatagtttc tctagc 26

<210> 26

<211> 1725

<212> ADN

<213> Santalum album

ES 2 626 456 T3

<400> 26

actaatggat tcttccaccg ccaccgccat gacagctcca ttcattgatc ctactgatca	60
tgtgaatctc aaaactgata ctgatgcctc agagaatcga aggatgggaa attataaacc	120
cagcatttgg aattatgatt ttttacaatc acttgcaact catcacaata ttgtggaaga	180
gaggcatcta aagctagctg agaagctgaa gggccaagtg aagtttatgt ttggggcacc	240
aatggagccg ttagcaaagc tggagcttgt ggatgtggtt caaaggcttg ggctaaacca	300
cctattttgag acagagatca aggaagcgct gtttagtatt tacaaggatg ggagcaatgg	360
atgggtggttt ggccaccttc atgcgacatc tctccgattt aggctgctac gacagtgtgg	420
gctttttatt ccccaagatg tgtttaaaac gttccaaaac aaaactgggg aatttgatat	480
gaaactgtgg gacaacgtaa aagggctgct gagcttatat gaagcttcat acttgggatg	540
gaagggtgaa aacatcctag atgaagccaa ggccttcacc accaagtgct tgaaaagtgc	600
atgggaaaat atatccgaaa agtgggttagc caaaagagtg aagcatgcat tggctttgcc	660
tttgcatggg agagtccctc gaatcgaagc tagatggttc attgaggtat atgagcaaga	720
agcgaatatg aaccaaacac tactcaaact cgcaaaatta gactttaata tggtgcaatc	780
aattcatcag aaagagattg ggaattagc aaggtggtggt gtgactactg gcttggataa	840
gtagactttt gctaggaata atttactgca gagctatatg tggagctgcc cgattgcttc	900
cgacccgaag ttcaaacttg ctagagaaac tattgtcgaa atcggaagtg tactcacagt	960
tgttgacgat ggatatgacg tctatggttc aatggacgaa cttgatctct acacaagctc	1020
cgttgaaagg tggagctgtg tgaaaattga caagttgcc aacacgttaa aattaatttt	1080
tatgtctatg ttcaacaaga ccaatgaggt tggcttccga gtccagcatg agcgaggcta	1140
caatagcatc cctactttta tcaaagcgtg ggttgaacag tgtaaatcat accagaaaga	1200
agcaagatgg ttccacgggg gacacacgcc tccattggaa gaatatagct tgaatggact	1260
tgtttccata ggattccctc tcttgtaat cacaggctac gtggcaatcg ctgagaacga	1320
ggctgcactg gataaagtgc acccccttcc tgatcttctg cactactcct cctccttag	1380
tcgccctcatc aatgatatag gaacgtctcc ggatgagatg gcaagaggcg ataatctgaa	1440
gtcaatccat tgttacatga acgaaactgg ggcttccgag gaagttgctc gtgagcacat	1500
aaagggagta atcgaggaga attggaaaat actgaatcag tgctgctttg atcaatctca	1560
gtttcaggag cttttataa ctttcaattt gaactctggt cgaggggtctc atttcttcta	1620
tgaatttggg gatggctttg gggtgacgga tagctggaca aaggttgata tgaagtccgt	1680
tttgatcgac cctattcctc tcggcgagga gtagtaagct cgaag	1725

<210> 27

<211> 569

5 <212> PRT

ES 2 626 456 T3

<213> Santalum album

<400> 27

Met Asp Ser Ser Thr Ala Thr Ala Met Thr Ala Pro Phe Ile Asp Pro  
 1 5 10 15  
 Thr Asp His Val Asn Leu Lys Thr Asp Thr Asp Ala Ser Glu Asn Arg  
 20 25 30  
 Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln  
 35 40 45  
 Ser Leu Ala Thr His His Asn Ile Val Glu Glu Arg His Leu Lys Leu  
 50 55 60  
 Ala Glu Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met  
 65 70 75 80  
 Glu Pro Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly  
 85 90 95  
 Leu Asn His Leu Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile  
 100 105 110  
 Tyr Lys Asp Gly Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr  
 115 120 125  
 Ser Leu Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln  
 130 135 140  
 Asp Val Phe Lys Thr Phe Gln Asn Lys Thr Gly Glu Phe Asp Met Lys  
 145 150 155 160  
 Leu Trp Asp Asn Val Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Tyr  
 165 170 175  
 Leu Gly Trp Lys Gly Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Thr  
 180 185 190

