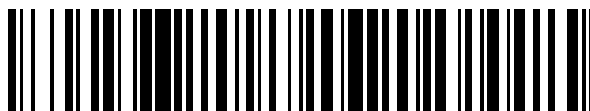


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 607 122**

51 Int. Cl.:

**C12N 9/88** (2006.01)

**C12P 7/22** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.10.2003** **E 10177152 (5)**

97 Fecha y número de publicación de la concesión europea: **14.09.2016** **EP 2365070**

54 Título: **Sesquiterpeno sintasas y procedimientos de uso**

30 Prioridad:

**04.10.2002 US 415765 P**  
**02.12.2002 WO PCT/IB02/05070**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**29.03.2017**

73 Titular/es:

**FIRMENICH SA (100.0%)**  
**P.O. Box 239, 1, route des Jeunes**  
**1211 Geneva 8, CH**

72 Inventor/es:

**SCHALK, MICHEL y**  
**CLARK, ANTHONY**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 607 122 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

Sesquiterpeno sintasas y procedimientos de uso

La presente invención se refiere a ácidos nucleicos y polipéptidos que codifican cubebol sintasa. La invención también se refiere a vectores que comprenden los ácidos nucleicos que codifican cubebol sintasa, y células huéspedes u organismos no humanos modificados para albergar los ácidos nucleicos de la invención, así como procedimientos para preparar dichas células huéspedes recombinantes. La invención también se refiere a procedimientos para producir polipéptidos de la invención. Además, la invención se refiere a procedimientos para preparar sesquiterpeno sintasa y procedimientos para preparar un terpenoide.

**Antecedentes de la invención**

Los compuestos de terpeno representan una amplia serie de moléculas naturales con una gran diversidad de estructura. El reino vegetal contiene la mayor diversidad de monoterpenos y sesquiterpenos. Con frecuencia desempeñan un papel en la defensa de las plantas contra patógenos, insectos y herbívoros y para atracción de insectos polinizantes.

La biosíntesis de terpenos se ha estudiado exhaustivamente en muchos organismos. El precursor común de los terpenos es isopentenil pirofosfato (IPP) y muchas de las enzimas que catalizan las etapas que conducen a IPP se han caracterizado. Se conocen en la actualidad dos rutas definidas para biosíntesis de IPP (Figura 1). La ruta de mevalonato se encuentra en el citosol de las plantas y en levadura y la ruta no de mevalonato (o ruta de desoxixilulosa-5-fosfato (DXP)) se encuentra en los plastidios de plantas y en *E. coli*.

Por ejemplo, el IPP se isomeriza a dimetilalil difosfato por la IPP isomerasa y estos dos compuestos C5 pueden condensarse mediante prenil transferasas para formar los precursores de pirofosfato terpeno acíclicos para cada clase de terpenos, es decir geranyl-pirofosfato (GPP) para los monoterpenos, farnesil-pirofosfato (FPP) para los sesquiterpenos, geranylgeranyl-pirofosfato (GGPP) para los diterpenos (Figura 2). Las enzimas que catalizan la etapa de ciclación de los precursores acíclicos se denominan terpeno ciclasas o terpeno sintasas, que se denominan terpeno sintasas en el presente documento.

Estas enzimas pueden ser capaces de catalizar ciclación compleja de múltiples etapas para formar el esqueleto de carbono de un compuesto de terpeno o sesquiterpeno. Por ejemplo, la etapa inicial de la ciclación catalizada puede ser la ionización del grupo difosfato para formar un catión alílico. El sustrato experimenta después isomerizaciones y reordenaciones que pueden controlarse por el sitio activo de la enzima. El producto puede ser, por ejemplo, acíclico, mono, di o tricíclico. Puede después liberarse un protón del carbocatión o el carbocatión reacciona con una molécula de agua y se libera el hidrocarburo de terpeno o alcohol. Algunas terpeno sintasas producen un único producto, pero muchas producen múltiples productos.

Se encuentran en la naturaleza una gran diversidad de estructuras de terpeno y sesquiterpeno sintasas. Se han clonado varios ADNc o genes que codifican sesquiterpeno sintasa y se han caracterizado de diferentes fuentes vegetales, por ejemplo, 5-epi-aristolóqueno sintasas de *Nicotiana tabacum* (Facchini, P.J. y Chappell, J. (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 11088-11092.) y de *Capsicum annum* (Back, K., y col. (1998) Plant Cell Physiol. 39 (9), 899-904.), una vetispiradieno sintasa de *Hyoscyamus muticus* (Back, K. y Chappell, J. (1995) J. Biol. Chem. 270 (13), 7375-7381.), una (E)- $\beta$ -farneseno sintasa de *Mentha piperita* (Crock, J., y col. (1997) Proc. Natl. Acad. Sci. U.S.A. 94 (24), 12833-12838.), una  $\delta$ -selineno sintasa y una  $\gamma$ -humuleno sintasa de *Abies grandis* (Steele, C.L., y col. (1998) J. Biol. Chem. 273 (4), 2078-2089.),  $\delta$ -cadineno sintasas de *Gossypium arboreum* (Chen, X.Y., y col. (1995) Arch. Biochem. Biophys. 324 (2), 255-266; Chen, X.Y., y col. (1996) J. Nat Prod. 59, 944-951.), una E- $\alpha$ -bisaboleno sintasa de *Abies grandis* (Bohlmann, J., y col. (1998) Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6756-6761.), una germacreno C sintasa de *Lycopersicon esculentum* (Colby, S.M., y col. (1998) Proc. Natl. Acad. Sci. U.S.A. 95 (5), 2216-2221.), una epi-cedrol sintasa y una amorfa-4,11-dieno sintasa de *Artemisia annua* (Mercke, P., y col. (1999) Arch. Biochem. Biophys. 369 (2), 213-222; Mercke, P., y col. (2000) Arch. Biochem. Biophys. 381 (2), 173-180.) y germacreno A sintasas de *Lactuca sativa*, de *Cichorium intybus* y de *Solidago canadensis* (Bennett, M.H., y col. (2002) Phytochem. 60, 255-261; Boow-meester, H.J., y col. (2002) Plant Physiol. 129 (1), 134-144; Prosser I, y col. (2002) Phytochem. 60, 691-702). Takuro y col. desvelan una sesquiterpeno sintasa de Yuzu (*Citrus junus*) (Takuro, M., y col. (2001) Biol Pharm Bull. 24 (10) 1171-1175).

Muchos compuestos de sesquiterpeno se usan en perfumería (por ejemplo pachulol, nootkatona, santalol, vetivona, sinensal) y muchos se extraen de plantas. Como resultado, su disponibilidad y sus precios pueden estar sujetos a fluctuación relacionada con la disponibilidad de las plantas y la estabilidad de los países productores. La disponibilidad de un sistema independiente de plantas para la producción de sesquiterpenos puede ser por lo tanto de interés. Debido a la complejidad estructural de algunos sesquiterpenos, su síntesis química a un coste aceptable puede no ser siempre factible.

**Sumario de la invención**

En un primer aspecto, la invención se refiere a un ácido nucleico aislado que codifica un polipéptido al menos 90 % idéntico a SEQ ID NO: 1; en el que el polipéptido codificado por dicho ácido nucleico es una cubebol sintasa.

En una realización, el ácido nucleico aislado se caracteriza porque el ácido nucleico comprende la secuencia de nucleótidos SEQ ID NO: 6.

En una realización, el ácido nucleico aislado se caracteriza porque el polipéptido codificado consiste en SEQ ID NO: 1.

- 5 En un segundo aspecto, la invención se refiere a un polipéptido de cubebol sintasa aislado codificado por el ácido nucleico.

En una realización, el polipéptido de cubebol sintasa aislado se caracteriza porque dicho polipéptido es SEQ ID NO: 1.

En un tercer aspecto, la invención se refiere a un vector que comprende el ácido nucleico.

- 10 En una realización, el vector se caracteriza porque es un vector viral o plásmido.

En un cuarto aspecto, la invención se refiere a una célula huésped o un organismo no humano modificado para albergar el ácido nucleico.

En una realización, la célula huésped u organismo no humano se caracteriza porque dicha célula huésped es una célula vegetal y dicho organismo no humano es una planta o un microorganismo.

- 15 En un quinto aspecto, la invención se refiere a un procedimiento para preparar una célula huésped recombinante que comprende introducir el vector en una célula huésped.

En un sexto aspecto, la invención se refiere a un procedimiento para producir un polipéptido que comprende cultivar la célula huésped u organismo no humano en condiciones para producir dicho polipéptido.

- 20 En una realización, el procedimiento para producir un polipéptido que comprende cultivar la célula huésped u organismo no humano en condiciones para producir dicho polipéptido se caracteriza porque la célula huésped se elige de células bacterianas, células de levadura, células vegetales y células animales.

En una realización, el procedimiento para producir un polipéptido que comprende cultivar la célula huésped u organismo no humano en condiciones para producir dicho polipéptido se caracteriza porque la célula huésped se elige de células *E. coli* y células de levadura.

- 25 En un séptimo aspecto, la invención se refiere a un procedimiento para preparar al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos una secuencia de ácido nucleico que codifica un polipéptido de cubebol sintasa al menos 90 % idéntico a SEQ ID NO: 1 en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa.

- 30 En una realización, el procedimiento para preparar al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos una secuencia de ácido nucleico que codifica un polipéptido de cubebol sintasa al menos 90 % idéntico a SEQ ID NO: 1 en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa se caracteriza porque dicho huésped es una planta o un microorganismo.

- 35 En una realización, el procedimiento para preparar al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos una secuencia de ácido nucleico que codifica un polipéptido de cubebol sintasa al menos 90 % idéntico a SEQ ID NO: 1 en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa se caracteriza porque dicho huésped se elige de células bacterianas, células de levadura, células vegetales y células animales.

En un octavo aspecto, la invención se refiere a un procedimiento para preparar al menos un terpenoide que comprende:

- 40 A) poner en contacto al menos un precursor de pirofosfato terpeno acíclico con al menos un polipéptido codificado por un ácido nucleico de acuerdo con el primer aspecto  
B) aislar al menos un terpenoide producido en A).

En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque dicho al menos un terpenoide se elige de sesquiterpenos.

- 45 En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque dicho al menos un precursor de pirofosfato terpeno acíclico es farnesil-pirofosfato.

En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque el al menos un sesquiterpeno es cubebol.

- 50 En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque el al menos un polipéptido se produce cultivando una célula huésped que comprende el ácido nucleico, en el que la célula huésped

se elige de células procariotas, células de levadura, células vegetales y células animales.

En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque dicho al menos un polipéptido se produce cultivando una célula de *E. coli* que comprende el ácido nucleico aislado.

5 En una realización, el organismo no humano se caracteriza porque dicho microorganismo es un procariota o una levadura.

En una realización, el organismo no humano se caracteriza porque dicho procariota se selecciona del grupo de *E. coli*, *Bacillus subtilis*, especie del género *Pseudomonas*, o especie del género *Streptomyces*, y en el que dicha levadura es una especie de un género seleccionado del grupo de *Saccharomyces*, *Pichia* o *Kluyveromyces*.

10 En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque dichos sesquiterpenos son cubebol y  $\alpha$ -cubebeno.

En una realización, el procedimiento para preparar al menos un terpenoide se caracteriza porque el pH en el momento del contacto es 7 o menos.

15 También se desvelan en el presente documento ácidos nucleicos aislados que codifican sesquiterpeno sintasas. Como se usa en el presente documento, una sesquiterpeno sintasa también puede referirse por al menos un compuesto producido por la enzima tras el contacto con un precursor de pirofosfato terpeno acíclico tal como farnesil-pirofosfato. Por ejemplo, una sesquiterpeno sintasa capaz de producir bicilogermacreno como uno de sus productos puede denominarse bicilogermacreno sintasa. Usando esta convención, los ejemplos de ácidos nucleicos desvelados en el presente documento incluyen ADNc que codifican cubebol sintasa (GFTpsC) (SEQ ID NO: 1);  $\delta$ -cadineno sintasa (GFTpsE) (SEQ ID NO: 2); bicilogermacreno sintasa (GFTpsB) (SEQ ID NO: 3); valenceno sintasa (GFTpsD1 y GFTpsD2) (SEQ ID NO: 4 y SEQ ID NO: 5).

20 Se desvela también un ácido nucleico aislado seleccionado de: (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa. En una realización, las condiciones definidas son condiciones de rigurosidad moderada y en una realización adicional condiciones de alta rigurosidad. También se desvela en el presente documento: un polipéptido codificado por un ácido nucleico desvelado en el presente documento; una célula huésped que comprende un ácido nucleico desvelado en el presente documento; un organismo no humano modificado para albergar un ácido nucleico desvelado en el presente documento; y procedimientos para producir un polipéptido que comprenden cultivar células huéspedes desveladas en el presente documento.

25 También se desvela en el presente documento un polipéptido aislado que comprende una secuencia de aminoácidos sustancialmente como se expone en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5.

35 También se desvela en el presente documento un vector que comprende al menos un ácido nucleico elegido de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa. También se desvelan en el presente documento procedimientos para preparar una célula huésped recombinante que comprenden introducir un vector desvelado en el presente documento en una célula huésped.

40 También se desvela en el presente documento un procedimiento para preparar al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos una secuencia de ácido nucleico en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa. En una realización, el al menos un ácido nucleico se selecciona de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa. El huésped puede seleccionarse de, por ejemplo, plantas, microorganismos, células bacterianas, células de levadura, células vegetales y células animales.

45 También se desvela en el presente documento un procedimiento para preparar al menos un terpenoide que comprende 1) poner en contacto al menos un precursor de pirofosfato terpeno acíclico con al menos un polipéptido codificado por un ácido nucleico. En una realización, el ácido nucleico se selecciona de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto

en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa, 2) aislar al menos un terpenoide producido en (1). En una realización, el al menos un terpenoide se elige de sesquiterpenos. En una realización adicional, el al menos un precursor de pirofosfato terpeno acíclico es farnesil-pirofosfato. Los sesquiterpenos producidos por los procedimientos desvelados en el presente documento incluyen, pero sin limitación, biciclogermacreno, cubebol, valenceno,  $\alpha$ -cubebeno, germacreno D, y  $\delta$ -cadineno (Figura 3).

Debe entenderse que tanto la descripción general anterior como la descripción detallada a continuación son ejemplares y explicativas. Se hará ahora referencia en detalle a realizaciones ejemplares de la presente invención.

## 10 **Breve descripción de las figuras**

Figura 1: Rutas ejemplares para la biosíntesis de isopentenil pirofosfato (ruta de mevalonato (A) y ruta de desoxixilulosa (B)).

Figura 2: Biosíntesis de terpenos ejemplares de isopentenil difosfato.

Figura 3: Estructura de compuestos de sesquiterpeno ejemplares.

15 Figura 4: Parte central de los alineamientos de las secuencias de aminoácidos de dos grupos de sesquiterpeno sintasa (germacreno C *L. esculentum*  $\delta$ -cadineno sintasa (SEQ ID NO: 11); (E)-beta-farneseno *M. piperita* (SEQ ID NO: 12); delta-selineno *A. grandis* (SEQ ID NO: 13); sesquiterpeno sintasa *C. junos* (SEQ ID NO: 14); 5-epi-aristoloqueno *N. tabacum* (SEQ ID NO: 15); 5-epi-aristoloqueno *C. annuum* (SEQ ID NO: 16); vetispiradieno *S. tuberosum* (SEQ ID NO: 17); vetispiradieno *H. muticus* (SEQ ID NO: 18); delta-cadineno *G. arboreum* (SEQ ID NO: 19); amorfa-4,11-dieno *A. annua* (SEQ ID NO: 20); epi-cedrol *A. annua* (SEQ ID NO: 21); delta-humuleno *A. grandis* (SEQ ID NO: 22)) y la secuencia de los cebadores degradados deducidos (TpsVFI (SEQ ID NO: 23); TpsVF2 (SEQ ID NO: 24); TpsVR3 (SEQ ID NO: 25); TpsCF1 (SEQ ID NO: 26); TpsCF2 (SEQ ID NO: 27); TpsCR3 (SEQ ID NO: 28)). Las flechas debajo de cada alineamiento muestran las regiones del alineamiento usado para diseñar los cebadores degradados y su orientación.

20 Figura 5: Alineamiento de las secuencias de aminoácidos deducidas a partir de los productos de amplificación obtenidos mediante RT-PCR en ARN total de pomelo (GFTpsA (SEQ ID NO: 29); GFTpsB (SEQ ID NO: 30); y GFTpsC (SEQ ID NO: 31)).

Figura 6: Alineamiento de secuencias de aminoácidos de las sesquiterpeno sintasas GFTpsA (clon parcial) (SEQ ID NO: 32), GFTpsB (SEQ ID NO: 3), GFTpsC (SEQ ID NO: 1), GFTpsD1 (SEQ ID NO: 4), GFTpsD2 (SEQ ID NO: 5), y GFTpsE (SEQ ID NO: 2) de *C. paradisi*.

Figura 7: Perfiles de GC de sesquiterpenos producidos mediante sesquiterpeno sintasas de pomelo recombinantes.

Figura 8: Las secuencias de aminoácidos y nucleótidos de (a) GFTpsA (SEQ ID NO: 32 y SEQ ID NO: 33), (b) GFTpsB (SEQ ID NO: 3 y SEQ ID NO: 8), (c) GFTpsC (SEQ ID NO: 1 y SEQ ID NO: 6), (d) GFTpsD1 (SEQ ID NO: 4 y SEQ ID NO: 9), (e) GFTpsD2 (SEQ ID NO: 5 y SEQ ID NO: 10), y (f) GFTpsE (SEQ ID NO: 2 y SEQ ID NO: 7).

## **Descripción de la invención**

Un terpeno es un hidrocarburo insaturado basado en una unidad de isopreno (C<sub>5</sub>H<sub>8</sub>) que puede ser acíclico o cíclico. Los derivados de terpeno incluyen pero sin limitación alcanfor, mentol, terpineol, y borneol, geraniol. Los terpenos o terpenoides, como se usa en el presente documento incluyen terpenos y derivados de terpenos, incluyendo compuestos que han experimentado una o más etapas de funcionalización tales como hidroxilaciones, isomerizaciones, óxido-reducciones o acilaciones. Como se usa en el presente documento, un sesquiterpeno es un terpeno basado en una estructura C<sub>15</sub> e incluye sesquiterpenos y derivados de sesquiterpenos, incluyendo compuestos que han experimentado una o más etapas de funcionalización tales como hidroxilaciones, isomerizaciones, óxido-reducciones o acilaciones.

Como se usa en el presente documento, un derivado es cualquier compuesto obtenido de un compuesto conocido o hipotético y que contiene elementos esenciales de la sustancia parental.

Como se usa en el presente documento, sesquiterpeno sintasa es cualquier enzima que cataliza la síntesis de un sesquiterpeno.

50 La expresión "idéntica", "sustancialmente idéntica" o "sustancialmente como se expone", significa que una secuencia relevante es al menos 70 %, 75 %, 80 %, 85 %, 90 %, 92 %, 95 %, 96 %, 97 %, 98 % o 99 % idéntica a una secuencia dada. Como ejemplo, dichas secuencias pueden ser variantes alélicas, secuencias derivadas de diversas especies, o pueden derivar de la secuencia dada por truncamiento, supresión, sustitución o adición de aminoácidos. Para polipéptidos, la longitud de secuencias de comparación será generalmente de al menos 20, 30, 50, 100 o más aminoácidos. Para ácidos nucleicos, la longitud de secuencias de comparación será generalmente de al menos 50, 100, 150, 300, o más nucleótidos. El porcentaje de identidad entre dos secuencias se determina mediante algoritmos de alineamiento convencionales tales como, por ejemplo, herramienta de alineamiento local básica (BLAST) descrita en Altschul y col. (1990) *J. Mol. Biol.*, 215: 403-410, el algoritmo de Needleman y col. (1970) *J. Mol. Biol.*, 48: 444-453, o el algoritmo de Meyers y col. (1988) *Comput. Appl. Biosci.*, 4: 11-17.

También se desvela en el presente documento un ácido nucleico aislado seleccionado de: (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa. En una realización, las condiciones definidas son condiciones de rigurosidad moderada y en una realización adicional condiciones de alta rigurosidad.

Como se usa en el presente documento, se determina si un polipéptido codificado por un ácido nucleico desvelado en el presente documento es una sesquiterpeno sintasa por el ensayo de caracterización enzimática descrito en los ejemplos en el presente documento.

Como se usa en el presente documento, se pretende que el término hibridación o hibrida en ciertas condiciones describa condiciones para hibridación y lavado en las que las secuencias de nucleótidos que son significativamente idénticas u homologas entre sí permanezcan unidas entre sí. Las condiciones pueden ser tales que las secuencias, que son al menos aproximadamente 70 %, tal como al menos aproximadamente 80 % y tal como al menos aproximadamente 85-90 % idénticas, permanezcan unidas entre sí. Se proporcionan en el presente documento definiciones de condiciones de hibridación de baja rigurosidad, rigurosidad moderada y alta rigurosidad.

Las condiciones de hibridación apropiadas pueden seleccionarse por los expertos en la materia con mínima experimentación como se ejemplifica en Ausubel y col. (1995), *Current Protocols in Molecular Biology*, John Wiley & Sons, secciones 2, 4 y 6. Adicionalmente, las condiciones de rigurosidad se describen en Sambrook y col. (1989) *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Press, capítulos 7, 9 y 11. Como se usa en el presente documento, las condiciones definidas de baja rigurosidad son las siguientes. Se pretratan filtros que contienen ADN durante 6 h a 40 °C en una solución que contiene formamida al 35 %, SSC 5x, Tris-HCl 50 mM (pH 7,5), EDTA 5 mM, PVP 0,1 %, Ficoll 0,1 %, BSA 1 %, y ADN de esperma de salmón desnaturalizado 500 µg/ml. Se llevan a cabo hibridaciones en la misma solución con las siguientes modificaciones: PVP 0,02 %, Ficoll 0,02 %, BSA 0,2 %, ADN de esperma de salmón 100 µg/ml, sulfato de dextrano 10 % (p/vol), y se usan 5-20x10<sup>6</sup> sondas marcadas con 32P. Se incuban filtros en mezcla de hibridación durante 18-20 h a 40 °C, y después se lavan durante 1,5 h a 55 °C en una solución que contiene SSC 2x, Tris-HCl 25 mM (pH 7,4), EDTA 5 mM, y SDS 0,1 %. La solución de lavado se reemplaza con solución nueva y se incuba durante 1,5 h adicionales a 60 °C. Los filtros se secan con material absorbente y se exponen para autorradiografía.

Como se usa en el presente documento, las condiciones definidas de rigurosidad moderada son las siguientes. Se pretratan filtros que contienen ADN durante 7 h a 50 °C en una solución que contiene formamida al 35 %, SSC 5x, Tris-HCl 50 mM (pH 7,5), EDTA 5 mM, PVP 0,1 %, Ficoll 0,1 %, BSA 1 %, y ADN de esperma de salmón desnaturalizado 500 µg/ml. Se llevan a cabo hibridaciones en la misma solución con las siguientes modificaciones: PVP 0,02 %, Ficoll 0,02 %, BSA 0,2 %, ADN de esperma de salmón 100 µg/ml, sulfato de dextrano 10 % (p/vol), y se usan 5-20x10<sup>6</sup> sondas marcadas con 32P. Los filtros se incuban en mezcla de hibridación durante 30 h a 50 °C, y después se lavan durante 1,5 h a 55 °C en una solución que contiene SSC 2x, Tris-HCl 25 mM (pH 7,4), EDTA 5 mM, y SDS 0,1 %. La solución de lavado se reemplaza con solución nueva y se incuba durante 1,5 h adicionales a 60 °C. Los filtros se secan mediante material absorbente y se exponen para autorradiografía.

Como se usan en el presente documento, las condiciones definidas de alta rigurosidad son las siguientes. Se lleva a cabo prehibridación de filtros que contienen ADN durante 8 h hasta durante una noche a 65 °C en tampón compuesto de SSC 6x, Tris-HCl 50 mM (pH 7,5), EDTA 1 mM, PVP 0,02 %, Ficoll 0,02 %, BSA 0,02 % y ADN de esperma de salmón desnaturalizado 500 µg/ml. Los filtros se hibridan durante 48 h a 65 °C en la mezcla de prehibridación que contiene ADN de esperma de salmón desnaturalizado 100 µg/ml y and 5-20x10<sup>6</sup> cpm de sonda marcada con 32P. Se realiza lavado de los filtros a 37 °C durante 1 h en una solución que contiene SSC 2x, PVP 0,01 %, Ficoll 0,01 % y BSA 0,01 %. Esto se sigue de un lavado en SSC 0,1x a 50 °C durante 45 minutos.

Pueden usarse otras condiciones de rigurosidad baja, moderada y alta bien conocidas en la técnica (por ejemplo, como se emplea para hibridaciones entre especies) si las condiciones anteriores son inapropiadas (por ejemplo, como se emplea para hibridaciones entre especies).

En una realización, un ácido nucleico y/o polipéptido desvelado en el presente documento se aísla de un cítrico, tal como, por ejemplo, un pomelo o una naranja. En una realización particular, la divulgación se refiere a ciertas secuencias de nucleótidos aisladas incluyendo las que están sustancialmente libres de material endógeno contaminante. Las expresiones "ácido nucleico" o "molécula de ácido nucleico" se refieren a un polímero desoxirribonucleotídico o ribonucleotídico en forma mono o bicatenaria, y a no ser que se limite de otro modo, abarcarían análogos conocidos de nucleótidos naturales que pueden actuar de una manera similar a nucleótidos de origen natural. Una "secuencia de nucleótidos" también se refiere a una molécula polinucleotídica o molécula oligonucleotídica en forma de un fragmento separado o como un componente de un ácido nucleico mayor. La secuencia de nucleótidos o molécula también puede denominarse "sonda de nucleótidos". Algunas de las moléculas de ácido nucleico desveladas en el presente documento derivan de ADN o ARN aislado al menos una vez en forma sustancialmente pura y en una cantidad o concentración que permite la identificación, manipulación y recuperación de su secuencia de nucleótidos componente mediante procedimientos bioquímicos convencionales. Se desvelan

ejemplos de dichos procedimientos, incluyendo procedimientos para protocolos de PCR que pueden usarse en el presente documento, en Sambrook y col., *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989), *Current Protocols in Molecular Biology* editado por F.A. Ausubel y col., John Wiley y Sons, Inc. (1987), e Innis, M. y col., eds., *PCR Protocols: A Guide to Methods and Applications*, Academic Press (1990).

