

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 540 791**

21 Número de solicitud: 201331820

51 Int. Cl.:

C12N 9/88 (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

13.12.2013

43 Fecha de publicación de la solicitud:

13.07.2015

Fecha de la concesión:

29.04.2016

45 Fecha de publicación de la concesión:

09.05.2016

56 Se remite a la solicitud internacional:

PCT/ES2014/070916

73 Titular/es:

**CONSEJO SUPERIOR DE INVESTIGACIONES
CIENTÍFICAS (CSIC) (75.0%)**

SERRANO, 117

28006 MADRID (Madrid) ES y

**INSTITUTO VALENCIANO DE INVESTIGACIONES
AGRARIAS (IVIA) (25.0%)**

72 Inventor/es:

RUIZ RIVERO, Omar José;

CARRERA BERGUA, Esther;

RODRIGO ESTEVE, María Jesús;

ZACARÍAS GARCÍA, Lorenzo Ángel y

TALÓN CUBILLO, Manuel

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

54 Título: **Nuevas sesquiterpeno sintasas aisladas de la piel de naranja.**

57 Resumen:

Nuevas sesquiterpeno sintasas aisladas de la piel de naranja.

La presente invención se relaciona con un enzima natural sesquiterpeno sintasa de naranjo (CsTPS3) con actividad δ-cadineno sintasa, así como con las variantes de dicha enzima modificadas por la sustitución de al menos un aminoácido que altera su actividad catalítica, transformándolas preferentemente en enzimas con actividad elemol sintasa o germacreno D sintasa. Son asimismo objeto de la presente invención, el ácido nucleico que codifica el enzima natural CsTPS3 así como las versiones modificadas derivadas de éste, además de las construcciones génicas, vectores de expresión y cepas bacterianas que los comprenden.

La invención también se refiere al uso de estas enzimas para sintetizar sesquiterpenos, preferentemente δ-cadineno, elemol y germacreno D, y proporciona un método para la obtención de los mismos.

ES 2 540 791 B1

DESCRIPCIÓN

NUEVAS SESQUITERPENO SINTASAS AISLADAS DE LA PIEL DE NARANJA

5 **Sector de la técnica**

La presente invención se relaciona con un enzima natural sesquiterpeno sintasa de naranjo (CsTPS3) con actividad δ -cadineno sintasa. Además, la invención se relaciona con una molécula natural de ácido nucleico aislada de la piel de frutos de naranjo (*Citrus sinensis*)
10 que codifica la sesquiterpeno sintasa natural CsTPS3, con moléculas de ácidos nucleicos modificadas *in vitro* que derivan de la secuencia natural *Cstps3*, con sus respectivos enzimas codificadas sesquiterpeno sintasas modificadas, con métodos para preparar sesquiterpeno sintasas, y con las células bacterianas huésped que expresan los ácidos nucleicos que codifican los polipéptidos de la invención. La presente invención se relaciona
15 también con métodos para sintetizar sesquiterpenos. Por ejemplo, las sesquiterpenos sintasas objeto de la invención pueden usarse para convertir el precursor farnesil-pirofosfato (FPP) en varios sesquiterpenos bicíclicos que incluyen δ -cadineno, β -cubebeno, α -copaeno, germacreno D, α -cadineno y cubebol, y el sesquiterpeno monocíclico elemol con una alta actividad repelente de mosquitos.

20

Estado de la técnica anterior

Los terpenoides (isoprenoides) son importantes productos naturales, de los cuales se han identificado más de 30,000 compuestos en el reino vegetal. Entre ellos, los monoterpenos
25 (C10) y sesquiterpenos (C15) son los principales constituyentes de los aceites esenciales de origen vegetal (Dewick, [2009] En *Medicinal Natural Products*, John Wiley & Sons Ltd, Chichester-West Sussex, pp 187-310; Fraga, [2006] *Nat Prod Rep.* **23**, 943-972). Los sesquiterpenos son usualmente compuestos volátiles que pueden actuar como moléculas señales para la atracción de polinizadores y dispersores de semillas, depredadores de
30 herbívoros o repelentes de plagas, además de tener un papel importante en la comunicación entre plantas (revisado por Gershenzon y Dudareva, [2007] *Nat Chem Biol.* **3**, 408-414).

Las terpeno sintasas (TPS) son los enzimas responsables para la formación de la gran diversidad de terpenos identificados en el reino vegetal (revisado por Chen y col., [2011]
35 *Plant J.* **66**, 212-219). Las sesquiterpeno sintasas convierten el substrato farnesil-pirofosfato (FPP) en diversos sesquiterpenos que pueden tener diferentes estructuras químicas. La

mayoría de las TPS vegetales que se han caracterizado bioquímicamente son enzimas multicatalíticas y multiproducto con una alta promiscuidad y plasticidad de aminoácidos. Al parecer, varios aminoácidos localizados en el sitio activo y capas circundantes de las TPS multifuncionales actúan colectivamente para controlar la cascada en reacción que conduce a la formación de mezclas complejas de sesquiterpenos (Miller y Allemann, [2012] *Nat Prod Rep.* **29**, 60-71).

Aunque han sido muchos los monoterpenos y sesquiterpenos que se han identificado en varias especies del género *Citrus* (revisado por Tadeo y col., [2008] *Ad Bot Res* **47**, 147-223), pocos han sido los genes y enzimas monoterpeno y sesquiterpeno sintasas que se han aislado y caracterizado bioquímicamente. Con respecto a las sesquiterpeno sintasas, el gen *Cstps1* de naranjo (*Citrus sinensis*) el cual codifica una valenceno sintasa (CVS) (Sharon-Asa y col., [2003] *Plant J.* **36**, 664-676), fue patentado por el Ministerio de Agricultura de Israel (WO 2005/021705, 10 de Marzo de 2005). Hasta la fecha, se han patentado otras monoterpeno y sesquiterpeno sintasas vegetales tales como: la *E*- α -bisaboleno sintasa, δ -selineno sintasa y γ -humuleno sintasa del abeto gigante *Abies grandis* (US 6451576, *Washington State University Research Foundation*, 17 de Septiembre de 2002); γ -curcumeno sintasa, (+)-germacreno A sintasa, (-)-germacreno D sintasas y pachoulol sintasas de pachouli (*Pogostemon cablin*) (WO 2005/052163, Firmenich S.A., 9 de June de 2005); sesquiterpeno sintasas multifuncional de *Vetiveria zizanioides*, que catalizan la formación de ciclocopacanfeno, (+)-*epi*- β -santaleno, *trans*- α -bergamoteno, *cis*- α -bergamoteno, β -bisaboleno y/o *trans*- γ -bisaboleno (WO 2006/134523, Firmenich S.A., 21 de Diciembre de 2006); germacreno D sintasa multifuncional del fruto del kiwi (WO2004/058814, *The Horticulture and Food Research Institute of New Zealand Limited*, 15 de Julio de 2004); santaleno sintasa (SSy) de las especies *Santalum album* (SaSSy), *S. austrocaledonicum* (SauSSy) y *S. spicatum* (SspiSSy) (WO 2011/000026, 6 de Junio de 2011, *The University of Western Australia, Forest Products Commission, University of British Columbia*, y Joerg Bohlmann); y valenceno sintasas modificadas (WO 2012/058636, Allylix INC, 3 de Mayo de 2012).

La comercialización de los terpenos vegetales a nivel mundial alcanza los 20 billones de dólares anuales. La creciente demanda de terpenos de origen vegetal por la industria cosmética y perfumera, ha forzado la búsqueda de nuevas fuentes y la subsecuente aceptación de equivalentes sintéticos con menos olor y calidad. La síntesis química de terpenos (usando sustancias pesadas), resulta en mezclas altamente complejas (productos colaterales e isómeros), los cuales requieren posteriormente de laboriosas y costosas

etapas de purificación. Hasta la fecha no existe ningún método de síntesis química de terpenos que sea respetuoso con el medio ambiente y que además sea notoriamente eficiente. Por el contrario, existen diversos ejemplos que muestran como la biotecnología puede ser una estrategia económica y fiable frente a la síntesis química o la destilación convencional a partir de materias primas de origen vegetal (revisado por Krings y Berger, 5 [2010] *Nat Prod Commun.* **5**, 1507-1522). La tecnología del DNA recombinante y herramientas de biología molecular, la disponibilidad de base de datos de ácidos nucleicos y proteínas, así como el diseño de programas informáticos para el modelaje *in silico* y acoplamiento molecular de proteínas, han contribuido al rápido progreso de la biosíntesis de 10 terpenos.

Además de su importancia biológica como señales químicas, los terpenos de origen vegetal tienen un inmenso valor comercial debido a su uso como fragancias en cosmética y perfumería, como aditivos organolépticos de alimentos y bebidas o para adicionar fragancias deseables a productos de limpieza. El uso de fitoaceites esenciales, así como de sus 15 componentes en forma aislada (principalmente monoterpenos y sesquiterpenos), como repelentes de mosquitos transmisores de ciertas enfermedades tropicales infecciosas de alto riesgo, tales como: paludismo, dengue hemorrágico, fiebre amarilla, filariasis o fiebre del Nilo Occidental (*West Nile Virus Infection: WNV*), les confiere además un gran interés 20 médico y farmacéutico (revisado por Pohlit y col., [2011] *Planta Medica* **77**, 598-617). En septiembre de 2012, el mosquito doméstico del norte (*Culex pipens*), causó 2.636 casos de infección del WNV en el este de los Estados Unidos de América, produciendo 118 defunciones. Del total de casos detectados, el 53% mostraron complicaciones neuroinvasivas (meningitis y encefalitis) (<http://www.cdc.gov/ncidod/dvbid/westnile/index.htm>).

25 Los fitoaceites esenciales extraídos de citronela (*Cymbopogon spp.*), eucalipto (*Eucalyptus L'Hér. spp.*), y albahaca (*Ocimum spp.*), se han utilizado tradicionalmente como repelentes de mosquitos en la fabricación de velas, lociones, geles, aspersores y toallas higiénicas. No obstante, la eficiencia de estos productos como repelentes de insectos es cuestionable. 30 Recientemente, 70 fitoaceites esenciales utilizados en diferentes patentes de formulas repelentes de mosquitos han sido objeto de revisiones exhaustivas (Nerio y col., [2010] *Bioresour Technol.* **101**, 372-378; Pohlit y col., [2011] *Planta Medica* **77**, 618-630). En contraparte a los repelentes de insectos de origen vegetal, el ingrediente sintético activo presente en la mayoría de las formulas farmacéuticas que se comercializan como repelentes 35 de mosquito hematófagos, es el DEET (*N,N*-dietil-*m*-toluamida). Sin embargo, su uso tiene ciertos inconvenientes, ya que tiene un olor desagradable y su aplicación cutánea puede

resultar irritante, además de conferírsele ciertos efectos perjuicios para la salud, especialmente en niños. En este sentido, los repelentes e insecticidas de origen vegetal se consideran como una prometedora alternativa frente a los repelentes sintéticos, debido a su baja toxicidad y seguridad medioambiental (Katz y col., [2008] *J Am Acad Dermatol.* **58**, 865-871). Los consumidores, cada vez más conscientes de temas relacionados con la salud pública y seguridad ambiental, demandan formulas repelentes de insectos que no representen un peligro para su salud ni para el medio ambiente. Es por ello que las industrias farmacéuticas continúan trabajando en la búsqueda de formulas repelentes de insectos inocuas para la salud pública, y en mejorar las ya existentes mediante la incorporación de terpenos vegetales con una alta capacidad y amplio espectro de acción repelente.

Los fitoaceites ricos en sesquiterpenos, así como sus componentes purificados, que contienen radicales alcohol, aldehído, cetona o ácido, tienen un mayor interés científico debido a su mayor eficacia como repelentes naturales alternativos, que aquellos sintéticos que contienen DEET. Así, por ejemplo, el elemol y el β -eudesmol además de ser efectivos como larvicidas exhiben también una alta repelencia contra hembras mosquitos adultas (Coats y col., [2003] US Patente N° 6.524.605; [2005] US Patente N° 6.524.043; Schultz y col., [2004] *Environ Entomol.* **33**, 1562-1569; Palauch y col., [2009] *J Agri Food Chem.* **57**, 7618- 7625; Zhu y col., [2008] *J Am Mosq Control Assoc.* **24**,161-168). Recientemente, se ha demostrado que el elemol es un potente repelente de ninfas de la garrapa de patas negras (*Ixodes scapularis*), y de la garrapata estrella solitaria (*Amblyomma americanum*), transmisores de la enfermedad de Lyme y la Ehrlichiosis monocitotrópica humana, respectivamente (Carrol y col., [2010] *Exp Appl Acarol.* **51**, 383-392). Ensayos *in vitro* e *in vivo* han demostrado que concentraciones relativamente bajas de elemol son suficientes para repeler garrapatas y mosquitos (hembras adultas y larvas). Además, el hecho de que el elemol sea tan efectivo o más que la DEET (el repelente estándar utilizado para la comparación de la eficacia de los demás repelentes), apoya fuertemente su uso como un repelente eficaz para ser utilizado en formulas comerciales. En este sentido, existe un registro de una patente internacional (WO 2007/109736, Iowa State University Research Foundation, 27th September 2007), relacionada con el diseño de mezclas de fitoaceites esenciales que contienen elemol, o fitoaceites y elemol, o elemol mezclado con estabilizadores, para ser utilizadas como repelentes del mosquito del norte (*Culex pipens*), vector de la fiebre amarilla (*A. aegypti*), y de la cucaracha común (*Blattella germanica*). En la actualidad, el elemol que se utiliza con fines comerciales se destila del aceite de citronela (*C. winterianus*) (<http://gbffchina.en.alibaba.com/>), y su comercio mundial es de 1 - 10

toneladas métricas por año (revisado por Bahtia y col., [2008] *Food & Chemical Toxicol.* **46**, S147-S148.).

Un objetivo de la presente invención es proveer nuevas sesquiterpenos sintasas (TPS) capaces de catalizar la transformación del FPP en sesquiterpenos cíclicos basados en la estructura química del germacreno o elemeno. Es además un objetivo de esta invención, el proveer TPS modificadas capaces de sintetizar sesquiterpenos para los cuales no se ha aislado ninguna enzima de cítricos. En particular, es del objetivo de la presente invención proveer la forma de sintetizar el sesquiterpeno monocíclico oxigenado elemol, el cual tiene una potente actividad repelente de mosquitos, a través de la expresión de TPS modificadas *in vitro* derivadas de la TPS natural δ -cadineno sintasa aislada de la piel de frutos de naranja, objeto de esta invención.