ES 2 626 456 T3

Thr Lys Cys Leu Lys Ser Ala Trp Glu Asn Ile Ser Glu Lys Trp Leu  
 195 200 205  
 Ala Lys Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val  
 210 215 220  
 Pro Arg Ile Glu Ala Arg Trp Phe Ile Glu Val Tyr Glu Gln Glu Ala  
 225 230 235 240  
 Asn Met Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met  
 245 250 255  
 Val Gln Ser Ile His Gln Lys Glu Ile Gly Glu Leu Ala Arg Trp Trp  
 260 265 270  
 Val Thr Thr Gly Leu Asp Lys Leu Asp Phe Ala Arg Asn Asn Leu Leu  
 275 280 285  
 Gln Ser Tyr Met Trp Ser Cys Pro Ile Ala Ser Asp Pro Lys Phe Lys  
 290 295 300  
 Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val  
 305 310 315 320  
 Asp Asp Gly Tyr Asp Val Tyr Gly Ser Met Asp Glu Leu Asp Leu Tyr  
 325 330 335  
 Thr Ser Ser Val Glu Arg Trp Ser Cys Val Lys Ile Asp Lys Leu Pro  
 340 345 350  
 Asn Thr Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu  
 355 360 365  
 Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Asn Ser Ile Pro Thr  
 370 375 380  
 Phe Ile Lys Ala Trp Val Glu Gln Cys Lys Ser Tyr Gln Lys Glu Ala  
 385 390 395 400  
 Arg Trp Phe His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu  
 405 410 415  
 Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr  
 420 425 430  
 Val Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu  
 435 440 445  
 Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp  
 450 455 460

ES 2 626 456 T3

Ile Gly Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser  
465 470 475 480

Ile His Cys Tyr Met Asn Glu Thr Gly Ala Ser Glu Glu Val Ala Arg  
485 490 495

Glu His Ile Lys Gly Val Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln  
500 505 510

Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn  
515 520 525

Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly  
530 535 540

Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu  
545 550 555 560

Ile Asp Pro Ile Pro Leu Gly Glu Glu  
565

## REIVINDICACIONES

1. Un polipéptido que comprende una secuencia de aminoácidos codificada por una secuencia de nucleótidos que tiene al menos el 60 % de identidad con la SEC ID N.º: 2, o un fragmento polipeptídico de la misma, que tiene una actividad β-santaleno sintasa y en el que el β-santaleno es al menos el 20 % de sesquiterpenos producidos.
- 5 2. Un polipéptido de acuerdo con la reivindicación 1, en el que la secuencia de nucleótidos tiene al menos el 90 % de identidad con la SEC ID N.º: 2.
3. Un ácido nucleico aislado que comprende un nucleótido que tiene al menos el 60 % de identidad con la SEC ID N.º: 2, o un fragmento del mismo, que codifica un polipéptido que tiene actividad β-santaleno sintasa y en el que el β-santaleno es al menos el 20 % de sesquiterpenos producidos; o un ácido nucleico complementario del mismo.
- 10 4. Un ácido nucleico aislado de acuerdo con la reivindicación 3, en el que la secuencia de nucleótidos tiene al menos el 90 % de identidad con la SEC ID N.º: 2.
5. Un procedimiento de producción de β-santaleno que comprende
- 15 a) poner en contacto FPP con un polipéptido que comprende una secuencia de aminoácidos codificada por una secuencia de nucleótidos que tiene al menos el 50 % de identidad con la SEC ID N.º: 2, o un fragmento polipeptídico de la misma, que tiene una actividad β-santaleno sintasa y en el que el β-santaleno es al menos una parte de la mezcla de sesquiterpenos producida;
- b) opcionalmente, aislar el β-santaleno producido en la etapa a).
6. Un procedimiento de acuerdo con la reivindicación 5, en el que la etapa a) comprende cultivar un organismo huésped no humano o una célula huésped no humana capaz de producir FPP y transformada para expresar el polipéptido en condiciones que conducen a la producción de β-santaleno.
- 20 7. Un procedimiento de acuerdo con la reivindicación 6, en el que el procedimiento comprende adicionalmente, antes de la etapa a), transformar un organismo huésped no humano o una célula huésped no humana capaz de producir FPP con al menos un ácido nucleico que tiene al menos el 50 % de identidad con la SEC ID N.º: 2, o un fragmento polipeptídico de la misma, de forma que el organismo exprese un polipéptido que tenga una actividad β-santaleno sintasa.
- 25 8. Un procedimiento de acuerdo con la reivindicación 6 o la reivindicación 7, en el que el organismo huésped no humano es una planta, un procarionte o un hongo.
9. Un procedimiento de acuerdo con la reivindicación 8, en el que el organismo huésped no humano es un microorganismo.
- 30 10. Un procedimiento de acuerdo con la reivindicación 9, en el que la bacteria es *E. coli* y la levadura es *Saccharomyces cerevisiae*.
11. Un vector de expresión que comprende el ácido nucleico de la reivindicación 3 o la reivindicación 4.
12. Un organismo no humano o célula huésped no humana transformado(a) para albergar al menos un ácido nucleico de acuerdo con la reivindicación 3 o la reivindicación 4, de tal forma que exprese o sobreexpresa heterológamente al menos un polipéptido de acuerdo con la reivindicación 1 o la reivindicación 2.
- 35 13. Un organismo huésped no humano de acuerdo con la reivindicación 12, en el que el organismo huésped no humano es una planta, un procarionte o un hongo.
14. Un organismo huésped no humano de acuerdo con la reivindicación 13, en el que el organismo huésped no humano es un microorganismo.
- 40 15. Un organismo huésped no humano de acuerdo con la reivindicación 14, en el que la bacteria es *E. coli* y la levadura es *Saccharomyces cerevisiae*.
16. Un procedimiento de producción de al menos un polipéptido de acuerdo con la reivindicación 1 o la reivindicación 2 que comprende
- 45 a) cultivar un organismo no humano o célula huésped no humana transformado(a) con el vector de expresión de la reivindicación 11, de forma que albergue un ácido nucleico de la reivindicación 3 o la reivindicación 4 y exprese o sobreexpresa un polipéptido de la reivindicación 1 o la reivindicación 2;
- b) aislar el polipéptido del organismo huésped no humano o de la célula huésped no humana cultivado(a) en la etapa a).

17. Un procedimiento de acuerdo con la reivindicación 16, que comprende adicionalmente, antes de la etapa a), transformar un organismo huésped no humano o célula huésped no humana con el vector de expresión de la reivindicación 11, de forma que albergue un ácido nucleico de acuerdo con la reivindicación 3 o la reivindicación 4 y exprese o sobreexpresa un polipéptido de la reivindicación 1 o la reivindicación 2.
- 5 18. Un procedimiento de preparación de un polipéptido variante que tiene una actividad  $\beta$ -santaleno sintasa que comprende las etapas de:
- (a) seleccionar un ácido nucleico que tiene al menos el 50 % de identidad con la SEC ID N.º: 2, o un fragmento de la misma;
  - (b) modificar el ácido nucleico seleccionado para obtener al menos un ácido nucleico mutante;
  - 10 (c) transformar células u organismos unicelulares huéspedes con la secuencia de ácidos nucleicos mutantes para expresar un polipéptido codificado por la secuencia de ácidos nucleicos mutantes;
  - (d) cribar el polipéptido para obtener al menos una propiedad modificada que incluye comprobar la actividad  $\beta$ -santaleno sintasa; y
  - 15 (e) opcionalmente, si el polipéptido no tiene actividad  $\beta$ -santaleno sintasa variante deseada, repetir las etapas del procedimiento (a) a (d) hasta que se obtenga un polipéptido con una actividad  $\beta$ -santaleno sintasa variante deseada;
  - (f) opcionalmente, si un polipéptido que tiene una actividad  $\beta$ -santaleno sintasa variante deseada se identificó en la etapa (d), aislar el ácido nucleico mutante correspondiente obtenido en la etapa (c).