Como se describe en el presente documento, las moléculas de ácido nucleico desveladas en el presente documento incluyen ADN en forma tanto monocatenaria como bicatenaria, así como el complemento de ARN del mismo. El ADN incluye, por ejemplo, ADNc, ADN genómico, ADN sintetizado químicamente, ADN amplificado por PCR y combinaciones de los mismos. El ADN genómico, incluyendo regiones traducidas, no traducidas y de control, puede aislarse mediante técnicas convencionales, por ejemplo, usando uno cualquiera de los ADNc desvelados en el presente documento, o fragmentos adecuados de los mismos, como una sonda, para identificar un trozo de ADN genómico que puede después clonarse usando procedimientos conocidos habitualmente en la técnica. En general, las moléculas de ácido nucleico desveladas en el presente documento incluyen secuencias que hibridan con secuencias desveladas en el presente documento en condiciones de hibridación y lavado descritas anteriormente y de 5°, 10°, 15°, 20°, 25° o 30° por debajo de la temperatura de fusión de la doble cadena de ADN de secuencias desveladas en el presente documento, incluyendo cualquier serie de condiciones subsumidas dentro de estos intervalos.

En otra realización, los ácidos nucleicos desvelados en el presente documento comprenden una secuencia sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10. En una realización, los ácidos nucleicos son al menos 85 %, al menos 90 % o al menos 95 % idénticos a los nucleótidos SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10. En una realización el ácido nucleico comprende la secuencia de nucleótidos SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10. En una realización adicional, el ácido nucleico codifica una proteína que es una sesquiterpeno sintasa, como se demuestra, por ejemplo, en el ensayo enzimático descrito en los ejemplos. También se proporcionan ácidos nucleicos que comprenden regiones conservadas entre diferentes especies.

En otra realización más, el ácido nucleico comprende un tramo contiguo de al menos 50, 100, 250, 500, 750 nucleótidos contiguos de SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10. Dichos fragmentos contiguos de estos nucleótidos también pueden contener al menos una mutación siempre que la secuencia mutante conserve la funcionalidad de la secuencia original y la capacidad de hibridar con estos nucleótidos en condiciones de rigurosidad baja o alta, tales como por ejemplo, condiciones de rigurosidad moderada o alta. Dicho fragmento puede derivar, por ejemplo, de los nucleótidos (nt) nt 200 a nt 1600, de nt 800 a nt 1600, de nt 1000 a nt 1600, de nt 200 a nt 1000, de nt 200 a nt 800, de nt 400 a nt 1600, o de nt 400 a nt 1000 de SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10.

Como se ha descrito anteriormente, los polipéptidos codificados por los ácidos nucleicos desvelados en el presente documento también se desvelan en el presente documento. Los ácidos nucleicos aislados desvelados en el presente documento pueden seleccionarse de un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5. En una realización, los polipéptidos son al menos 85 %, al menos 90 % o al menos 95 % idénticos a SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5.

En una realización, un polipéptido desvelado en el presente documento comprende una secuencia de aminoácidos como se expone en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5. En otra realización, el polipéptido comprende una secuencia de aminoácidos sustancialmente como se expone en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5. En otra realización más, el polipéptido comprende una secuencia de aminoácidos que es al menos 85 % idéntica, al menos 90 % o al menos 95 % idéntica a SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5. En una realización, el polipéptido es una sesquiterpeno sintasa, como se demuestra, por ejemplo, en el ensayo enzimático descrito posteriormente.

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 el mismo polipéptido. Dichas secuencias de ADN variantes pueden resultar de la deriva genética o manipulación artificial (por ejemplo, aparecen durante la amplificación por PCR o como el producto de mutagénesis deliberada de una secuencia nativa). También se desvela en el presente documento cualquier ácido nucleico capaz de codificar una proteína derivada de las SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10 o variantes de las mismas.

Puede llevarse a cabo mutagénesis deliberada de una secuencia nativa usando numerosas técnicas bien conocidas en este campo. Por ejemplo, pueden emplearse procedimientos de mutagénesis específica de sitio dirigida a oligonucleótidos, particularmente cuando se desee mutar un gen de modo que se alteren nucleótidos de restricción predeterminados o codones mediante sustitución, supresión o inserción. Se desvelan procedimientos ejemplares para realizar dichas alteraciones en Walder y col. (*Gene* 42: 133, 1986); Bauer y col. (*Gene* 37: 73, 1985); Craik (*BioTechniques*, enero 12-19, 1985); Smith y col. (*Genetic Engineering: Principles and Methods*, Plenum Press, 1981); Kunkel (*Proc. Natl. Acad. Sci. USA* 82: 488, 1985); Kunkel y col. (*Methods in Enzymol.* 154: 367, 1987); y patentes de Estados Unidos n.º 4.518.584 y 4.737.462.

También se desvelan en el presente documento polipéptidos aislados. Como se usa en el presente documento, el término "polipéptidos" se refiere a un género de polipéptido o fragmentos peptídicos que abarca las secuencias de aminoácidos identificadas en el presente documento, así como fragmentos más pequeños. Como alternativa, un polipéptido puede definirse con respecto a su relación antigénica con cualquier péptido codificado por las secuencias de ácido nucleico desveladas en el presente documento. Por tanto, en una realización, un polipéptido desvelado en el presente documento se define como una secuencia de aminoácidos que comprende un epítipo lineal o tridimensional compartido con cualquier péptido codificado por las secuencias de ácido nucleico desveladas en el presente documento.

Como alternativa, un polipéptido desvelado en el presente documento se reconoce por un anticuerpo que reconoce específicamente cualquier péptido codificado por las secuencias de ácido nucleico desveladas en el presente documento. Se define que los anticuerpos se unen específicamente si se unen con polipéptidos desvelados en el presente documento con una  $K_a$  de más de o igual a aproximadamente  $10^7 M^{-1}$ , tal como más de o igual a  $10^8 M^{-1}$ .

Una "variante" polipeptídica como se indica en el presente documento significa un polipéptido sustancialmente homólogo de un polipéptido nativo, pero que tiene una secuencia de aminoácidos diferente de la codificada por cualquiera de las secuencias de ácido nucleico desveladas en el presente documento debido a una o más supresiones, inserciones o sustituciones.

Las variantes pueden comprender secuencias sustituidas de forma conservativa, lo que significa que un resto de aminoácido dado se reemplaza por un resto que tiene características fisicoquímicas similares. Los ejemplos de sustituciones conservativas incluyen sustitución de un resto alifático por otro, tales como Ile, Val, Leu o Ala entre sí, o sustituciones de un resto polar por otro, tal como entre Lys y Arg; Glu y Asp; o Gin y Asn. Véase Zubay, *Biochemistry*, Addison-Wesley Pub. Co., (1983). Los efectos de dichas sustituciones pueden calcularse usando matrices de puntuación de sustitución tales como PAM-120, PAM-200 y PAM-250 como se analiza en Altschul, (*J. Mol. Biol.* 219: 555-65, 1991). Se conocen bien otras de dichas sustituciones conservativas, por ejemplo, las sustituciones de regiones completas que tienen características de hidrofobicidad similares.

Se desvelan en el presente documento también variantes peptídicas de origen natural. Son ejemplos de dichas variantes proteínas que resultan de acontecimientos de corte y empalme de ARNm alternativos o de escisión proteolítica de los polipéptidos descritos en el presente documento. Las variaciones atribuibles a proteólisis incluyen, por ejemplo, diferencias en el extremo N o C terminal tras la expresión en diferentes tipos de células huéspedes, debido a la retirada proteolítica de uno o más aminoácidos terminales de los polipéptidos codificados por las secuencias desveladas en el presente documento.

Las variantes de las sesquiterpeno sintasas desveladas en el presente documento pueden usarse para obtener actividad enzimática potenciada o reducida deseada, regioquímica o estereoquímica modificadas, o utilización de sustrato o distribución de producto alteradas. Una variante o un mutante directo de sitio pueden realizarse mediante cualquier procedimiento conocido en la técnica.

Como se ha indicado anteriormente, se desvelan en el presente documento polipéptidos recombinantes y no recombinantes, aislados y purificados, tales como de plantas de cítricos. Pueden obtenerse variantes y derivados de polipéptidos nativos aislando variantes de origen natural o la secuencia de nucleótidos de variantes, de otras líneas o especies vegetales o las mismas, o programando artificialmente mutaciones de secuencias de nucleótidos que codifiquen polipéptidos de cítricos nativos. Pueden conseguirse alteraciones de la secuencia de aminoácidos nativa por cualquiera de varios procedimientos convencionales. Pueden introducirse mutaciones en loci particulares sintetizando oligonucleótidos que contienen una secuencia mutante, flanqueada por sitios de restricción que permiten el ligamento a fragmentos de la secuencia nativa. Después del ligamento, la secuencia reconstruida resultante codifica un análogo que tiene la inserción, sustitución o supresión de aminoácidos deseada. Como alternativa, pueden emplearse procedimientos de mutagénesis específica de sitio dirigida a oligonucleótido para proporcionar un gen alterado en el que pueden alterarse los codones predeterminados por sustitución, supresión o inserción.

En una realización, la divulgación contempla: vectores que comprenden los ácidos nucleicos desvelados en el presente documento. Por ejemplo, un vector que comprende al menos un ácido nucleico seleccionado de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa.

Un vector como se usa en el presente documento incluye cualquier vector recombinante incluyendo pero sin limitación vectores virales, bacteriófagos y plásmidos.

Pueden prepararse vectores de expresión recombinantes que contienen una secuencia de ácido nucleico desvelada en el presente documento usando procedimientos bien conocidos. En una realización, los vectores de expresión incluyen una secuencia de ADNc que codifica el polipéptido unido operativamente con secuencias de nucleótidos



reguladoras de la transcripción o la traducción adecuadas, tales como las derivadas de un gen de mamífero, microbiano, vírico o de insecto. Los ejemplos de secuencias reguladoras incluyen promotores de la transcripción, operadores o potenciadores, sitios de unión a ribosomas de ARNm y secuencias apropiadas que controlan el inicio y la terminación de la transcripción y la traducción.

5 Son secuencias de nucleótidos "unidas operativamente" en las que la secuencia reguladora está relacionada funcionalmente con la secuencia de ADNc desvelada en el presente documento. Por lo tanto, una secuencia de nucleótidos promotora está unida operativamente con una secuencia de ADNc si la secuencia de nucleótidos promotora controla la transcripción de la secuencia de ADNc. Puede incorporarse adicionalmente en el vector de expresión la capacidad de replicar en las células huéspedes deseadas, conferida habitualmente por un origen de replicación, y un gen de selección por el que se identifican transformantes.

10 Además, pueden incorporarse en vectores de expresión secuencias que codifican péptidos señal apropiados que no están asociados de forma natural con los polipéptidos desvelados en el presente documento. Por ejemplo, puede fusionarse una secuencia de ADN para un péptido señal (líder secretor) en fase con una secuencia de nucleótidos desvelada en el presente documento de modo que los polipéptidos desvelados en el presente documento se traduzcan inicialmente como una proteína de fusión que comprende el péptido señal. Un péptido señal que es funcional en las células huéspedes pretendidas potencia la secreción extracelular del polipéptido expresado. El péptido señal puede escindirse del polipéptido tras secreción de la célula.

15 Pueden usarse fusiones de secuencias peptídicas adicionales en los extremos amino y carboxilo terminales de los polipéptidos desvelados en el presente documento para potenciar la expresión de los polipéptidos o ayudar en la purificación de la proteína.

20 También se desvela en el presente documento una célula huésped que comprende un ácido nucleico desvelado en el presente documento.

También se desvela en el presente documento un procedimiento para preparar una célula huésped recombinante que comprende introducir los vectores desvelados en el presente documento, en una célula huésped. En una realización adicional, se contempla un procedimiento para producir un polipéptido que comprende cultivar las células huéspedes desveladas en el presente documento en condiciones para producir el polipéptido. En una realización, el polipéptido se recupera. Los procedimientos desvelados en el presente documento incluyen procedimientos para preparar al menos una sesquiterpeno sintasa desvelada en el presente documento que comprenden cultivar una célula huésped que comprende un ácido nucleico desvelado en el presente documento, y recuperar la sesquiterpeno sintasa acumulada.

30 Las células huéspedes adecuadas para expresión de polipéptidos desvelados en el presente documento incluyen procariotas, células de levadura o eucariotas superiores. Se describen vectores de clonación y expresión apropiados para uso con huéspedes celulares bacterianos, fúngicos, de levadura y de mamífero, por ejemplo, en Pouwels y col., Cloning Vectors: A Laboratory Manual, Elsevier, Nueva York, (1985). También podrían emplearse sistemas de traducción sin células para producir los polipéptidos desvelados usando ARN derivados de construcciones de ADN desveladas en el presente documento.

35 Los procariotas incluyen organismos gram negativos o gram positivos, por ejemplo, *E. coli* o bacilos. Las células huéspedes procariotas adecuadas para transformación incluyen, por ejemplo, *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium*, y diversas otras especies dentro de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*. En una célula huésped procariota, tal como *E. coli*, los polipéptidos pueden incluir un resto de metionina N terminal para facilitar la expresión del polipéptido recombinante en la célula huésped procariota. La metionina N terminal puede escindirse del polipéptido recombinante expresado.

40 Los ejemplos de vectores de expresión útiles para células huéspedes procariotas incluyen los derivados de plásmidos disponibles en el mercado tales como los plásmidos de vector de clonación pET (Novagen, Madison, WI, Estados Unidos) o incluso pBR322 (ATCC 37017). pBR322 contiene genes para resistencia a ampicilina y tetraciclina y por lo tanto proporciona un medio sencillo para identificar células transformadas. Para construir un vector de expresión usando pBR322, se insertan un promotor apropiado y una secuencia de ADN que codifica uno o más de los polipéptidos desvelados en el presente documento en el vector pBR322. Otros vectores disponibles en el mercado incluyen, por ejemplo, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Suecia) y pGEM-1 (Promega Biotec, Madison, WI, Estados Unidos). Otros vectores disponibles en el mercado incluyen los que se diseñan específicamente para la expresión de proteínas; estos incluirían vectores pMAL-p2 y pMAL-c2 que se usan para la expresión de proteínas fusionadas con proteínas de unión a maltosa (New England Biolabs, Beverly, MA, Estados Unidos).

45 Las secuencias promotoras habitualmente usadas para vectores de expresión de células huéspedes procariotas recombinantes incluyen promotor de bacteriófago T7 (Studier F.W. y Moffatt B.A., J. Mol. Biol. 189: 113, 1986),  $\beta$ -lactamasa (penicilinas), sistema de promotor de lactosa (Chang y col., Nature 275: 615, 1978; y Goeddel y col., Nature 281: 544, 1979), sistema promotor de triptófano (*trp*) (Goeddel y col., Nucl. Acids Res. 8: 4057, 1980; y documento EP-A-36776), y promotor de *tac* (Maniatis, MoLecular Cloning: A Laboratory Manual, Cold Spring Harbor

Laboratory, p. 412, 1982). Un sistema de expresión de células huéspedes procariotas útil emplea un promotor PL de fago  $\lambda$  y una secuencia represora termolábil de cl857ts. Los vectores plasmídicos disponibles de la Colección Americana de Cultivos Tipo ("ATCC"), que incorporan derivados del promotor PL, incluyen el plásmido pHUB2 (residente en la cepa de *E. coli* JMB9 (ATCC 37092)) y pPLc28 (residente en *E. coli* RR1 (ATCC 53082)).

5 Los polipéptidos desvelados en el presente documento también pueden expresarse en células huéspedes de levadura, preferentemente del género *Saccharomyces* (por ejemplo, *S. cerevisiae*). También pueden emplearse otros géneros de levadura, tales como *Pichia* o *Kluyveromyces* (por ejemplo *K. lactis*). Los vectores de levadura contendrán con frecuencia una secuencia de origen de replicación de un plásmido de levadura  $2\mu$ , una secuencia de replicación autónoma (ARS), una región promotora, secuencias para poliadenilación, secuencias para terminación de la transcripción y un gen marcador seleccionable. Las secuencias promotoras adecuadas para vectores de levadura incluyen, entre otros, promotores para metalotionina, 3-fosfoglicerato quinasa (Hitzeman y col., J. Biol. Chem. 255: 2073, 1980), u otras enzimas glucolíticas (Hess y col., J. Adv. Enzyme Reg. 7: 149, 1968; y Holland y col., Biochem. 17: 4900, 1978), tales como enolasa, gliceraldehído-3-fosfato deshidrogenasa, hexoquinasa, piruvato descarboxilasa, fosfofructoquinasa, glucosa-6-fosfato isomerasa, 3-fosfoglicerato mutasa, piruvato quinasa, triosafosfato isomerasa, fosfoglucosa isomerasa y glucoquinasa. Otros vectores y promotores adecuados para su uso en expresión de levadura se describen adicionalmente en Hitzeman, documento EPA-73.657 o en Fleer y col., Gene, 107: 285-195 (1991); y van den Berg y col., Bio/Technology, 8: 135-139 (1990). Otra alternativa es el promotor de ADH2 reprimible por glucosa descrito en Russell y col. (J. Biol. Chem. 258: 2674, 1982) y Beier y col. (Nature 300: 724, 1982). Pueden construirse vectores lanzadera replicables tanto en levadura como en *E. coli* insertando secuencias de ADN de pBR322 para selección y replicación en *E. coli* (gen de Ampr y origen de replicación) en los vectores de levadura anteriormente descritos.

También se desvela en el presente documento un organismo no humano modificado para albergar un ácido nucleico desvelado en el presente documento. El organismo no humano y/o la célula huésped pueden modificarse por cualquier procedimiento conocido en la técnica para transferencia génica incluyendo, por ejemplo, el uso de dispositivos de suministro tales como lípidos y vectores virales, ADN desnudo, electroporación y transferencia génica mediada por partículas. En una realización, el organismo no humano es una planta, un insecto o un microorganismo.

Por ejemplo, se desvela en el presente documento un procedimiento para preparar al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos un ácido nucleico en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa en el que dicho al menos un ácido nucleico se selecciona de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa.

En una realización adicional, el huésped es una planta tal como tabaco, animal o microorganismo incluyendo también pero sin limitación células bacterianas, células de levadura, células vegetales y células animales. Como se usa en el presente documento, las células vegetales y células animales incluyen el uso de plantas y animales como un huésped. Por ejemplo, en algunas realizaciones, la expresión es en un organismo no humano modificado genéticamente.

En una realización, se emplean sistemas de cultivo de células huéspedes de mamífero o de insecto para expresar polipéptidos recombinantes desvelados en el presente documento. Se revisan sistemas de baculovirus para producción de proteínas heterólogas en células de insectos en Luckow y Summers, Bio/Technology 6: 47 (1988). También pueden emplearse líneas celulares establecidas de origen mamífero. Los ejemplos de líneas de células huéspedes de mamífero adecuadas incluyen la línea COS-7 de células de riñón de mono (ATCC CRL 1651) (Gluzman y col., Cell 23: 175, 1981), células L, células C127, células 3T3 (ATCC CCL 163), células de ovario de hámster chino (CHO), células HeLa y líneas celulares BHK (ATCC CRL 10), y la línea celular CV-1/EBNA-1 (ATCC CRL 10478) derivada de la línea celular de riñón de mono verde africano CV1 (ATCC CCL 70) como se describe en McMahan y col. (EMBO J. 10: 2821, 1991).

50 Se han descrito procedimientos establecidos para introducir ADN en células de mamífero (Kaufman, R.J., Large Scale Mammalian Cell Culture, 1990, pp. 15-69). Pueden usarse protocolos adicionales usando reactivos disponibles en el mercado, tales como Lipofectamine (Gibco/BRL) o Lipofectamine-Plus, para transfectar células (Feigner y col., Proc. Natl. Acad. Sci. USA 84: 7413-7417, 1987). Además, puede usarse electroporación para transfectar células de mamífero usando procedimientos convencionales, tales como los de Sambrook y col. Molecular Cloning: A Laboratory Manual, 2 ed. Vol. 1-3, Cold Spring Harbor Laboratory Press, 1989). Puede realizarse selección de transformantes estables usando resistencia a fármacos citotóxicos como un procedimiento de selección. Kaufman y col., Meth. in Enzymology 185: 487-511, 1990, describe varios esquemas de selección, tales como resistencia a dihidrofolato reductasa (DHFR). Una cepa huésped adecuada para selección de DHFR puede ser la cepa de CHO DX-B11, que es deficiente en DHFR (Urlaub y Chasin, Proc. Natl. Acad. Sci. USA 77: 4216-4220, 1980). Puede introducirse un plásmido que expresa el ADNc de DHFR en la cepa DX-B11, y solamente células que contienen el plásmido pueden crecer en los medios selectivos apropiados.

- Pueden escindirse secuencias de control de la transcripción y la traducción para vectores de expresión de células huéspedes de mamífero de genomas virales. Secuencias promotoras y secuencias potenciadoras habitualmente usadas derivan de virus del polio, adenovirus 2, virus de simio 40 (SV40) y citomegalovirus humano. Pueden usarse secuencias de ADN derivadas del genoma viral de SV40, por ejemplo, origen de SV40, promotor temprano y tardío, potenciador, sitios de corte y empalme y de poliadenilación para proporcionar otros elementos genéticos para expresión de una secuencia génica estructural en una célula huésped de mamífero. Los promotores virales temprano y tardío son particularmente útiles porque ambos se obtienen fácilmente de un genoma viral como un fragmento, que también puede contener un origen de replicación viral (Fiers y col., Nature 273: 113, 1978; Kaufman, Meth. in Enzymology, 1990).
- Existen varios procedimientos conocidos en la técnica para la creación de plantas transgénicas. Estos incluyen, pero sin limitación: electroporación de protoplastos vegetales, transformación mediada por liposomas, transformación mediada por polietilenglicol, microinyección de células vegetales y transformación usando virus.
- En una realización, se utiliza transferencia génica directa por bombardeo de partículas.
- La transferencia génica directa por bombardeo de partículas proporciona un ejemplo para transformar tejido vegetal. En esta técnica se dispara una partícula, o un microproyectil, revestido con ADN a través de las barreras físicas de la célula. Puede usarse bombardeo de partículas para introducir ADN en cualquier tejido diana que sea penetrable por partículas revestidas con ADN, pero para transformación estable, es imperativo usar células regenerables. Típicamente, las partículas están hechas de oro o wolframio. Las partículas se revisten con ADN usando procedimientos de precipitación con  $\text{CaCl}_2$  o etanol que se conocen habitualmente en la técnica.
- Se disparan partículas revestidas con ADN de una pistola de partículas. Puede obtenerse una pistola de partículas adecuada de Bio-Rad Laboratories (Hercules, CA). La penetración de partículas se controla por diversos parámetros tales como la intensidad de la ráfaga explosiva, el tamaño de las partículas, o la distancia que las partículas deben viajar para alcanzar el tejido diana.
- El ADN usado para revestir las partículas puede comprender un casete de expresión adecuado para conducir la expresión del gen de interés que comprenderá un promotor unido operativamente con el gen de interés.
- Se desvelan procedimientos para realizar transferencia génica directa por bombardeo de partículas en la patente de Estados Unidos 5.990.387 de Tomes y col.
- En una realización, los ADNc desvelados en el presente documento pueden expresarse de tal manera que produzcan ARN con sentido o antisentido. El ARN antisentido es ARN que tiene una secuencia que es el complemento inverso del ARNm (ARN con sentido) codificado por un gen. Un vector que conducirá la expresión de ARN antisentido es uno en el que el ADNc se coloca en "orientación inversa" con respecto al promotor de modo que se transcribe la cadena no codificante (en lugar de la cadena codificante). La expresión de ARN antisentido puede usarse para modular negativamente la expresión de la proteína codificada por el ARNm para el que el ARN antisentido es complementario. Podrían usarse vectores que producen ARN antisentido para realizar plantas transgénicas, como se ha descrito anteriormente.
- En una realización, el ADN transfectado se integra en un cromosoma de un organismo no humano de modo que se produzca un sistema recombinante estable. Puede usarse cualquier procedimiento de integración cromosómica conocido en la técnica en la práctica de la invención, incluyendo pero sin limitación intercambio de casete mediado por recombinasa (RMCE), inserción cromosómica específica de sitio viral, adenovirus e inyección pronuclear.
- También se desvelan en el presente documento procedimientos para preparar terpenoides y compuestos de sesquiterpeno, por ejemplo, usando los nucleótidos y polipéptidos desvelados en el presente documento. Los ejemplos incluyen procedimientos para preparar al menos un terpenoide que comprenden poner en contacto al menos un precursor de pirofosfato terpeno acíclico con al menos un polipéptido codificado por un ácido nucleico seleccionado de (a) un ácido nucleico que comprende la secuencia de nucleótidos sustancialmente como se expone en SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 o SEQ ID NO: 10; (b) un ácido nucleico que codifica el polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (c) un ácido nucleico que hibrida con el ácido nucleico de (a) o (b) en condiciones de baja rigurosidad, en el que el polipéptido codificado por dicho ácido nucleico es una sesquiterpeno sintasa, y aislar al menos un terpenoide producido. Otro ejemplo es un procedimiento para preparar al menos un terpenoide que comprende poner en contacto al menos un precursor de pirofosfato terpeno acíclico con al menos un polipéptido sustancialmente expuesto en SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5 y aislar al menos un terpenoide producido.
- Como se usa en el presente documento un precursor de pirofosfato terpeno acíclico es cualquier compuesto de pirofosfato acíclico que sea un precursor para la producción de al menos un terpeno incluyendo pero sin limitación geranyl-pirofosfato (GPP), farnesil-pirofosfato (FPP) y geranylgeranyl-pirofosfato (GGPP).
- En una realización, el al menos un terpenoide se selecciona de sesquiterpenos. En una realización, el al menos un precursor de pirofosfato terpeno acíclico es farnesil-pirofosfato. En una realización adicional, el al menos un

sesquiterpeno se selecciona de bicilogermacreno ((E,E)-3,7,11,11-tetrametil-biciclo[8.1.0]undeca-2,6-dieno), cubebol ((1R,4S,5R,6R,7S,10R)-7-isopropil-4,10-dimetil-triciclo[4.4.0.0(1,5)]decan-4-ol, valenceno ((+)-(1R)-1,2,3,5,6,7,8,8A-octahidro-7-isopropenil-1 @,8A@-dimetilnaftaleno),  $\alpha$ -cubebeno, germacreno D, y  $\delta$  - cadineno (rel-(1R,8AS)-1,2,3,5,6,8A-hexahidro-1-isopropil-4,7-dimetilnaftaleno) (Figura 2). Los terpenoides desvelados en el presente documento pueden aislarse por cualquier procedimiento usado en la técnica incluyendo pero sin limitación cromatografía, extracción y destilación.