Breve descripción de la invención

La gran mayoría de los sesquiterpenos que se utilizan en cosmética y perfumería se extraen de plantas en forma de aceites esenciales (fitoaceites). El precio y disponibilidad de los fitoaceites depende en gran medida de la distribución geográfica de las especies vegetales utilizadas como materia prima para su extracción, así como de la abundancia y rendimiento de los mismos. Debido a la complejidad de su estructura, la síntesis química de los terpenos está limitada por sus laboriosos procesos y altos costes de producción. La baja abundancia de las terpeno sintasas vegetales y la elevada dificultad de purificarlas a partir de las mezclas de resinas y compuestos fenólicos, hace casi imposible aislar estos enzimas a partir de los tejidos que expresan los genes que las codifican. El constante incremento del conocimiento de las rutas de biosíntesis de terpenos, así como la caracterización bioquímica de la actividad de sus enzimas biosintéticos, ha abierto nuevas perspectivas para la producción biotecnológica de terpenos y garantizar la sostenibilidad del sector. El diseño de procesos bioquímicos para la producción de terpenos es, por lo tanto, de gran interés, pero requiere del aislamiento y manipulación mediante técnicas de ingeniería genética de las secuencias nucleotídicas que codifican terpenos sintasas.

Específicamente, la presente invención contempla: (i) moléculas naturales de ácidos nucleicos aisladas de la piel de naranjas que codifican una sesquiterpeno sintasa que cataliza la síntesis de sesquiterpenos cíclicos, (ii) sesquiterpeno sintasas recombinantes, (iii) vectores de expresión recombinantes en los cuales se han clonado las secuencias polinucleotídicas que codifican las sesquiterpeno sintasas objeto de la invención, (iv) células

bacterianas que han sido transformadas con los vectores recombinantes objeto de la invención, y (v) métodos para la producción *in vitro* de sesquiterpenos cíclicos, incluyendo las etapas de introducción de los vectores recombinantes en las células bacterianas, en condiciones que aseguran la expresión de las secuencias nucleotídicas objeto de la invención, así como la inducción de la síntesis de las proteínas codificadas en las células bacterianas transformadas. En la presente invención también se describen los métodos utilizados para la modificación de la sesquiterpeno sintasa natural de naranjo objeto de esta invención, así como las versiones modificadas derivadas de ésta. Moléculas de DNA representativas de aquella TPS natural aislada del flavedo (piel) de naranjas objeto de la presente invención, fueron clonadas y secuenciadas, se dedujo su secuencia de aminoácidos y se ensayó *in vitro* su actividad enzimática. Específicamente, la secuencia nucleotídica natural aislada del flavedo de frutos de naranjo (*Citrus sinensis*) codifica el enzima δ -cadineno sintasa SEQ ID NO:2 (CsTPS3).

Además del enzima natural δ -cadineno sintasa de naranjo objeto de la presente invención, también se describe la obtención de versiones modificadas derivadas de ésta, las cuales están contempladas dentro de la protección de la invención, salvo en la medida en que estuviesen limitadas por el estado de la técnica. Los enzimas modificados objeto de esta invención, tienen una alterada especificidad y racionalización de productos. Ejemplo de estos, son aquellos enzimas modificados que producen mayores niveles relativos de germacreno D y pequeñas cantidades de β -elemeno y α -gurjuneno, o aquellas que producen mayores niveles de elemol en comparación con los correspondientes niveles producidos por la versión natural CsTPS3. Cabe destacar que el elemol es un potente repelente de mosquitos transmisores de enfermedades infecciones consideradas de alto riesgo para la salud pública, además de ser un ingrediente fragante muy utilizado en cosmética y perfumería.

En otros aspectos, está patente provee las células bacterianas que fueron transformadas con los vectores de expresión recombinantes en los cuales se clonaron las secuencias de DNA objeto de la invención. En este sentido, y de forma específica se proveen los plásmidos recombinantes que expresan la sesquiterpeno sintasa natural δ -cadineno sintasa (CsTPS3) de naranjo, así como las versiones modificadas que se derivan de ésta, en las cuales se han introducidos cambios de aminoácidos que alteran su actividad catalítica y/o especificidad de producto. Por lo tanto, microorganismos (células bacterianas) o plantas transgénicas que expresen las secuencias nucleotídicas que codifican las sesquiterpeno sintasas objeto de la presente invención, pueden usarse como biofactorías para la producción, aislamiento y

purificación de cantidades significativas de los enzimas naturales δ -cadineno sintasa (CsTPS3) y de sus versiones modificadas, así como de los sesquiterpenos producto de su actividad catalítica.

5 Descripción de las figuras

Figura 1. Secuencia de DNA del gen *Cstps3* (SEQ ID NO:1), y de las versiones mutantes derivadas de éste, y secuencia deducida de aminoácidos de los polipéptidos codificados. **(A)** Secuencia de DNA del gen *Cstps3* (SEQ ID NO:1), y del polipéptido codificado (CsTPS3, SEQ ID NO:2). **(B)** Secuencia de DNA de la versión *Cstps3-VB2* (SEQ ID NO:3), del polipéptido codificado (CsTPS3-VB2, SEQ ID NO:4). **(C)** Secuencia de DNA de la versión *Cstps3-VB5* (SEQ ID NO:5), y del polipéptido codificado (CsTPS3-VB5, SEQ ID NO:6). **(D)** Secuencia de DNA de la versión *Cstps3-VC1* (SEQ ID NO:7), y del polipéptido codificado (CsTPS3-VC2, SEQ ID NO:8). **(E)** Secuencia de DNA de la versión *Cstps3-VC2* (SEQ ID NO:9), y del polipéptido codificado (CsTPS3-VC2, SEQ ID NO:10).

Figura 2. Estructura química de los sesquiterpenos volátiles emitidos por el flavedo de naranjas Navel. Algunos de ellos corresponden a los sesquiterpenos identificados como productos de la actividad catalítica de la sesquiterpeno sintasa natural CsTPS3 y sus versiones modificadas objetos de la invención.

Figura 3. Alineamiento múltiple de la secuencia deducida de aminoácidos de los polipéptidos CsTPS3 (SEQ ID NO:2), y de las versiones mutantes VB2 (SEQ ID NO:4), VB5 (SEQ ID NO: 6), VC1 (SEQ ID NO:8), y VC2 (SEQ ID NO:10). La secuencia deducida de aminoácidos del polipéptido CSTPS3 se muestra en la parte superior del alineamiento. Los aminoácidos substituidos en cada una de las versiones mutantes están indicados. Los puntos representan aminoácidos idénticos entre los polipéptidos variantes y la secuencia natural CsTPS3.

Figura 4. Alineamiento múltiple de los ESTs truncados que comparten similitud con la secuencia nucleotídica *Cstps3* que codifica la sesquiterpeno sintasa natural CsTPS3 de naranjo (*C. sinensis* var. Washington Navel) objeto de la invención. Se indican los oligonucleotidos utilizados como cebadores en ensayos de PCR para la amplificación de la secuencia codificante completa del gen *Cstps3*. Las letras blancas sobre fondo negro representan nucleótidos idénticos en la misma posición. Las letras sobre fondo gris

representan nucleótidos diferentes en la misma posición. Los guiones representan huecos (*gaps*) en la secuencia de nucleótidos.

Figura 5. Alineamiento múltiple de la secuencia deducida de aminoácidos del polipéptido natural CsTPS3 (SEQ ID NO:2) y otras germacreno sintasas vegetales (Ger) filogenéticamente relacionadas. El alineamiento se realizó con el programa informático Expresso 3D Coffee (<http://www.tcoffee.org/>), utilizando como modelo la estructura cristalográfica de la 5-*epi*-aristoloqueno sintasa de tabaco (TEAS. PDB: 5EAT [DOI:10.2210/pdb5eat/pdb]). Los aminoácidos absolutamente conservados se muestran sobre fondo negro, mientras que los residuos altamente conservados se muestran sobre fondo gris (pálido u oscuro). Los motivos conservados RRX8W, RXR, GVYFEP, DDX2D, NSE/DTE de las sesquiterpenos sintasas vegetales se indican sobre el alineamiento. AdGerD (*Actinidia deliciosa* GerD, AAAX16121), CcGerB (*Cistus creticus* GerB, ACF94469), Nt5EAS (*Nicotiana tabacum* TEAS, Q40577); PtdGerD (*Populus trichocarpa x deltoides* GerD, AAR99061), RhGerD (*Rosa hybrida* GerD, BQ105086), y VvGerD (*Vitis vinifera* GerD, AAS66357). (Δ) Señala aminoácidos involucrados en la catálisis de la TEAS. Los residuos señalados con (▼) se mutagenizaron para conocer su papel en la actividad catalítica del enzima CsTPS3. (§) Señala aminoácidos que, en la misma posición, son diferentes en el polipéptido CsTPS3 en comparación al resto de los polipéptidos con actividad TPS utilizadas en el alineamiento.

Figura 6. Árbol filogenético de sesquiterpeno sintasas vegetales de la subfamilia TPS-a, incluyendo el polipéptido CsTPS3. El árbol se construyó mediante análisis de máxima verosimilitud (*maximum likelihood analysis*), utilizando la secuencia de aminoácidos de enzimas con actividad cubebeno (Cub), cadineno (Cad) y germacreno (Ger) sintasa. AaGerA (*Artemisia annua* GerA, AAX16121), AdGerD (*Actinidia deliciosa* GerD, AAAX16121), CcGerB (*Cistus creticus* GerB, ACF94469), CiGerA (*Cichorium intybus* GerA, AAM21658), CmGerD (*Cucumis melo* GerD, ABX83200), CsTPS3 (*Citrus sinensis* TPS3), Ga(+)-δ-Cad (*Gossypium arboreum* (+)-δ-Cad, CAA77191), Gh(+)-δ-Cad (*Gossypium hirsutum* (+)-δ-Cadineno, AAX44033), HaCad (*Helianthus annuus* Cad, AAY41422), IdGerA (*Ixeris dentate* GerA, AAL92481), LhGerB (*Lycopersicum hirsutum* GerB, AAG41891), LhGerD (*L. hirsutum* GerD, AAG41892), LsGerA (*Lactuca sativa* GerA, AAM11627), Mg-β-Cub (*Magnolia grandiflora* β-Cub, ACC66281), MtTPS5 (*Medicago truncatula* TPS5, ABB01625), Ob(+)-GerD (*Ocimum basilicum* (+)-GerD, AAV63786), PcGerA (*Pogostemon cablin* GerA, AAS86321), Pc(-)-GerD (*P. cablin* (-)-GerD, AAS86320), Pt(-)-GerD (*Populus trichocarpa x deltoids* (-)-GerD, AAR99061), RhGerD (*Rosa hybrida* GerD, BQ105086),

SaGerD (*Santalum album* GerD, ACF24768), ScGerA (*Solidago canadensis* GerA, CAC36896), Sc(+)-GerD (*S. canadensis* (+)-GerD, AAR3114), Sc(-)-GerD (*S. canadensis* GerD, AAR3115), SIGerC (*Solanum lycopersicon* GerC, AAC39431), Vv(-)-GerD (*Vitis vinifera* (-)-GerD, AAS66357), y Zo(+)-GerD (*Zingiber officinale* (+)-GerD, AAX40665).

5

Figura 7. Perfil de sesquiterpenos volátiles producidos por la proteína CsTPS3, y cada una de sus versiones mutantes, en las condiciones del *ensayo enzimático I*. Los sesquiterpenos se identificaron mediante análisis de cromatografía de gases acoplada a la espectrofotometría de masas. Se muestran los sesquiterpenos volátiles identificados en ensayos enzimáticos en los cuales el sustrato FPP se expuso a: **(A)** el polipéptido CsTPS3 (SEQ ID NO:2); **(B)** el polipéptido modificado CsTPS3-VB2 (SEQ ID NO:4); **(C)** el polipéptido modificado CsTPS3-VB5 (SEQ ID NO:6); **(D)** el polipéptido modificado CsTPS3-VC1 (SEQ ID NO:8); **(E)** el polipéptido modificado CsTPS3-VC2 (SEQ ID NO:10). El número sobre cada pico del cromatograma de gases indica el sesquiterpeno volátil identificado mediante espectrofotometría de masas. **(F)** Espectro de masas de cada uno de los sesquiterpenos identificados. Para cada pico se muestra el espectro de masas, el nombre común y el índice de retención calculado. Se muestra también, con fines comparativos y de autenticidad de los sesquiterpenos identificados, el espectro de masas de cada uno de los estándares de la librería de sesquiterpenos volátiles utilizada.

20

Figura 8. Perfil de sesquiterpenos volátiles producidos por la proteína CsTPS3, y cada una de sus versiones mutantes, en las condiciones del *ensayo enzimático II*. Los sesquiterpenos se identificaron mediante análisis de cromatografía de gases acoplada a la espectrofotometría de masas. Se muestran los sesquiterpenos volátiles identificados en ensayos enzimáticos en los cuales el sustrato FPP se expuso a: **(A)** el polipéptido CsTPS3 (SEQ ID NO:2); **(B)** el polipéptido modificado CsTPS3-VB2 (SEQ ID NO:4); **(C)** el polipéptido modificado CsTPS3-VB5 (SEQ ID NO:6); **(D)** el polipéptido modificado CsTPS3-VC1 (SEQ ID NO:8); **(E)** el polipéptido modificado CsTPS3-VC2 (SEQ ID NO:10). El número sobre cada pico del cromatograma de gases indica el sesquiterpeno volátil identificado mediante espectrofotometría de masas. **(F)** Espectro de masas de cada uno de los sesquiterpenos identificados. Para cada pico se muestra el espectro de masas, el nombre común y el índice de retención calculado. Se muestra también, con fines comparativos y de autenticidad de los sesquiterpenos identificados, el espectro de masas de cada uno de los estándares de la librería de sesquiterpenos volátiles utilizada.

35

Figura 9. Mecanismos de ciclización del FPP propuestos para explicar la especificidad de los sesquiterpenos cíclicos producidos por la actividad enzimática del polipéptido CsTPS3 (SEQ ID NO:2). Cada ruta de ciclización involucra la transformación del *Z,E*-germacradienil catión **(A)** o *E,E*-germacradienil catión **(B)**. Se muestra el nombre de los intermediarios catiónicos macrocíclicos propuestos para cada ruta de ciclización. Las posibles rutas que conducen a la formación de múltiple productos se indican con letras minúsculas en negritas y entre paréntesis. Los productos finales de cada ruta se resaltan en negrita. (Δ) Calentamiento en el puerto de inyección. Las estructuras químicas se dibujaron con el programa informático ChemBioDraw Ultra 11.0.1 (CambridgeSoft).

10

Figura 10. Estructura tridimensional (3D) del polipéptido CsTPS3. La simulación de la estructura de la proteína CsTPS3 se realizó utilizando como patrón la estructura 3D de la 5-*epi*-aristoloqueno sintasa de tabaco obtenida mediante cristalografía de rayos X (TEAS, PDB código: 5EAT) (Starks y col., [1997] *Science* **277**, 1815-1820). Los recuadros muestran un detalle del sitio activo la proteína CsTPS3. Las cadenas laterales de los aminoácidos que forman la cavidad catalítica, y los enlaces vecinos se representan con líneas. Asimismo, se indica la posición de los aminoácidos involucrados en la catálisis. La estructura 3D del polipéptido CsTPS3 se muestra impuesta sobre aquella de la TEAS formando un complejo con el substrato análogo farnesil hidroxifosfato (FHP). Las imágenes se obtuvieron mediante modelaje *in silico* utilizando el programa informático Discovery Studio v3.5 (Accelrys Software Inc.).