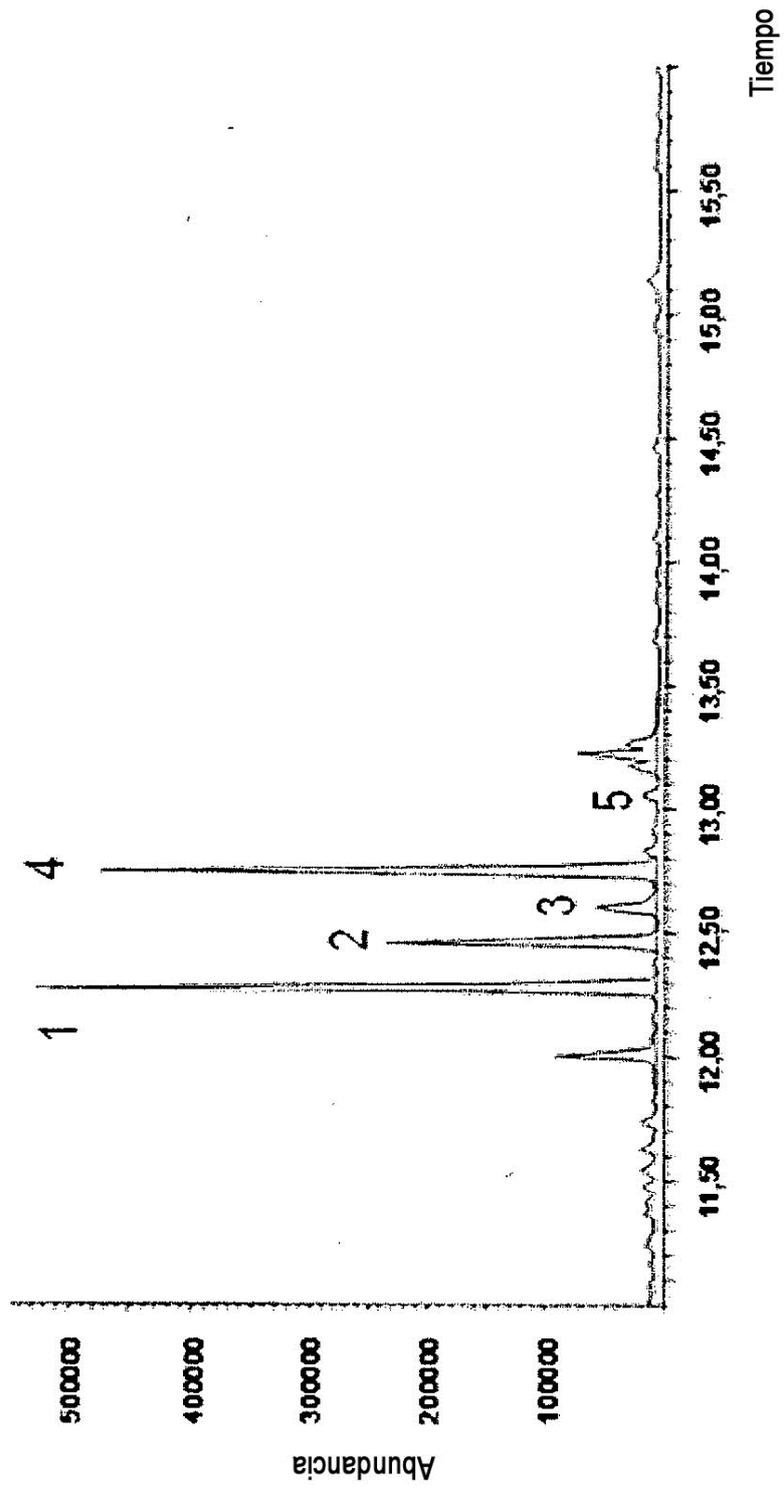


Figura 1, parte 1