En una realización, la distribución de productos o los productos reales formados pueden alterarse variando el pH al que la sintasa entra en contacto con el precursor de pirofosfato terpeno acíclico, tal como, por ejemplo farnesil-pirofosfato. En una realización, el pH es 7. En una realización adicional el pH es menor de 7, tal como, por ejemplo, 6, 5, 4 y 3.

También se desvela en el presente documento un organismo (por ejemplo, microorganismo o planta) que se usa para construir una plataforma para producción a alto nivel de un sustrato de sesquiterpenos sintasas (por ejemplo, FPP) y la introducción de un ácido nucleico desvelado en el presente documento en el organismo. Por ejemplo, se incorpora al menos un ácido nucleico desvelado en el presente documento que codifica una sesquiterpeno sintasa en un organismo no humano que produce FPP efectuando de este modo conversión de FPP a un sesquiterpeno, y la producción metabólica posterior de sesquiterpeno. En una realización, esto da como resultado una plataforma para la producción a alto nivel de sesquiterpenos.

En una realización, los ácidos nucleicos desvelados en el presente documento se usan para crear otros ácidos nucleicos que codifican sesquiterpenos sintasas. Por ejemplo, se desvela en el presente documento un procedimiento para identificar sesquiterpenos sintasas que comprende construir una biblioteca de ADN usando los ácidos nucleicos desvelados en el presente documento, explorar la biblioteca con respecto a ácidos nucleicos que codifiquen al menos una sesquiterpeno sintasa. La biblioteca de ADN usando los ácidos nucleicos desvelados en el presente documento puede construirse por cualquier proceso conocido en la técnica en el que se crean secuencias de ADN usando los ácidos nucleicos desvelados en el presente documento como un punto de partida, incluyendo, pero sin limitación redistribución de ADN. En dicho procedimiento, la biblioteca puede explorarse con respecto a sesquiterpeno sintasas usando un ensayo funcional para encontrar un ácido nucleico diana que codifique una sesquiterpeno sintasa. La actividad de una sesquiterpeno sintasa puede analizarse usando, por ejemplo, los procedimientos descritos en el presente documento. En una realización, se utiliza exploración de alto rendimiento para analizar la actividad de los polipéptidos codificados.

Como se usa en el presente documento una "sonda nucleotídica" se define como un oligonucleótido o polinucleótido capaz de unirse con un ácido nucleico diana de secuencia complementaria mediante uno o más tipos de enlaces químicos, mediante formación de pares de bases complementarias o mediante formación de enlaces de hidrógeno. Como se ha descrito anteriormente, la sonda oligonucleotídica puede incluir bases naturales (es decir A, G, C, o T) o modificadas (7-desazaguanosina, inosina, etc.). Además, las bases en una sonda de nucleótidos pueden unirse por un enlace distinto de un enlace fosfodiéster, siempre que no eviten la hibridación. Por lo tanto, las sondas oligonucleotídicas pueden tener bases constituyentes unidas por enlaces peptídicos en lugar de enlaces fosfodiéster.

Un "ácido nucleico diana" en el presente documento se refiere a un ácido nucleico con el que puede hibridar específicamente la sonda o molécula de nucleótidos. La sonda se diseña para determinar la presencia o ausencia del ácido nucleico diana, y la cantidad de ácido nucleico diana. El ácido nucleico diana tiene una secuencia que es complementaria de la secuencia de ácido nucleico de la sonda correspondiente dirigida a la diana. Como se reconoce por un experto en la materia, la sonda también puede contener ácidos nucleicos adicionales u otros restos, tales como marcadores, que pueden no hibridar específicamente con la diana. La expresión ácido nucleico diana puede referirse a la secuencia de nucleótidos específica de un ácido nucleico mayor al que se dirige la sonda o a la secuencia general (por ejemplo, gen o ARNm). Un experto en la materia reconocerá la utilidad completa en diversas condiciones.

Además del ejemplo operativo, o cuando se indique de otro modo, todos los números que expresan cantidades de ingredientes, condiciones de reacción y así sucesivamente usados en la memoria descriptiva y reivindicaciones deben entenderse como modificados en todos los casos por el término "aproximadamente". En consecuencia, a no ser que se indique lo contrario, los parámetros numéricos expuestos en la memoria descriptiva y reivindicaciones son aproximaciones que pueden variar dependiendo de las propiedades deseadas que se busca obtener por la presente invención. Como mínimo, y no como un intento de limitar la aplicación de la doctrina de equivalentes al alcance de las reivindicaciones, cada parámetro numérico debería interpretarse a la luz del número de dígitos significativos y enfoques de redondeo ordinarios.

A pesar de que los intervalos numéricos y parámetros que exponen el amplio alcance de la invención con aproximaciones, los valores numéricos expuestos en los ejemplos específicos se presentan con tanta precisión como sea posible. Cualquier valor numérico, sin embargo, contiene inherentemente ciertos errores necesariamente resultantes de la desviación típica hallada en sus mediciones de ensayo respectivas. Se pretende que los siguientes ejemplos ilustren la divulgación sin limitar el alcance como resultado. Los porcentajes se proporcionan en peso.

## Ejemplos

### Material

Una piel externa de pomelo (*Citrus paradisi*) se usó como material de partida para los experimentos descritos en el presente documento. La piel externa (parte coloreada externa de la piel de la fruta) contiene las glándulas oleosas que son el sitio de biosíntesis de terpenos. La piel externa del pomelo se preparó en Simone Gatto (Sicilia) a partir de frutos en maduración recién recolectado. La piel externa (3-4 mm de grosor) se cortó, se congeló inmediatamente y se mantuvo congelada durante el transporte y todas las etapas posteriores para evitar la degradación.

### Ejemplo 1: Aislamiento de ADNc de sesquiterpeno sintasa usando RT-PCR

Las secuencias de aminoácidos deducidas de sesquiterpeno sintasas vegetales se alinearon para identificar regiones conservadas y diseñar oligonucleótidos específicos de sesquiterpeno sintasas vegetales. Para obtener mejor homología de secuencia, las secuencias se separaron en dos grupos (Figura 4). El primer grupo contenía las secuencias de la germacreno C sintasa de *Lycopersicon esculentum* cv. VFNT cherry (Colby y col, 1998), la (E)- $\beta$ -farneseno sintasa de *Mentha x piperita* (Crock y col, 1997), la  $\delta$ -selineno sintasa de *Abies grandis* (Steele y col, 1998), una sesquiterpeno sintasa de *Citrus junos* (n.º de referencia de GenBank AF288465), las 5-epi aristoloqueno sintasas de *Nicotiana tabacum* (Facchini y Chappell, 1992) y de *Capsicum annuum* (Back y col, 1998), las vetispiradieno sintasas de *Solanum tuberosum* y de *Hyoscyamus muticus* (Back y Chappel, 1995). El segundo grupo contenía secuencias de las (+)- $\delta$ -cadineno sintasas de *Gossypium arboreum* (Chen y col. 1995), la amorfa-4,11-dieno sintasa (Mercke y col, 2000) y la epi-cedrol sintasa (Merck y col, 1999) de *Artemisia annua* y la  $\gamma$ -humuleno sintasa de *Abies grandis* (Steele y col, 1998). La mayor homología de secuencia se encontró en la parte central de las secuencias. Se seleccionaron tres regiones que contenían aminoácidos suficientemente conservados y se diseñaron oligonucleótidos degradados específicos para estas regiones (es decir dos cebadores directos y uno inverso se dedujeron de cada alineamiento) (Figura 4).

Se realizó RT-PCR usando ARN total de piel externa de pomelo preparada por la técnica de borato caliente y la diferente combinación de los cebadores degradados directos e inversos. El procedimiento de borato caliente para extracción de ARN total se adaptó de Wan y Wilkins (Wan y col. (1994) Anal. Biochem. 223, 7-12.). Los tejidos se trituraron hasta obtener un polvo fino en nitrógeno líquido usando un mortero. El polvo se añadió a 5 ml de tampón de extracción (borato 200 mM, EGTA 30 mM, DTT 10 mM, SDS 1 %, NA desoxicolato 1 %, PVP 2 %, nonidet NP-40 0,5 %) precalentado a 80 °C. La mezcla se homogeneizó usando un homogeneizador Ultraturax TM y se filtró a través de dos capas de filtro MiraclothTM (Calbiochem). La mezcla se incubó 90 minutos a 42 °C en presencia de proteinasa K 0,5 mg/ml (para digestión de proteína/ribonucleasa). Después se añadió KCl a una concentración final 1 M y, después de 1 hora de incubación en hielo, la mezcla se colocó en una centrífuga durante 10 min a 10.000 g y 4 °C. El sobrenadante se recuperó y se añadió 1/3 del volumen de LiCl 8 M. La mezcla se incubó durante una noche a 4 °C y se colocó en una centrífuga durante 20 minutos a 10.000 g y 4 °C.

El sobrenadante se descartó y el sedimento se lavó con LiCl 2 M y se centrifugó de nuevo como anteriormente. El sedimento se resuspendió después en agua destilada sin ARNasa y se añadieron 0,15 volúmenes de solución de acetato potásico 2 M y 2,5 volúmenes de etanol absoluto. El ARN se precipitó incubando 2 horas a -20 °C y se sedimentó mediante centrifugación como anteriormente. El sedimento de ARN se lavó con etanol al 80 % y se resuspendió en agua destilada sin ARNasa.

La concentración de ARN se estimó a partir de la DO a 260 nm. La integridad del ARN se evaluó en un gel de agarosa verificando la integridad de las bandas de ARN ribosómico.

Se realizó RT-PCR usando el kit de RT-PCR OneStep de Qiagen y un termociclador de gradiente Mastercycler Eppendorf. Las mezclas de reacción típicas contenían 10  $\mu$ l de tampón RT-PCR OneStep de Qiagen 5X, 400  $\mu$ M de cada dNTP, 400 nM de cada cebador, 2  $\mu$ l de mezcla de enzimas de RT-PCR OneStep de Qiagen, 1  $\mu$ l de inhibidor de ribonucleasa RNasin® (Promega Co.) y 1  $\mu$ g de ARN total en un volumen final de 50  $\mu$ l. Las condiciones del termociclador fueron: 30 min a 50 °C (transcripción inversa); 15 min a 95 °C (activación de ADN polimerasa); 40 ciclos de 45 s a 94 °C, 10 s a 42 °C hasta 45 °C (dependiendo del cebador usado), 45 s a 90 s (dependiendo del tamaño del fragmento de ADN para amplificar) a 72 °C; y 10 min a 72 °C.

El tamaño de los productos de PCR se evaluó en un gel de agarosa al 1 %. Las bandas correspondientes al tamaño esperado se escindieron del gel, se purificaron usando el kit de extracción en gel QIAquick® (Qiagen) y se clonaron en el vector pCR®2.1- TOPO usando el kit de clonación TOPO TA (Invitrogen). Se sometieron después ADNc insertados a secuenciación de ADN y la secuencia se comparó con respecto a la base de datos de proteínas no redundante de GenBank (NCBI) usando el algoritmo BLASTX (Altschul y col 1990).

El análisis por secuenciación de ADN y la búsqueda de Blast de los productos de PCR con longitud esperada (de 80 a 240 pb de tamaño) reveló fragmentos de tres sesquiterpeno sintasas diferentes, que se denominaron GFTpsA, GFTpsB y GFTpsC. El fragmento de GFTpsA de 180 pb se amplificó con los cebadores TpsVF1 y TpsVR3, el fragmento de GFTpsB de 115 pb se amplificó con los cebadores TpsVF2 y TpsVR3, y el fragmento de GFTpsA de 115 pb se amplificó con los cebadores TpsVF1 y TpsVR3. Las secuencias de aminoácidos deducidas revelaron diferencias significativas entre estos tres fragmentos (Figura 5).

**Ejemplo 2: Aislamiento de ADNc de sesquiterpeno sintasa usando 5'/3'-RACE**

Para aislar las secuencias de longitud completa de las sesquiterpeno sintasas, se usó en primer lugar un enfoque de 5'/3'-RACE. Se sintetizó ADNc usando kit de amplificación de ADNc Marathon™ (Clontech) y partiendo de 1 µg de ARNm purificado a partir de ARN total preparado con la técnica de borato caliente. La calidad del ADNc sintetizado era escasa, esencialmente se obtuvieron ADNc de tamaño pequeño (tamaño promedio de 0,5 Kb). Sin embargo, usando este ADNc, se obtuvo el extremo 3' de los GFTpsB y GFTpsC usando los cebadores específicos de genes GFTpsBRF1 y GFTpsBRF2 para GFTpsB y los cebadores específicos de genes GFTpsCRF1 y GFTpsCRF2 para GFTpsC (véase Tabla 1). El ARNm preparado por el procedimiento de extracción de tiocianato de guanidinio/fenol proporcionó ADNc de mayor calidad con un tamaño promedio de 2 Kb y permitió el aislamiento de la secuencia del extremo 5' de GFTpsC usando los cebadores específicos de genes GFTpsCRR1 y GFTpsRR2 (véase Tabla 1), completando la secuencia de longitud completa de este clon. El extremo 5' de GFTpsB y el extremo 5' y extremo 3' de GFTpsA no pudieron obtenerse por este enfoque.

El procedimiento de extracción de tiocianato de guanidinio /fenol se basa en la técnica descrita por Chomczynski y Sacchi usando la solución RNAClean™ de ThermoHybaid (Chomczynski, P., y Sacchi, N., (1987) Anal. Biochem. 162, 156-159.). Brevemente, se trituraron 2 g de tejidos congelados hasta obtener un polvo fino en nitrógeno líquido usando un mortero. El polvo se transfirió a 20 ml de solución RNAClean™ y la suspensión se homogeneizó usando un homogeneizador Ultraturax™ y se filtró a través de un filtro Miracloth™ (Calbiochem). Después de la adición de 0,1 volúmenes de cloroformo, el tubo se colocó en hielo y se centrifugó durante 20 min a 12.000 g y 4 °C. La fase acuosa superior se recuperó y se añadió un volumen de isopropanol.

Después de 20 min de incubación a -20 °C, la muestra se centrifugó durante 20 min. a 12.000 g y 4 °C. El sedimento blanco grande obtenido se lavó con etanol 70 % y se secó a temperatura ambiente. Este sedimento, que contenía el ARN total, se sometió inmediatamente a purificación de ARNm mediante cromatografía de afinidad de oligodT-celulosa usando el kit de aislamiento de ARNm FastTrack® 2.0 (Invitrogen). Se siguió el protocolo del fabricante excepto que después de la resuspensión de ARN total en el tampón de lisis, la muestra se calentó a 100 °C en lugar de a 65 °C.

Se realizó amplificación rápida 3' y 5' de los extremos de ADNc (RACE) usando el kit de amplificación de ADNc Marathon™ (Clontech). El procedimiento comenzó con síntesis de ADNc de primera cadena partiendo de ARNm usando un cebador oligo(dT). Después de síntesis de segunda cadena, se ligaron adaptadores específicos a los extremos de ADNc bicatenario (ADNc bc). Este procedimiento proporciona una biblioteca no clonada de ADNc bc ligado a adaptadores. Se usó ARNm de pomelo purificado (1 µg) como material de partida. La calidad y cantidad de ADNc se evaluó en un gel de agarosa.

El extremo 3' o 5' de los ADNc específicos se amplificó con mezcla de polimerasa Advantage® 2 usando una combinación de oligonucleótidos específicos de gen y adaptador. Las mezclas de reacción RACE típicas contenían, en un volumen final de 50 µl, 5 µl de tampón de reacción de PCR de ADNc 10X (clontech), 200 nM de cada dNTP, 1 µl de mezcla de polimerasa Advantage® 2, cebador específico de adaptador 200 µM (Clontech), cebador específico de gen 200 µM (véase Tabla 1) y 5 µl de ADNc ligado a adaptador diluido de 50 a 250 veces. Se realizó amplificación en un termociclador de gradiente Mastercycler Eppendorf. Las condiciones de termociclación fueron las siguientes: 1 min a 94 °C, 5 ciclos de 30 s a 94 °C y de 2 a 4 min a 72 °C, 5 ciclos de 30 s a 94 °C y de 2 a 4 min a 70 °C, 20 ciclos de 30 s a 94 °C y de 2 a 4 min a 68 °C. Cuando fue necesario se realizó un segundo ciclo de amplificación usando un cebador específico de adaptador anidado (Clontech) y un cebador específico de gen anidado.

Los productos de amplificación se evaluaron, se subclonaron y la secuencia se analizó como anteriormente con respecto a los productos de RT-PCR.

**TABLA 1:** Los siguientes cebadores se usaron en los experimentos de 3'/5'-RACE

Nombre	Descripción	Secuencia (5' a 3').
GFTpsARF1 (SEQ ID NO: 34)	Cebador directo de 3'-RACE.	CTGGGAGTGTCTATGAGCT CAATTTGG
GFTpsARF2 (SEQ ID NO: 35)	Cebador anidado directo de 3'-RACE GFTpsA.	GTAGAATTTTTGCATCCAAAG TGGTGTGC
GFTpsARR1 (SEQ ID NO: 36)	Cebador inverso de 5'-RACE GFTpsA.	CACACCACTTTGGATGCAA AATTCATC
GFTpsARR2 (SEQ ID NO: 37)	Cebador inverso anidado de 5'-RACE GFTpsA.	CCAAATTTGGGCTCATAGGAC ACTCCCAG
GFTpsBRF1 (SEQ ID NO: 38)	Cebador directo de 3'-RACE GFTpsB.	TAGGGACGTATTTTGAACCA AAGTAC
GFTpsBRF2 (SEQ ID NO: 39)	Cebador anidado directo de 3'-RACE GFTpsB.	AAATAATGACCAAACAATTT ACACGG

(continuación)

Nombre	Descripción	Secuencia (5' a 3').
GFTpsBRR1 (SEQ ID NO: 40)	Cebador inverso de 5'-RACE GFTpsB.	GCACTTTTCATGTATTCTGGAA G
GFTpsBRR2 (SEQ ID NO: 41)	Cebador inverso anidado de 5'-RACE GFTpsB.	GTTTGAGCTCTTCAAAGAAA CC
GFTpsCRF1 (SEQ ID NO: 42)	Cebador directo de 3'-RACE GFTpsC.	AATGGGAGTGATTTTGAGC CTCGATACTCC
GFTpsCRF2 (SEQ ID NO: 43)	Cebador anidado directo de 3'-RACE GFTpsC.	GATATTTTCCAAAGTAATTGC AATGGCATCC
GFTpsCRR1 (SEQ ID NO: 44)	Cebador inverso de 5'-RACE GFTpsC.	GTTGCTAATATCCCACCTTTT GATAGC
GFTpsCRR2 (SEQ ID NO: 45)	Cebador inverso anidado de 5'-RACE GFTpsC.	AAGTGTGCCATAGGCGTCGT AGG
GFTpsDRR1 (SEQ ID NO: 46)	Cebador inverso de 5'-RACE GFTpsD.	CTGTTCCGCAAGCTTAGGGG TTACATG
GFTpsDRR2 (SEQ ID NO: 47)	Cebador inverso anidado de 5'-RACE GFTpsD.	CTGAGCTACCAATGACTTCA GGTGAGTGG
GFTpsERR1 (SEQ ID NO: 48)	Cebador inverso de 5'-RACE GFTpsE.	CAATTTTGCCATACACATCAT AGATATCATC
GFTpsERR2 (SEQ ID NO: 49)	Cebador inverso anidado de 5'-RACE GFTpsE.	AACAGAAGTCATGGAGATCA CTTTCGTC
GFTpsERR3 (SEQ ID NO: 50)	Cebador inverso de 5'-RACE GFTpsE.	CGCAAGAGATGTTTTAAAGTT CCCATCC
GFTpsERR4 (SEQ ID NO: 51)	Cebador inverso anidado de 5'-RACE GFTpsE.	TGAACATCAGCGGAAATTTTA TAGCC

El ADNc parcial de GFTpsB obtenido mediante 3'-RACE reveló alta identidad de secuencia con un ADNc de terpeno sintasa potencial hallado en las bases de datos públicas (n.º de referencia de GeneBank AF288465) y aislado de *Citrus junos*. Se diseñaron cebadores específicos para la secuencia de esta terpeno sintasa: un par de cebadores directos e inversos (junosF1 y junosR1) diseñados contra la región no codificante de la terpeno sintasa de *C. junos* y un par de cebadores directo e inverso (junosF2 y junosR2) diseñados en la región codificante incluyendo los codones de inicio y terminación. La PCR en la biblioteca de ADNc Marathon™ de pomelo usando los cebadores junosF1 y junosR1 no produjo amplicón. Usando los cebadores junosF2 y junosR2 se amplificó un fragmento de 1,6 Kb. La secuenciación de ADN confirmó que este producto de PCR era una sesquiterpeno sintasa de longitud completa de 1669 pb del clon GFTpsB (Figura 6).

### Ejemplo 3: Exploración de biblioteca de ADNc y secuenciación de EST

Para obtener el ADNc de longitud completa de los clones parciales obtenidos por el enfoque de PCR, se construyó una biblioteca de ADNc preparada a partir de ARNm de piel externa del pomelo.

Se realizó construcción de biblioteca y síntesis de ADNc usando el kit de construcción de biblioteca Uni-ZAP® XR (Stratagene) de acuerdo con el protocolo del fabricante partiendo de 7,5 µg de ARNm de piel externa del pomelo (preparado por el procedimiento de tiocianato de guanidinio/fenol). El título original de la biblioteca fue de  $2 \times 10^7$  UFP (unidades formadoras de placas) y el tamaño de inserto promedio fue de 1,1 Kb.

Se usaron dos enfoques para aislar ADNc que codificaban sesquiterpeno sintasas de la biblioteca de ADNc: secuenciación de EST y exploración usando una sonda de ADN. Para secuenciación de EST (marcador de secuencia expresado), se usó una fracción de la biblioteca para escindir el fagémido pBluescript del vector Uni-ZAP XR de acuerdo con el protocolo de escisión de masa de Stratagene. Se seleccionaron aleatoriamente colonias bacterianas transformadas resultantes (576) y el plásmido se purificó. Se realizó una reacción de secuenciación para cada clon usando el cebador T3. Las secuencias se editaron en primer lugar para retirar las secuencias de vector y se compararon con la base de datos de proteínas no redundante de GenBank (NCBI) usando el algoritmo BLASTX (Altschul y col 1990).

La biblioteca se exploró con sondas de ADN marcadas con digoxigenina (DIG). Las sondas se sintetizaron mediante incorporación de DIG-dUTP en un fragmento de ADN mediante PCR usando el kit de síntesis de sondas PCR DIG (digoxigenina) (Roche Diagnostics). Se amplificó una sonda de ADN derivada de GFTpsA de 149 pb de una alícuota de la biblioteca usando el cebador directo GFTpsAproF (SEQ ID NO: 52) (5'-ACGATTTAGGCTTCCCTAAAAAGG-3') y el cebador inverso GFTpsAproR (SEQ ID NO: 53) (5'-TATTATGGAATATTATGCACACCAC-3'). Se amplificó una sonda derivada de GFTpsB de 1036 pb del plásmido pET-GFTpsB2-2 usando el cebador directo junosF2 (SEQ ID NO: 54) (5'-AAATGTCCGCTCAAGTTCTAGCAACGG-3') y el cebador inverso GFTpsBRR2 (SEQ ID NO: 41). Se amplificó una sonda de ADN derivada de GFTpsC de 1008 pb a partir del plásmido pET-GFTpsC usando el cebador directo GFTpsCpetF1 (SEQ ID NO: 55) (5'-ATGGCACTTCAAGATTCAGAAAGTCC-3') y el cebador inverso

GFTpsCRR1 (SEQ ID NO: 44).

Las sondas se hibridaron con transferencias de placas, preparadas a partir de placas que contenían de 5.000 a 25.000 UFP, a 40 °C en solución de hibridación Dig Easy Hyb (Roche Diagnostics). La detección de los híbridos de sonda-diana de las membranas se realizó mediante quimioluminiscencia usando alcalina fosfatasa antidigoxigenina (Roche Diagnostics) y sustrato de fosfato alcalino CDP-Star (Roche Diagnostics), y la visualización se realizó en un sistema de captura de imágenes VersaDoc (BioRad).

Se extrajeron fagos de señales positivas de las placas de agar y se sometieron a exploración secundaria y si fuera necesario terciaria para conseguir señales positivas de placas individuales. Los aislados positivos se escindieron *in vivo* y el inserto se secuenció como se ha descrito anteriormente.

La biblioteca se exploró usando sondas de ADN preparadas a partir de GFTpsA, GFTpsB y GFTpsC. Este enfoque produjo tres ADNc de sesquiterpeno sintasa diferentes. Dos de ellos eran clones previamente hallados: un clon, denominado GF2-30-1, era una forma truncada de 300 pb de extremo 5' del ADNc de GFTpsC; el segundo clon, denominado 9-13-6, era un clon parcial de GFTpsA que estaba truncado en su extremo 5' en aproximadamente 168 pb como se valoró por comparación con otras sesquiterpeno sintasas. El clon 9-13-6 proporcionó información de secuencia 5' adicional de 618 pb en comparación con el clon de GFTpsA parcial anteriormente mencionado obtenido mediante RT-PCR. En su extremo 3', el clon 9-13-6 también estaba incompleto. Faltaban aproximadamente 680 nucleótidos y se reemplazaron por un fragmento de 600 pb sin función definida. El tercer ADNc que codificaba sesquiterpeno sintasa obtenido por exploración, GF2-5-11, codificaba una nueva sesquiterpeno sintasa que se denominó GFTpsD. Este clon estaba truncado en el extremo 5', pero los 414 pb ausentes se recuperaron mediante 5'-RACE usando los cebadores descritos en la Tabla 1. El GFTpsD de longitud completa se amplificó después a partir de la biblioteca de ADNc Marathon™ y se clonó en el vector de expresión bacteriana como se ha descrito anteriormente.