20

Figura 11. Detalle de la estructura 3D de la cavidad catalítica en el sitio activo del polipéptido CsTPS3, en comparación con el correspondiente sitio activo de la proteína 5-*epi*-aristoloqueno sintasa (TEAS, PDB código: 5EAT). La imposición del polipéptido CsTPS3 sobre la estructura 3D de la TEAS muestra las diferencias de identidad y localización espacial de las cadenas laterales de los aminoácidos dentro de la cavidad del sitio activo. La estructura 3D de la TEAS se muestra formando un complejo con el substrato análogo farnesil hidroxifosfato (FHP). Las flechas señalan los residuos de aminoácidos que difieren en identidad o localización espacial. Las imágenes se obtuvieron mediante modelaje *in silico* utilizando el programa informático Discovery Studio v3.5 (Accelrys Software Inc.).

30

Figura 12. Detalle de la estructura de la cavidad catalítica en el sitio activo de los polipéptidos modificados CsTPS3-VB2 (SEQ ID NO: 4) y CsTPS3-VC1 (SEQ ID NO: 8), en comparación con el correspondiente sitio activo de la proteína 5-*epi*-aristoloqueno sintasa (TEAS, PDB código: 5EAT). La imposición de los polipéptidos CsTPS3-VB2 **(A)** o CsTPS3-

35

VC1 (B) sobre la estructura 3D de la TEAS muestra las diferencias de identidad y localización espacial de las cadenas laterales de los aminoácidos reemplazados dentro de la cavidad del sitio activo. Las cadenas laterales de los aminoácidos que forman la cavidad catalítica, y los enlaces vecinos se representan con líneas. Los aminoácidos seleccionados para la mutagénesis dirigida del polipéptido CsTPS3 se representan en gris claro, mientras que los residuos substituidos en los polipéptidos modificados se muestran indicados con flechas. La estructura de los polipéptidos modificados se muestra solapada con la TEAS (aunque no se muestra en la imagen) formando complejo con el substrato análogo farnesil hidroxifosfato (FHP). Las imágenes se obtuvieron mediante modelaje *in silico* utilizando el programa informático Discovery Studio v3.5 (Accelrys Software Inc.).

Descripción detallada

A continuación se describe de forma minuciosa aspectos específicos de la presente invención, tales como los términos técnicos usados, las técnicas empleadas, los experimentos llevados a cabo y los resultados obtenidos. Asimismo, en el texto se hace referencia a las figuras adjuntas, las cuales forman a su vez parte de la invención, y tienen por finalidad ilustrar los conceptos discutidos y/o resultados obtenidos. Se entiende que aunque se realicen modificaciones o cambios estructurales en algunos aspectos de la invención, ello no implica la pérdida de la novedad ni de la protección de la invención. Por lo tanto, la descripción detallada de la invención, y el rango de protección de la presente patente no debe ser considerada en un sentido limitado. El rango de protección de la presente patente se subyuga a las reivindicaciones adjuntas, protegiendo en su totalidad todos y cada uno de los equivalentes a los cuales tales reivindicaciones dan derecho.

Primero se definen los términos científicos y técnicos empleados en la presente patente. A continuación, se describe la secuencia nucleotídica natural correspondiente a la región codificante del gen *Cstps3* aislado del flavedo de frutos de naranjo, así como el polipéptido codificado (CsTPS3). Posteriormente, se describen las secuencias nucleotídicas modificadas *in vitro* que derivan de la secuencia natural *Cstps3*, así como los correspondientes polipéptidos variantes codificados. Por último se describe también de forma detallada los materiales y métodos y resultados obtenidos, como ejemplos específicos de la presente invención.

Definición de términos

Un "terpeno" es una molécula hidrocarbonada con una estructura química básica de cinco átomos de carbono llamada isopreno (C_5H_8), la cual puede ser cerrada (cíclica) o abierta (acíclica). Los terpenos incluyen entre otros δ -cadineno, α -copaeno, β -cubebeno, germacreno D, α -gurjuneno, β -elemeno y elemol. En este sentido, los términos "terpenos" y "terpenoides", se usan aquí para referirse a terpenos y terpenos derivados, incluyendo aquellos que han sufrido una o más etapas de modificación (hidroxilación, isomerización, oxido-reducciones, metilación o acilación). Tal como se usa aquí, el término "sesquiterpeno" se refiere a un terpeno formado por cinco unidades de isoprenos (C_{15}), incluyendo aquellos que han sufrido una o varias etapas de modificación y/o funcionalización.

10

Tal como se usa aquí, el término "sesquiterpeno sintasa" se refiere a un enzima que cataliza, en presencia de magnesio y oxígeno molecular, la transformación del precursor acíclico farnesil-pirofosfato (FPP) en uno o más sesquiterpenos cíclicos o acíclicos. En este sentido, si una sesquiterpeno sintasa multifuncional (aquella que es capaz de producir más de un sesquiterpeno), produce β -copaeno como principal producto, el enzima recibe en consecuencia el nombre de β -copaeno sintasa. Según la invención, el producto principal o sesquiterpeno mayoritario puede obtenerse en función de las condiciones empleadas con una abundancia relativa de entre un 25% y un 100%, preferentemente de entre un 25% y un 90%, más preferentemente de entre un 45% y un 90%, aún más preferentemente de entre un 65% y un 90%, y aún todavía más preferentemente de entre un 75% y un 90%, respecto a la cantidad total de sesquiterpenos producidos. Del mismo modo, cuando la sesquiterpeno sintasa produce más de un sesquiterpeno, el sesquiterpeno o sesquiterpenos minoritarios se obtienen con una abundancia relativa inferior al 25%, preferentemente de entre un 1% y un 25%, más preferentemente de entre un 2,5% y un 25%, más preferentemente de entre un 4% y un 23%, aún más preferentemente de entre un 8% y un 23%, y aún todavía más preferentemente de entre un 11% y un 20%, con respecto a la cantidad total de sesquiterpenos producido.

30

El término "ácido nucleico" o "molécula de ácido nucleico" hace referencia a polímeros de deoxirribonucleótidos (nucleótidos) en la forma de una molécula de doble cadena (DNA: *deoxyribonucleic acid*). Tal como usa aquí, el término "nucleótido" se refiere a la unidad monomérica del DNA, la cual está formada por una molécula de azúcar de cinco átomos de carbono (pentosa), un grupo fosfato y una base nitrogenada (adenina [A], guanina [G], citosina [C] y timina [T]). El orden de los nucleótidos en la cadena principal del DNA determina el orden de los aminoácidos a lo largo de la cadena polipeptídica o proteína. Por lo tanto, la secuencia de DNA codifica la secuencia de aminoácidos de una proteína. Tal

35

como se usa aquí, el término “secuencia de DNA codificante” (CDS: *Coding DNA Sequence*), se refiere a la porción de un gene (incluyendo únicamente los exones) que codifica una proteína integra (*full length*). La CDS integra de cualquier gen comprende la secuencia de nucleótidos entre el codón de inicio (ATG), en su extremo 5', y el codón de
5 parada (TAG, TAA o TGA), en el extremo 3'.

Tal como se usa aquí, el término “aminoácido” se refiere a moléculas naturales ($H_2NCHR\text{COOH}$) que contienen un grupo amida ($-NH_2$) y uno carboxílico ($-COOH$), además de una cadena lateral (**R**) que es específica de cada aminoácido. Los veinte (20)
10 aminoácidos naturales combinados en la forma de cadenas polipeptídicas, forman las unidades básicas del conjunto de proteínas (proteoma) de un organismo. Los aminoácidos se identifican mediante una nomenclatura internacional establecida por la IUPAC (*International Union of Pure and Applied Chemistry*), tal como aparecen en la **Tabla 1**.

TABLA 1

Aminoácido	Código de una letra	Código de tres letras	Aminoácido	Código de una letra	Código de tres letras
Alanina	A	Ala	Metionina	M	Met
Cisteína	C	Cys	Asparagina	N	Asn
Ácido aspártico	D	Asp	Prolina	P	Pro
Ácido Glutámico	E	Glu	Glutamina	Q	Gln
Fenilalanina	F	Phe	Arginina	R	Arg
Glicina	G	Gly	Serina	S	Ser
Histidina	H	His	Treonina	T	Thr
Isoleucina	I	Ile	Valina	V	Val
Lisina	K	Lys	Tirosina	Y	Tyr
Leucina	L	Leu	Triptofano	W	Trp

Tal como se utiliza aquí, el término “oligonucleotidos” se refiere a secuencias cortas o
15 polideoxiribonucleotidos de cadena sencilla, que se utilizan como cebadores para la amplificación de fragmentos específicos de DNA mediante ensayos de PCR (*Polymerase Chain Reaction*), ampliamente conocidos en estado de la técnica. Los oligonucleotidos se sintetizan de forma química, y se purifican mediante diferentes métodos, los cuales incluyen el uso de cartuchos de fase inversa, HPLC (*High Performance Liquid Chromatography*), o
20 geles de poliacrilamida.

Tal como se utiliza aquí, el término “porcentaje de identidad” se refiere al porcentaje de aminoácidos que en una comparación de dos o más proteínas ocupan la misma posición relativa a lo largo del alineamiento. El término porcentaje de similaridad usado aquí es una medida estadística del grado de parentesco de dos o más secuencias de proteínas, basadas en las propiedades químicas de cada grupo de aminoácidos comparados a lo largo de las secuencias. Ambos parámetros, el porcentaje de identidad y el de similaridad, se calculan mediante programas informáticos que asignan un valor numérico a cada grupo de aminoácidos comparados a lo largo de la secuencia de proteínas alineadas. La similaridad deja margen para la variación conservativa, esto es la sustitución de un residuo hidrofóbico tal como isoleucina, valina, leucina o metionina aminoácidos por otro, o la sustitución de un residuo polar por otro, tales como arginina por lisina, ácido glutámico por aspártico, or glutamina por asparagina. La sustitución intencionada de un aminoácido puede hacerse en base a similitud en polaridad, carga, solubilidad, hidrofobicidad y/o la naturaleza anfipática de los residuos siempre que se conserve la actividad biológica del polipéptido. En una realización preferente, los porcentajes de homología indicados en el presente documento se refieren a porcentajes de identidad de secuencia. Por ejemplo la identidad de secuencia de aminoácidos en tanto por ciento (%) con respecto a una secuencia de referencia particular, puede ser el porcentaje de residuos de aminoácido en una secuencia candidata que son idénticos a los residuos de aminoácido en la secuencia de referencia. Si fuese necesario, pueden introducirse huecos en las secuencias alineadas para conseguir el máximo porcentaje de identidad de secuencia, sin que por ello sean considerados sustituciones conservativas de la identidad de secuencias.

El término “molécula de ácido nucleico modificada” tal como se usa aquí, se refiere a una molécula de DNA que tienen diferencias en su secuencia de deoxiribonucleotidos (deleciones, inserciones, transiciones y transversiones) con respecto a la molécula natural o salvaje (*wild type*). En consecuencia, tales cambios o “mutaciones” son transmitidos a las proteínas codificadas en la forma de aminoácidos substituidos, deleccionados o añadidos (inserciones). La mutación puede no tener efecto sobre la actividad enzimática de la proteína mutada, o por el contrario puede incrementarla, reducirla o aniquilarla por completo. En otros casos, el aminoácido remplazado corresponde a una substitución conservativa o no conservativa. Por ejemplo, el cambio de un aminoácido hidrofóbico por uno hidrofílico (cargado o polar) en el sitio activo de una proteína o en capas circundantes podría afectar la actividad de la proteína. Tales modificaciones pueden introducir en una molécula de ácidos nucleicos mediante técnicas conocidas con el nombre de mutagénesis dirigida (*site-directed mutagenesis*). Para tal fin, existen diversos kits comerciales basados en técnicas de PCR,

tales como los sistemas de mutagénesis dirigida simple o múltiple de Stratagen. Estos sistemas permiten generar mutaciones simples o múltiples de forma simultánea en una secuencia de nucleótidos dada, la cual se ha clonado previamente en un plásmido o vector de expresión, sin la necesidad de llevar a cabo posteriores subclonajes o la generación de fagémidos de cadena sencilla. Ambos métodos requieren el uso de oligonucleótidos específicos que contienen la mutación deseada y varios nucleótidos a cada lado que aparen perfectamente con la secuencia diana. De esta forma el oligonucleótido que lleva la mutación, y el cual hace las funciones de cebador, puede hibridar con la secuencia de DNA diana que se utiliza como molde para la síntesis mediante una reacción de PCR de secuencias de DNA que llevaran la mutación deseada en la posición seleccionada.

El término “vector de expresión” tal como se utiliza aquí, se refiere a un plásmido (molécula circular de DNA de doble cadena capaz de replicarse de forma autónoma e independiente del DNA genómico de la cepa bacteriana huésped), el cual contiene elementos reguladores (promotores, operadores y/o potenciadores) que controlan la expresión de una secuencia de DNA foránea que codifica el polipéptido de interés. Ya que los vectores de expresión son por lo general plásmidos relativamente pequeños que pueden estar hasta cientos de copias por célula bacteriana (plásmidos multicopia), la inducción de la expresión de la secuencia de DNA foráneo garantiza la rápida acumulación de altos niveles de transcritos (moléculas de RNA mensajero) y por ende de la proteína codificada.

Como se usa aquí, el término “célula transformada”, “transformada” y “transformación” se refieren a la introducción por diferentes métodos (químico o electroporación) bien conocidos en el estado de la técnica, de plásmidos o plásmidos recombinantes en una célula aquí llamada “célula huésped”. Esta puede ser una célula procariota tal como las diferentes cepas de la bacteria *Escherichia coli* (TOP10, Novablue, BL21[DE3]pLysS, etc).

Descripción detallada de la invención

La presente invención describe una secuencia nucleotídica natural correspondiente a la secuencia codificante de un gen aquí aislado del flavedo (piel) de frutos de naranjo (*Citrus sinensis*) de la variedad Navel. Dicha secuencia codifica una TPS multiproducto que cataliza la síntesis varios sesquiterpenos cíclicos. Moléculas de cDNA representativas del mencionado gen fueron secuenciadas y se dedujo la secuencia de aminoácidos del polipéptido codificado. En este sentido, la presente invención se relaciona en primera instancia con:

- (a) Una secuencia polinucleotídica correspondiente a un gen de naranjo, referido en lo sucesivo como *Cstps3*, que comprende la secuencia mostrada en la **Fig. 1A** (SEQ ID NO:1), la cual codifica un polipéptido, en adelante CsTPS3, con una actividad mayoritaria δ -cadineno sintasa, que tiene la secuencia deducida de aminoácidos mostrada en la **Fig. 1B** (SEQ ID NO:2).