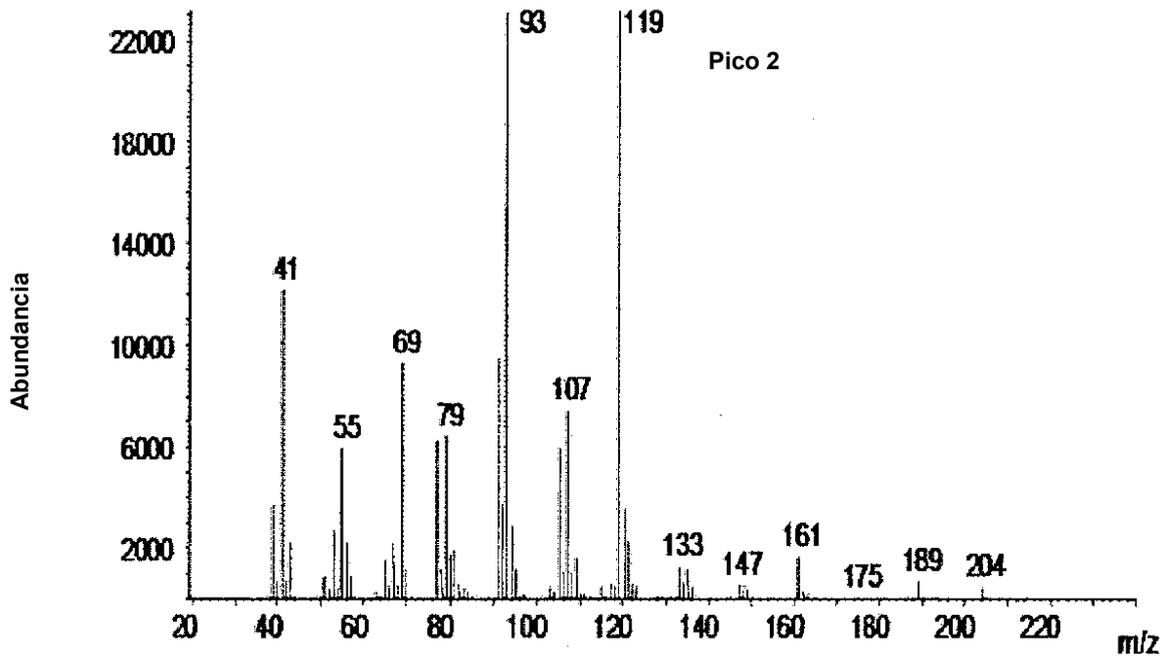
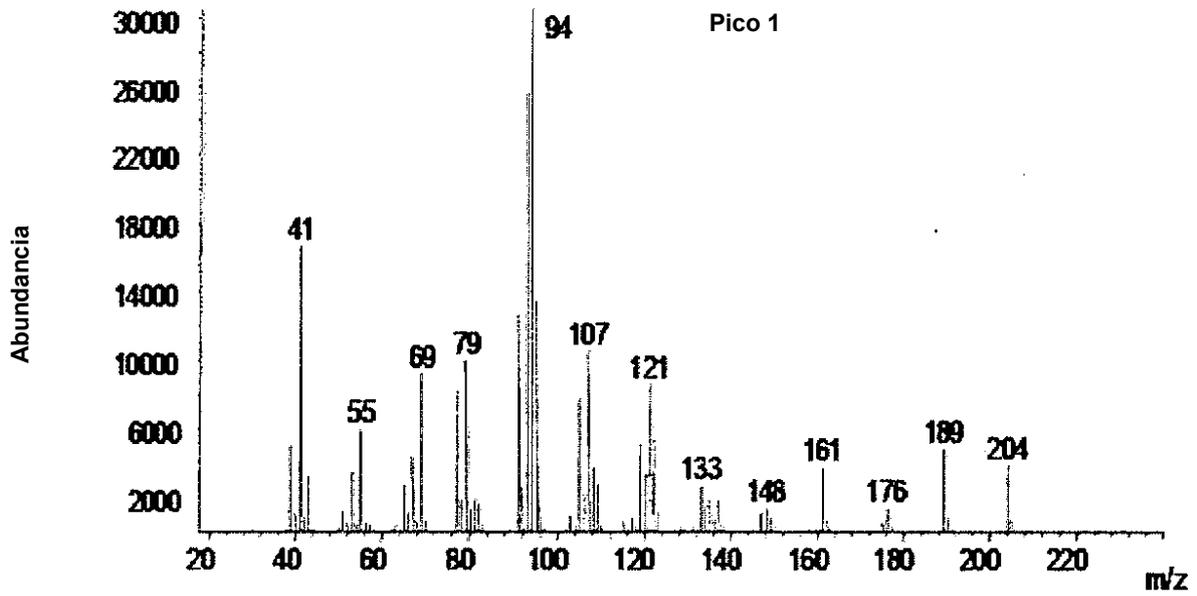