El análisis de la secuencia de ADN de varios ADNc de GFTpsD de longitud completa reveló que estaban presentes dos sesquiterpeno sintasas estrechamente relacionadas con diferencias de secuencia menores, se denominaron GFTpsD1 y GFTpsD2 (Figura 6). La secuencia de aminoácidos deducida de estas dos variantes difirió entre sí en cuatro restos.

Para el enfoque de secuenciación de EST, se secuenciaron 576 clones. Entre ellos, solamente tres clones codificaban proteínas de tipo terpeno-sintasas potenciales y, más exactamente, dos proteínas diferentes de tipo sesquiterpeno sintasa y una de tipo monoterpene sintasa. Uno de los dos clones de tipo sesquiterpeno sintasa (clon GF002-G3) era el clon previamente identificado GFTpsD1. El segundo (clon GF006-G7) era un ADNc truncado que codificaba una nueva sesquiterpeno sintasa que se denominó GFTpsE. Este clon estaba de nuevo truncado en el extremo 5' (en aproximadamente 800 pb) y el fragmento ausente podría amplificarse mediante 5'-RACE en dos estadios de la siguiente manera. Un primer 5'-RACE usando los cebadores GFTpsERR1 y GFTpsERR2 (véase Tabla 1) proporcionó 500 nucleótidos del extremo 5' adicionales y un segundo 5'-RACE usando los cebadores GFTpsERR3 y GFTpsERR4 (véase Tabla 1) proporcionó la secuencia de ADN del extremo 5' ausente para reconstituir el ADNc de GFTpsE de longitud completa (Figura 6).

#### Ejemplo 4: Construcción de plásmidos y expresión de enzimas

##### Construcción de plásmidos de expresión

El ADNc se subclonó en el plásmido de expresión pET11a (Novagen) para expresión funcional de las sesquiterpeno sintasas. Para GFTpsB, GFTpsD y GFTpsE, los ADNc de longitud completa se amplificaron mediante PCR para introducir un sitio NdeI en el extremo 5' incluyendo el codón de inicio y un sitio BamHI en el extremo 3' inmediatamente después del codón de terminación. Para GFTpsB, se usaron el cebador directo GFTpsBNdeI (SEQ ID NO: 56) 5'-GCATGTTCCATATGTCCGCTCAAGTTCTAGCAACGGTTTCC-3' (sitio NdeI en cursiva, codón de inicio subrayado) y el cebador inverso GFTpsBBam (SEQ ID NO: 57) 5'-CGCGGATCCTCAGATGGTAACAGGGTCTCTGAGCACTGC-3' (sitio BamHI en cursiva, codón de terminación subrayado). De la misma manera, se usaron el cebador directo GFTpsDNdeI (SEQ ID NO: 58) 5'-GCATGTTCCATATGTCTGAGAAACATTTCTGTC-3' y el cebador inverso GFTpsDBam (SEQ ID NO: 59) 5'-CGCGGATCCTCAAATGGAACGTGGTCTCTAG-3' para GFTpsD y se usaron el cebador directo GFTpsENdeI (SEQ ID NO: 60) 5'-GCATGTTCCATATGTCTTTGGAAGTTTCAGCCTCTCCTG-3' y el cebador inverso GFTpsEBam (SEQ ID NO: 61) 5'-CGCGGATCCTCATATCGGCACAGGATTAATAACAAGAAGC-3' para GFTpsE.

Las amplificaciones se realizaron usando la ADN polimerasa Pfu (Promega) en un volumen final de 50 µl que contenía 5 µl de ADN polimerasa Pfu tampón 10X, 200 µM de cada dNTP, 0,4 µM de cada cebador directo e inverso, 2,9 unidades de ADN polimerasa Pfu y 5 µl de ADNc diluido 100 veces (preparado como se ha descrito anteriormente usando el kit de amplificación de ADNc Marathon™ (Clontech)). Las condiciones de termociclación fueron las siguientes: 2 min a 95 °C; 25 ciclos de 30 s a 95 °C, 30 s a 55 °C y 4 min a 72 °C; y 10 min a 72 °C. El producto de PCR se purificó en un gel de agarosa, se diluyó usando el kit de extracción en gel QIAquick® (Qiagen), se digirió con NdeI y BamHI y se ligó en el plásmido pET11a digerido de forma similar.



Para construcción del vector de expresión GFTpsC-pET11, se emplearon dos PCR separadas para generar el inserto flanqueado con los extremos cohesivos NdeI y BamHI. Se realizó una primera PCR en la misma condición que se ha descrito anteriormente, usando el cebador directo GFTpsCpETF1 (SEQ ID NO: 62) 5'-TAATGGCACTTCAAGATTCAGAAGTTCCTC-3' y el cebador inverso GFTpsCpETR1 (SEQ ID NO: 63) 5'-AAAAGGGAACAGGCTTCTCAAGCAATG-3' y se realizó una segunda PCR usando el cebador directo GFTpsCpETF2 (acortado por AT en el extremo 5' en comparación con GFTpsCpETF1) (SEQ ID NO: 64) 5'-ATGGCACTTCAAGATTCAGAAGTTCCTC-3' y el cebador inverso GFTpsCpETR2 (extendido por GATC en el extremo 5' en comparación con GFTpsCpETR1) (SEQ ID NO: 65) 5'-GATCAAAGGGAACAGGCTTCTCAAGCAATG-3'. Los dos productos de PCR se purificaron como se ha descrito ante, se combinaron, se desnaturalizaron mediante hervido de 5 min y se enfriaron 5 min en hielo. El ADN resultante se usó directamente para ligamiento en el plásmido pET11 a.

Los productos de ligamientos se transformaron inicialmente en células *E. coli* JM109 y las construcciones se verificaron mediante digestión de restricción y secuenciación de ADN.

#### *Expresión de sesquiterpeno sintasas*

Para expresión de proteínas, los plásmidos pET11 a que contenían los ADNc de sesquiterpeno sintasa así como el plásmido pET11 a vacío se transformaron en las células *E. coli* BL21 (DE3) (Novagen). Se usaron colonias individuales para inocular 5 ml de medio LB. Después de 5 a 6 horas de incubación a 37 °C, los cultivos se transfirieron a un incubador a 20 °C y se dejaron durante 1 h para equilibrado. Después se indujo la expresión de la proteína mediante la adición de IPTG 0,5 mM y el cultivo se incubó durante una noche a 20 °C.

Al día siguiente, las células se recogieron por centrifugación, se resuspendieron en 0,5 ml de tampón de extracción (MOPSO 50 mM pH 7, EDTA 5 mM, EDTA 5 mM, glicerol 10 %) y se sonicó 3 veces durante 30 s. Los residuos celulares se sedimentaron por centrifugación durante 30 min a 18.000 g y el sobrenadante que contenía las proteínas solubles se recuperó. La expresión de las sesquiterpeno sintasas se evaluó mediante separación del extracto proteico en un SDS-PAGE (electroforesis en gel de poliacrilamida SDS) y se tiñó con azul de coomassie y comparación con el extracto de proteína obtenido de células transformadas con el plásmido vacío.

Se observó una banda definida con el peso molecular calculado esperado para todas las construcciones y la banda no estaba presente en las proteínas solubles de *E. coli* transformadas con el plásmido vacío.

#### **Ejemplo 5: ensayo de función enzimática**

Los ensayos enzimáticos se realizaron en tubos de vidrio sellados usando de 50 a 100 µl de extracto proteico en un volumen final de tampón de extracción complementado con MgCl<sub>2</sub> 15 mM y FPP de 100 a 250 µM (Sigma). El medio se superpuso con 1 ml de pentano y los tubos se incubaron durante una noche a 30 °C. La fase de pentano, que contenía los sesquiterpenos, se recuperó y el extracto de medio con un segundo volumen de pentano. Las fracciones de pentano combinadas se concentraron con nitrógeno y se analizaron mediante cromatografía de gases en un sistema GC de serie 6890 de Hewlett-Packard usando un diámetro interno de 0,25 mm mediante columna capilar SPB-1 de 30 m (Supelco). El gas transportador era He a flujo constante de 1,6 ml/min. Se realizó inyección a una relación de división de 2:1 con la temperatura de inyector establecida en 200 °C y el horno programado de 80 °C (0 min de mantenimiento) a 7,5 °C/min hasta 200 °C (0 min de mantenimiento) seguido de 20 °C/min hasta 280 °C (2 min de mantenimiento). Se realizó detección con un detector de ionización de llama. La identificación de compuestos se basó en la identidad del tiempo de retención con patrones auténticos cuando estén disponibles. Para confirmación de las identidades de productos, se analizaron muestras mediante GC-MS capilar combinada usando un sistema detector selectivo de masas cuadrupolo GC 6890 de Hewlett-Packard, equipado con un diámetro interno de 0,25 mm por una columna capilar de SPB-1 de 30 m (Supelco). El horno se programó de 80 °C (0 min de mantenimiento) a 280 °C a 7,5 °C a un flujo constante de 1,5 ml/min de He. Los espectros se registraron a 70 eV con una tensión multiplicadora de electrones de 2.200 V.

La actividad enzimática de las diferentes enzimas recombinantes se evaluó en este ensayo usando la fracción de proteína soluble y farnesil-difosfato como sustrato. La actividad sesquiterpeno sintasa se obtuvo para todos los clones ensayados y los productos formados se caracterizaron por tiempo de retención y GC-MS (Figura 7).

El ADNc de GFTpsB codificó una sesquiterpeno sintasa que producía biciclo-germacreno (Figura 3) como un producto principal en al menos 15 olefinas de sesquiterpeno menores o sesquiterpenos oxigenados tales como delta-cadineno (Figura 3) (biciclo-elemento (Figura 3), resultante del reordenamiento por calor de biciclo-germacreno también se observó en la traza de GC). El ADNc de GFTpsC codificaba un producto múltiple que formaba sesquiterpeno sintasa identificándose un producto importante como cubebol (Figura 3). La enzima también produjo 3 sesquiterpenos adicionales en una proporción relativamente grande, (-)-α-cubebeno y sesquiterpenos 2 oxigenados y en cantidades pequeñas, al menos 11 olefinas de sesquiterpeno o sesquiterpenos oxigenados. GFTpsD codificaba una sesquiterpeno sintasa que produce valenceno como un producto importante. Se identificaron también otros diez picos menores como olefinas de sesquiterpeno o alcoholes. Entre ellos, β-elemento es un producto de degradación por calor de germacreno A. GFTpsE codificaba una sesquiterpeno sintasa produciendo una mezcla compleja de compuestos de sesquiterpeno siendo δ-cadineno (Figura 3) ligeramente la más abundante. También se identificaron

cubebol y germacreno D (Figura 3) en la mezcla de productos formados por GFTpsE.

Las sesquiterpeno sintasas aisladas demostraron múltiples propiedades formadoras de producto en los experimentos descritos. Por ejemplo, la cubebol sintasa y el δ-cadineno producen 3 o 5 productos en proporción relativamente grande. La bicilogermacreno sintasa produjo un sesquiterpeno importante y varios productos secundarios en cantidades traza.

5

LISTADO DE SECUENCIAS

<110> Firmenich SA Michel, Schalk Anthony, Clark

10 <120> Sesquiterpeno sintasas y procedimientos de uso

<130> 5770-EUR-D1

15 <140> PCT/IB/03/005072

<141> 02-10-2003

<150> US 60/415.765

<151> 04-10-2002

20 <150> PCT/IB02/05070

<151> 02-12-2002

<160> 65

25 <170> PatentIn versión 3.1

<210> 1

<211> 556

<212> PRT

30 <213> *Citrus paradisi*

<400> 1

Met Ala Leu Gln Asp Ser Glu Val Pro Ser Ser Ile Leu Asn Ala Thr  
 1 5 10 15

Ala Gly Asn Arg Pro Thr Ala Ser Tyr His Pro Thr Leu Trp Gly Gly  
 20 25 30

Lys Phe Leu Asp Tyr Ser Ser Val Asp Asp Ser Glu Ala Met Asp Ala  
 35 40 45

ES 2 607 122 T3

Thr Ile Asp Gln Asp Glu Phe Glu Ala Leu Lys Gln Lys Ile Lys Asn  
 50 55 60  
 Met Leu Ile Ser Pro Thr Asp Lys Ser Phe Gln Lys Leu Asn Leu Ile  
 65 70 75 80  
 Asp Ala Val Gln Arg Leu Gly Val Ala Tyr His Phe Glu Arg Glu Ile  
 85 90 95  
 Glu Asp Glu Leu Glu Lys Leu Ser Pro Asp Glu Tyr Asp Gly Asn Asp  
 100 105 110  
 Val His Ser Val Ala Leu Arg Phe Arg Leu Leu Arg Gln Gln Gly Tyr  
 115 120 125  
 Arg Ile Ser Cys Asp Ile Phe Gly Gly Phe Lys Asp Asp Arg Gly Lys  
 130 135 140  
 Phe Lys Val Ser Leu Ile Asn Asp Val Thr Gly Met Leu Ser Leu Tyr  
 145 150 155 160  
 Glu Ala Ala His Leu Arg Ile Arg Gly Glu Asp Ile Leu Asp Glu Ala  
 165 170 175  
 Leu Ala Phe Thr Thr Ser His Leu Glu Ser Met Val Thr Gln Val Ser  
 180 185 190  
 Pro Gln Leu Ser Asp Glu Ile Leu His Ala Leu Asn Arg Pro Ile Arg  
 195 200 205  
 Arg Gly Leu Pro Arg Leu Glu Ala Val Tyr Tyr Ile Asp Leu Tyr Ser  
 210 215 220  
 Arg Asp Asp Ser Lys Asp Lys Ala Ile Leu Leu Lys Phe Ala Lys Leu  
 225 230 235 240  
 Asp Phe Cys Met Leu Gln Val Ile His Arg Lys Glu Leu Ser Ile Ile  
 245 250 255  
 Thr Glu Trp Trp Lys Asn Leu Asp Val Glu Ile Asn Leu Pro Tyr Ala  
 260 265 270  
 Arg Asn Arg Val Val Glu Cys Tyr Phe Trp Ala Met Gly Val Tyr Phe  
 275 280 285  
 Glu Pro Arg Tyr Ser Phe Ala Arg Lys Ile Leu Ser Lys Val Ile Ala  
 290 295 300  
 Met Ala Ser Ile Leu Asp Asp Thr Tyr Asp Ala Tyr Gly Thr Leu Glu  
 305 310 315 320

ES 2 607 122 T3

Glu Leu Glu Leu Phe Thr Asn Ala Ile Lys Arg Trp Asp Ile Ser Asn  
 325 330 335  
 Ile Asp Val Leu Pro Lys Tyr Met Lys Leu Ile Tyr Gln Gly Leu Leu  
 340 345 350  
 Asp Val Phe Gly Glu Ala Glu Glu Ile Ser Lys Glu Gly Gln Thr  
 355 360 365  
 Tyr Cys Met Ser Tyr Val Ile Gln Ala Val Lys Lys Val Val Gln Ala  
 370 375 380  
 Tyr Phe Glu Glu Ala Lys Trp Cys Ser Glu Gly Tyr Phe Pro Lys Val  
 385 390 400  
 Glu Glu Tyr Met Gln Val Ser Leu Val Thr Thr Cys Tyr His Met Leu  
 405 410 415  
 Ala Thr Ala Ser Phe Leu Gly Met Gly Lys Ile Ala Asp Lys Gln Ala  
 420 425 430  
 Phe Glu Trp Ile Ser Asn Tyr Pro Lys Thr Val Lys Ala Ser Gln Val  
 435 440 445  
 Ile Cys Arg Leu Met Asp Asp Ile Val Ser His Glu Phe Glu Gln Lys  
 450 455 460  
 Arg Lys His Val Ala Ser Gly Ile Glu Cys Tyr Met Lys Gln His Gly  
 465 470 475 480  
 Val Ser Asp Glu Glu Val Ile Lys Val Phe Arg Lys Gln Ile Ser Asn  
 485 490 495  
 Gly Trp Lys Asp Val Asn Glu Gly Phe Met Lys Pro Thr Glu Val Ala  
 500 505 510  
 Met Pro Leu Leu Glu Arg Ile Leu Asn Leu Ala Arg Val Ile Asp Val  
 515 520 525  
 Ile Tyr Lys Asp Asp Asp Gly Tyr Thr Asn Ser Tyr Val Ile Lys Asp  
 530 535 540  
 Tyr Ile Ala Thr Leu Leu Glu Lys Pro Val Pro Phe  
 545 550 555

<210> 2  
 <211> 562  
 <212> PRT  
 <213> *Citrus paradisi*  
 <400> 2

5

ES 2 607 122 T3

Met Ser Leu Glu Val Ser Ala Ser Pro Ala Lys Val Ile Gln Asn Ala  
1 5 10 15

Gly Lys Asp Ser Thr Arg Arg Ser Ala Asn Tyr His Pro Ser Ile Trp  
20 25 30

Gly Asp His Phe Leu Gln Tyr Thr Cys Asp Thr Gln Glu Thr Asp Asp  
35 40 45

Gly Ser Asn Val Lys His Leu Glu Leu Lys Lys Glu Ile Arg Arg Met  
50 55 60

Leu Lys Ala Asp Asn Lys Pro Ser Arg Thr Leu Gln Leu Ile Asp Ala  
65 70 75 80

Ile Gln Arg Leu Gly Val Ser Tyr His Phe Glu Ser Glu Ile Asp Glu  
85 90 95

Ile Leu Gly Lys Met His Lys Ala Ser Gln Asp Ser Asp Leu Cys Asp  
100 105 110

Asn Glu Asn Asp Glu Leu Tyr Tyr Ile Ser Leu His Phe Arg Leu Leu  
115 120 125

Arg Gln Asn Gly Tyr Lys Ile Ser Ala Asp Val Phe Lys Lys Phe Lys  
130 135 140

Asp Thr Asp Gly Asn Phe Lys Thr Ser Leu Ala Lys Asp Val Arg Gly  
145 150 155 160

Met Leu Ser Leu Tyr Glu Ala Thr His Leu Gly Val His Glu Glu Asp  
165 170 175

Ile Leu Asp Glu Ala Leu Ala Phe Thr Thr Ser His Leu Glu Ser Ile  
180 185 190

Ala Thr His Gln Ile Arg Ser Pro Leu Val Glu Gln Val Lys His Ala  
195 200 205

Leu Val Gln Pro Ile His Arg Gly Phe Gln Arg Leu Glu Ala Arg Gln  
210 215 220

Tyr Ile Pro Ile Tyr Gln Glu Glu Ser Pro His Asn Glu Ala Leu Leu  
225 230 235 240

Thr Phe Ala Lys Leu Asp Phe Asn Lys Leu Gln Lys Pro His Gln Lys  
245 250 255

Glu Leu Gly Asp Ile Ser Arg Trp Trp Lys Glu Leu Asp Phe Ala His  
260 265 270

ES 2 607 122 T3

Lys Leu Pro Phe Ile Arg Asp Arg Val Ala Glu Cys Tyr Phe Trp Ile  
 275 280 285  
 Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Phe Ala Arg Arg Ile Leu  
 290 300  
 Thr Lys Val Ile Ser Met Thr Ser Val Ile Asp Asp Ile Tyr Asp Val  
 305 310 315 320  
 Tyr Gly Lys Ile Glu Glu Leu Glu Leu Phe Thr Ser Ala Ile Glu Arg  
 325 330 335  
 Trp Asp Ile Ser Ala Ile Asp Gln Leu Pro Glu Tyr Met Lys Leu Cys  
 340 345 350  
 Tyr Arg Ala Leu Leu Asp Val Phe Ser Glu Ala Glu Lys Asp Leu Ala  
 355 360 365  
 Pro Gln Gly Lys Ser Tyr Arg Leu Tyr Tyr Ala Lys Glu Ala Met Lys  
 370 375 380  
 Asn Met Val Lys Asn Tyr Phe Tyr Glu Ala Lys Trp Cys Leu Gln Asn  
 385 390 395 400  
 Tyr Val Pro Thr Val Asp Glu Tyr Met Thr Val Ala Leu Val Thr Ser  
 405 410 415  
 Gly Ser Pro Met Leu Ser Thr Thr Ser Phe Val Gly Met Gly Asp Ile  
 420 425 430  
 Val Thr Lys Glu Ser Phe Glu Trp Leu Phe Ser Asn Pro Arg Phe Ile  
 435 440 445  
 Arg Ala Ser Ser Ile Val Cys Arg Leu Met Asp Asp Ile Val Ser His  
 450 455 460  
 Lys Phe Glu Gln Ser Arg Gly His Val Ala Ser Ser Val Glu Cys Tyr  
 465 470 475 480  
 Met Lys Gln His Gly Ala Thr Glu Glu Glu Ala Cys Asn Glu Phe Arg  
 485 490 495  
 Lys Gln Val Ser Asn Ala Trp Lys Asp Ile Asn Glu Asp Cys Leu Arg  
 500 505 510  
 Pro Thr Val Val Pro Met Pro Leu Leu Met Arg Ile Leu Asn Leu Thr  
 515 520 525  
 Arg Val Ile Asp Val Ile Tyr Lys Tyr Glu Asp Gly Tyr Thr His Ser  
 530 535 540

ES 2 607 122 T3

Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val  
 545 550 555 560

Pro Ile

5

<210> 3  
 <211> 555  
 <212> PRT  
 <213> *Citrus paradisi*

<400> 3

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

Val Arg Pro Ile Ala Gly Phe His Pro Asn Leu Trp Gly Asp Tyr Phe  
 20 25 30

Leu Thr Leu Ala Ser Asp Cys Lys Thr Asp Asp Thr Thr His Gln Glu  
 35 40 45

Glu Tyr Glu Ala Leu Lys Gln Glu Val Arg Ser Met Ile Thr Ala Thr  
 50 55 60

Ala Asp Thr Pro Ala Gln Lys Leu Gln Leu Val Asp Ala Val Gln Arg  
 65 70 75 80

Leu Gly Val Ala Tyr His Phe Glu Gln Glu Ile Glu Asp Ala Met Glu  
 85 90 95

Lys Ile Tyr His Asp Asp Phe Asp Asn Asn Asp Asp Val Asp Leu Tyr  
 100 105 110

Thr Val Ser Leu Arg Phe Arg Leu Leu Arg Gln Gln Gly Phe Lys Val  
 115 120 125

Pro Cys Asp Val Phe Ala Lys Phe Lys Asp Asp Glu Gly Lys Phe Lys  
 130 135 140

Ala Ser Leu Val Arg Asp Val His Gly Ile Leu Ser Leu Tyr Glu Ala  
 145 150 155 160

Gly His Leu Ala Ile Arg Gly Glu Gly Ile Leu Asp Glu Ala Ile Ala  
 165 170 175

Phe Thr Arg Thr His Leu Gln Ser Met Val Ser Gln Asp Val Cys Pro  
 180 185 190

10

ES 2 607 122 T3

Asn Asn Leu Ala Glu Gln Ile Asn His Thr Leu Asp Cys Pro Leu Arg  
 195 200 205

Arg Ala Leu Pro Arg Val Glu Thr Arg Phe Phe Leu Ser Val Tyr Pro  
 210 215 220

Arg Asp Asp Lys His Asp Lys Thr Leu Leu Lys Phe Ser Lys Leu Asp  
 225 230 240

Phe Asn Leu Val Gln Arg Ile His Gln Lys Glu Leu Ser Ala Ile Thr  
 245 250 255

Arg Trp Trp Lys Asp Leu Asp Phe Thr Thr Lys Leu Pro Tyr Ala Arg  
 260 265 270

Asp Arg Ile Val Glu Leu Tyr Phe Trp Ile Val Gly Thr Tyr Phe Glu  
 275 280 285

Pro Lys Tyr Thr Leu Ala Arg Lys Ile Met Thr Lys Thr Ile Tyr Thr  
 290 295 300

Ala Ser Ile Ile Asp Asp Thr Phe Asp Ala Tyr Gly Phe Phe Glu Glu  
 305 310 315 320

Leu Lys Leu Phe Ala Glu Ala Val Gln Arg Trp Asp Ile Gly Ala Met  
 325 330 335

Asp Ile Leu Pro Glu Tyr Met Lys Val Leu Tyr Lys Ala Leu Leu Asp  
 340 345 350

Thr Phe Asn Glu Ile Glu Gln Asp Leu Ala Lys Glu Gly Arg Ser Ser  
 355 360 365

Tyr Leu Pro Tyr Gly Lys Glu Lys Met Gln Glu Leu Val Gln Met Tyr  
 370 375 380

Phe Val Gln Ala Lys Trp Phe Ser Glu Gly Tyr Val Pro Thr Trp Asp  
 385 390 395 400

Glu Tyr Tyr Pro Val Gly Leu Val Ser Cys Gly Tyr Phe Met Leu Ala  
 405 410 415

Thr Asn Ser Phe Leu Gly Met Cys Asp Val Ala Asn Glu Glu Ala Phe  
 420 425 430

Glu Trp Ile Ser Lys Asp Pro Lys Ile Ser Thr Ala Ser Ser Val Ile  
 435 440 445

Cys Arg Leu Arg Asn Asp Ile Val Ser His Gln Phe Glu Gln Lys Arg  
 450 455 460

Gly His Ile Ala Ser Gly Phe Glu Cys Tyr Ile Lys Gln Tyr Gly Val





ES 2 607 122 T3

Leu Ser Leu Tyr Glu Ala Ala Tyr Met Ala Val Arg Gly Glu His Ile  
 145 150 155 160  
 Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val  
 165 170 175  
 Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala  
 180 185 190  
 Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr  
 195 200 205  
 Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Tyr Asn Lys Thr  
 210 215 220  
 Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Leu His  
 225 230 235 240  
 Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe  
 245 250 255  
 Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe  
 260 265 270  
 Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys  
 275 280 285  
 Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr  
 290 295 300  
 Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val  
 305 310 315 320  
 Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys  
 325 330 335  
 Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp  
 340 345 350  
 Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu  
 355 360 365  
 Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser  
 370 375 380  
 Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr  
 385 390 395 400  
 Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly  
 405 410 415