En la presente invención, también se describen moléculas de ácidos nucleicos mutadas *in vitro* (variantes), que codifican polipéptidos modificados que contienen al menos un cambio de aminoácido con respecto a la secuencia polipeptídica natural de la proteína CsTPS3 (SEQ ID NO:2) (**Tabla 2**). En este sentido, la presente invención provee moléculas de ácidos nucleicos modificadas que contienen:

- (b) Cambios en la secuencia de nucleótidos de la molécula natural de DNA identificada como SEQ ID NO:1 (*Cstps3*), que se traducen en cambios en la secuencia de aminoácidos del polipéptido codificado. Ejemplo de tales moléculas de DNA modificadas son aquellas identificadas como SEQ ID NO:3 (**Fig. 1B**), SEQ ID NO:5 (**Fig. 1C**), SEQ ID NO:7 (**Fig. 1D**), y SEQ ID NO:9 (**Fig. 1E**).

La invención además provee sesquiterpeno sintasas modificadas que contienen cambios de aminoácidos con respecto a la secuencia polipeptídica natural del polipéptido CsTPS3. Entre los polipéptidos modificados se proveen aquellos que contienen:

- (c) Cambios (substituciones) simples o múltiples de aminoácidos con respecto a la secuencia polipeptídica natural del polipéptido CsTPS3 identificado con la referencia SEQ ID NO:2.

En un sentido más específico, las sesquiterpeno sintasas modificadas codificadas por las moléculas de ácidos nucleicos mutadas *in vitro* provista aquí son aquellos polipéptidos modificados que:

- (d) Contienen el reemplazamiento C⁴⁵⁵G (en lo sucesivo versión VB2, identificada como SEQ ID NO:4), o una combinación de las substituciones S⁴¹⁸H y C⁴⁵⁵G (S⁴¹⁸H/C⁴⁵⁵G) (versión VB5, identificada como SEQ ID NO:6), o una combinación de los reemplazamientos de aminoácidos S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y (en lo sucesivo versión VC1, identificada como SEQ ID NO:8), o una combinación de los reemplazamientos de aminoácidos S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y/S⁴²²A (en lo sucesivo VC2, identificada como SEQ ID NO:10), numerados de acuerdo a las posiciones relativas de la SEQ ID NO:2. Por ejemplo para la versión modificada CsTPS3-VB2 la molécula de ácido nucleico

provista aquí codifica una CsTPS3 modificada que contiene el reemplazamiento de aminoácido C⁴⁵⁵ que es C⁴⁵⁵G (Fig. 3).

TABLA 2

Lista de reemplazamientos simples y múltiples introducidos en el polipéptido natural CsTPS3 (SEQ ID NO:2), como consecuencia de substituciones de pares de base (transiciones y/o transversiones) introducidas en la secuencia de nucleótidos de la región codificante integra del gen *Cstps3* (SEQ ID NO:1)

Versión mutante	Mutaciones introducidas en CsTPS3	SEQ ID NO mutada	
		cDNA	Polipéptido
B2	C ⁴⁵⁵ G	3	4
B5	S ⁴¹⁶ H/C ⁴⁵⁵ G	5	6
VC1	S ⁴¹⁶ T/G ⁴¹⁷ T/S ⁴¹⁸ Y	7	8
VC2	S ⁴¹⁶ T/G ⁴¹⁷ T/S ⁴¹⁸ Y/S ⁴²² A	9	10

Las sesquiterpenos sintasas modificadas provistas aquí tienen una actividad catalítica alterada. El efecto de los cambios de aminoácidos introducidos en la secuencia polipeptídica de las proteínas modificadas fue tal que alteró la especificidad y/o la distribución/racionalización de los productos (es decir, alteración de las cantidades y/o tipos de sesquiterpenos producidos), comparado a la actividad de la sesquiterpeno sintasa natural CsTPS3 identificada como SEQ ID NO:2.

Por tanto, un primer aspecto de la invención se refiere a un polipéptido, en la presente memoria referido como polipéptido de la invención, que presenta actividad sesquiterpeno sintasa y comprende una secuencia con al menos un 90% de identidad, preferentemente con al menos un 95%, y más preferentemente con al menos un 98%, y aun más preferentemente con al menos un 99% de identidad, a SEQ ID NO:14. Dicha actividad sesquiterpeno sintasa del polipéptido SEQ ID No:14 es, preferentemente, al menos una seleccionada del grupo compuesto por: actividad δ -cadineno sintasa, actividad elemol sintasa y actividad germacreno D sintasa.

SEQ ID NO:14 consiste en una secuencia polipeptídica en donde:

- el aminoácido en posición 416 se selecciona independientemente entre Thr o Ser;
- el aminoácido en posición 417 se selecciona independientemente entre Thr o Gly;

- el aminoácido en posición 418 se selecciona independientemente entre Ser, Hys o Tyr, y preferentemente es Ser o Tyr;
 - el aminoácido en posición 422 se selecciona independientemente entre Ser o Ala, y preferentemente es Ser; y
- 5 - el aminoácido en posición 455 se selecciona independientemente entre Cys o Gly;
- y comprende la secuencia de la proteína CsTPS3 aislada de naranjo o de uno de sus mutantes con actividad sesquiterpeno sintasa.

En particular, dicha proteína CsTPS3 aislada de naranjo tiene actividad mayoritaria δ -cadineno sintasa, y consiste en la secuencia SEQ ID NO:2, que corresponde a la

10 δ -cadineno sintasa, y consiste en la secuencia SEQ ID NO:2, que corresponde a la secuencia SEQ ID NO:14 en donde: el aminoácido en posición 416 es Ser, el aminoácido en posición 417 es Gly, el aminoácido en posición 418 es Ser, el aminoácido en posición 422 es Ser, y el aminoácido en posición 455 es Cys. No obstante, siendo considerada una δ -cadineno sintasa y que es capaz de producir preferentemente entre un 25% y un 50% de δ -cadineno en determinadas condiciones, respecto a la cantidad total de sesquiterpenos

15 producidos por dicha sintasa, el polipéptido CsTPS3 (SEQ ID NO:2) corresponde a un enzima multiproducto que puede utilizarse también para la preparación de otros sesquiterpenos cíclicos tales como el α -copaeno, β -cubebeno, germacreno D, y elemol, además del δ -cadineno (**Fig. 2**).

20 En una realización preferida, el polipéptido comprende una secuencia que tiene al menos un 90% de identidad con SEQ ID NO:2, o con la SEQ ID NO:2 modificada con al menos una sustitución en un aminoácido que altera la actividad sesquiterpeno sintasa y/o especificidad de producto. Dicha sustitución se selecciona preferentemente entre al menos

25 una del grupo compuesto por:

- a. una sustitución de la cisteína en posición 455 por glicina (C⁴⁵⁵G), y
- b. una triple sustitución de la serina en posición 416 por treonina, de la glicina en posición 417 por treonina, y de la serina en posición 418 por tirosina (S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y).

30 No obstante, aparte de la sustitución anterior, la SEQ ID NO:2 modificada opcionalmente puede comprender además otra sustitución en un aminoácido de su secuencia, también para alterar la especificidad y/o la distribución/racionalización de los sesquiterpenos que produce, y que se selecciona entre:

- 35 a. una sustitución de la serina en posición 418 por histidina (S⁴¹⁸H), cuando la SEQ ID NO:2 comprende la sustitución C⁴⁵⁵G, o

- b. una sustitución de la serina en posición 422 por alanina (S⁴²²A), preferentemente cuando la SEQ ID NO:2 comprende la triple sustitución S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y.

Se ha comprobado que cuando el polipéptido de la invención comprende la secuencia SEQ ID NO:2 modificada con la sustitución C⁴⁵⁵G, presenta mayoritariamente actividad elemol sintasa, que puede incrementarse aún más cuando la sustitución adicional S⁴¹⁸H. Así, en un ejemplo preferido y en determinadas condiciones, es capaz de producir elemol con una abundancia relativa, preferentemente en un intervalo comprendido entre el 65% y el 90%, y más preferentemente entre el 65% y el 85% ó entre el 75% y el 90%, respecto a la cantidad total de los distintos sesquiterpenos producidos con dicho polipéptido.

Además, también se comprobó que cuando el polipéptido de la invención comprende la secuencia SEQ ID NO:2 modificada con la triple sustitución S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y, éste presenta mayoritariamente actividad germacreno D sintasa. Y de hecho, en un ejemplo preferido y en determinadas condiciones, es capaz de producir germacreno D con una abundancia relativa, preferentemente en un intervalo comprendido entre el 75% y el 80%, y más preferentemente entre el 75% y el 80% ó entre el 80% y el 85%, respecto a la cantidad total de los distintos sesquiterpenos producidos con dicho polipéptido.

En otra realización preferida de este aspecto de la invención o de cualquiera de sus realizaciones anteriores, y dependiendo de la actividad catalítica deseada, el polipéptido de la invención comprende una secuencia que se selecciona del grupo compuesto por: SEQ ID NO:2, SEQ ID NO:16 y SEQ ID NO:18.

El polipéptido de secuencia SEQ ID NO:16 tiene actividad elemol sintasa, es decir, es una elemol sintasa, que consiste en la secuencia polipeptídica natural SEQ ID NO:2 del polipéptido CsTPS3 modificada al menos con la sustitución de la Cys en posición 455 por Gly (C⁴⁵⁵G), y en el que el aminoácido 418 de su secuencia es Ser o His, y más preferentemente es His.

El polipéptido de secuencia SEQ ID NO:18 tiene actividad germacreno D sintasa, es decir, es una germacreno D sintasa, que consiste en la secuencia polipeptídica natural SEQ ID NO:2 del polipéptido CsTPS3 modificada al menos con la triple sustitución S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y, y en el que el aminoácido 422 de su secuencia es Ser o Ala, y más preferentemente es Ser.

En otra realización preferida de este aspecto de la invención o de cualquiera de sus realizaciones anteriores, el polipéptido es una sesquiterpeno sintasa seleccionada del grupo compuesto por: la δ -cadineno sintasa CsTPS3 que consiste en SEQ ID NO:2, la elemol sintasa VB2 que consiste en SEQ ID NO:4, la elemol sintasa VB5 que consiste en SEQ ID NO:6, la germacreno D sintasa VC1 que consiste en SEQ ID NO:8 y la germacreno D sintasa VC2 que consiste en SEQ ID NO:10.

Otro aspecto de la invención se refiere a un ácido nucleico o polinucleótido que comprende una secuencia que codifica un polipéptido de la invención, según uno cualquiera de los definidos anteriormente. A lo largo de la presente memoria, dicho ácido nucleico o polinucleótido, así como en cualquiera de sus realizaciones preferidas que se indican a continuación, puede también encontrarse indistintamente referido como "ácido nucleico de la invención" o "polinucleótido de la invención". Preferentemente, el ácido nucleico de la invención codifica un polipéptido de secuencia SEQ ID NO: 13. Y más preferentemente, codifica un polipéptido de secuencia SEQ ID NO:2 ó SEQ ID NO:2 modificada con al menos una sustitución en un aminoácido que altera la actividad sesquiterpeno sintasa y/o especificidad de producto, como se definieron anteriormente.

En una realización preferida, cuando el ácido nucleico codifica un polipéptido que comprende la secuencia SEQ ID NO:2 correspondiente a la sesquiterpeno sintasa natural CsTPS3, consiste en un polinucleótido que comprende SEQ ID NO:1.

En otra realización preferida, cuando el ácido nucleico codifica un polipéptido que comprende la secuencia SEQ ID NO:2 modificada, consiste en un polinucleótido que comprende SEQ ID NO:1 con al menos una mutación en su secuencia seleccionada del grupo compuesto por:

- a. una transversión T→G en la posición +1363 con respecto al codón de inicio (ATG) de SEQ ID NO:1; y
- b. un grupo de sustituciones de los nucleótidos T→A/G→A/AGC→TAT en las posiciones +1246/1249/1252-1254 con respecto al codón de inicio (ATG) de SEQ ID NO:1.

En una realización preferida de la anterior, el ácido nucleico comprende otra mutación adicional en el polinucleótido de secuencia SEQ ID NO:1 que se selecciona entre:

- a. una transversión/transición AG→CA en las posiciones +1252/1253 con respecto al codón de inicio de SEQ ID NO:1, cuando SEQ ID NO:1 comprende la transversión T→G; y
- b. una transversión T→A en la posición +1264 con respecto al codón de inicio de SEQ ID NO:1, cuando SEQ ID NO:1 comprende el grupo de sustituciones T→A/G→A/AGC→TAT.

Y en una realización más preferida de la anterior, el ácido nucleico comprende la secuencia SEQ ID NO:15, que comprende dicha transversión T→G en posición +1363 y dicha transversión/transición AG→CA en posiciones +1252/1253; o la secuencia SEQ ID NO:17, que comprende dicho grupo de sustituciones T→A/G→A/AGC→TAT en posiciones +1246/1249/1252-1254 y dicha transversión T→A en posición +1264.

En otra realización preferida de este aspecto de la invención, el ácido nucleico comprende al menos una secuencia polinucleotídica seleccionada del grupo formado por: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 y SEQ ID NO:8. Y preferentemente, consiste en SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 o SEQ ID NO:8.

Otro aspecto de la invención se refiere a una construcción génica que comprende al menos un ácido nucleico de la invención, como cualquiera de los definidos anteriormente.

En la presente invención también se proveen los vectores que contienen las moléculas de ácidos nucleicos descritas aquí, constituyendo así otro aspecto de la presente invención. Los vectores incluyen plásmidos (moléculas circulares de DNA de doble cadena autoreplicables) para la expresión heteróloga de proteína eucariotas en células bacterianas. También se proveen en otro aspecto de la invención, las células o cepas bacterianas que contienen dichos vectores. Las células bacterianas provistas aquí pueden cultivarse para inducir la expresión de la secuencia natural *Cstps3*, o en su defecto de alguno de sus derivados modificados (mutados), y por consiguiente la síntesis del polipéptido natural CTPS3, o en su defecto del polipéptido modificado correspondiente (CsTPS3-VB2, CsTPS3-VB5, CsTPS3-VC1 o CsTPS3-VC2).