Figura 1, parte 2

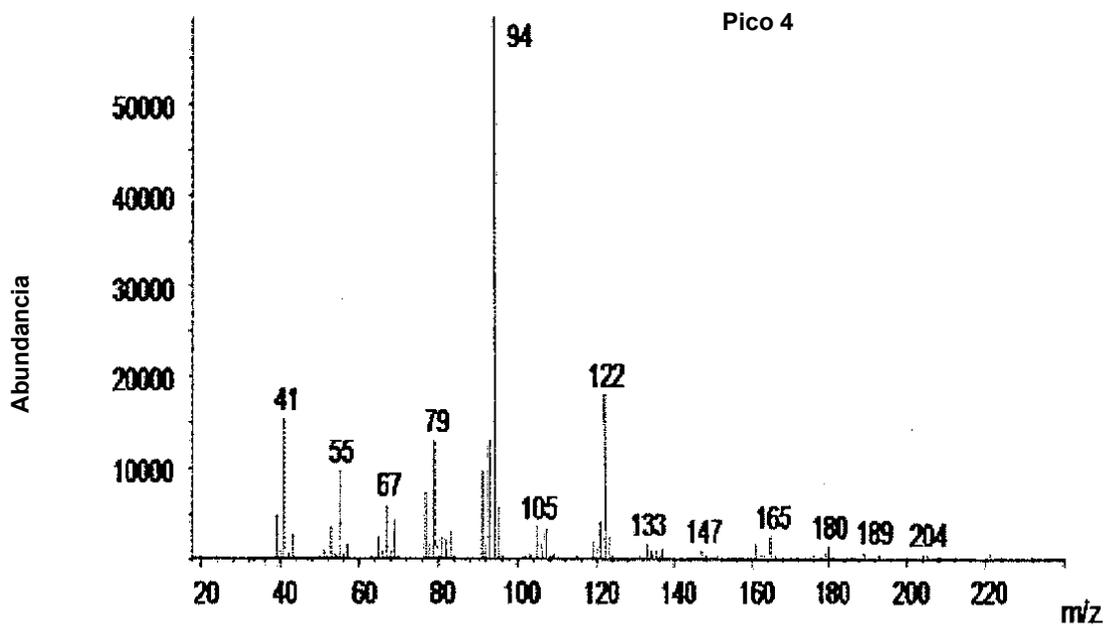
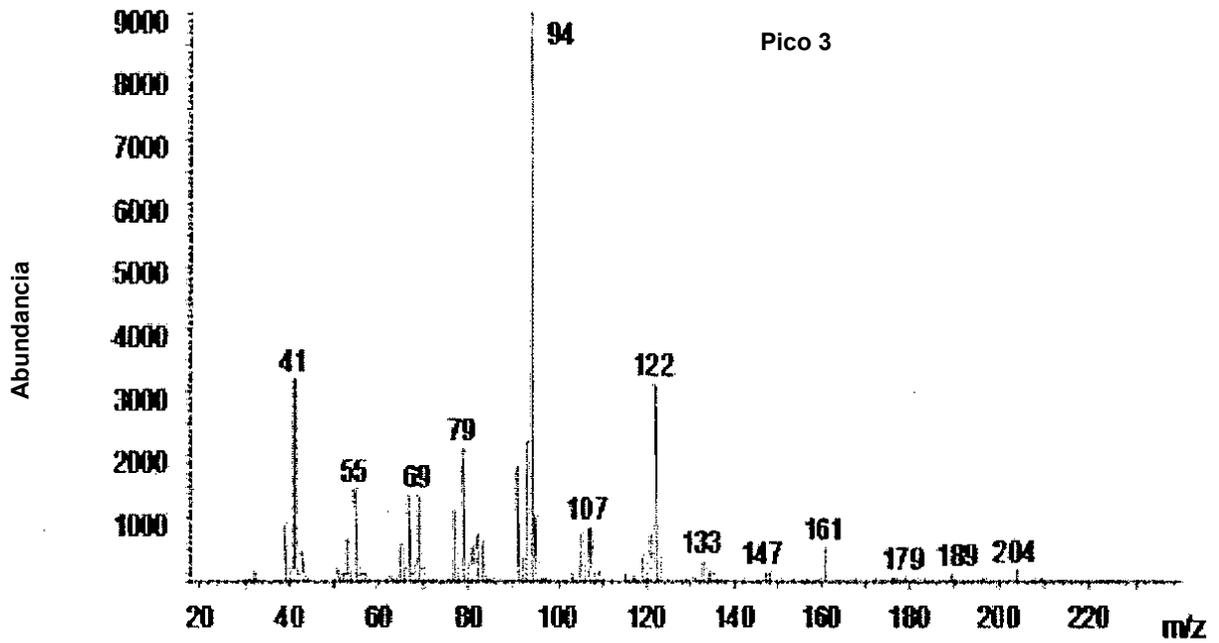