ES 2 607 122 T3

Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys  
 420 425 430  
 Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln  
 435 440 445  
 Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu  
 450 455 460  
 Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met  
 465 470 475 480  
 Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu  
 485 490 495  
 Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu  
 500 505 510  
 Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr  
 515 520 525  
 Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp  
 530 535 540

His Val Pro Phe

545

<210> 5  
 <211> 548  
 <212> PRT  
 <213> *Citrus paradisi*  
 <400> 5

5

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

10

ES 2 607 122 T3

Glu Ile Glu Asp Ala Ile Gln Lys Leu Cys Pro Asn Tyr Ile His Ser  
 85 90 95  
 Asn Ser Pro Asp Leu His Thr Val Ser Leu His Phe Arg Leu Leu Arg  
 100 105 110  
 Gln Gln Gly Ile Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys Asp  
 115 120  
 Asp Glu Gly Arg Phe Lys Ser Ser Leu Ile Asn Asp Val Gln Gly Met  
 130 135  
 Leu Ser Leu Tyr Glu Ala Ala Tyr Met Ala Val Arg Gly Glu His Ile  
 145 150 155 160  
 Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val  
 165 170 175  
 Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala  
 180 185 190  
 Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr  
 195 200 205  
 Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Tyr Asn Lys Thr  
 210 215 220  
 Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Leu His  
 225 230 235 240  
 Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe  
 245 250 255  
 Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Leu Val Glu Leu Tyr Phe  
 260 265 270  
 Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys  
 275 280 285  
 Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr  
 290 295 300  
 Asp Ala Tyr Gly Thr Leu Glu Glu Leu Ser Leu Phe Thr Glu Ala Val  
 305 310 315 320  
 Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys  
 325 330 335  
 Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp  
 340 345 350  
 Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu



ES 2 607 122 T3

gatgccgtcc aacgcttagg agtggcttac cattttgaga gggagataga agatgaacta 300  
gaaaaactat ctctgatga gtatgatggc aacgatgtac actccgttgc tcttcgattt 360  
cggttactca gacaacaagg atatcgcata tcatgcgata tttttggcgg tttcaaagat 420  
gatcgaggaa agttcaaggt atccttaatt aatgatgtga ccggcatgct aagtttgtat 480  
gaggctgcac atcttcgcat tcgcggggaa gatatacctgg atgaagccct agctttcact 540  
acttctcacc tggaatcaat ggttactcaa gtaagccctc agctttctga tgaataactt 600  
catgccttga ataggccaat ccgcagaggc ttaccaaggc tggaggcagt ctattacatc 660  
gatctctact cacgagatga ttcaaaggat aaagcaatat tactaaagtt tgcaaaacta 720  
gatttttgca tgcttcaagt aattcaccgt aaggagttaa gtatcatcac agagtgggtg 780  
aaaaatthag atgttgaaat aaatctccca tatgctagaa acagagttgt agaatgctat 840  
ttttgggcaa tgggagtgta ttttgagcct cgatactcct ttgcaagaaa gatattgtcc 900  
aaagtaattg caatggcatc catttttagat gatacctacg acgcctatgg cacacttgaa 960  
gaacttgagc tctttacaaa tgctatcaaa aggtgggata ttagcaacat agatgtactt 1020  
ccgaagtaca tgaaactgat ttatcaagga ctcttggatg tttttggtga agctgaggag 1080  
gaaatctcaa aggaaggaca gacatattgc atgtcatatg tcatacaagc ggtgaagaaa 1140  
gtagtccaag cctactttga ggaagccaag tgggtgcagtg aaggttattt tccaaaagtg 1200  
gaggagtata tgcaagtttc acttgtgaca acttgctatc atatgctggc aacggcttct 1260  
tttcttggca tgggaaagat tgctgataag caggcctttg aatggatctc caattaccct 1320  
aaaactgtga aagcctccca agttatttgc agacttatgg atgatatagt gtctcacgag 1380  
tttgaacaaa aaagaaagca tgttgccctg ggtattgaat gttacatgaa gcagcatggc 1440  
gtctctgatg aagaggtaat taaagtattc cgcaaacaaa tatcaaatgg atggaaagat 1500  
gtaaatgaag gattcatgaa gccaacagaa gtggcaatgc ctctccttga gcgcattctc 1560  
aatcttgac gagtgataga tgttatttac aaggatgatg atggctacac caactcttat 1620  
gtgatcaaag actacatcgc cacattgctt gagaagcctg ttccttttg a 1671

<210> 7  
<211> 1689  
<212> ADN  
<213> *Citrus paradisi*

<400> 7

atgtctttgg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct 60  
actcgtcgtc ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120  
tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180  
attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240

5

10

ES 2 607 122 T3

attcagcgtt taggagtgtc ttaccatfff gaaagtgaga ttgatgaaat attgggaaag 300  
 atgcataagg cttcccaaga ctctgatcct tgtgataatg aaaatgatga gctctattat 360  
 atctctcttc attttcgatt acttagacaa aatggctata aaatttccgc tgatgtgttc 420  
 aaaaagtca aagacacgga tgggaacttt aaaacatctc ttgcgaaaga tgttcgagga 480  
 atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540  
 gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat caggtctcca 600  
 cttgttgaac aagtcaaaca tgccttagtt cagcctatcc acaggggctt ccaaaggctt 660  
 gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720  
 acttttgcaa agttagatff taacaaattg caaaagcctc accagaagga actcggtgat 780  
 atttcaaggt ggtggaaaga attagactff gcacataagc tacctttcat aagagataga 840  
 gttgcggagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca 900  
 agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960  
 tatggcaaaa ttgaagaact tgagctffff acttcagcta ttgagaggtg ggatatcagt 1020  
 gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgtffff 1080  
 agtgaagcag agaaggatff ggcccccaa ggaaaatcat accgcctcta ttatgcaaaa 1140  
 gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200  
 tatgtaccta cagtggatga gtacatgacg gttgcattag ttacatctgg ctccccaatg 1260  
 ttgtcaacca catcctttgt tggcatggga gacattgtaa ctaaagaatc ttttgagtgg 1320  
 ttattcagca atcctagatt tattagggct tcttctatag tttgccgact catggatgac 1380  
 atagtgtcac acaagtttga acaaagcaga gggcacgttg cctcaagcgt tgagtgttac 1440  
 atgaaacaac atggagcaac agaagaggaa gcatgcaatg agtttcggaa acaagtttca 1500  
 aatgcctgga aggatataaa tgaggactgc ctacgccccaa cggttgtgcc aatgccactt 1560  
 ctgatgcgaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc 1620  
 tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680  
 ccgatatga 1689

<210> 8  
 <211> 1668  
 <212> ADN  
 <213> *Citrus paradisi*

5

<400> 8

atgtccgctc aagttctagc aacggtttcc agttcgacag aaaaaactgt tcgtcccatt 60  
 gctggtttcc atcctaactt atggggagac tatttctga ccctcgcttc tgattgcaag 120  
 acagatgata ctacgcacca agaggaatac gaagcgctga agcaagaagt cagaagcatg 180  
 ataacggcta cggcagatag acctgcccag aagttgcaat tggttgatgc agtccaacga 240

10

ES 2 607 122 T3

ttgggtgtgg cctatcactt cgaacaggag atagaagatg caatggaaaa gatttatcac 300  
 gatgactttg ataataacga tgatgtcgat ctctacactg tttctcttcg ttttcgactg 360  
 cttaggcagc aaggatttaa ggttccgtgt gatgtgttcg cgaagttcaa agatgatgaa 420  
 ggtaaattca aggcattcatt ggtgcgggat gttcatggca ttctaagttt gtatgaggca 480  
 ggacacttgg ccattcgcgg agaaggata ttagatgaag ccattgcttt cactagaact 540  
 caccttcagt caatggtatc tcaggatgta tgccctaata atcttgctga acaaattaat 600  
 catactctcg actgtcctct ccgcagagcc cttccaagag tggagacaag attttcttg 660  
 tcggtctatc caagagatga taaacacgat aaaactttgt taaagttttc aaagttagac 720  
 tttaaccttg tgcaaagaat acatcagaag gaattaagtg ccatcacacg gtggtggaaa 780  
 gatttagact tcaactacaaa gctaccttat gcaagagaca gaatcgtaga gttgtatttt 840  
 tggattgtag ggacgtattht tgaaccaaag tacactttag caagaaaaat aatgaccaa 900  
 acaatttaca cggcatctat catagatgac actttcgacg cttatggttt ctttgaagag 960  
 ctcaaaactct ttgcagaagc agtccagagg tgggacattg gagccatgga tatacttcca 1020  
 gaatacatga aagtgcctta taaggccctt ttagatactt tcaatgaaat tgagcaagac 1080  
 ttggccaagg aaggaagatc gtcctactta ctttatggca aagaaaagat gcaagagctt 1140  
 gttcaaagtgt actttgttca agccaagtgg ttcagtgaag gttatgttcc gacatgggac 1200  
 gaatattatc cggttggact tgtaagttgc ggctacttca tgcttgcgac aaactccttc 1260  
 cttggcatgt gtgatgttgc aaacgaggaa gcttttgaat ggatatcaa ggaccctaag 1320  
 atttcaacag cgtcatcagt tatctgcaga cttaggaatg acattgtttc ccaccagttt 1380  
 gaacagaaga gaggacatat tgcctcagga tttgaatgct acattaagca gtatggtggt 1440  
 tcagaagaag aggtagttag agtttttact gaagaagttg agaatgcatg gaaagatag 1500  
 aatgaggaat tcctgaaacc aactgctttt cctgtggctt tgattgagag acctttcaat 1560  
 atcgcacgtg tgattgaatt tctaaacaag aagggtgatt ggtacactca ttctcatgcg 1620  
 attaaagacc agattgccgc agtgctcaga gaccctgtta ccatctga 1668

<210> 9  
 <211> 1647  
 <212> ADN  
 <213> *Citrus paradisi*

5

<400> 9

atgtcgtctg gagaacatt tcgtcctact gcagatttcc atcctagttt atggagaaac 60  
 catttctca aaggtgcttc tgatttcaag acagttgatc atactgcaac tcaagaacga 120  
 cacgaggcac tgaaagaaga ggtaaggaga atgataacag atgctgaaga taagcctggt 180  
 cagaagttac gcttgattga tgaagtacaa cgcctggggg tggcttatca ctttgagaaa 240

10



ES 2 607 122 T3

gaaatagaag atgcaataca aaaattatgt ccaatctata ttgacagtaa tagagctgat 300  
 ctccacaccg tttcccttca ttttcgattg cttaggcagc aaggaatcaa gatttcatgt 360  
 gatgtgtttg agaagttcaa agatgatgag ggtagattca agtcatcggt gataaacgat 420  
 gttcaaggga tgtaagttt gtacgaggca gcatacatgg cagttcgcgg agaacatata 480  
 ttagatgaag ccattgcttt cactaccact cacctgaagt cattggtagc tcaggatcat 540  
 gtaacccta agcttgcgga acagataaat catgctttat accgtcctct tcgtaaaacc 600  
 ctaccaagat tagaggcgag gtattttatg tccatgatca attcaacaag tgatcattta 660  
 tacaataaaa ctctgctgaa ttttgcaaag ttagatttta acatattgct agagctgcac 720  
 aaggaggaac tcaatgaatt aacaaagtgg tggaaagatt tagacttcac taaaaacta 780  
 ccttatgcaa gagacagatt agtggagtta tatttttggg atttagggac atacttcgag 840  
 cctcaatatg catttgggag aaagataatg acccaattaa attacatatt atccatcata 900  
 gatgatactt atgatgcgta tggtagactt gaagaactca gcctctttac tgaagcagtt 960  
 caaagatgga atattgaggc cgtagatatg cttccagaat acatgaaatt gatttacagg 1020  
 acactcttag atgcttttaa tgaaattgag gaagatatgg ccaagcaagg aagatcacac 1080  
 tgcgtacgtt atgcaaaaaga ggagaatcaa aaagtaattg gagcatactc tgttcaagcc 1140  
 aatgggttca gtgaagggtta cgttccaaca attgaggagt atatgcctat tgcactaaca 1200  
 agttgtgctt acacattcgt cataacaaat tccttccttg gcatgggtga ttttgcaact 1260  
 aaagaggttt ttgaatggat ctccaataac cctaagggtg taaaagcagc atcagttatc 1320  
 tgcagactca tggatgacat gcaaggatc gagtttgagc agaagagagg acatggttgcg 1380  
 tcagctattg aatgttacac gaagcagcat ggtgtctcta aggaagaggc aattaaatg 1440  
 tttgaagaag aagttgcaa tgcatggaaa gatattaacg aggagttgat gatgaagcca 1500  
 accgtcgttg cccgaccact gctcgggacg attcttaatc ttgctcgtgc aattgatttt 1560  
 atttacaag aggacgacgg ctatacgcat tcttacctaa ttaaagatca aattgcttct 1620  
 gtgctaggag accacgttcc attttga 1647

<210> 10  
 <211> 1647  
 <212> ADN  
 <213> *Citrus paradisi*

5

<400> 10

atgtcgtctg gagaacatt tcgtcctact gcagatttcc atcctagttt atggagaac 60  
 catttcctca aagggtgctt tgatttcaag acagttgatc atactgcaac tcaagaacga 120  
 cacgaggcac tgaagaaga ggtaaggaga atgataacag atgctgaaga taagcctgtt 180  
 cagaagttac gcttgattga tgaagtacaa cgcctggggg tggcttatca ctttgagaaa 240  
 gaaatagaag atgcaataca aaaattatgt ccaactata ttcacagtaa tagccctgat 300

10

ES 2 607 122 T3

cttcacaccg tttctcttca ttttcgattg cttaggcagc aaggaatcaa gatttcatgt 360  
gatgtgtttg agaagttcaa agatgatgag ggtagattca agtcatcggt gataaacgat 420  
gttcaaggga tgtaagttt gtacgaggca gcatacatgg cagttcgcgg agaacatata 480  
ttagatgaag ccattgcttt cactaccact cacctgaagt cattggtagc tcaggatcat 540  
gtaaccctta agcttgcgga acagataaat catgctttat accgtcctct tcgtaaaacc 600  
ctaccaagat tagaggcgag gtattttatg tccatgatca attcaacaag tgatcattta 660  
tacaataaaa ctctgctgaa ttttgcaaag ttagatttta acatattgct agagctgcac 720  
aaggaggaac tcaatgaatt aacaaagtgg tggaaagatt tagacttcac tacaaaacta 780  
ccttatgcaa gagacagatt agtggagtta tatttttggg attagggac atacttcgag 840  
cctcaatatg catttgggag aaagataatg acccaattaa attacatatt atccatcata 900  
gatgatactt atgatgcgta tggtagactt gaagaactca gcctctttac tgaagcagtt 960  
caaagatgga atattgaggc cgtagatatg cttccagaat acatgaaatt gatttacagg 1020  
acactcttag atgcttttaa tgaaattgag gaagatatgg ccaagcaagg aagatcacac 1080  
tgcgtacggt atgcaaaaaga ggagaatcaa aaagtaattg gagcatactc tgttcaagcc 1140  
aaatggttca gtgaaggta cgttccaaca attgaggagt atatgcctat tgcactaaca 1200  
agttgtgctt acacattcgt cataacaaat tccttccttg gcatgggtga ttttgcaact 1260  
aaagaggttt ttgaatggat ctccaataac cctaaggttg taaaagcagc atcagttatc 1320  
tgcagactca tggatgacat gcaaggatc gatgttgagc agaagagagg acatgtttgcg 1380  
tcagctattg aatgttacac gaagcagcat ggtgtctcta aggaagaggc aattaaatg 1440  
tttgaagaag aagttgcaa tgcatggaaa gatattaacg aggagttgat gatgaagcca 1500  
accgtcgttg cccgaccact gctcgggacg attcttaatc ttgctcgtgc aattgatttt 1560  
atttacaaag aggacgacgg ctatacgcat tcttacctaa ttaaagatca aattgcttct 1620  
gtgctaggag accacgttcc attttga 1647

<210> 11  
<211> 74  
<212> PRT  
<213> *Lycopersicon esculentum*

5

<400> 11

Phe His Gln Arg Glu Leu Ser Asp Leu Thr Arg Trp Trp Lys Asp Leu  
1 5 10 15

Asp Phe Ala Asn Lys Tyr Pro Tyr Ala Arg Asp Arg Leu Val Glu Cys  
20 25 30

Tyr Phe Trp Ile Leu Gly Val Tyr Phe Glu Pro Lys Tyr Ser Arg Ala

10

ES 2 607 122 T3

35 40 45

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

Thr Phe Asp Ala Tyr Ala Thr Phe Asp Glu  
65 70

5  
<210> 12  
<211> 74  
<212> PRT  
<213> *M. piperita*  
<400> 12

Leu Tyr Lys Glu Glu Leu Ser Gln Leu Ser Arg Trp Trp Asn Thr Trp  
1 5 10 15

Asn Leu Lys Ser Lys Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Ala  
20 25 30

Tyr Val Trp Gly Val Gly Tyr His Tyr Glu Pro Gln Tyr Ser Tyr Val  
35 40 45

Arg Met Gly Leu Ala Lys Gly Val Leu Ile Cys Gly Ile Met Asp Asp  
50 55 60

Thr Tyr Asp Asn Tyr Ala Thr Leu Asn Glu  
65 70

10  
15  
<210> 13  
<211> 73  
<212> PRT  
<213> *A. grandis*  
<400> 13

Lys His His Lys Glu Ile Gln Phe Ile Thr Arg Trp Trp Arg Asp Ser  
1 5 10 15

Gly Ile Ser Gln Leu Asn Phe Tyr Arg Lys Arg His Val Glu Tyr Tyr  
20 25 30

Ser Trp Val Val Met Cys Ile Phe Glu Pro Glu Phe Ser Glu Ser Arg  
35 40 45

Ile Ala Phe Ala Lys Thr Ala Ile Leu Cys Thr Val Leu Asp Asp Leu  
50 55 60

Tyr Asp Thr His Ala Thr Leu His Glu  
65 70

20

ES 2 607 122 T3

<210> 14  
 <211> 74  
 <212> PRT  
 <213> *Citrus junos*

5

<400> 14

Ile His Gln Lys Glu Leu Ser Ala Ile Thr Arg Trp Trp Lys Asp Leu  
 1 5 10 15  
 Asp Phe Thr Thr Lys Leu Pro Tyr Ala Arg Asp Arg Ile Val Glu Leu  
 20 25 30  
 Tyr Phe Trp Ile Val Gly Thr Tyr Phe Glu Pro Lys Tyr Thr Leu Ala  
 35 40 45  
 Arg Lys Ile Met Thr Lys Thr Ile Tyr Thr Ala Ser Ile Ile Asp Asp  
 50 55 60  
 Thr Phe Asp Ala Tyr Gly Phe Phe Glu Glu  
 65 70

10 <210> 15  
 <211> 73  
 <212> PRT  
 <213> *Nicotiana tabacum*

15 <400> 15

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

20 <210> 16  
 <211> 74  
 <212> PRT  
 <213> *C. annuum*

25 <400> 16

ES 2 607 122 T3

Leu His Lys Gln Glu Leu Ala Glu Val Ser Arg Trp Trp Lys Asp Leu  
 1 5 10 15

Asn Phe Val Asn Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys  
 20 25 30

Tyr Phe Trp Ala Leu Gly Val Tyr Tyr Glu Pro Gln Tyr Ser Gln Ala  
 35 40 45

Arg Val Met Leu Val Lys Thr Ile Ala Met Ile Ser Ile Val Asp Asp  
 50 55 60

Thr Tyr Asp Ala Tyr Gly Thr Val Asp Glu  
 65 70

5  
 <210> 17  
 <211> 74  
 <212> PRT  
 <213> *Solanum tuberosum*  
 <400> 17

Leu His Lys His Glu Leu Ser Glu Val Ser Arg Trp Trp Lys Asp Leu  
 1 5 10 15

Asp Phe Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Ala Val Glu Cys  
 20 25 30

Tyr Phe Trp Thr Met Gly Val Tyr Ala Glu Pro Gln Tyr Ser Gln Ala  
 35 40 45

Arg Val Ile Leu Ala Lys Thr Ile Ala Met Ile Ser Ile Val Asp Asp  
 50 55 60

10  
 Thr Phe Asp Ala Tyr Gly Ile Val Lys Glu  
 65 70

15  
 <210> 18  
 <211> 74  
 <212> PRT  
 <213> *H. muticus*  
 <400> 18

ES 2 607 122 T3

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

5  
 <210> 19  
 <211> 86  
 <212> PRT  
 <213> *G. arboreum*  
 <400> 19

Leu Glu Phe Ala Lys Ile Asp Phe Asn Met Val Gln Leu Leu His Arg  
 1 5 10 15  
 Lys Glu Leu Ser Glu Ile Ser Arg Trp Trp Lys Asp Leu Asp Phe Gln  
 20 25 30  
 Arg Lys Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Gly Tyr Phe Trp  
 35 40 45  
 Ile Ser Gly Val Tyr Phe Glu Pro Gln Tyr Ser Leu Gly Arg Lys Met  
 50 55 60  
 Leu Thr Lys Val Ile Ala Met Ala Ser Ile Val Asp Asp Thr Tyr Asp  
 65 70 75 80  
 Ser Tyr Ala Thr Tyr Glu  
 85

10  
 15  
 <210> 20  
 <211> 86  
 <212> PRT  
 <213> *A. annua*  
 <400> 20

ES 2 607 122 T3

Leu Lys Leu Ala Lys Leu Glu Phe Asn Leu Leu Gln Ser Leu His Lys  
 1 5 10 15  
 Glu Glu Leu Ser His Val Cys Lys Trp Trp Lys Ala Phe Asp Ile Lys  
 20 25 30  
 Lys Asn Ala Pro Cys Leu Arg Asp Arg Ile Val Glu Cys Tyr Phe Trp  
 35 40 45  
 Gly Leu Gly Ser Gly Tyr Glu Pro Gln Tyr Ser Arg Ala Arg Val Phe  
 50 55 60  
 Phe Thr Lys Ala Val Ala Val Ile Thr Leu Ile Asp Asp Thr Tyr Asp  
 65 70 75 80  
 Ala Tyr Gly Thr Tyr Glu  
 85

5  
 <210> 21  
 <211> 86  
 <212> PRT  
 <213> *A. annua*  
 <400> 21

Leu Lys Leu Ala Lys Leu Gly Phe Asn Gln Leu Gln Ser Leu His Lys  
 1 5 10 15  
 Lys Glu Leu Ser Ile Ile Ser Lys Trp Trp Lys Ser Phe Asp Val Ala  
 20 25 30  
 Asn Asn Leu Pro Tyr Ala Arg Asn Arg Pro Val Glu Cys Tyr Phe Trp  
 35 40 45  
 Ala Leu Ala Val Tyr Phe Glu Pro Gln Tyr Ser Glu Ser Arg Val Phe  
 50 55 60  
 Leu Ser Arg Phe Phe Ser Ile Gln Thr Phe Leu Asp Asp Thr Tyr Asp  
 65 70 75 80  
 Ala Tyr Gly Thr Tyr Glu  
 85

10  
 15  
 <210> 22  
 <211> 85  
 <212> PRT  
 <213> *A. grandis*  
 <400> 22

ES 2 607 122 T3

Leu Glu Leu Ala Lys Leu Asp Phe Asn Ile Leu Gln Cys Thr His Gln  
 1 5 10  
 Lys Glu Leu Gln Ile Ile Ser Arg Trp Phe Ala Asp Ser Ser Ile Ala  
 20 25 30  
 Ser Leu Asn Phe Tyr Arg Lys Cys Tyr Val Glu Phe Tyr Phe Trp Met  
 35 40 45  
 Ala Ala Ala Ile Ser Glu Pro Glu Phe Ser Gly Ser Arg Val Ala Phe  
 50 55 60  
 Thr Lys Ile Ala Ile Leu Met Thr Met Leu Asp Asp Leu Tyr Asp Thr  
 65 70 75 80  
 His Gly Thr Leu Asp  
 85

- 5
  - <210> 23
  - <211> 18
  - <212> ADN
  - <213> Secuencia artificial
- 10
  - <220>
  - <223> Cebador deducido (Ejemplo 1)
- 15
  - <400> 23
  - hhvthwcmmg gtggtgga 18
- 20
  - <210> 24
  - <211> 18
  - <212> ADN
  - <213> Secuencia artificial
- 25
  - <220>
  - <221 > misc\_feature
  - <222> (3)..(3)
  - <223> puede ser G, A, C o T
- 30
  - <400> 24
  - gtngarkbbt atktttg 18
- 35
  - <210> 25
  - <211> 22
  - <212> ADN
  - <213> Secuencia artificial
- 40
  - <220>
  - <223> Cebador deducido (Ejemplo 1)
- 40
  - <400> 25
  - crtrkrtrc gwadgkgtcr tc 22
- 40
  - <210> 26
  - <211> 25
  - <212> ADN
  - <213> Secuencia artificial



ES 2 607 122 T3

<220>  
 <223> Cebador deducido (Ejemplo 1)

5 <400> 26  
 gckaagwtvg gkttcaathw kbtcd 25

<210> 27  
 <211> 29  
 <212> ADN  
 10 <213> Secuencia artificial

<220>  
 <223> Cebador deducido (Ejemplo 1)

15 <220>  
 <221 > misc\_feature  
 <222> (9)..(9)  
 <223> puede ser a, t, g o c

20 <400> 27  
 gggawwgnw bgttgaakkt tattttgg 29

<210> 28  
 <211> 23  
 25 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Cebador deducido (Ejemplo 1)