En otro aspecto, la invención hace referencia a un método para obtener un polipéptido de la invención con actividad sesquiterpeno sintasa, el cual comprende cultivar una célula bacteriana transformada con un ácido nucleico de la invención, o con una construcción génica o un vector recombinante que lo comprenda, según se definieron anteriormente, en

condiciones adecuadas para favorecer o inducir la expresión de dicho ácido nucleico. De manera general, el método de inducción utilizado depende en cierta medida del tipo de vector de expresión utilizado, que como regla general en el caso de vectores comerciales es indicado por el fabricante. Asimismo, con la finalidad de conseguir las mejores condiciones de cultivo y de inducción posibles para cada tipo de proteína que se quiera expresar, se pueden modificar otras variables tales como temperatura, tiempo de inducción, medio de cultivo utilizado, etc. En lo que se refiere a la expresión de proteínas del tipo sesquiterpeno sintasa en bacterias, las condiciones de crecimiento e inducción son bastante conocidas y están disponibles para el experto medio en el estado de la técnica, y pueden variarse a juicio y deseo del especialista o técnico en cuestión.

Otro aspecto de la invención hace referencia al uso del polipéptido de la invención para la síntesis, preferentemente *in vitro*, de sesquiterpenos. Según la presente invención, el polipéptido de la invención permite sintetizar un sesquiterpeno que se selecciona entre al menos uno del grupo compuesto por: α -cubebeno, α -copaeno, β -cubebeno, β -elemeno, α -gurjuneno, *E*- β -cariofileno, α -humuleno, germacreno D, biciclogermacreno, α -muuroleno, α -bulneseno, cubebol, δ -cadineno, elemol, germacreno D-4-ol, bulnesol, guaiol y una combinación cualquiera de dos o más de los anteriores. Más preferentemente, el sesquiterpeno que se sintetiza es al menos uno seleccionado del grupo que consiste en: δ -cadineno, elemol y germacreno D, o una combinación cualquiera de dos o más de los anteriores.

En particular, y como se mencionó anteriormente, la δ -cadineno sintasa CsTPS3 (SEQ ID NO:2) por una parte puede utilizarse en la síntesis δ -cadineno y también de otros sesquiterpenos cíclicos tales como α -copaeno, β -cubebeno, germacreno D, y/o elemol. Por otra parte, las TPS modificadas producto de la invención pueden usarse para mejorar la producción y/o la biosíntesis de elemol, germacreno D, α -gurjuneno, biciclogermacreno, guaiol, bulnesol y β -elemeno, incrementando la distribución de estos sesquiterpenos en comparación a la cantidad relativa de δ -cadineno, β -cubebeno, cubebol y α -copaeno, productos principales de la actividad catalítica de la proteína natural de referencia CsTPS3. Este hecho puede resultar en métodos para incrementar la producción y/o la pureza del elemol, germacreno D, α -gurjuneno y biciclogermacreno, guaiol, bulnesol y β -elemeno, incrementando la recuperación de estos sesquiterpenos cíclicos del medio de reacción. Ejemplo de ello, son aquellos polipéptidos modificados que generan como producto principal elemol, además de una pequeña parte de guaiol y bulnesol, y solo trazas de α -copaeno,

germacreno D, δ -cadineno, en comparación con la abundancia relativa de estos sesquiterpenos producidos por polipéptido natural CsTPS3 (SEQ ID NO:2) bajo las mismas condiciones. Ejemplo específico de tales polipéptidos modificados es el identificado con la referencia SEQ ID NO:4 (CsTPS3-VB2), el cual contiene el reemplazamiento C⁴⁵⁵G. La introducción del cambio S⁴¹⁸H en el polipéptido CsTPS3-VB2, tiene un efecto sinérgico sobre la producción de elemol. Ejemplo de tal polipéptido modificado es el identificado con la referencia SEQ ID NO:6 (CsTPS3-VB5). La invención también provee un polipéptido modificado que produce una mayor abundancia relativa de germacreno D, y una disminución del porcentaje de otros sesquiterpenos minoritarios, en comparación con la abundancia relativa de este sesquiterpeno producido en las mismas condiciones de reacción por el polipéptido natural CsTPS3 (SEQ ID NO:2). Por ejemplo, sustituciones simultáneas de aminoácidos en las posiciones 416, 417 y 418 causan un incremento del porcentaje de germacreno D, en comparación con la abundancia relativa de este sesquiterpeno producido en las mismas condiciones por el polipéptido natural CsTPS3. Notoriamente, los reemplazamientos S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y incrementan la producción de β -elemeno y α -gurjuneno en detrimento de la síntesis de α -copaeno, β -cubebeno y δ -cadineno. Ejemplo de tal polipéptido modificado es aquel identificado con la referencia SEQ ID NO:8 (CsTPS3-VC1).

Los polipéptidos modificados de la invención pueden utilizarse para la preparación de elemol (**Fig. 2**), un sesquiterpeno cíclico presente en algunos fitoaceites esenciales. En diversos artículos científicos publicados recientemente, se ha descrito que el elemol es un potente agente repelente de diferentes especies de mosquitos transmisores de enfermedades contagiosas, e incluso un efectivo repelente de cucarachas. Este hecho ha convertido al elemol en un prometedor y potente repelente natural que puede utilizarse en formulas farmacológicas diseñadas para repeler insectos transmisores de graves enfermedades infectocontagiosas. Así, en una realización preferida de uso, el polipéptido de la invención es utilizado para la preparación o síntesis, preferentemente *in vitro*, de elemol, preferentemente cuando dicho polipéptido comprende la secuencia SEQ ID NO:16, más preferentemente cuando comprende la secuencia SEQ ID NO:4 ó SEQ ID NO:6, y aún más preferentemente cuando su secuencia es SEQ ID NO:4 ó SEQ ID NO:6.

Considerando el uso anterior, referente a la síntesis de sesquiterpenos con el polipéptido de la invención, otro aspecto de la invención se refiere a un procedimiento para sintetizar al menos uno de dichos sesquiterpenos, en adelante procedimiento de la invención, que comprende:

- a. incubar el substrato farnesil-pirofosfato con al menos un polipéptido de la invención, pudiendo utilizarse un único polipéptido o una combinación de polipéptidos según su capacidad o actividad para sintetizar distintos sesquiterpenos, y en función del sesquiterpeno o combinación de sesquiterpenos que se desea preparar, y
- 5 b. dejar reaccionar hasta obtener un producto volátil que comprende al menos un sesquiterpeno.

De este modo, también se proveen los métodos para la síntesis de sesquiterpenos cíclicos en los cuales el precursor acíclico farnesil-pirofosfato (FPP) se añade a la mezcla de
10 reacción que contiene la proteína natural CsTPS3, o en su defecto alguno de los polipéptidos modificados (CsTPS3-VB2, CsTPS3-VB5, CsTPS3-VC1 o CsTPS3-VC2), bajo condiciones de reacción que favorecen la formación de sesquiterpenos cíclicos.

El experto medio puede identificar y tener a su disposición fácilmente las condiciones del
15 procedimiento de la invención que favorecen la formación de sesquiterpenos cíclicos, al ser ampliamente conocidas y habituales para cualquier experto del campo de las enzimas (p. ej. tampones, cofactores, temperatura, disolvente, tiempo de inducción, etc) y en particular aquellas referidas a enzimas con actividad sesquiterpeno sintasa, y que a modo de ejemplo comprenderían entre otras las especificadas de forma explícita el ejemplo 4 de la presente
20 invención. Dichas condiciones serán adecuadas, por ejemplo, para producir la unión del substrato al centro activo responsable de la actividad sesquiterpeno sintasa, para conservar la función enzimática, etc.

En un apartado de los métodos descritos aquí, la etapa de contacto del FPP con las
25 sesquiterpeno sintasas se realiza *in vitro*, adicionando el extracto proteico crudo que contiene la TPS natural CsTPS3 o alguno de los polipéptidos modificados a la mezcla de reacción que contiene el substrato FPP y magnesio como cofactor. Los sesquiterpenos producidos por los métodos de la invención incluyen, pero no se limitan al δ -cadineno, α -copaeno, β -copaeno, β -cubebeno, germacreno D y elemol (**Fig. 2**). En algunos casos, al
30 menos uno de los sesquiterpenos cíclicos mayoritarios es el δ -cadineno, el germacreno D o el elemol.

Se entiende que la tanto la anterior descripción general y la siguiente descripción detallada constituyen ejemplos y explicación de la invención, y que por tanto no son restrictivas de las
35 reivindicaciones de la invención. A continuación se realizará una descripción detallada y explicativa de los materiales y métodos, así como de los resultados frutos de la invención.

EJEMPLO 1. Análisis de sesquiterpenos extraídos del flavedo de naranjas

En los frutos cítricos los aceites esenciales ricos en terpenos se acumulan en el flavedo (piel), y de forma específica en el interior de las glándulas de aceites (estructuras secretoras multicelulares). El flavedo es la parte externa y coloreada de la piel de los frutos cítricos (exocarpo), que consta de una cutícula protectora y una epidermis formada por células parenquimatosas que rodean las glándulas de aceites. Los aceites esenciales de cítricos son una mezcla de diferentes compuestos volátiles que se caracterizan por tener un alto contenido de terpenos hidrocarbonados y cantidades relativamente bajas de terpenos oxigenados (Högnadóttir y Rouseff, [2003] *J. Chromatogr. A.* **998**, 201-211). Las propiedades organolépticas (sabor y aroma) de los aceites esenciales de cítricos son el resultado de la combinación de todos ellos.

Materiales y Métodos: La composición de sesquiterpenos del flavedo de naranjas de la variedad Navel se estudió a diferentes estadios del desarrollo del fruto (verde-inmaduro, verde-maduro, y frutos maduros coloreados), mediante análisis de cromatografía de gases acoplada a la espectrofotometría de masas (GC-MS) (Rodríguez y col. [2011]. *Plant Physiol.* **156**, 793-802). Para ello se mondó el flavedo con el uso de un cuchillo afilado y se congeló inmediatamente en nitrógeno (N₂) líquido. El flavedo congelado se trituro en N₂ líquido con la ayuda de una mano de mortero hasta formar un polvo muy fino. Doscientos miligramos (200 mg) de flavedo triturado se mezclaron con 3ml de pentano frío y 25 µg de 2-octanol (Sigma-Aldrich) en un tubo Pyrex con tapa de rosca. El tejido se homogenizó en hielo durante 30 s con un Politrón (PT 10-35 GT), se agitó durante 15 s con la ayuda de un vortex, se le adicionaron 3 ml de agua Milli-Q y se agitó nuevamente en el vortex durante 30 s. La suspensión se clarificó mediante centrifugación a 1800 g durante 10 min a 4°C. La fase orgánica se extrajo con una pipeta Pasteur y se colocó en un tubo limpio. La fase acuosa se extrajo dos veces con 3 ml de pentano, recuperando cada vez la fase orgánica en el mismo tubo. Una alícuota de 2 µl de la fase orgánica se inyectó directamente en el GC-MS para el análisis de volátiles. Para cada muestra se realizaron al menos dos extracciones y dos inyecciones independientes. Para el análisis de los volátiles se utilizó un cromatógrafo de gases Thermo Trace GC Ultra acoplado a un espectrofotómetro de masas Thermo DSQ con un método de ionización de electrones (EI) ajustado a 70eV. La fuente de iones y la línea de transferencia se ajustaron a 200 y 260°C, respectivamente. Los compuestos volátiles se separaron en una columna HP-INNOWax (Agilent J&C Columns) de 30 m de longitud x 0.25 mm de diámetro interno x 0.25 µm de espesor de fase. El calentamiento de la columna se programó a 40°C durante 5 min, incrementándose a 150°C a razón de 5°C/min, y luego a 250°C a razón de 20°C/min, manteniéndose a 250°C durante 2 min. La temperatura del

inyector fue de 220°C. Como gas portador (*carrier*) se utilizó helio a un flujo de 1.5 ml/min y la muestra se inyectó sin fragmentación. El espectro de masas producido por la impacto de electrones se registró a una velocidad de 0.5/scan. Los compuestos volátiles se identificaron por aproximación del espectro de masas obtenido con el correspondiente al de los compuestos presentes en las librerías de referencia (Wiley6 y NIST), o por su identidad con el compuesto estándar. La abundancia relativa de los sesquiterpenos volátiles identificados se calculó mediante la integración del área de los picos del cromatograma total de iones (TIC) y normalizando los valores con la tasa de recuperación del estándar utilizado como control interno (2-octanol).

10

Resultados: Los sesquiterpenos volátiles identificados en los extractos del flavedo de naranjas Navel a diferentes estadios del desarrollo se muestran en la **Tabla 3**. Los resultados muestran diferencias significativas en el tipo de compuestos y sus abundancias relativas entre frutos maduros e inmaduros. Un total de diez sesquiterpenos cíclicos y cinco acíclicos dan cuenta de menos del 0.7% del contenido total de terpenos volátiles identificados en el flavedo. Estos resultados están en concordancia con resultados previos de otros autores, que demuestran que el aceite del flavedo de naranjas contiene niveles relativamente bajos de sesquiterpenos hidrocarbonados (Shaw PE, [1979] *J Agr Food Chem.* **27**, 246-257; Sawamura M, [2000] En *Recent Research Development in Agricultural and Food Chemistry*, Pandalai SG [Ed], Research Signpost, Trivandrum – India, Vol. 4, pp 131-164; Mitiku y col., [2000] *Flav Frag J.* **15**, 240-244). Además, nuestros resultados demuestran que la abundancia relativa de cada sesquiterpeno identificado, así como la abundancia relativa de la suma de todos ellos, cambia a través del desarrollo del fruto. Con la única excepción del valenceno y el α -farneseno, todos los demás sesquiterpenos que aparecen listados en la **Tabla 3** fueron identificados tanto en el flavedo de frutos verdes (inmaduros y maduros), como en frutos maduros que habían sufrido el cambio de color. El α -farneseno se identificó solamente en frutos verdes inmaduros, mientras que el valenceno – un indicador de la madurez del fruto (Elston y col., [2004] *Flav Frag J.* **20**, 381-386) – solo se identificó en frutos maduros que habían sufrido el cambio de color. En este sentido, la acumulación del valenceno en frutos maduros de naranja que comienzan a sufrir el cambio de color, se ha correlacionado con la inducción de la expresión del gen *Cstps1*, el cual codifica una ciclasa con actividad valenceno sintasa (Sharon-Asa y col., [2003] *Plant J.* **36**, 664-674). Como era de esperar, nuestros resultados también demuestran que los monoterpenos forman más del 98% del contenido total de volátiles del flavedo de las naranjas Navel. Aunque el limoneno es el monoterpeno hidrocarbonado más abundante (>91%) (dato no mostrado), se ha sugerido que su aparente actividad aromática es debido a la co-elución y contaminación con otros compuestos tales como alcoholes

35

oxygenados. Es sabido que el aroma más intenso del aceite esencial de las naranjas se debe a la presencia de terpenos cuyas concentraciones son relativamente muy bajas. El característico olor dulce y fresco de la piel de las naranjas se debe principalmente al α - y β -sinensal (0.02%), cuyo umbral de detección es tan solo de 0.05 ppb (Högnadóttir y Rouseff, [2003] *J. Chromatogr. A.* **998**, 201-211).