Figura 1, parte 3

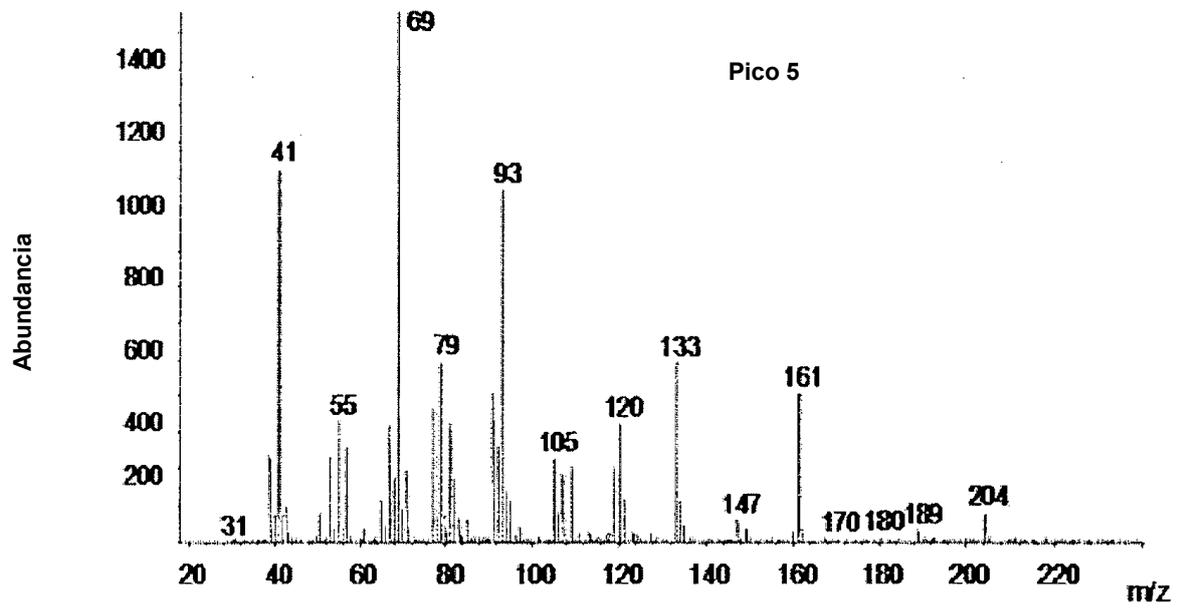
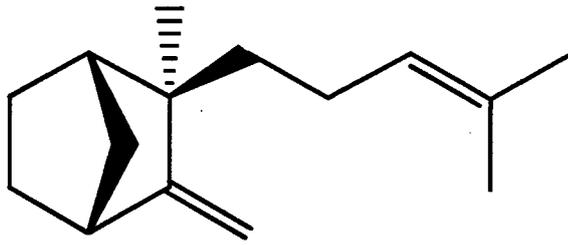
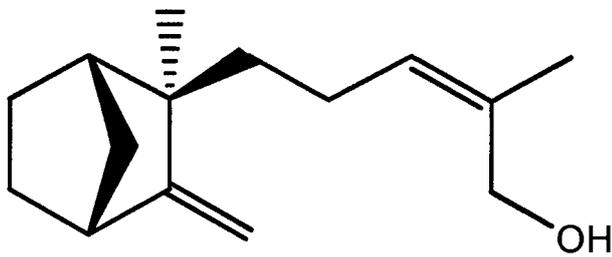


Figura 1, parte 4



Beta Santaleno

Figura 2A



Beta Santalol

Figura 2B