30 <220>  
 <221 > misc\_feature  
 <222> (8)..(8)  
 <223> puede ser a, t, g o c

35 <400> 28  
 gtwscgtng h gtcgtahgtc atc 23

<210> 29  
 <211> 49  
 <212> PRT  
 <213> *Citrus paradisi*

40 <400> 29

45

Asp Leu Gly Phe Pro Lys Lys Val Pro Tyr Ala Arg Asp Arg Val Val  
 1 5 10 15

Glu Thr Tyr Ile Trp Met Leu Leu Gly Val Ser Tyr Glu Pro Asn Leu  
 20 25 30

Ala Phe Gly Arg Ile Phe Ala Ser Lys Val Val Cys Ile Ile Ser Ile  
 35 40 45

Ile

<210> 30  
 <211> 27  
 <212> PRT  
 <213> *Citrus paradisi*

50

ES 2 607 122 T3

<400> 30

Ile Val Gly Thr Tyr Phe Glu Pro Lys Tyr Thr Leu Ala Arg Lys Ile  
1 5 10 15

Met Thr Lys Thr Ile Tyr Thr Ala Ser Ile Ile  
20 25

5 <210> 31  
<211> 26  
<212> PRT  
<213> *Citrus paradisi*

10 <400> 31

Met Gly Val Tyr Phe Glu Pro Arg Tyr Ser Phe Ala Arg Lys Ile Leu  
1 5 10 15

Ser Lys Val Ile Ala Met Ala Ser Ile Leu  
20 25

15 <210> 32  
<211> 260  
<212> PRT  
<213> *Citrus paradisi*

20 <400> 32

ES 2 607 122 T3

Gln Lys Leu His Met Ile Asp Ala Ala Gln Arg Leu Gly Val Ala Tyr  
1 5 10 15

His Phe Glu Lys Glu Ile Glu Asp Glu Leu Gly Lys Val Ser His Asp  
20 25 30

Leu Asp Ser Asp Asp Leu Tyr Val Val Ser Leu Arg Phe Arg Leu Phe  
35 40 45

Arg Gln Gln Gly Val Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys  
50 55 60

Asp Asp Glu Gly Lys Phe Lys Glu Ser Leu Ile Asn Asp Ile Arg Gly  
65 70 75 80

Met Ser Ser Leu Tyr Glu Ala Ala Tyr Leu Ala Ile Arg Gly Glu Asp  
85 90 95

Ile Leu Asp Glu Ala Ile Val Phe Thr Thr Thr His Leu Lys Ser Val  
100 105 110

Ile Ser Val Ser Asp His Ser His Val Asn Ser Asp Leu Ala Glu Gln  
115 120 125

Ile Arg His Ser Leu Gln Ile Pro Leu Arg Lys Ala Ala Ala Arg Leu  
130 135 140

Glu Ala Arg Tyr Phe Leu Asp Ile Tyr Ser Arg Asp Asp Leu His Asp  
145 150 155 160

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

Ala His Lys Lys Glu Ala Ser Ile Met Thr Arg Trp Trp Asn Asp Leu  
180 185 190

Gly Phe Pro Lys Lys Val Pro Tyr Ala Arg Asp Arg Val Val Glu Thr  
195 200 205

Tyr Ile Trp Met Leu Leu Gly Val Ser Tyr Glu Pro Asn Leu Ala Phe  
210 215 220

Gly Arg Ile Phe Ala Ser Lys Val Val Cys Ile Ile Ser Ile Ile Asp  
225 230 235 240

Asp Thr Phe Asp Ala Tyr Gly Thr Phe Glu Glu Leu Thr Leu Phe Thr  
245 250 255

Glu Ala Val Thr  
260

ES 2 607 122 T3

<210> 33  
 <211> 780  
 <212> ADN  
 <213> *Citrus paradisi*

5

<400> 33

```

caaaaactac acatgattga tgcagcacia cgattagggtg tcgcttatca ttttgaaaaa    60
gagattgaag atgaattggg aaaggtatct catgatcttg acagtgatga tctatacgtt    120
gtttctcttc gttttcgact ttttagacag caaggagtta agatttcatg tgatgtgttt    180
gagaagttca aagatgacga aggtaaattc aaggaatcat tgatcaacga tatacgaggc    240
atgtcgagtt tgtacgaggc agcataccta gcaattcggg gggaagacat tttagatgaa    300
gccattgttt tcaactaccac tcaccttaag tcagtaatat ctgtatctga tcattctcat    360
gtaaactctg atcttgctga acaaatacgt cattctctgc aaattcctct ccgtaaagcc    420

gcagcaaggt tagaggcaag gtatTTTTTg gatatctatt caagggatga tttgcatgat    480
gaaactttgc tcaagtttgc aaagttagac tttaatatat tacaagcagc acacaagaag    540
gaagcaagta tcatgaccag gtggtggaac gatttaggct tccctaaaaa ggtgccttat    600
gcaagagata gagtagtaga gacatatatt tggatgttgc tgggagtgtc ctatgagccc    660
aatttggcat ttggtagaat ttttgcattc aaagtgggtg gcataatata cataatagac    720
gacacatttg atgcttacgg tacttttgaa gagctcacac tttttactga agcagtcaca    780
    
```

10 <210> 34  
 <211> 28  
 <212> ADN  
 <213> Artificial

15 <400> 34  
 ctgggaggt cctatgagct caatttg 28

20 <210> 35  
 <211> 29  
 <212> ADN  
 <213> Artificial

25 <400> 35  
 gtagaatttt tgcattcaaa gtggtgtgc 29

30 <210> 36  
 <211> 28  
 <212> ADN  
 <213> Artificial

<400> 36  
 cacaccactt tggatgcaaa aattcatc 28

35 <210> 37  
 <211> 28  
 <212> ADN  
 <213> Artificial

<400> 37

	ccaaattggg ctcataggac actcccag	28
	<210> 38 <211> 26 <212> ADN <213> Artificial	
5		
	<400> 38 tagggacgta tttgaacca aagtac	26
10		
	<210> 39 <211> 27 <212> ADN <213> Artificial	
15		
	<400> 39 aaataatgac caaaacaatt tacacgg	27
20		
	<210> 40 <211> 22 <212> ADN <213> Artificial	
25		
	<400> 40 gcacttcat gtattctgga ag	22
30		
	<210> 41 <211> 22 <212> ADN <213> Artificial	
35		
	<400> 41 gttgagctc tcaaagaaa cc	22
40		
	<210> 42 <211> 31 <212> ADN <213> Artificial	
45		
	<400> 42 aatgggagtg tattttgagc ctcgatactc c	31
50		
	<210> 43 <211> 31 <212> ADN <213> Artificial	
55		
	<400> 43 gatatttcc aaagtaattg caatggcatc c	31
60		
	<210> 44 <211> 27 <212> ADN <213> Artificial	
65		
	<400> 44 gttgctaata tcccacctt tgatagc	27
70		
	<210> 45 <211> 23 <212> ADN <213> Artificial	
75		
	<400> 45 aagtgcca taggcgctg agg	23

ES 2 607 122 T3

	<210> 46	
	<211> 27	
	<212> ADN	
	<213> Artificial	
5		
	<400> 46	
	ctgtccgca agcttagggg ttacatg	27
10		
	<210> 47	
	<211> 29	
	<212> ADN	
	<213> Artificial	
15		
	<400> 47	
	ctgagctacc aatgactca ggtgagtgg	29
20		
	<210> 48	
	<211> 31	
	<212> ADN	
	<213> Artificial	
25		
	<400> 48	
	caattttgcc atacacatca tagatatcat c	31
30		
	<210> 49	
	<211> 28	
	<212> ADN	
	<213> Artificial	
35		
	<400> 49	
	aacagaagtc atggagatca cttcgtc	28
40		
	<210> 50	
	<211> 28	
	<212> ADN	
	<213> Artificial	
45		
	<400> 50	
	cgcaagagat gtttaaagt tcccatcc	28
50		
	<210> 51	
	<211> 26	
	<212> ADN	
	<213> Artificial	
55		
	<400> 51	
	tgaacatcag cggaatttt atagcc	26
60		
	<210> 52	
	<211> 24	
	<212> ADN	
	<213> Artificial	
65		
	<400> 52	
	acgatttagg cttccctaaa aagg	24
70		
	<210> 53	
	<211> 25	
	<212> ADN	
	<213> Artificial	
75		
	<400> 53	
	tattatggaa tattatgcac accac	25
80		
	<210> 54	
	<211> 27	

ES 2 607 122 T3

<212> ADN  
 <213> Artificial

5 <400> 54  
 aaatgtccgc tcaagtcta gcaacgg 27

10 <210> 55  
 <211> 26  
 <212> ADN  
 <213> Artificial

15 <400> 55  
 atggcacttc aagattcaga agttcc 26

20 <210> 56  
 <211> 41  
 <212> ADN  
 <213> Artificial

25 <400> 56  
 gcatgttcca tatgtccgct caagtctag caacggtttc c 41

30 <210> 57  
 <211> 39  
 <212> ADN  
 <213> Artificial

35 <400> 57  
 cgcggatcct cagatggtaa cagggtctct gagcactgc 39

40 <210> 58  
 <211> 37  
 <212> ADN  
 <213> Artificial

45 <400> 58  
 gcatgttcca tatgtcgtct ggagaaacat ttcgtcc 37

50 <210> 59  
 <211> 33  
 <212> ADN  
 <213> Artificial

55 <400> 59  
 cgcggatcct caaatggaa cgtggtctcc tag 33

60 <210> 60  
 <211> 39  
 <212> ADN  
 <213> Artificial

65 <400> 60  
 gcatgttcca tatgtcttg gaagttcag cctctctg 39

70 <210> 61  
 <211> 42  
 <212> ADN  
 <213> Artificial

75 <400> 61  
 cgcggatcct catatcggca caggattaat aaacaaagaa gc 42

80 <210> 62  
 <211> 30  
 <212> ADN  
 <213> Artificial

# ES 2 607 122 T3

	<400> 62 taatggcact tcaagattca gaagttcctc	30
5	<210> 63 <211> 27 <212> ADN <213> Artificial	
10	<400> 63 aaaaggaac aggcttctca agcaatg	27
15	<210> 64 <211> 28 <212> ADN <213> Artificial	
20	<400> 64 atggcacttc aagattcaga agttcctc	28
25	<210> 65 <211> 31 <212> ADN <213> Artificial	
	<400> 65 gatcaaaagg gaacaggctt ctcaagcaat g	31



**REIVINDICACIONES**

1. Un ácido nucleico aislado que codifica un polipéptido al menos 90 % idéntico a SEQ ID NO: 1; en el que el polipéptido codificado por dicho ácido nucleico es una cubebol sintasa.
- 5 2. El ácido nucleico aislado de la reivindicación 1, **caracterizado porque** el ácido nucleico comprende la secuencia de nucleótidos SEQ ID NO: 6.
3. El ácido nucleico aislado de la reivindicación 1 o reivindicación 2, **caracterizado porque** el polipéptido codificado consiste en SEQ ID NO: 1.
4. Un polipéptido de cubebol sintasa aislado codificado por el ácido nucleico de una cualquiera de las reivindicaciones precedentes.
- 10 5. Un polipéptido aislado de acuerdo con la reivindicación 4, **caracterizado porque** dicho polipéptido es SEQ ID NO: 1.
6. Un vector que comprende el ácido nucleico de una cualquiera de las reivindicaciones 1 a 3.
7. El vector de la reivindicación 6, **caracterizado porque** dicho vector es un vector viral o un plásmido.
- 15 8. Una célula huésped o un organismo no humano modificado para albergar el ácido nucleico de una cualquiera de las reivindicaciones 1 a 3.
9. La célula huésped u organismo no humano de la reivindicación 8, **caracterizado porque** dicha célula huésped es una célula vegetal y dicho organismo no humano es una planta o un microorganismo.
10. Un procedimiento de preparación de una célula huésped recombinante que comprende introducir el vector de las reivindicaciones 6 o 7 en una célula huésped.
- 20 11. Un procedimiento de producción de un polipéptido que comprende cultivar la célula huésped u organismo no humano de la reivindicación 8 en condiciones para producir dicho polipéptido.
12. El procedimiento de la reivindicación 11, **caracterizado porque** la célula huésped se elige de células bacterianas, células de levadura, células vegetales y células animales.
- 25 13. El procedimiento de la reivindicación 11, **caracterizado porque** la célula huésped se elige de células de *E. coli* y células de levadura.
14. Un procedimiento de preparación de al menos una sesquiterpeno sintasa que comprende cultivar un huésped modificado para contener al menos una secuencia de ácido nucleico que codifica un polipéptido de cubebol sintasa al menos 90 % idéntico a SEQ ID NO: 1 en condiciones que conducen a la producción de dicha al menos una sesquiterpeno sintasa.
- 30 15. El procedimiento de la reivindicación 14, **caracterizado porque** dicho huésped es una planta o un microorganismo.
16. El procedimiento de la reivindicación 14, **caracterizado porque** dicho huésped se elige de células bacterianas, células de levadura, células vegetales y células animales.
17. Un procedimiento de preparación de al menos un terpenoide que comprende
  - 35 A) poner en contacto al menos un precursor de pirofosfato terpeno acíclico con al menos un polipéptido codificado por un ácido nucleico de acuerdo con una cualquiera de las reivindicaciones 1 a 3;
  - B) aislar al menos un terpenoide producido en A).
18. El procedimiento de la reivindicación 17, **caracterizado porque** dicho al menos un terpenoide se elige de sesquiterpenos.
- 40 19. El procedimiento de la reivindicación 17 o 18, **caracterizado porque** dicho al menos un precursor de pirofosfato terpeno acíclico es farnesil-pirofosfato.
20. El procedimiento de la reivindicación 18, **caracterizado porque** dicho al menos un sesquiterpeno es cubebol.
21. El procedimiento de una cualquiera de las reivindicaciones 17 a 20, **caracterizado porque** el al menos un polipéptido se produce cultivando una célula huésped que comprende el ácido nucleico, en el que la célula huésped se elige de células procariotas, células de levadura, células vegetales y células animales.
- 45 22. El procedimiento de una cualquiera de las reivindicaciones 17 a 20, **caracterizado porque** dicho al menos un polipéptido se produce cultivando una célula de *E. coli* que comprende el ácido nucleico de una cualquiera de las

reivindicaciones 1 a 3.

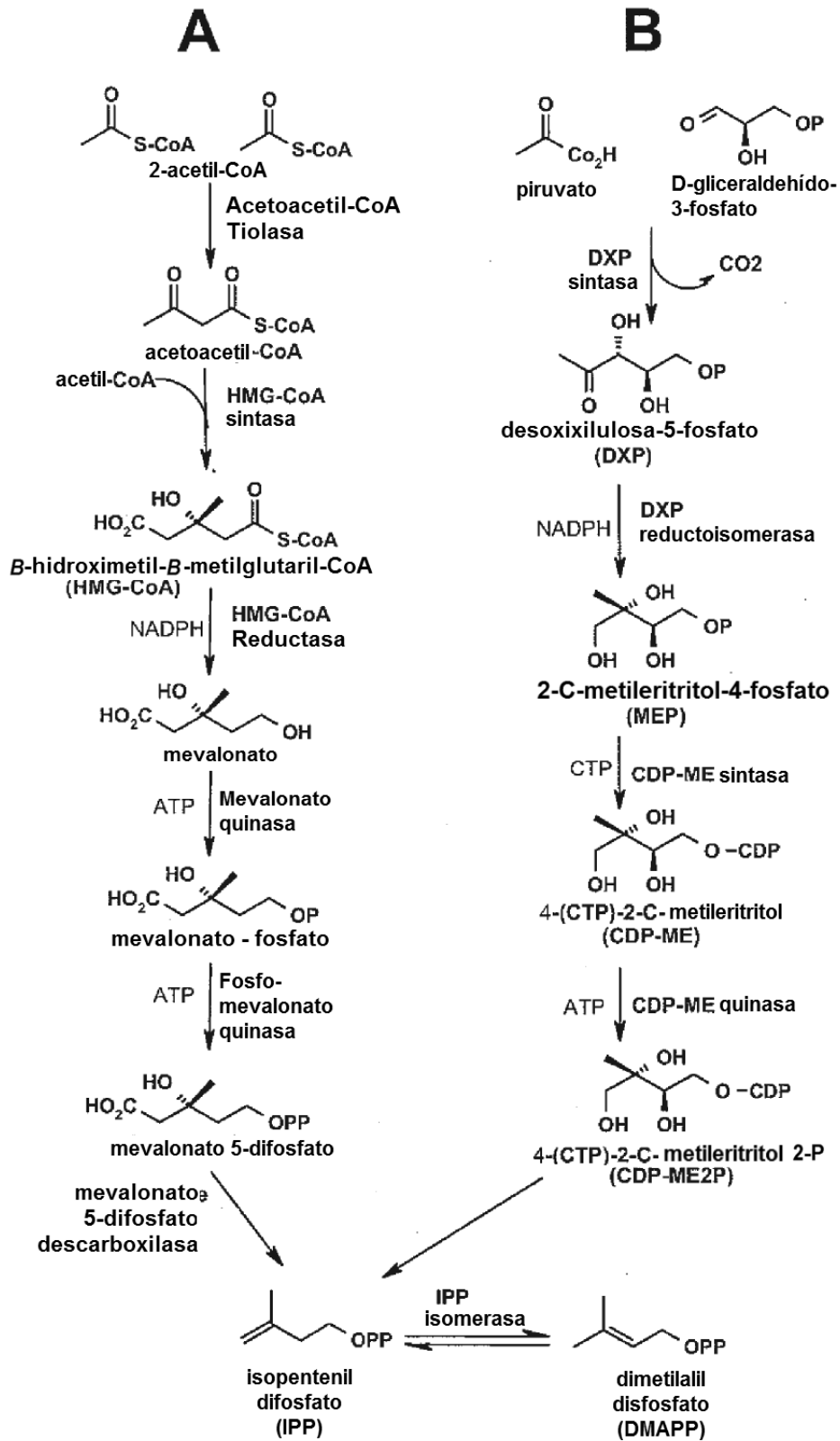
23. Un organismo no humano de acuerdo con la reivindicación 9, **caracterizado porque** dicho microorganismo es un procariota o levadura.

5 24. El organismo no humano de la reivindicación 9, **caracterizado porque** dicho procariota se selecciona del grupo de *E. coli*, *Bacillus subtilis*, especies del género *Pseudomonas*, o especies del género *Streptomyces*, y en el que dicha levadura es una especie de un género seleccionado del grupo de *Saccharomyces*, *Pichia* o *Kluyveromyces*.

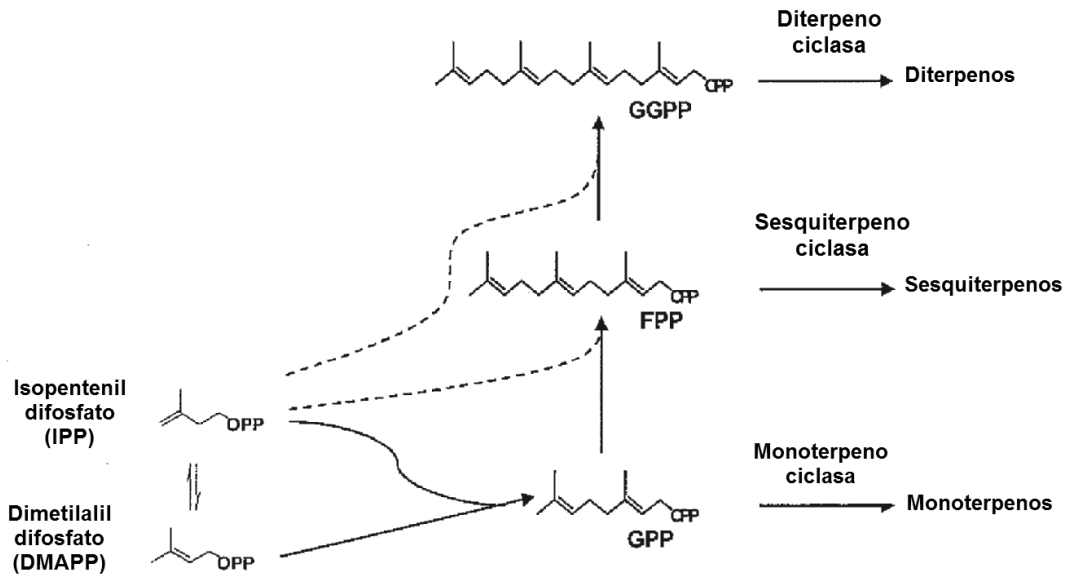
25. El procedimiento de la reivindicación 18, **caracterizado porque** dichos sesquiterpenos son cubebol y alfa-cubebeno.

26. El procedimiento de la reivindicación 17, **caracterizado porque** el pH en el momento de contacto es 7 o menos.

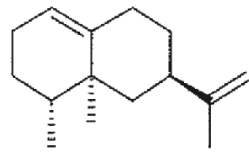
10



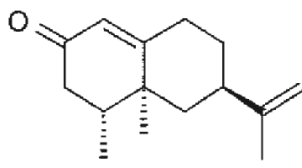
**FIG. 1**



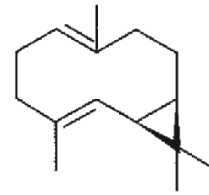
**FIG. 2**



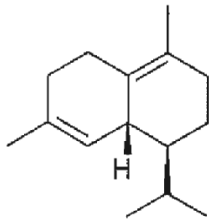
Valenceno



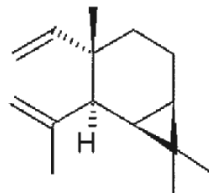
Notcatona



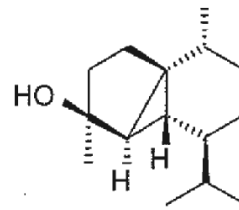
Bicyclogermacreno



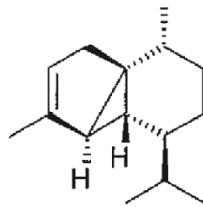
delta-Cadineno



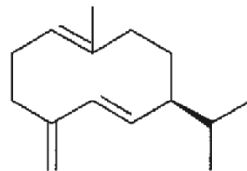
biciclo-elemeno



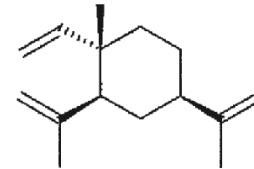
Cubebol



(-)-alfa-cubebeno



germacreno D



beta-elemeno

**FIG. 3**

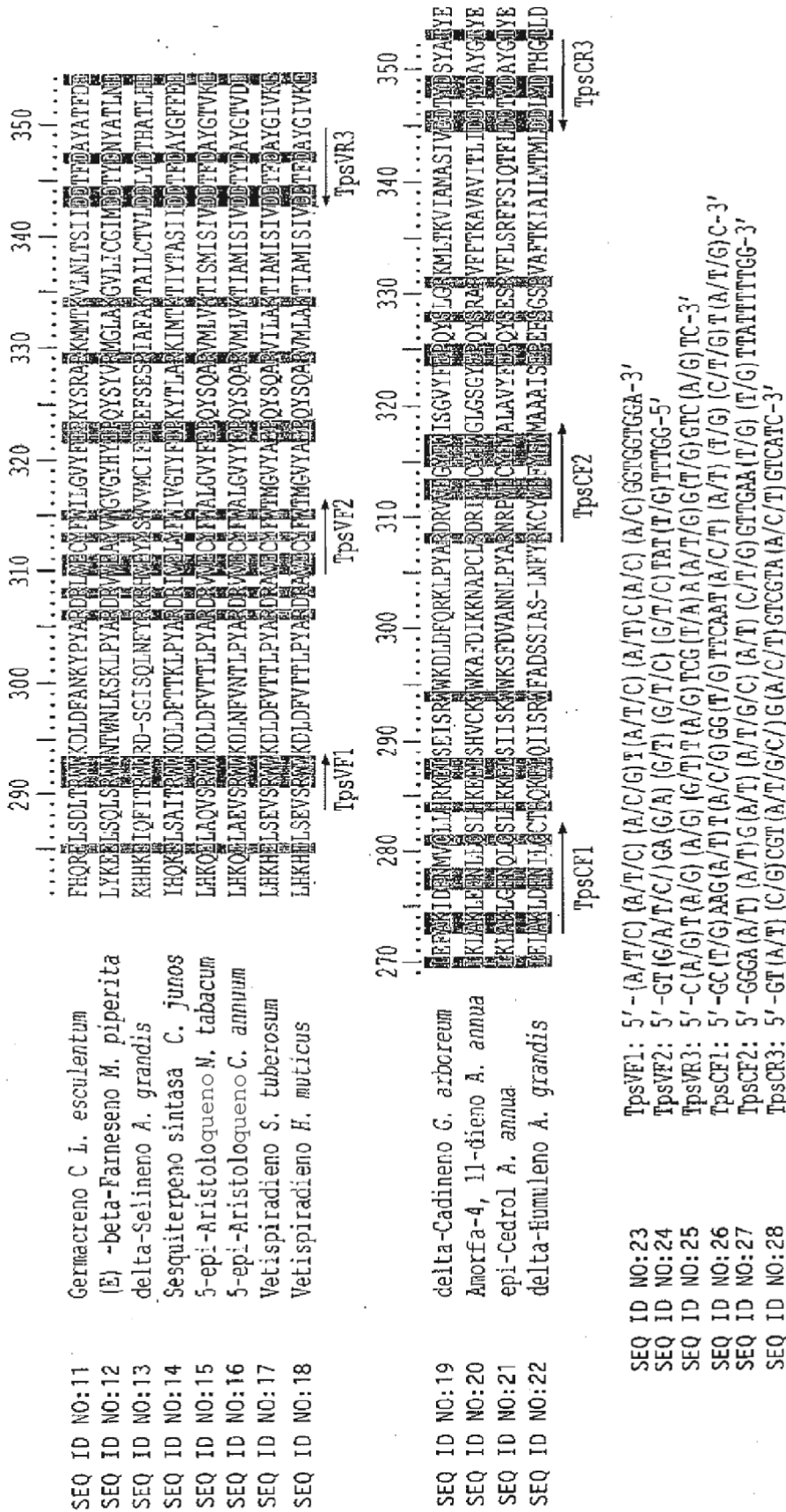


FIG. 4

SEQ ID NO:29	GFTpsA	10	20	30	40	
SEQ ID NO:30	GFTpsB	DI GF PK VPY	AR DR VV EYI	WMLL G W S Y L P	NL A T G R I F A S	K V V C I I S I I
SEQ ID NO:31	GFTpsC	-----	-----	-- I V G T Y P P P	K Y T L A R K K I M T	K T I Y T A S I I
		-----	-----	--- M G V M P P P	R Y S F A R K K I L S	K V I A M A S I I