TABLA 3

Composición de sesquiterpenos del flavedo de naranjas Navel. El análisis de sesquiterpenos se realizó en extractos de flavedo mediante GC-MS. Los sesquiterpenos se identificaron por coincidencia del tiempo de retención (RT) y el espectro de masas (MS) de los estándares disponibles, o por la coincidencia el espectro de masas de los compuestos presentes en las librerías de referencia. Los datos corresponden a las cantidades relativas de cada sesquiterpeno identificado, representados en la forma de porcentaje (%) de área respecto al área total de los picos del cromatograma. Los resultados corresponden al valor de la media de tres réplicas biológicas y sus correspondientes desviaciones estándar.

	Verde-inmaduro	Verde-maduro	Maduros
	%	%	%
Sesquiterpenos cíclicos			
α -Copaeno	0.0213 \pm 0,0013	0.0304 \pm 0.0036	0.0261 \pm 0.0010
β -Cubebeno	0.0169 \pm 0.0056	0.0226 \pm 0.0001	0.0180 \pm 0.0007
β -Elemeno	0.0424 \pm 0.0018	0.0422 \pm 0.0022	0.0485 \pm 0.0055
β -Copaeno	0.0029 \pm 0.0004	0.0245 \pm 0.0001	0.0294 \pm 0.0004
α -Humuleno	0.0122 \pm 0.0010	0.0111 \pm 0.0011	0.0054 \pm 0.0005
Germacreno D	0.0135 \pm 0.0022	0.0279 \pm 0.0018	0.0153 \pm 0.0003
α -Muuroleno	n.d.	0.0093 \pm 0.0024	0.0082 \pm 0.0010
Valenceno	n.d.	n.d.	0.0915 \pm 0.0049
δ -Cadineno	0.0340 \pm 0.0022	0.0542 \pm 0.0001	0.0425 \pm 0.0029
β -Sesquifelandreno	0.0093 \pm 0.0021	0.0115 \pm 0.0004	0.0055 \pm 0.0001
Elemol	0.0078 \pm 0.0020	0.0161 \pm 0.0004	0.0132 \pm 0.0012
Total relativo	0.1603	0.2497	0.3035
Sesquiterpenos acíclicos			
Z- β -Farneseno	0.0310 \pm 0.0004	0.0170 \pm 0.0003	0.0102 \pm 0.0034
α -Farneseno	0.0540 \pm 0.0081	n.d.	n.d.
Nerolidol	0.0022 \pm 0.0003	0.0065 \pm 0.0007	0.0053 \pm 0.0015
β -Sinensal	0.0781 \pm 0.0029	0.0463 \pm 0.0020	0.0277 \pm 0.0012
α -Sinensal	0.1073 \pm 0.0024	0.0468 \pm 0.0004	0.0295 \pm 0.0013
Total relativo	0.2726	0.1166	0.0727

Aunque muchos han sido los sesquiterpenos que se han identificado en aceites esenciales de cítricos, solo se han aislado y clonado dos genes de cítricos que codifican enzimas con actividad sesquiterpeno sintasa. El gen *CjFS* (Nº de acceso GenBank: AF374462), el cual codifica un polipéptido con actividad farneseno sintasa, se aisló de hojas jóvenes de kuzu
5 (*Citrus junos*) (Maruyama y col., [2001] *Biol Pharm Bull.* **24**, 1171-1175); y el gen *Cstps1* (Nº de acceso GenBank: AF441124) que codifica una valenceno sintasa, se aisló del flavedo de naranjas (Sharon-Asa y col., [2003] *Plant J.* **36**, 664-674; WO 2005/021705, Ministerio of Agricultura de Israel, 10 de Marzo de 2005). La gran diversidad de los sesquiterpenos con actividad aromática que se han identificado en cítricos y la escasa información de la que se
10 dispone sobre los enzimas TPS encargados de su biosíntesis, fue el motivo por el cual nos animamos a aislar y clonar nuevos genes sesquiterpeno sintasa de naranja.

EJEMPLO 2. *Aislamiento de la región codificante de un nuevo gen sesquiterpeno sintasa de naranja*

15

Materiales y Métodos. *Aislamiento de RNA total del flavedo de naranjas y síntesis de cDNA.* El flavedo (exocarpo) de naranjas Navel (*Citrus sinensis* var Washington Navel) se separó cuidadosamente del albedo (mesocarpo) para evitar su contaminación, y se congeló inmediatamente en N₂ líquido. A continuación el flavedo se trituró también en N₂ líquido con
20 la ayuda de un molinillo hasta formar un polvo muy fino, y el RNA total se aisló utilizando el sistema RiboPure™ (Ambion). El DNA genómico contaminante se removió del RNA total mediante digestión enzimática con DNase utilizando el sistema Turbo DNA-free™ (Ambion), siguiendo las recomendaciones del fabricante. La concentración de la solución de RNA total se determinó con un espectrofotómetro ND-1000 (NanoDrop Technologies, Inc.). La calidad
25 del RNA total se evaluó mediante el método estandarizado de electroforesis en gel desnaturalizante de agarosa al 3%, visualizando la integridad de las bandas de los RNA ribosomales de diferente tamaño. El RNA total del flavedo de naranjas se utilizó para sintetizar RNA poliadenilado (RNA poli(A)⁺) utilizando el sistema MicroPoly(A)Purist™ (Ambion), siguiendo las instrucciones del fabricante. La integridad del RNA total poli(A)⁺ se
30 verificó mediante electroforesis en un gel desnaturalizante de agarosa al 1.6%. La síntesis de la 1ª cadena de cDNA (1st cDNA) del RNA poli(A)⁺ se realizó utilizando el sistema RetroScript® (Ambion) y oligo(dT) como cebador para la Transcriptasa Inversa MMLV, siguiendo las instrucciones del fabricante. La solución de 1st cDNA se almacenó a -20°C hasta su uso.

35

Aislamiento de la región codificante del gen Cstps3 del flavedo de naranja: La solución de 1st cDNA preparada como se describe arriba se utilizó como molde para la amplificación de la región codificante (cDNA) entera de un gen de naranja (*Cstps3*) que codifica una nueva sesquiterpeno sintasa. Oligonucleotidos diseñados sobre secuencias de ESTs truncados que codifican un polipéptido con homología a sesquiterpenos sintasas se utilizaron como cebadores en reacciones de PCR. Para la amplificación del cDNA del gen *Cstps3* se utilizaron como cebadores los oligonucleotidos listados en la **Tabla 4 y** el sistema Advantage® -HF 2 PCR (Clontech), siguiendo las indicaciones del fabricante. Para facilitar el clonaje del fragmento amplificado, a ambos oligonucleotidos se les adicionó un sitio de restricción único en su extremo 5', ambos protegidos por dos nucleótidos adicionales. Al cebador directo (TPS3-F) se le adicionó la secuencia correspondiente al sitio de restricción de la endonucleasa *BamH I* (GGATCC), justo antes del codón **ATG**. Al cebador inverso (TPS3-R) se le adicionó el sitio de restricción de la endonucleasa *Sal I* (GTCGAC), justo después del codón de parada (**TGA**). El fragmento de DNA amplificado se purificó de un gel de agarosa al 1% utilizando el sistema PureLink™ Quick Gel Extraction (Invitrogen), siguiendo las indicaciones del fabricante. El fragmento de DNA eluido se concentró mediante centrifugación con el sistema Microcon® Centrifugal Filters (Millipore) hasta un volumen de 10 µl, siguiendo las instrucciones del fabricante. El fragmento de DNA purificado se clonó en el plásmido pCR®4Blunt-TOPO utilizando el sistema Zero Blunt® TOPO PCR Cloning Kit for Sequencing (Invitrogen). 2µl de la reacción de ligación se utilizaron para transformar células electrocompetentes de la cepa One Shot® TOP10 (Invitrogen) de la bacteria *Escherichia coli*, siguiendo las instrucciones del fabricante. El DNA plasmídico se aisló de los transformantes usando el sistema Wizard® Plus SV Minipreps DNA Purification System (Promega), y el fragmento de cDNA del gen *Cstps3* clonado se secuenció utilizando los cebadores T3 (5'-ATTAACCCTCACTAAAGG-3') y T7 (5'-TAATACGACTCACTATAGGG-3'). Las secuencias obtenidas se analizaron utilizando el programa informático Sequencher v4.9 (Genes Codes Corporation, Ann Harbor, Michigan).

Resultados: La expresión de un EST (*Expressed Sequenced Tag*) del híbrido mandarina x naranja *Citrus reticulata* var. Clementina (Nº de acceso GenBank: CX292608), con homología a terpeno ciclasas de plantas de la Clase I (NCBI CDD: cd00868), se determinó que estaba inducida en el flavedo de frutos verdes-maduros de naranja Navel, con respecto a frutos maduros que habían sufrido el cambio de color. La secuencia nucleotídica de este fragmento de DNA se utilizó como cebo para buscar secuencias homólogas de naranja en la base de datos de ESTs del GenBank, mediante el algoritmo BLASTN (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). El análisis permitió la identificación de varios ESTs

truncados de naranjo Navel: CK935484, CK935772, CK937172, EY703037, EY708586, EY731065, EY713717, EY722034, y EY732650, los cuales solapaban el EST de clementino utilizado como cebo. El alineamiento de los ESTs identificados mediante análisis de CLUSTALW (Thompson y col., [1994] *Nucleic Acids Res.* **22**, 4673-4680) (**Fig. 4**), reveló un

5 alto grado de similitud entre de las regiones solapadas. Este hecho permitió el ensamblaje de las secuencias solapadas en una secuencia consenso de tamaño completo (desde el codon ATG al codon de parada de lectura). El ensamblaje de la secuencia consenso se realizó con el algoritmo "Assamby" del programa informático Sequencher v4.8 software (Gene Codes Corporation, Ann Harbor, Michigan). La secuencia consenso predijo un cDNA

10 con un marco de lectura abierto (ORF: *Open Reading Frame*) de 1695 nucleótidos de longitud, codificando un polipéptido de 564 aminoácidos. La comparación de la secuencia deducida de aminoácidos de este polipéptido con las secuencias de proteínas de la base de datos no redundantes del GenBank (<http://www.ncbi.nlm.nih.gov/Blast.cgi>), mediante el algoritmo BLASTP, reveló que el dicho polipéptido tenía homología con sesquiterpeno

15 sintasas vegetales. Ya que no existía evidencia previa de que la secuencia nucleotídica aquí ensamblada pudiese efectivamente corresponder al cDNA de un gen de naranjo que codificase una nueva sesquiterpeno sintasa, se diseñaron cebadores sobre la secuencia consenso con la finalidad de amplificar el ORF aquí propuesto. Utilizando como cebadores los oligonucleotidos listados en la **Tabla 4** y como templado el cDNA de cadena sencilla

20 preparado a partir del RNA total aislado del flavedo de naranjas Navel, se amplificó un fragmento de DNA del tamaño esperado (1695 pb). El tamaño del fragmento amplificado se confirmó mediante electroforesis en un gel de agarosa (1%). El producto de PCR se clonó en el vector pCR[®]4Blunt-TOPO (Invitrogen), y se determinó la secuencia nucleotídica del inserto de clones recombinantes. El análisis de secuencia confirmó que el fragmento de

25 cDNA clonado correspondía efectivamente al ORF completo del gen *Cstps3* aislado del flavedo de naranjas Navel, el cual se identificó con el número de referencia SEQ ID NO:1. El análisis de BLASTN de la secuencia deducida de aminoácidos del polipéptido codificado por la secuencia SEQ ID NO:1, aquí identificado como SEQ ID NO:2, reveló que la proteína codificada tenía homología a sesquiterpeno sintasas vegetales de la clase Tps-a (NCBI

30 *Conserved Domain Database*: cd00684). La secuencia de nucleótidos del ORF aquí aislado (SEQ ID NO:1), así como la secuencia de aminoácidos de la proteína codificada (SEQ ID NO:2) se muestra en las **Figs. 1A y 2A**. El fragmento de cDNA aquí aislado fue nombrado como *Cstps3* (SEQ ID NO:1), siguiendo la nomenclatura de Sharon-Asa y col., [2003] *Plant J.* **36**, 664-674. Estos autores aislaron el primer gen sesquiterpeno sintasa de naranjo, al

35 cual denominaron *Cstps1* (Solicitud de Patente Internacional WO 2005/021705, Ministerio de Agricultura de Israel, 10 de Marzo de 2005).

TABLA 4

Lista de los oligonucleotidos utilizados como cebadores para la amplificación de la región codificante del gen *Cstps3* del flavedo de frutos de naranja Navel.

SEQ ID NO	Oligonucleotido	Secuencia (5'→3')
11	TPS3-F	CGGGATCCCATGTCTTTGGAAGTTTCAGCCTC
12	TPS3-R	CGGTCGACTCATACACATATCGGCACAGG

EJEMPLO 3. *Análisis de secuencia y filogenético del cDNA de naranjo que codifica la sesquiterpeno sintasa CsTPS3 (SEQ ID NO:2)*

Resultados: En general las sesquiterpeno sintasa vegetales tienen un peso molecular de 50 – 100 KDa, con una longitud entre 550 – 580 aminoácidos (Bohlman y col., [1998] *Proc. Natl. Acad. Sci. USA.* **95**, 4126-4133). El cDNA del gen *Cstps3* (SEQ ID NO:1) aislado a partir del flavedo de naranjas Navel objeto de la invención, codifica una proteína de 564 aminoácidos (SEQ ID NO:2) con un peso molecular (PM) de 65.38 KDa. El PM estimado de la proteína CsTPS3 está en concordancia con el estimado para otras sesquiterpeno sintasas vegetales, lo cual apoya nuestra hipótesis inicial de que el fragmento de PCR aquí aislado corresponde a un OFR que codifica una proteína completa. El análisis de BLASTP reveló que la secuencia deducida de aminoácidos del polipéptido CsTPS3 (SEQ ID NO:2) tiene similitud a germacreno D sintasas (GerD) de otras especies de Rósidas, así como a la GerD de kiwi (Asterácea). El polipéptido CsTPS3 mostró entre un 54 – 66% de identidad a sesquiterpeno sintasas de la subfamilia Tps-a (Bohlman y col., [1997] *J. Biol. Chem.* **272**, 21784-21792), incluyendo la GerD de uva (AAS66357), rosa (BQ105086), álamo de California (AAR99061), y kiwi (AAAX16121), y la GerB de *Cistus creticus* (ACF94469). Las sesquiterpenos sintasas de la subfamilia Tps-a contienen dos dominios estructurales: (i) un dominio N-terminal con actividad glicosil hidrolasa, y (ii) un dominio C-terminal que comprende el sitio activo del enzima. Los dominios conservados DDXXD, NSE/DTE y RXR que se encuentran presentes en todas las sesquiterpenos sintasas vegetales (revisado por Degenhardt y col., [2009] *Phytochemistry* **70**, 1621-1637), también se encuentran conservados en el polipéptido CsTPS3 (Fig. 5). Los dominios DDXXD y NSE/DTE están involucrados en el enlace cooperativo de tres átomos de Mg⁺² que estabilizan el grupo pirofosfato del substrato FPP. Por su parte, el dominio RXR está implicado en el secuestro del grupo difosfato, de tal forma que lo mantiene alejado de la cavidad hidrofóbica a la vez que favorece la formación de diferentes carbocationes altamente reactivos que se forman durante la catálisis. Además, la comparación de la secuencia de los polipéptidos analizados

reveló que otros residuos de aminoácidos altamente conservados y que son esenciales para la actividad de la proteína modelo 5-*epi*-aristolóqueno sintasa de tabaco (TEAS) (Starks y col., [1997] *Science* **277**, 1815-1820), están también conservados en polipéptido CsTPS3. Interesantemente, solo el primer residuo de treonina de la tétada ⁴⁰¹TTY⁴⁰⁵ involucrada en la actividad catalítica de TEAS, está conservado en la secuencia homóloga ⁴¹⁵TSGS⁴¹⁸ del polipéptido CsTPS3 (**Fig. 5**). El residuo G⁴¹⁷ está conservado en las GerD de populus y kiwi, así como en la GerB de *C. creticus*. Por otra parte, la última posición de la tétada está ocupada por un residuo de serina (S⁴¹⁸) en el polipéptido CsTPS3, mientras que en el resto de las proteínas presentes en el alineamiento tienen un residuo de tirosina (Y) conservado en dicha posición (**Fig. 5**). Estas observaciones sugieren que los aminoácidos que conforman la tétada ⁴¹⁵TSGS⁴¹⁸ podrían estar involucrados en la especificidad de producto de la proteína CsTPS3 (SEQ ID NO:2) de naranjo objeto de la invención. El residuo de tirosina en la posición 278 (Y²⁷⁸) de TEAS, el cual está conservado en todas las sesquiterpenos sintasas cíclicas, también lo está en el polipéptido CsTPS3. Este hecho apoya nuestra hipótesis de que la proteína CsTPS3 corresponde a una sesquiterpeno sintasa cíclica. Se ha sugerido que el residuo Y²⁷⁸ estabiliza la carga positiva del intermediario carbocatiónico que precede a la etapa de ciclización del FPP (Maruyama y col., [2001] *Biol Pharm Bull.* **24**, 1171-1175; Chang y col., [2005] *Biotechnol. Lett.* **27**, 285-288).

El análisis filogenético de la secuencia deducida de aminoácidos de la proteína CsTPS3 (SEQ ID NO:4) objeto de la invención, con el de otras sesquiterpeno sintasas con probada actividad germacreno D y cadineno (Cad) sintasa, agrupa a la CsTPS3 con GerD de otras especies de Rósidas (*Vitis vinefera*, *Rosa hybrid*, *Populus trichorcapa*, y *Cistus creticus*,) y con la asterácea *Actinia deliciosa* (**Fig. 6**). La GerD de la monocotiledónea *Zingiber officinale* y la cubebeno sintasa (Cub) de la gimnosperma *Magnolia grandiflora* agrupan separadamente de las sesquiterpeno ciclasas de dicotiledóneas.

EJEMPLO 4. *Caracterización bioquímica de la actividad Sesquiterpeno Sintasa: Expresión heteróloga de la proteína CsTPS3 en células bacterianas, Preparación de extractos proteicos libres de células y ensayo in vitro de la actividad Sesquiterpeno Sintasa*

La actividad catalítica de la proteína CsTPS3 (SEQ ID NO:2) se ensayó *in vitro* utilizando extractos proteicos preparados a partir de células de la cepa bacteriana *Escherichia coli* BL21(DE3) portadoras de la región codificante del gen *Cstps3* (SEQ ID NO:1). que había sido previamente clonado en el vector de expresión pET45b(+). Extractos proteicos

preparados a partir de células de *E. coli* BL21(DE3) transformadas con el vector pET45b(+) vacío se utilizaron como control negativo. La caracterización funcional de la nueva sesquiterpeno sintasa de naranjo CsTPS3 (SEQ ID NO:2) aquí aislada, reveló que dicha proteína es una sesquiterpeno sintasas multifuncional. La biosíntesis de múltiples sesquiterpenos es una característica de la gran mayoría de sesquiterpenos sintasas aisladas de diferentes especies vegetales, y se encuentra bien documentado en el estado de la técnica (Lesburg y otros, [1997] *Science* **277**, 1820-1824; Starks y col., [1997] *Science* **277**, 1815-1820; Steel y col., [1998] *J. Biol. Chem.* **273**, 2078-2089; Caruthers y col., [2000] *J. Biol. Chem.* **275**, 25533-25539; Rynkiewicz y col., [2001] *Proc. Natl. Acad. Sci. USA.* **98**, 13543-13548; Arimura y col., [2004] *Plant J.* **37**, 603-616; Lücker y col., [2004] *Photochemistry* **65**, 2649-2659; Deguerry y col., [2006] *Arch. Biochem. Biophys.* **454**, 123-136; Picaud y col., [2006] *Arch. Biochem. Biophys.* **452**, 17-28; Jones y col., [2008] *Arch Biochem Biophys.* **477**, 121-130; Sungbeom y Chappell [2008] *Plant Physiol.* **147**, 1017-1033; Nieuwenhuizen y col., [2009] *J Exp Bot.* **60**, 3203-3219; Dunner y col., [2011] *Phytochemistry* **72**, 897-908).

Materiales y Métodos: El fragmento de cDNA de naranjo con el correcto ORF que codifica un polipéptido con homología a sesquiterpeno sintasas, se subclonó en fase entre los sitios de restricción *Bam*H I y *Sal* I del vector de expresión pET45b(+) (Novagen), usando el sistema Rapid DNA Ligation Kit (Roche). Células competentes NovaBlue (Novagen) se transformaron con una pequeña alícuota de la reacción de ligación, siguiendo las instrucciones del fabricante. Los plásmidos recombinantes pET45b(+) portadores del DNA codificante de la nueva sesquiterpeno sintasa de naranjo CsTPS3 (SEQ ID NO:2) se seleccionaron en medio Luria-Bertani (LB) suplementado con ampicilina (50 µg/ml). El DNA plasmídico extraído de clones recombinantes se secuenció para verificar la presencia de la región codificante del gen *Cstps3* (SEQ ID NO:1) en la orientación y el marco lectura correcto, antes de transformar un huésped de expresión apropiado. Para la producción de la proteína CsTPS3, así como de sus respectivas formas mutantes, los plásmidos recombinantes se transformaron en las células de expresión BL21(DE3) (Novagen). Esta cepa es un lisógeno del bacteriófago λDE3, un derivado de bacteriófago λ que lleva la región de inmunidad del bacteriófago 21 y un fragmento de DNA del gen *lacI*, el promotor *lacUV5*, y el gen de la RNA polimerasa del transposón T7. Este fragmento insertado dentro del gen *int*, evita la integración o escisión del lisógeno DE3 del cromosoma en ausencia de un bacteriófago ayudante. El único promotor que dirige la transcripción del gen de la RNA T7 polimerasa es el promotor *lacUV5*, el cual se induce al añadir IPTG al medio de cultivo. La expresión del gene diana fusionado en fase al marco de lectura abierto del promotor T7 es

activado tras la inducción de la RNA T7polimerasa. Tras la producción de la proteína de interés en el citoplasma de las células bacterianas recombinantes, se utilizaron dos métodos para la preparación de los extractos proteicos libres de células y el posterior ensayo *in vitro* de actividad sesquiterpeno sintasa.

5

Ensayo enzimático I: Se inoculó una colonia de la cepa bacteriana (lisógeno λ D3) portadora del vector pET45 recombinante en 10 ml de medio LB suplementado con 50 μ g/ml carbenicilina, 34 μ g/ml cloranfenicol y 0.5% glucosa. El pre-inóculo se incubó durante toda la noche a 37°C con agitación (250 rpm). El cultivo se utilizó para inocular 100 ml de medio LB suplementado como se describe arriba, y se creció a 37°C hasta una DO_{600} de 0.6. La inducción se llevó a cabo tras añadir al cultivo 1 mM de IPTG. La incubación se continuó durante toda la noche a 18°C. Las células se precipitaron por centrifugación y se resuspendieron en 1ml del tampón SSAB (100 mM fosfato de sodio pH 7.0, 10 mM $MgCl_2$, 1 mM DTT, 10% v/v glicerol; Prosser y col., [2002] *Phytochemistry* **60**, 691-702). Las células se lisaron mediante tres/cinco ciclos de congelación/descongelación, seguido por la adición de 7.5 KU de rLisozima (Novagen) por cada ml del tampón SSAB. La suspensión celular se incubó a temperatura ambiente en una plataforma con agitación a baja velocidad durante 15 min. El DNA presente en el lisado se digirió con 25 U de Benzonasa Nucleasa (Novagen) por cada ml del tampón SSAB. La suspensión celular se incubó a temperatura ambiente y a baja velocidad durante 20 min hasta que la viscosidad desapareció por completo. En este momento los tubos se incubaron en hielo. El debrís celular se eliminó mediante centrifugación a 13.000 rpm durante 30 min a 4°C. Los extractos proteicos crudos se utilizaron para el ensayo *in vitro* de la actividad sesquiterpeno sintasa, utilizando FPP (16 μ g FPP/ml de extracto proteico) como sustrato. Los ensayos enzimáticos se llevaron a cabo en viales de vidrio con tapón de PTFE utilizando para ello 500 μ l de lisado crudo suplementado con 16 μ g de FPP. La reacción se cubrió con 500 μ l de pentano y los tubos se incubaron toda la noche a 30°C. Tras el período de incubación, los volátiles se recuperaron de la capa de pentano y se extrajeron dos veces de la fase acuosa utilizando 1 ml de pentano cada vez. Las fracciones de pentano se combinaron y se concentraron hasta 50 μ L bajo atmósfera de nitrógeno, y finalmente se analizaron por cromatografía de gases acoplada a espectrofotometría de masas (GC-MS). Un equipo Thermo Trace GC se utilizó acoplado a un espectrofotómetro de masas con modo de ionización de electrones (EI) a 70 eV. Los compuestos volátiles se separaron en una columna de HP-INNOWax (30 m x 0.25 mm i.d. x 0.25 μ m de película) (Agilent J&C Columns). La inyección de la muestra se realizó sin fragmentación y el horno se programó como sigue: 40°C durante 5 min, posteriormente

35

la temperatura se incrementó hasta los 150°C a razón de 5°C/min, y finalmente a 250°C durante 2 min (a razón de 20°C/min). La temperatura del inyector fue de 220°C, y se utilizó helio como el gas portador a un flujo constante de 1.5 ml/min. Tras el impacto de electrones, los espectros de masas se registraron en el rango de 30-400 m/z con una velocidad de exploración de 0,5 barridos/s. La identificación de los compuestos se realizó por coincidencia de los espectros de masas adquiridos y los índices de retención (RI) con aquellos correspondientes al de los patrones auténticos disponibles, o con aquellos sesquiterpenos presentes en la base de datos de referencia (Wiley y NIST).

10 Ensayo enzimático II: Las células BL21(DE3) portadoras del vector pET45b recombinante que expresa la sesquiterpeno sintasa CsTPS3 o cualquiera de sus versiones modificadas se crecieron a 37°C en medio Terrific Broth (TB) suplementado con 50 µg/ml carbenicilina hasta una DO₆₀₀ de 0.6. Las células se indujeron con 1 mM IPTG durante 16 h a 18 °C. Los cultivos bacterianos crecidos durante toda la noche y que expresan las proteínas recombinantes se centrifugaron para recolectar las células. Las células se resuspendieron en tampón de lavado (20 mM Tris·HCl, pH 7.0, 50 mM KCl) mediante agitación en vortex, y posteriormente se centrifugaron. Las células se lavaron con 1 ml de tampón de extracción de proteínas (50 mM 3-(N-morfolino)-ácido 2-hidroxipropanosulfónico, pH 7.0, 10% [v/v] glicerol, 5 mM MgCl₂, 5 mM DTT, 5 mM ascorbato de sodio, 0.5 mM fluoruro de fenilmetilsulfonilo), y posteriormente se agitaron, sonicaron y centrifugaron. Tras recuperar el sobrenadante se realizó el cambio de tampón de extracción por el tampón de ensayo enzimático (10 mM de ácido 3-(N-morfolino)-2-hidroxipropanosulfónico, pH 7.0, 10% [v/v] glicerol, 1 mM DTT) usando columnas PD-10 (GE Healthcare Life Sciences). Los ensayos enzimáticos se llevaron a cabo en viales de vidrio con tapa de rosca PTFE, usando 500 µl de lisado crudo suplementado con 10 µg FPP, MgCl₂ (20 mM), e inhibidores de fosfatasas (0.2 mM NaWO₄, 0.1 mM NaF). Las mezclas de reacción se cubrieron con 500 µl de hexano y los tubos se incubaron durante toda la noche a 30°C. Posteriormente, la fase superior de hexano se agitó con la fase acuosa y tras centrifugar se utilizó para el análisis de metabolitos. Las fracciones de hexano se concentraron hasta 50 µl bajo una atmósfera de nitrógeno y se analizaron por cromatografía de gases acoplada a espectrofotometría de masas (GC-MS). El equipo utilizado fue un Hewlett Packard 6890 GC acoplado a un detector selectivo de masas (MSD) 5973 provisto de un inyector automático múltiple 7683, y equipado con una columna capilar HP-5MS. La inyección se realizó sin fragmentación y el horno se programó de la siguiente manera: de 50°C (5 min de espera) hasta 250°C (5 min de espera) a razón de 5°C/min a. La temperatura del inyector fue de 200°C y se utilizó helio como gas portador con un flujo constante de 1.5 ml/min. La temperatura de la fuente de

iones y del cuádrupolo era de 250 y 150°C, respectivamente. El espectro de masas tras el impacto de electrones se registró en el rango de 40-300 m/z. La identificación de los compuestos se realizó mediante coincidencia de los espectros de masas adquiridos y los índices de retención (RI) con aquellos correspondientes a los de los patrones auténticos disponibles, o con aquellos sesquiterpenos presentes en las bases de datos de referencia NIST y MassFinder. La abundancia de cada producto detectado se cuantificaron mediante integración de las áreas de cada pico utilizando el programa informático Enhanced Chemstation version B.01.00 (Agilent Technologies).