**FIG. 5**

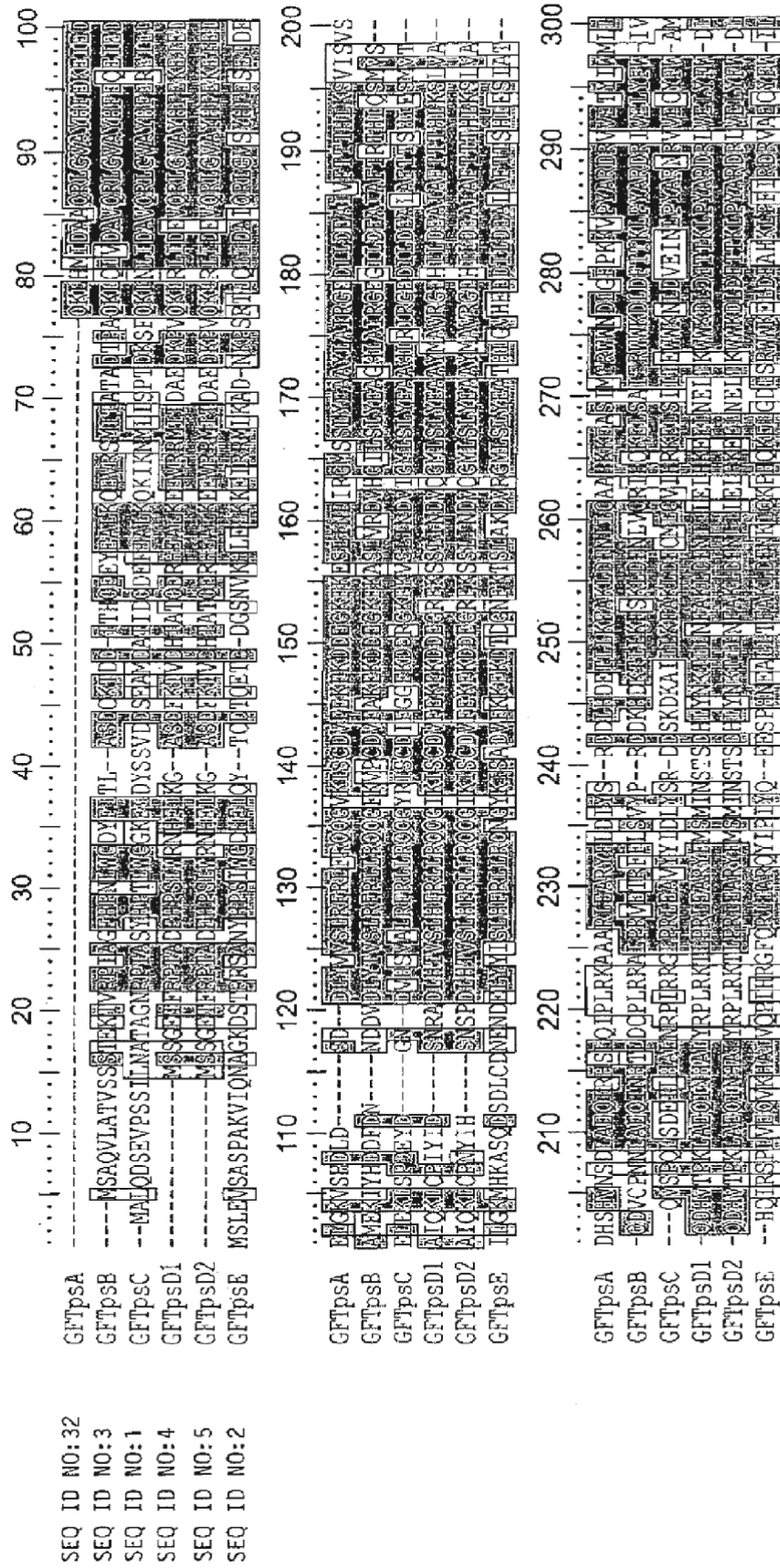


FIG. 6A



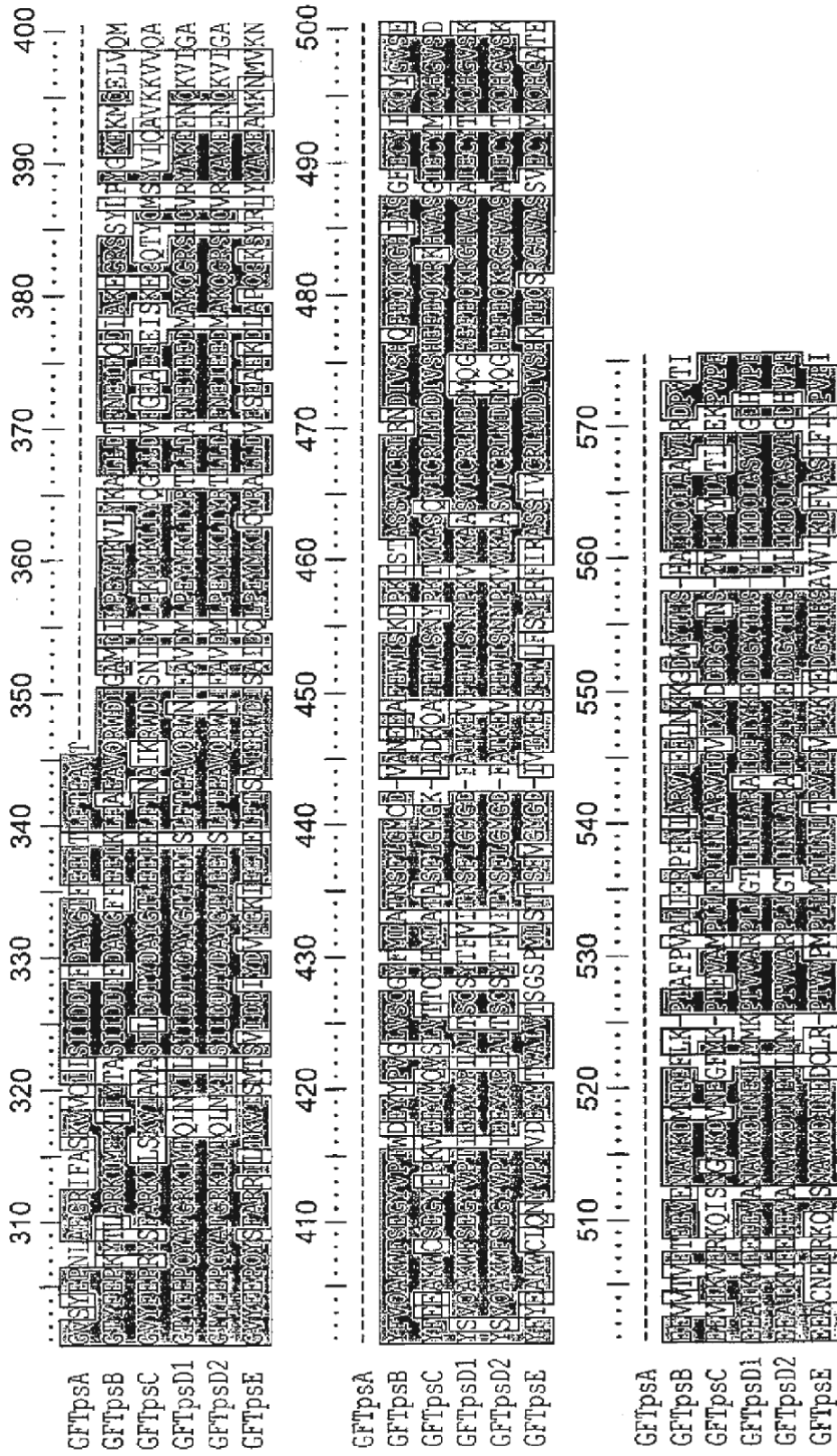
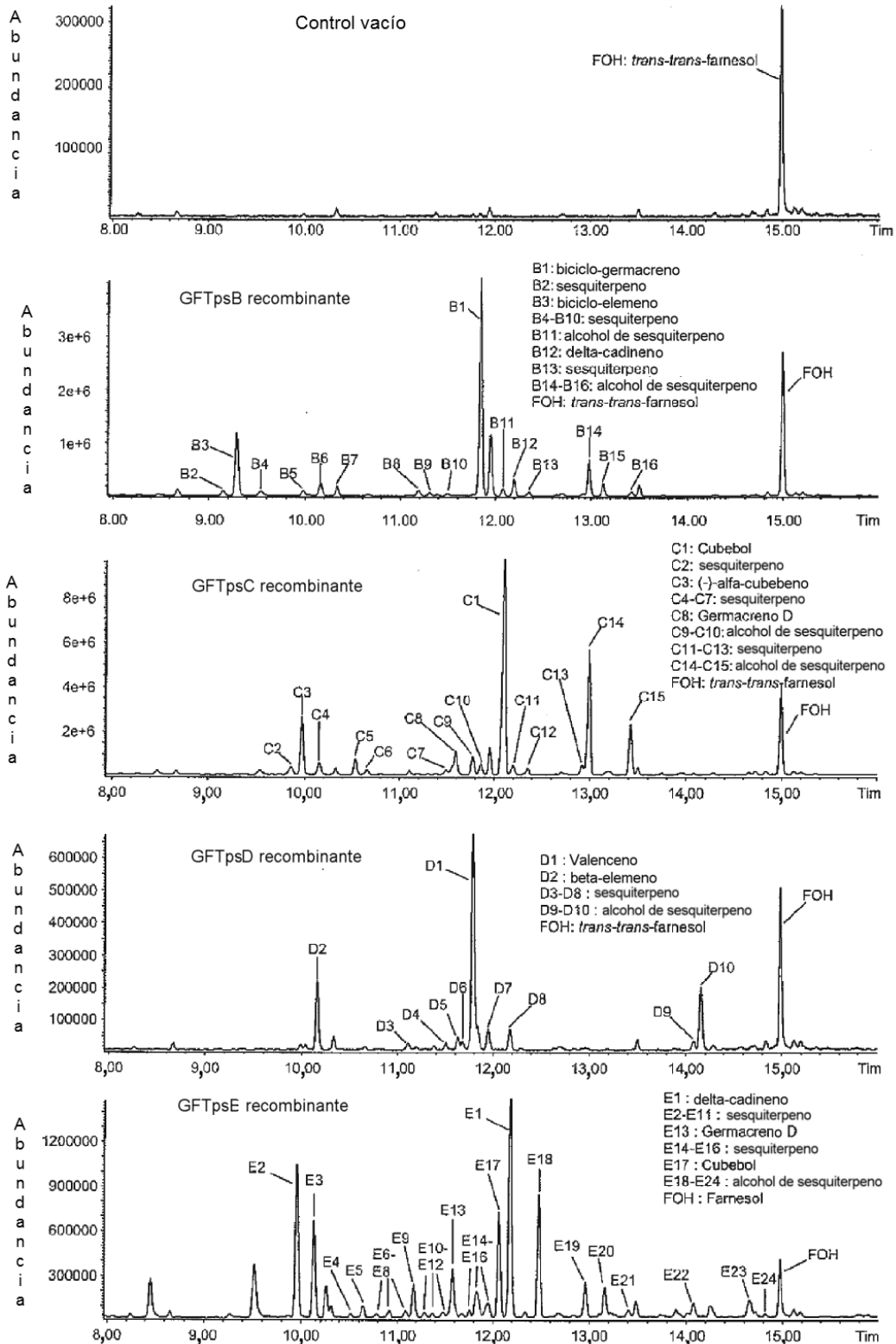


FIG. 6B



**FIG. 7**

ES 2 607 122 T3

		GFTpsA															
SEQ ID NO:33	1	CAA	AAA	CTA	CAC	ATG	ATT	GAT	GCA	GCA	CAA	CGA	TTA	GGT	GTC	GCT	45
SEQ ID NO:32	1	Q	K	L	H	M	I	D	A	A	Q	R	L	G	V	A	15
	46	TAT	CAT	TTT	GAA	AAA	GAG	ATT	GAA	GAT	GAA	TTG	GGA	AAG	GTA	TCT	90
	16	Y	H	F	E	K	E	I	E	D	E	L	G	K	V	S	30
	91	CAT	GAT	CTT	GAC	AGT	GAT	GAT	CTA	TAC	GTT	GTT	TCT	CTT	CGT	TTT	135
	31	H	D	L	D	S	D	D	L	Y	V	V	S	L	R	F	45
	136	CGA	CTT	TTT	AGA	CAG	CAA	GGA	GTT	AAG	ATT	TCA	TGT	GAT	GTG	TTT	180
	46	R	L	F	R	Q	Q	G	V	K	I	S	C	D	V	F	60
	181	GAG	AAG	TTC	AAA	GAT	GAC	GAA	GGT	AAA	TTC	AAG	GAA	TCA	TTG	ATC	225
	61	E	K	F	K	D	D	E	G	K	F	K	E	S	L	I	75
	226	AAC	GAT	ATA	CGA	GGC	ATG	TCG	AGT	TTG	TAC	GAG	GCA	GCA	TAC	CTA	270
	76	N	D	I	R	G	M	S	S	L	Y	E	A	A	Y	L	90
	271	GCA	ATT	CGG	GGG	GAA	GAC	ATT	TTA	GAT	GAA	GCC	ATT	GTT	TTC	ACT	315
	91	A	I	R	G	E	D	I	L	D	E	A	I	V	F	T	105
	316	ACC	ACT	CAC	CTT	AAG	TCA	GTA	ATA	TCT	GTA	TCT	GAT	CAT	TCT	CAT	360
	106	T	T	H	L	K	S	V	I	S	V	S	D	H	S	H	120
	361	GTA	AAC	TCT	GAT	CTT	GCT	GAA	CAA	ATA	CGT	CAT	TCT	CTG	CAA	ATT	405
	121	V	N	S	D	L	A	E	Q	I	R	H	S	L	Q	I	135
	406	CCT	CTC	CGT	AAA	GCC	GCA	GCA	AGG	TTA	GAG	GCA	AGG	TAT	TTT	TTG	450
	136	P	L	R	K	A	A	A	R	L	E	A	R	Y	F	L	150
	451	GAT	ATC	TAT	TCA	AGG	GAT	GAT	TTG	CAT	GAT	GAA	ACT	TTG	CTC	AAG	495
	151	D	I	Y	S	R	D	D	L	H	D	E	T	L	L	K	165
	496	TTT	GCA	AAG	TTA	GAC	TTT	AAT	ATA	TTA	CAA	GCA	GCA	CAC	AAG	AAG	540
	166	F	A	K	L	D	F	N	I	L	Q	A	A	H	K	K	180
	541	GAA	GCA	AGT	ATC	ATG	ACC	AGG	TGG	TGG	AAC	GAT	TTA	GGC	TTC	CCT	585
	181	E	A	S	I	M	T	R	W	W	N	D	L	G	F	P	195
	586	AAA	AAG	GTG	CCT	TAT	GCA	AGA	GAT	AGA	GTA	GTA	GAG	ACA	TAT	ATT	630
	196	K	K	V	P	Y	A	R	D	R	V	V	E	T	Y	I	210
	631	TGG	ATG	TTG	CTG	GGA	GTG	TCC	TAT	GAG	CCC	AAT	TTG	GCA	TTT	GGT	675
	211	W	M	L	L	G	V	S	Y	E	P	N	L	A	F	G	225
	676	AGA	ATT	TTT	GCA	TCC	AAA	GTG	GTG	TGC	ATA	ATA	TCC	ATA	ATA	GAC	720
	226	R	I	F	A	S	K	V	V	C	I	I	S	I	I	D	240
	721	GAC	ACA	TTT	GAT	GCT	TAC	GGT	ACT	TTT	GAA	GAG	CTC	ACA	CTT	TTT	765
	241	D	T	F	D	A	Y	G	T	F	E	E	L	T	L	F	255
	766	ACT	GAA	GCA	GTC	ACA	780										
	256	T	E	A	V	T											

**FIG. 8(a)**

ES 2 607 122 T3

		GFTpsB															
SEQ ID NO:8	1	ATG	TCC	GCT	CAA	GTT	CTA	GCA	ACG	GTT	TCC	AGT	TCG	ACA	GAA	AAA	45
SEQ ID NO:3	1	M	S	A	Q	V	L	A	T	V	S	S	S	T	E	K	15
	46	ACT	GTT	CGT	CCC	ATT	GCT	GGT	TTC	CAT	CCT	AAC	TTA	TGG	GGA	GAC	90
	16	T	V	R	P	I	A	G	F	H	P	N	L	W	G	D	30
	91	TAT	TTC	CTG	ACC	CTC	GCT	TCT	GAT	TGC	AAG	ACA	GAT	GAT	ACT	ACG	135
	31	Y	F	L	T	L	A	S	D	C	K	T	D	D	T	T	45
	136	CAC	CAA	GAG	GAA	TAC	GAA	GCG	CTG	AAG	CAA	GAA	GTC	AGA	AGC	ATG	180
	46	H	Q	E	E	Y	E	A	L	K	Q	E	V	R	S	M	60
	181	ATA	ACG	GCT	ACG	GCA	GAT	ACA	CCT	GCC	CAG	AAG	TTG	CAA	TTG	GTT	225
	61	I	T	A	T	A	D	T	P	A	Q	K	L	Q	L	V	75
	226	GAT	GCA	GTC	CAA	CGA	TTG	GGT	GTG	GCC	TAT	CAC	TTC	GAA	CAG	GAG	270
	76	D	A	V	Q	R	L	G	V	A	Y	H	F	E	Q	E	90
	271	ATA	GAA	GAT	GCA	ATG	GAA	AAG	ATT	TAT	CAC	GAT	GAC	TTT	GAT	AAT	315
	91	I	E	D	A	M	E	K	I	Y	H	D	D	F	D	N	105
	316	AAC	GAT	GAT	GTC	GAT	CTC	TAC	ACT	GTT	TCT	CTT	CGT	TTT	CGA	CTG	360
	106	N	D	D	V	D	L	Y	T	V	S	L	R	F	R	L	120
	361	CTT	AGG	CAG	CAA	GGA	TTT	AAG	GTT	CCG	TGT	GAT	GTG	TTC	GCG	AAG	405
	121	L	R	Q	Q	G	F	K	V	P	C	D	V	F	A	K	135
	406	TTC	AAA	GAT	GAT	GAA	GGT	AAA	TTC	AAG	GCA	TCA	TTG	GTG	CGG	GAT	450
	136	F	K	D	D	E	G	K	F	K	A	S	L	V	R	D	150
	451	GTT	CAT	GCC	ATT	CTA	AGT	TTG	TAT	GAG	GCA	GGA	CAC	TTG	GCC	ATT	495
	151	V	H	G	I	L	S	L	Y	E	A	G	H	L	A	I	165
	496	CGC	GGA	GAA	GGG	ATA	TTA	GAT	GAA	GCC	ATT	GCT	TTC	ACT	AGA	ACT	540
	166	R	G	E	G	I	L	D	E	A	I	A	F	T	R	T	180
	541	CAC	CTT	CAG	TCA	ATG	GTA	TCT	CAG	GAT	GTA	TGC	CCT	AAT	AAT	CTT	585
	181	H	L	Q	S	M	V	S	Q	D	V	C	P	N	N	L	195
	586	GCT	GAA	CAA	ATT	AAT	CAT	ACT	CTC	GAC	TGT	CCT	CTC	CGC	AGA	GCC	630
	196	A	E	Q	I	N	H	T	L	D	C	P	L	R	R	A	210
	631	CTT	CCA	AGA	CTC	GAG	ACA	AGA	TTT	TTC	TTG	TCG	GTC	TAT	CCA	AGA	675
	211	L	P	R	V	E	T	R	F	F	L	S	V	Y	P	R	225
	676	GAT	GAT	AAA	CAC	GAT	AAA	ACT	TTG	TTA	AAG	TTT	TCA	AAG	TTA	GAC	720
	226	D	D	K	H	D	K	T	L	L	K	F	S	K	L	D	240
	721	TTT	AAC	CTT	GTG	CAA	AGA	ATA	CAT	CAG	AAG	GAA	TTA	AGT	GCC	ATC	765
	241	F	N	L	V	Q	R	I	H	Q	K	E	L	S	A	I	255
	766	ACA	CGG	TGG	TGG	AAA	GAT	TTA	GAC	TTC	ACT	ACA	AAG	CTA	CCT	TAT	810
	256	T	R	W	W	K	D	L	D	F	T	T	K	L	P	Y	270

**FIG. 8(b)-1**

811	GCA AGA GAC AGA ATC GTA GAG TTG TAT TTT TGG ATT GTA GGG ACG	855
271	A R D R I V E L Y F W I V G T	285
856	TAT TTT GAA CCA AAG TAC ACT TTA GCA AGA AAA ATA ATG ACC AAA	900
286	Y F E P K Y T L A R K I M T K	300
901	ACA ATT TAC ACG GCA TCT ATC ATA GAT GAC ACT TTC GAC GCT TAT	945
301	T I Y T A S I I D D T F D A Y	315
946	GGT TTC TTT GAA GAG CTC AAA CTC TTT GCA GAA GCA GTC CAG AGG	990
316	G F F E E L K L F A E A V Q R	330
991	TGG GAC ATT GGA GCC ATG GAT ATA CTT CCA GAA TAC ATG AAA GTG	1035
331	W D I G A M D I L P E Y M K V	345
1036	CTT TAT AAG GCC CTT TTA GAT ACT TTC AAT GAA ATT GAG CAA GAC	1080
346	L Y K A L L D T F N E I E Q D	360
1081	TTG GCC AAG GAA GGA AGA TCG TCC TAC TTA CCT TAT GGC AAA GAA	1125
361	L A K E G R S S Y L P Y G K E	375
1126	AAG ATG CAA GAG CTT GTT CAA ATG TAC TTT GTT CAA GCC AAG TGG	1170
376	K M Q E L V Q M Y F V Q A K W	390
1171	TTC CGT GAA GGT TAT GTT CCG ACA TGG GAC GAA TAT TAT CCG GTT	1215
391	F S E G Y V P T W D E Y Y P V	405
1216	GGA CTT GTA AGT TGC GGC TAC TTC ATG CTT GCG ACA AAC TCC TTC	1260
406	G L V S C G Y F M L A T N S F	420
1261	CTT GGC ATG TGT GAT GTT GCA AAC GAG GAA GCT TTT GAA TGG ATA	1305
421	L G M C D V A N E E A F E W I	435
1306	TCC AAG GAC CCT AAG ATT TCA ACA GCG TCA TCA GTT ATC TGC AGA	1350
436	S K D P K I S T A S S V I C R	450
1351	CTT AGG AAT GAC ATT GTT TCC CAC CAG TTT GAA CAG AAG AGA GGA	1395
451	L R N D I V S H Q F E Q K R G	465
1396	CAT ATT GCC TCA GGA TTT GAA TGC TAC ATT AAG CAG TAT GGT GTT	1440
466	H I A S G F E C Y I K Q Y G V	480
1441	TCA GAA GAA GAG GTA GTT ACA GTT TTT ACT GAA GAA GTT GAG AAT	1485
481	S E E E V V T V F T E E V E N	495
1486	GCA TGG AAA GAT ATG AAT GAG GAA TTC CTG AAA CCA ACT GCT TTT	1530
496	A W K D M N E E F L K P T A F	510
1531	CCT GTG GCT TTG ATT GAG AGA CCT TTC AAT ATC GCA CGT GTG ATT	1575
511	P V A L I E R P F N I A R V I	525
1576	GAA TTT CTA AAC AAG AAG GGT GAT TGG TAC ACT CAT TCT CAT GCG	1620
526	E F L N K K G D W Y T H S H A	540
1621	ATT AAA GAC CAG ATT GCC GCA GTG CTC AGA GAC CCT GTT ACC ATC	1665
541	I K D Q I A A V L R D P V T I	555

**FIG. 8(b)-2**

ES 2 607 122 T3

		GFTpsC																
SEQ ID NO:6	1	ATG	GCA	CTT	CAA	GAT	TCA	GAA	GTT	CCT	TCT	TCC	ATT	CTG	AAT	GCT	45	
SEQ ID NO:7	1	M	A	L	Q	D	S	E	V	P	S	S	I	L	N	A	15	
	46	ACA	GCC	GGC	AAC	CGT	CCC	ACA	GCT	AGT	TAT	CAT	CCC	ACC	CTC	TGG	90	
	16	T	A	G	N	R	P	T	A	S	Y	H	P	T	L	W	30	
	91	GGG	GGA	AAA	TTC	CTT	GAC	TAT	TCT	TCT	GTT	GAC	GAC	TCT	GAG	GCA	135	
	31	G	G	K	F	L	D	Y	S	S	V	D	D	S	E	A	45	
	136	ATG	GAT	GCC	ACA	ATT	GAT	CAA	GAC	GAA	TTT	GAA	GCA	CTT	AAG	CAA	180	
	45	M	D	A	T	I	D	Q	D	E	F	E	A	L	K	Q	60	
	181	AAA	ATA	AAG	AAC	ATG	TTA	ATC	TCA	CCA	ACC	GAT	AAG	TCT	TTT	CAA	225	
	61	K	I	K	N	M	L	I	S	P	T	D	K	S	F	Q	75	
	226	AAA	TTG	AAC	TTG	ATT	GAT	GCC	GTC	CAA	CGC	TTA	GGA	GTG	GCT	TAC	270	
	76	K	L	N	L	I	D	A	V	Q	R	L	G	V	A	Y	90	
	271	CAT	TTT	GAG	AGG	GAG	ATA	GAA	GAT	GAA	CTA	GAA	AAA	CTA	TCT	CCT	315	
	91	H	F	E	R	E	I	E	D	E	L	E	K	L	S	P	105	
	316	GAT	GAG	TAT	GAT	GGC	AAC	GAT	GTA	CAC	TCC	GTT	GCT	CTT	CGA	TTT	360	
	106	D	E	Y	D	G	N	D	V	H	S	V	A	L	R	F	120	
	361	CGG	TTA	CTC	AGA	CAA	CAA	GGA	TAT	CGC	ATA	TCA	TGC	GAT	ATT	TTT	405	
	121	R	L	L	R	Q	Q	G	Y	R	I	S	C	D	I	F	135	
	406	GGC	GGT	TTC	AAA	GAT	GAT	CGA	GGA	AAG	FTC	AAG	GTA	TCC	TTA	ATT	450	
	136	G	G	F	K	D	D	R	G	K	F	K	V	S	L	I	150	
	451	AAT	GAT	GTG	ACC	GGC	ATG	CTA	AGT	TTG	TAT	GAG	GCT	GCA	CAT	CTT	495	
	151	N	D	V	T	G	M	L	S	L	Y	E	A	A	H	L	165	
	496	CGC	ATT	CGC	GGG	GAA	GAT	ATC	CTG	GAT	GAA	GCC	CTA	GCT	TTC	ACT	540	
	166	R	I	R	G	E	D	I	L	D	E	A	L	A	F	T	180	
	541	ACT	TCT	CAC	CTG	GAA	TCA	ATG	GTT	ACT	CAA	GTA	AGC	CCT	CAG	CTT	585	
	181	T	S	H	L	E	S	M	V	T	Q	V	S	P	Q	L	195	
	586	TCT	GAT	GAA	ATA	CTT	CAT	GCC	TTG	AAT	AGG	CCA	ATC	CGC	AGA	GGC	630	
	196	S	D	E	I	L	H	A	L	N	R	P	I	R	R	G	210	
	631	TTA	CCA	AGG	CTG	GAG	GCA	GTC	TAT	TAC	ATC	GAT	CTC	TAC	TCA	CGA	675	
	211	L	P	R	L	E	A	V	Y	Y	I	D	L	Y	S	R	225	
	676	GAT	GAT	TCA	AAG	GAT	AAA	GCA	ATA	TTA	CTA	AAG	TTT	GCA	AAA	CTA	720	
	226	D	D	S	K	D	K	A	I	L	L	K	F	A	K	L	240	
	721	GAT	TTT	TGC	ATG	CTT	CAA	GTA	ATT	CAC	CGT	AAG	GAG	TTA	AGT	ATC	765	
	241	D	F	C	M	L	Q	V	I	H	R	K	E	L	S	I	255	
	766	ATC	ACA	GAG	TGG	TGG	AAA	AAT	TTA	GAT	GTT	GAA	ATA	AAT	CTC	CCA	810	
	256	I	T	E	W	W	K	N	L	D	V	E	I	N	L	P	270	

**FIG. 8(c)-1**

811	TAT GCT AGA AAC AGA GTT GTA GAA TGC TAT TTT TGG GCA ATG GGA	855
271	Y A R N R V V E C Y F W A M G	285
856	GTG TAT TTT GAG CCT CGA TAC TCC TTT GCA AGA AAG ATA TTG TCC	900
286	V Y F E P R Y S F A R K I L S	300
901	AAA GTA ATT GCA ATG GCA TCC ATT TTA GAT GAT ACC TAC GAC GCC	945
301	K V I A M A S I L D D T Y D A	315
946	TAT GGC ACA CTT GAA GAA CTT GAG CTC TTT ACA AAT GCT ATC AAA	990
316	Y G T L E E L E L F T N A I K	330
991	AGG TGG GAT ATT AGC AAC ATA GAT GTA CTT CCG AAG TAC ATG AAA	1035
331	R W D I S N I D V L P K Y M K	345
1036	CTG ATT TAT CAA GGA CTC TTG GAT GTT TTT GGT GAA GCT GAG GAG	1080
346	L I Y Q G L L D V F G E A E E	360
1081	GAA ATC TCA AAG GAA GGA CAG ACA TAT TGC ATG TCA TAT GTC ATA	1125
361	E I S K E G Q T Y C M S Y V I	375
1126	CAA GCG GTG AAG AAA GTA GTC CAA GCC TAC TTT GAG GAA GCC AAG	1170
376	Q A V K K V V Q A Y F E E A K	390
1171	TGG TGC AGT GAA GGT TAT TTT CCA AAA GTG GAG GAG TAT ATG CAA	1215
391	W C S E G Y F P K V E E Y M Q	405
1216	GTT TCA CTT GTG ACA ACT TGC TAT CAT ATG CTG GCA ACG GCT TCT	1260
406	V S L V T T C Y H M L A T A S	420
1261	TTT CTT GGC ATG GGA AAG ATT GCT GAT AAG CAG GCC TTT GAA TGG	1305
421	F L G M G K I A D K Q A F E W	435
1306	ATC TCC AAT TAC CCT AAA ACT GTG AAA GCC TCC CAA GTT ATT TGC	1350
436	I S N Y P K T V K A S Q V I C	450
1351	AGA CTT ATG GAT GAT ATA GTG TCT CAC GAG TTT GAA CAA AAA AGA	1395
451	R L M D D I V S H E F E Q K R	465
1396	AAG CAT GTT GCC TCG GGT ATT GAA TGT TAC ATG AAG CAG CAT GGC	1440
466	K H V A S G I E C Y M K Q H G	480
1441	GTC TCT GAT GAA GAG GTA ATT AAA GTA TTC CGC AAA CAA ATA TCA	1485
481	V S D E E V I K V F R K Q I S	495
1486	AAT GGA TGG AAA GAT GTA AAT GAA GGA TTC ATG AAG CCA ACA GAA	1530
496	N G W K D V N E G F M K P T E	510
1531	GTG GCA ATG CCT CTC CTT GAG CGC ATT CTC AAT CTT GCA CGA GTG	1575
511	V A M P L L E R I L N L A R V	525
1576	ATA GAT GTT ATT TAC AAG GAT GAT GAT GGC TAC ACC AAC TCT TAT	1620
526	I D V I Y K D D D G Y T N S Y	540
1621	GTG ATC AAA GAC TAC ATC GCC ACA TTG CTT GAG AAG CCT GTT CCC	1665
541	V I K D Y I A T L L E K P V P	555
1666	TTT TGA	1671
556	F *	

**FIG. 8(c)-2**

ES 2 607 122 T3

GFTpsD1

SEQ ID NO: 9	1	ATG TCG TCT GGA GAA ACA TTT CGT CGT ACT GCA GAT TTC CAT CCT	45
SEQ ID NO: 4	1	M S S G E T F R P T A D F H P	15
	46	AGT TTA TGG AGA AAC CAT TTC CTC AAA GGT GCT TCT GAT TTC AAG	90
	16	S L W R N H F L K G A S D F K	30
	91	ACA GTT GAT CAT ACT GCA ACT CAA GAA CGA CAC GAG GCA CTG AAA	135
	31	T V D H T A T Q E R H E A L K	45
	136	GAA GAG GTA AGG AGA ATG ATA ACA GAT GCT GAA GAT AAG CCT GTT	180
	46	E E V R R M I T D A E D K P V	60
	181	CAG AAG TTA CGC TTG ATT GAT GAA GTA CAA CGC CTG GGG GTG GCT	225
	61	Q K L R L I D E V Q R L G V A	75
	226	TAT CAC TTT GAG AAA GAA ATA GAA GAT GCA ATA CAA AAA TTA TGT	270
	76	Y H F E K E I E D A I Q K L C	90
	271	CCA ATC TAT ATT GAC AGT AAT AGA GCT GAT CTC CAC ACC GTT TCC	315
	91	P I Y I D S N R A D L H T V S	105
	316	CTT CAT TTT CGA TTG CTT AGG CAG CAA GGA ATC AAG ATT TCA TGT	360
	106	L H F R L L R Q Q G I K I S C	120
	361	GAT GTG TTT GAG AAG TTC AAA GAT GAT GAG GGT AGA TTC AAG TCA	405
	121	D V F E K F K D D E G R F K S	135
	406	TCG TTG ATA AAC GAT GTT CAA GGG ATG TTA AGT TTG TAC GAG GCA	450
	136	S L I N D V Q G M L S L Y E A	150
	451	GCA TAC ATG GCA GTT CGC GGA GAA CAT ATA TTA GAT GAA GCC ATT	495
	151	A Y M A V R G E H I L D E A I	165
	496	GCT TTC ACT ACC ACT CAC CTG AAG TCA TTG GTA GCT CAG GAT CAT	540
	166	A F T T T H L K S L V A Q D H	180
	541	GTA ACC CCT AAG CTT GCG GAA CAG ATA AAT CAT GCT TTA TAC CGT	585
	181	V T P K L A E Q I N H A L Y R	195
	586	CCT CTT CGT AAA ACC CTA CCA AGA TTA GAG GCG AGG TAT TTT ATG	630
	196	P L R K T L P R L E A R Y F M	210
	631	TCC ATG ATC AAT TCA ACA AGT GAT CAT TTA TAC AAT AAA ACT CTG	675
	211	S M I N S T S D H L Y N K T L	225
	676	CTG AAT TTT GCA AAG TTA GAT TTT AAC ATA TTG CTA GAG CTG CAC	720
	226	L N F A K L D F N I L L E L H	240
	721	AAG GAG GAA CTC AAT GAA TTA ACA AAG TGG TGG AAA GAT TTA GAC	765
	241	K E E L N E L T K W W K D L D	255
	766	TTC ACT ACA AAA CTA CCT TAT GCA AGA GAC AGA TTA GTG GAG TTA	810
	256	F T T K L P Y A R D R L V E L	270

**FIG. 8(d)-1**



811	TAT TTT TGG GAT TTA GGG ACA TAC TTC GAG CCT	
271	Y F W D L G T Y F E P	
856	GGG AGA AAG ATA ATG ACC CAA TTA AAT TAC ATA TTA TCC ATC ATA	900
286	G R K I M T Q L N Y I L S I I	300
901	GAT GAT ACT TAT GAT GCG TAT GGT ACA CTT GAA GAA CTC AGC CTC	945
301	D D T Y D A Y G T L E E L S L	315
946	TTT ACT GAA GCA GTT CAA AGA TGG AAT ATT GAG GCC GTA GAT ATG	990
316	F T E A V Q R W N I E A V D M	330
991	CTT CCA GAA TAC ATG AAA TTG ATT TAC AGG ACA CTC TTA GAT GCT	1035
331	L P E Y M K L I Y R T L L D A	345
1036	TTT AAT GAA ATT GAG GAA GAT ATG GCC AAG CAA GGA AGA TCA CAC	1080
346	F N E I E E D M A K Q G R S H	360
1081	TGC GTA CGT TAT GCA AAA GAG GAG AAT CAA AAA GTA ATT GGA GCA	1125
361	C V R Y A K E E N Q K V I G A	375
1126	TAC TCT GTT CAA GCC AAA TGG TTC AGT GAA GGT TAC GTT CCA ACA	1170
376	Y S V Q A K W F S E G Y V P T	390
1171	ATT GAG GAG TAT ATG CCT ATT GCA CTA ACA AGT TGT GCT TAC ACA	1215
391	I E E Y M P I A L T S C A Y T	405
1216	TTC GTC ATA ACA AAT TCC TTC CTT GGC ATG GGT GAT TTT GCA ACT	1260
406	F V I T N S F L G M G D F A T	420
1261	AAA GAG GTT TTT GAA TGG ATC TCC AAT AAC CCT AAG GTT GTA AAA	1305
421	K E V F E W I S N N P K V V K	435
1306	GCA GCA TCA GTT ATC TGC AGA CTC ATG GAT GAC ATG CAA GGT CAT	1350
436	A A S V I C R L M D D M Q G H	450
1351	GAG TTT GAG CAG AAG AGA GGA CAT GTT GCG TCA GCT ATT GAA TGT	1395
451	E F E Q K R G H V A S A I E C	465
1396	TAC ACG AAG CAG CAT GGT GTC TCT AAG GAA GAG GCA ATT AAA ATG	1440
466	Y T K Q H G V S K E E A I K M	480
1441	TTT GAA GAA GAA GTT GCA AAT GCA TGG AAA GAT ATT AAC GAG GAG	1485
481	F E E E V A N A W K D I N E E	495
1486	TTG ATG ATG AAG CCA ACC GTC GTT GCC CGA CCA CTG CTC GGG ACG	1530
496	L M M K P T V V A R P L L G T	510
1531	ATT CTT AAT CTT GCT CGT GCA ATT GAT TTT ATT TAC AAA GAG GAC	1575
511	I L N L A R A I D F I Y K E D	525
1576	GAC GGC TAT ACG CAT TCT TAC CTA ATT AAA GAT CAA ATT GCT TCT	1620
526	D G Y T H S Y L I K D Q I A S	540
1621	GTG CTA GGA GAC CAC GTT CCA TTT TGA	1647
541	V L G D H V P F *	

**FIG. 8(d)-2**

ES 2 607 122 T3

GFTpsD2

SEQ ID NO: 10	1	ATG TCG TCT GGA GAA ACA TTT CGT CCT ACT GCA GAT TTC CAT CCT	45
SEQ ID NO: 5	1	M S S G E T F R P T A D F H P	15
	46	AGT TTA TGG AGA AAC CAT TTC CTC AAA GGT GCT TCT GAT TTC AAG	90
	16	S L W R N H F L K G A S D F K	30
	91	ACA GTT GAT CAT ACT GCA ACT CAA GAA CGA CAC GAG GCA CTG AAA	135
	31	T V D H T A T Q E R H E A L K	45
	136	GAA GAG GTA AGG AGA ATG ATA ACA GAT GCT GAA GAT AAG CCT GTT	180
	46	E E V R R M I F D A E D K P V	60
	181	CAG AAG TTA CGC TTG ATT GAT GAA GTA CAA CGC CTG GGG CTG GCT	225
	61	Q K L R L I D E V Q R L G V A	75
	226	TAT CAC TTT GAG AAA GAA ATA GAA GAT GCA ATA CAA AAA TTA TGT	270
	76	Y H F E K E I E D A I Q K L C	90
	271	CCA AAC TAT ATT CAC AGT AAT AGC CCT GAT CTT CAC ACC GTT TCT	315
	91	P N Y I H S N S P D L H T V S	105
	316	CTT CAT TTT CGA TTG CTT AGG CAG CAA GGA ATC AAG ATT TCA TGT	360
	106	L H F R L L R Q Q G I K I S C	120
	361	GAT GTG TTT GAG AAG TTC AAA GAT GAT GAG GGT AGA TTC AAG TCA	405
	121	D V F E K F K D D E G R F K S	135
	406	TCG TTG ATA AAC GAT GTT CAA GGG ATG TTA AGT TTG TAC GAG GCA	450
	136	S L I N D V Q G M L S L Y E A	150
	451	GCA TAC ATG GCA GTT CGC GGA GAA CAT ATA TTA GAT GAA GCC ATT	495
	151	A Y M A V R G E H I L D E A I	165
	496	GCT TTC ACT ACC ACT CAC CTG AAG TCA TTG GTA GCT CAG GAT CAT	540
	166	A F T T T H L K S L V A Q D H	180
	541	GTA ACC CCT AAG CTT CGC GAA CAG ATA AAT CAT GCT TTA TAC CGT	585
	181	V T P K L A E Q I N H A L Y R	195
	586	CCT CTT CGT AAA ACC CTA CCA AGA TTA GAG GCG AGG TAT TTT ATG	630
	196	P L R K T L P R L E A R Y F M	210
	631	TCC ATG ATC AAT TCA ACA AGT GAT CAT TTA TAC AAT AAA ACT CTG	675
	211	S M I N S T S D H L Y N K T L	225
	676	CTG AAT TTT GCA AAG TTA GAT TTT AAC ATA TTG CTA GAG CTG CAC	720
	226	L N F A K L D F N I L L E L H	240
	721	AAG GAG GAA CTC AAT GAA TTA ACA AAG TGG TGG AAA GAT TTA GAC	765
	241	K E E L N E L T K W W K D L D	255
	766	TTC ACT ACA AAA CTA CCT TAT GCA AGA GAC AGA TTA GTG GAG TTA	810
	256	F T T K L P Y A R D R L V E L	270

**FIG. 8(e)-1**

811	TAT TTT TGG GAT TTA GGG ACA TAC TTC GAG CCT CAA TAT GCA TTT	855
271	Y F W D L G T Y F E P Q Y A F	285
856	GGG AGA AAG ATA ATG ACC CAA TTA AAT TAC ATA TTA TCC ATC ATA	900
286	G R K I M T Q L N Y I L S I I	300
901	GAT GAT ACT TAT GAT GCG TAT GGT ACA CTT GAA GAA CTC AGC CTC	945
301	D D T Y D A Y G T L E E L S L	315
946	TTT ACT GAA GCA GTT CAA AGA TGG AAT ATT GAG GCC GTA GAT ATG	990
316	F T E A V Q R W N I E A V D M	330
991	CTT CCA GAA TAC ATG AAA TTG ATT TAC AGG ACA CTC TTA GAT GCT	1035
331	L P E Y M K L I Y R T L L D A	345
1036	TTT AAT GAA ATT GAG GAA GAT ATG GCC AAG CAA GGA AGA TCA CAC	1080
346	F N E I E E D M A K Q G R S H	360
1081	TGC GTA CGT TAT GCA AAA GAG GAG AAT CAA AAA GTA ATT GGA GCA	1125
361	C V R Y A K E E N Q K V I G A	375
1126	TAC TCT GTT CAA GCC AAA TGG TTC AGT GAA GGT TAC GTT CCA ACA	1170
376	Y S V Q A K W F S E G Y V P T	390
1171	ATT GAG GAG TAT ATG CCT ATT GCA CTA ACA AGT TGT GCT TAC ACA	1215
391	I E E Y M P I A L T S C A Y T	405
1216	TTC GTC ATA ACA AAT TCC TTC CTT GGC ATG GGT GAT TTT GCA ACT	1260
406	F V I T N S F L G M G D F A T	420
1261	AAA GAG GTT TTT GAA TGG ATC TCC AAT AAC CCT AAG GTT GTA AAA	1305
421	K E V F E W I S N N P K V V K	435
1306	GCA GCA TCA GTT ATC TGC AGA CTC ATG GAT GAC ATG CAA GGT CAT	1350
436	A A S V I C R L M D D M Q G H	450
1351	GAG TTT GAG CAG AAG AGA GGA CAT GTT GCG TCA GCT ATT GAA TGT	1395
451	E F E Q K R G H V A S A I E C	465
1396	TAC ACG AAG CAG CAT GGT GTC TCT AAG GAA GAG GCA ATT AAA ATG	1440
466	Y T K Q H G V S K E E A I K M	480
1441	TTT GAA GAA GAA GTT GCA AAT GCA TGG AAA GAT ATT AAC GAG GAG	1485
481	F E E E V A N A W K D I N E E	495
1486	TTG ATG ATG AAG CCA ACC GTC GTT GCC CGA CCA CTG CTC GGG ACG	1530
496	L M M K P T V V A R P L L G T	510
1531	ATT CTT AAT CTT GCT CGT GCA ATT GAT TTT ATT TAC AAA GAG GAC	1575
511	I L N L A R A I D F I Y K E D	525
1576	GAC GGC TAT ACG CAT TCT TAC CTA ATT AAA GAT CAA ATT GCT TCT	1620
526	D G Y T H S Y L I K D Q I A S	540
1621	GTG CTA GGA GAC CAC GTT CCA TTT TGA	1647
541	V L G D H V P F *	

**FIG. 8(e)-2**

ES 2 607 122 T3

		GFTpsE															
SEQ ID NO:7	1	ATG	TCT	TTG	GAA	GTT	TCA	GCC	TCT	CCT	GCT	AAA	GTT	ATC	CAA	AAT	45
SEQ ID NO:2	1	M	S	L	E	V	S	A	S	P	A	K	V	I	Q	N	15
	46	GCT	GGG	AAA	GAT	TCT	ACT	CGT	CGC	TCT	GCA	AAT	TAT	CAT	CCA	AGC	90
	16	A	G	K	D	S	T	R	R	S	A	N	Y	H	P	S	30
	91	ATC	TGG	GGG	GAT	CAT	TTC	CTT	CAA	TAT	ACT	TGT	GAC	ACC	CAG	GAA	135
	31	I	W	G	D	H	F	L	Q	Y	T	C	D	T	Q	E	45
	136	ACT	GAT	GAT	GGC	AGC	AAT	GTA	AAG	CAT	CTA	GAG	CTG	AAG	AAA	GAA	180
	46	T	D	D	G	S	N	V	K	H	L	E	L	K	K	E	60
	181	ATT	AGA	AGA	ATG	CTA	AAA	GCT	GAT	AAC	AAG	CCT	TCA	CGT	ACA	CTT	225
	61	I	R	R	M	L	K	A	D	N	K	P	S	R	T	L	75
	226	CAA	TTG	ATT	GAT	GCA	ATT	CAG	CGT	TTA	GGA	GTG	TCT	TAC	CAT	TTT	270
	76	Q	L	I	D	A	I	Q	R	L	G	V	S	Y	H	F	90
	271	GAA	ACT	GAG	ATT	GAT	GAA	ATA	TTG	GGA	AAG	ATG	CAT	AAG	GCT	TCC	315
	91	E	S	E	I	D	E	I	L	G	K	M	H	K	A	S	105
	316	CAA	GAC	TCT	GAT	CTT	TGT	GAT	AAT	GAA	AAT	GAT	GAG	CTC	TAT	TAT	360
	106	Q	D	S	D	L	C	D	N	E	N	D	E	L	Y	Y	120
	361	ATC	TCT	CTT	CAT	TTT	CGA	TTA	CTT	AGA	CAA	AAT	GGC	TAT	AAA	ATT	405
	121	I	S	L	H	F	R	L	L	R	Q	N	G	Y	K	I	135
	406	TCC	GCT	GAT	GTG	TTC	AAA	AAG	TTC	AAA	GAC	ACG	GAT	GGG	AAC	TTT	450
	136	S	A	D	V	F	K	K	F	K	D	T	D	G	N	F	150
	451	AAA	ACA	TCT	CTT	GCG	AAA	GAT	GTT	CGA	GSA	ATG	TTA	AGC	TTG	TAT	495
	151	K	T	S	L	A	K	D	V	R	G	M	L	S	L	Y	165
	496	GAA	GCT	ACG	CAT	CTC	GGG	GTA	CAT	GAA	GAA	GAT	ATA	CTA	GAT	GAA	540
	166	E	A	T	H	L	G	V	H	E	E	D	I	L	D	E	180
	541	GCG	CTT	GCT	TTC	ACC	ACT	AGT	CAC	CTA	GAG	TCA	ATA	GCG	ACT	CAT	585
	181	A	L	A	F	T	T	S	H	L	E	S	I	A	T	H	195
	586	CAA	ATC	AGG	TCT	CCA	CTT	GTT	GAA	CAA	GTC	AAA	CAT	GCC	TTA	GTT	630
	196	Q	I	R	S	P	L	V	E	Q	V	K	H	A	L	V	210
	631	CAG	CCT	ATC	CAC	AGG	GGC	TTC	CAA	AGG	CTT	GAG	GCA	AGA	CAG	TAC	675
	211	Q	P	I	H	R	G	F	Q	R	L	E	A	R	Q	Y	225
	676	ATT	CCT	ATC	TAT	CAA	GAA	GAA	TCT	CCC	CAC	AAT	GAA	GCT	CTG	TTA	720
	226	I	P	I	Y	Q	E	E	S	P	H	N	E	A	L	L	240
	721	ACT	TTT	GCA	AAG	TTA	GAT	TTT	AAC	AAA	TTG	CAA	AAG	CCT	CAC	CAG	765
	241	T	F	A	K	L	D	F	N	K	L	Q	K	P	H	Q	255
	766	AAG	GAA	CTC	GGT	GAT	ATT	TCA	AGG	TGG	TGG	AAA	GAA	TTA	GAC	TTT	810
	256	K	E	L	G	D	I	S	R	W	W	K	E	L	D	F	270

**FIG. 8(f)-1**

811	GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA GTT GCG GAG TGC TAC	855
271	A H K L P F I R D R V A E C Y	285
856	TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA	900
286	F W I L G V Y F E P Q Y S F A	300
901	AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT	945
301	R R I L T K V I S M T S V I D	315
946	GAT ATC TAT GAT GTG TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT	990
316	D I Y D V Y G K I E E L E L F	330
991	ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT GCC ATA GAT CAA CTT	1035
331	T S A I E R W D I S A I D Q L	345
1036	CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT	1080
346	P E Y M K L C Y R A L L D V F	360
1081	AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC	1125
361	S E A E K D L A P Q G K S Y R	375
1126	CTC TAT TAT GCA AAA GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC	1170
376	L Y Y A K E A M K N M V K N Y	390
1171	TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT TAT GTA CCT ACA GTG	1215
391	F Y E A K W C L Q N Y V P T V	405
1216	GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA TCT GGC TCC CCA ATG	1260
405	D E Y M T V A L V T S G S P M	420
1261	TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA	1305
421	L S T T S F V G M G D I V T K	435
1306	GAA TCT TTT GAG TGG TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT	1350
436	E S F E W L F S N P R F I R A	450
1351	TCT TCT ATA GTT TGC CGA CTC ATG GAT GAC ATA GTG TCA CAC AAG	1395
451	S S I V C R L M D D I V S H K	465
1396	TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC	1440
466	F E Q S R G H V A S S V E C Y	480
1441	ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT	1485
481	M K Q H G A T E E E A C N E F	495
1486	CGG AAA CAA GTT TCA AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC	1530
496	R K Q V S N A W K D I N E D C	510
1531	CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT CTG ATG CGA ATT CTC	1575
511	L R P T V V P M P L L M R I L	525
1576	AAT CTT ACA CGC GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC	1620
526	N L T R V I D V I Y K Y E D G	540
1621	TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG	1665
541	Y T H S A V V L K D F V A S L	555
1666	TTT ATT AAT CCT GTG CCG ATA TGA	1689
556	F I N P V P I *	

**FIG. 8(f)-2**