10 Resultados: El ensayo *in vitro* de la actividad sesquiterpeno sintasa realizado en cualquiera de los *ensayos enzimáticos I y II*, demostraron que el cDNA del gen *Cstps3* codifica inequívocamente un polipéptido con actividad sesquiterpeno sintasa. Dicha actividad se determinó por la capacidad del polipéptido CsTPS3 de transformar el sustrato FPP en sesquiterpenos cíclicos. Ensayo enzimático I: Los resultados del análisis GS-MS mostraron que los lisados crudos de las células bacterianas que expresan la proteína recombinante CsTPS3, preparados según en *ensayo enzimático I*, forman como principal producto δ -cadineno (46,91%), además de otros sesquiterpenos cíclicos minoritarios (**Tabla 5** y **Fig. 7A**). La identificación de los picos más pequeños mostró que estos correspondían a los sesquiterpenos cíclicos β -cubebeno (14,55%), germacreno D (11,8%), α -copaeno (9,65%), cubebol (7,11%), elemol (5,93%), y biciclogermacreno (4,05%). A pesar de su actividad multifuncional, el polipéptido CsTPS3 (SEQ ID NO:4), codificado por el cDNA *Cstps3* (SEQ ID NO:2), se clasificó como una δ -cadineno sintasa, en función de la abundancia relativa del δ -cadineno. La denominación de las sesquiterpenos sintasas multifuncionales en función de su producto mayoritario es una norma general bien conocida en estado de la técnica. La estructura molecular de los productos formados por el polipéptido CsTPS3 se muestra en la **Fig. 2**. El control correspondiente a las células BL21(DE3) transformadas con el vector pET45b vacío, no produjo ninguno de los compuestos volátiles sintetizados por el enzima multifuncional CsTPS3, lo cual demuestra que los sesquiterpenos volátiles identificados en los extractos preparados a partir de las células que expresan la región codificante del gen *Cstps3* (SEQ ID NO:1), son producto de la actividad catalítica del enzima CsTPS3. Ensayo enzimático II: Cuando la actividad ciclasa del polipéptido CsTPS3 se probó bajo las condiciones del *ensayo enzimático II*, se identificó un mayor número de sesquiterpenos (**Tabla 6**). Ésta diferencia puede ser consecuencia de la mayor sensibilidad del equipo GC-MS utilizado en este ensayo. Los sesquiterpenos producidos por el polipéptido CsTPS3 en los *ensayos enzimáticos I y II* no pueden compararse entre sí, ya que no se cuantificó la cantidad de proteína recombinante añadida en cada tipo de ensayo. En el perfil de

sesquiterpenos volátiles producidos por el polipéptido CsTPS3 (SEQ ID NO:2) se identificaron varios sesquiterpenos, de los cuales el δ -cadineno fue el producto mayoritario (27,36%) (Table 6 y Fig. 8A). La identificación de los picos minoritarios reveló que correspondían a los sesquiterpenos cíclicos α -copaeno (19,60%), β -cubebeno (14,30%),
 5 germacreno D (9,36%), elemol (7,49%), cubeol (4,42%), biciclogermacreno (3,58%), α -muuroleno (1,92%), germacreno D-4-ol (1,90%), α -bulneseno (1,69%), *E*-cariofileno (1,48%), α -cubebeno (1,12%). Ya que el δ -cadineno fue el producto mayoritario de la actividad enzimática de la proteína CsTPS3 (SEQ ID NO:2), esta se clasificó como una δ -cadineno sintasa. La estructura molecular de los sesquiterpenos producidos por el polipéptido
 10 CsTPS3 se muestra en la Fig. 2.

TABLA 5

Abundancia relativa de los sesquiterpenos producidos *in vitro* por el polipéptido natural CsTPS3 y sus respectivas versiones modificadas. La actividad catalítica de los enzimas se ensayó según el método descrito en el *ensayo enzimático I*. Los datos corresponden a porcentajes (%) de área de cada sesquiterpeno con respecto al total del área de los picos detectados en el perfil de volátiles.

Sesquiterpeno	Polipéptidos modificados				
	CsTPS3	VB2	VB5	VC1	VC2
	%	%	%	%	%
α -Copaeno	9,65	-	-	-	-
α -Gurjuneno	-	-	-	13,98	10,59
β -Cubebeno	14,55	-	-	-	-
Germacreno D	11,8	-	-	<u>77,61</u>	<u>75,91</u>
Biciclogermacreno	4,05	-	-	8,4	13,5
δ -Cadineno	<u>46,91</u>	-	-	-	-
Cubebol	7,11	-	-	-	-
Elemol	5,93	<u>77,49</u>	<u>87,78</u>	-	-
Guaiol	-	22,51	12,22	-	-

TABLA 6

Abundancia relativa de los sesquiterpenos producidos *in vitro* por el polipéptido natural CsTPS3 y sus respectivas versiones modificadas. La actividad catalítica de los enzimas se ensayó según el método descrito en el *ensayo enzimático II*. Los datos corresponden a porcentajes (%) de área de cada sesquiterpeno con respecto al total del área de los picos detectados en el perfil de volátiles. Productos con abundancias relativas del 1% o menos se indica como trazas.

Sesquiterpeno	Polipéptidos modificados				
	CsTPS3	VB2	VB5	VC1	VC2
	%	%	%	%	%
α -cubebeno	1.12	trazas	trazas	trazas	trazas
α -copaeno	19.60	1.66	trazas	trazas	trazas
β -cubebeno	14.30	-	trazas	-	-
β -elemeno	-	2.40	trazas	3.43	3.96
α -gurjuneno	trazas	-	-	5.44	5.70
<i>E</i> - β -cariofileno	1.48	1.63	trazas	trazas	trazas
α -humuleno	3.58	2.71	trazas	-	-
germacreno D	9.36	2.01	trazas	84.46	83.10
biciclogermacreno	3.56	-	-	2.53	2.91
α -muuroleno	1.92	-	-	-	-
α -bulneseno	1.69	1.48	trazas	-	-
cubebol	4.42	1.07	trazas	-	-
δ -cadineno	27.36	3.19	1.15	trazas	trazas
elemol	7.49	66.70	83.86	-	-
germacreno D-4-ol	1.90	trazas	trazas	2.12	2.40
bulnesol	trazas	15.48	11.44	-	-

Se ha establecido que la formación de múltiples sesquiterpenos por enzimas sesquiterpeno sintetas multifuncionales se produce a través de la formación de diferentes estructuras carbocatiónicas intermediarias (revisado por Degenhardt y col., [2009] *Photochemistry* **70**, 1621-1637). La reacción en cascada se inicia con la ionización dependiente de iones metálicos divalentes y/o por la isomerización del precursor farnesil-pirofosfato (FPP). Los intermediarios cationicos resultantes sufren varios rearrreglos intramoleculares tales como: ciclaciones, cambios de grupos hidridos (migración de un átomo de hidrógeno en un carbocation al átomo de carbono con carga +1 desde un carbón adyacente), o transiciones conformacionales, hasta que la reacción culmina por la pérdida de protones o activada mediante la adición de una molécula nucleofílica (una especie química tal como un grupo hidroxilo que dona un par de electrones a una molécula electrofílica para formar un enlace químico). La diversidad estructural de los productos formados por las sesquiterpeno sintetas se debe al gran esqueleto carbonado (cola prenil) del FPP y a la presencia de tres enlaces

dobles. El esquema mecanístico propuesto para la formación de sesquiterpenos por el polipéptido natural CsTPS3 y que se muestra en las **Figs. 9A y 9B**, trata de explicar la relación de sus múltiples productos basados en la información de los mecanismos de reacción propuestos para otras sesquiterpenos sintasas vegetales (Deguerry y col., [2006] *Arch Biochem Biophys.* **454**, 123-136; Jones y col., [2008] *Arch Biochem Biophys.* **477**, 121-130; Lee y Chappell, [2008] *Plant Physiol.* **147**, 1017-1033; O'Maille y col., [2006] *Arch Biochem Biophys.* **448**, 73-82; Picaud y col., [2006] *Arch Biochem Biophys.* **452**, 17-28; Starks y col., [1997] *Science* **277**, 1815-1820; Yoshikuni y col., [2006], *Chem Biol.* **13**, 91-98). La formación de sesquiterpenos cíclicos a partir del precursor FPP por la sesquiterpeno sintasa CsTPS3 de naranjo producto de la invención, puede proceder a través de la formación de intermediarios carbocatiónicos altamente reactivos e inestables. Este hipotético mecanismo explicaría la formación de múltiples productos por la proteína CsTPS3 recombinante expresada en bacterias, los cuales fueron posteriormente identificados mediante análisis de GC-MS. Cuando el sustrato *E,E*-FPP se une al sitio activo de la enzima a través de los cationes divalentes de magnesio (Mg^{2+}), el grupo fosfato se libera formándose el *E,E*-farnesil catión (catión *transoid*-farnesil), el cual puede sufrir una etapa de ciclización (cierre de anillo 11,1; ruta **a** en **Fig. 9A**) para generar el C11 humulil catión macrocíclico. Éste puede deprotonarse para formar el α -humuleno (3,58% [**Tabla6**]), o sufrir un cierre de anillo 2,10, seguido por deprotonación del carbono C3 para formar β -cariofileno [**Tabla 6**]. Alternativamente, la isomerización del *E,E*-farnesil catión en el enlace C2 – C3 (ruta **b** en **Fig. 9A**) formaría el intermediario nerolidil catión necesario para la formación de sesquiterpenos tipo germacranos y cadinanos. Diferentes estudios relacionados con el uso de análogos radiactivos del FPP y del nerolidil difosfato apoyan la intermediación del nerolidil catión en el mecanismo de reacción catalizado por sesquiterpenos sintasas (revisado por Degenhardt y col., [2009] *Photochemistry* **70**, 1621-1637). En una posterior etapa de reacción, el *Z,E*-farnesil catión podría sufrir un cierre de anillo 10,1 para formar el *3Z,7E*-germacradienil catión. La deprotonación directa del *3Z,7E*-germacradienil catión formaría el producto minoritario germacreno D. Alternativamente, la hidroxilación del *3Z,7E*-germacradienil catión en el carbono C3 formaría el Germacreno D-4-ol (ruta **c** en **Fig. 9A**). Ya que los germacranos son moléculas flexibles, el *3Z,7E*-germacradienil catión podría fácilmente sufrir posteriores transformaciones a través de la interconversión de las conformaciones de silla de sus respectivos isómeros. La formación de diferentes esqueletos de carbón implicaría la subsecuente ciclización de los conformeros a través del ataque de carbocationes sobre cualquiera de los dos dobles enlaces. Así, los dos conformeros del *3Z,7E*-germacradienil catión darían lugar a la formación de los cationes cadinenil (ruta **d** en **Fig. 9A**) y muurolenil (ruta **e** en **Fig. 9A**), respectivamente. El cadinenil catión generado por

el enzima CsTPS3 podría sufrir deprotonación en C7 y la subsecuente formación del doble enlace C6 – C7 para generar δ -cadineno como producto mayoritario (46,91% según los resultados del *ensayo enzimático I* [Tabla 5], y 27,36% según el *ensayo enzimático II* [Tabla 6]). El β -cubebeno, uno de los dos productos secundarios más abundantes producidos por la CsTPS3 (14,55% [Tabla 5], y 14,30% [Tabla 6]), podría generarse mediante la migración del átomo de hidrógeno en C1 al C2 (migración de hidrógeno 1,2), seguido por cierre electrofílico en 1,2 y deprotonación del cadinenil catión. Alternativamente, el cadinenil cation podría sufrir también otros rearrreglos (migración de átomos de hidrógeno, cierre de anillo y oxidación) para formar cubebol como producto colateral minoritario (4,42% según el *ensayo enzimático II*). Por su parte, el muurolenil catión podría sufrir la migración de hidrógeno 1,2, seguido de la formación de un enlace entre C2 - C7 (cierre de anillo 2,7) para general un macrociclo transitorio del cual se formaría por deprotonación en C3 el producto colateral α -copaeno (9,65% [Tabla 5], y 19,60% [Tabla 6]). Alternativamente, el muurolenil catión podría deprotonarse para formar el α -muuroleno como producto colateral minoritario (1,92% [Tabla 6]). En una reacción mecanística adicional, el *3Z,7E*-germacradienil catión podría transformarse en germacreno D (11,8% [Tabla 5] y 9,36% [Tabla 6]), y biciclogermacreno (4,05% [Tabla 5], y 3,56% [Tabla 6]) (ruta f en Fig. 9A). La biosíntesis de sesquiterpenos derivados del elemano o gauiano por la proteína CsTPS3 podría ocurrir a través de la formación transitoria del intermediario germacreno A (Fig. 9B). En este escenario mecanístico, la ciclización del *E,E*-farnesyl catión entre los carbonos 10 y 1 (ruta g en Fig. 9B), formaría el *E,E*-germacradienyl catión. La posterior deprotonación del *E,E*-germacradienil catión generaría el intermediario neutro germacreno A. Éste no podría dejar el sitio activo del enzima CsTPS3, ya que su naturaleza hidrofóbica no permitiría que entrase de nuevo en dicha cavidad. Un mecanismo similar se ha sugerido para explicar la actividad catalítica de la SasesquiTPS1 del árbol del sándalo (*Santalum album*) (Jones y col., [2008], *Arch Biochem Biophys.* **447**:121-130). Los residuos de Tyr que ocupan las posiciones 520 (Y⁵²⁰) y 527 (Y⁵²⁷) en la secuencia de aminoácidos de la proteína TEAS, y que se han involucrado en la formación del germacreno A, están conservados en la sesquiterpeno sintasa CsTPS3 de naranjo (Fig. 5). El germacreno A se ha postulado como un producto intermediario en el mecanismo de reacción que da lugar a la formación del 5-*epi*-aristolocheno. El residuo Y⁵²⁷ de TEAS parece estar implicado en la protonación del germacreno A (Starks y col., [1997] *Science* **277**, 1815-1820). La mutagénesis del residuo Y⁵²⁰ (Y⁵²⁰F) de TEAS, genera una TEAS modificada que produce germacreno A y no 5-*epi*-aristolocheno como producto final (Rising y col., [2000] *J Am Chem Soc.* **122**, 1861-1866). Por lo tanto, el residuo Y⁵²⁰ evitaría la salida del intermediario germacreno A de la cavidad catalítica de la proteína TEAS, mientras que el residuo Y⁵²⁷ estaría implicado en la activación

del germacreno A necesaria para la formación del 5-*epi*-aristoloueno. La formación del producto colateral α -bulneseno (1,69% [Tabla 6]) por la CsTPS3, tendría lugar como consecuencia de la formación del guianil catión por la protonación del germacreno A y cierre de anillo 6,2. Los elemanos son considerados productos de degradación de los germacrenos debido a Cope rearrreglos (reacciones estereoespecíficas que proceden vía estados de transiciones de conformaciones de silla), inducidas por las elevadas temperaturas del inyector durante el análisis de GC-MS (revisado por Adio, [2009] *Tetrahedron* **65**, 1533-1552). Así, por ejemplo el Cope rearrreglo del germacreno A y germacreno B conducen a la formación de β -elemeno y γ -elemeno, respectivamente. También se ha sugerido que el elemol es probablemente el producto de ruptura térmica del hedicariol, ya que el aceite esencial rico en elemol de la morera australiana (*Hedycarya angustifolia* A. Cunn) produce principalmente hedicariol y solo trazas de elemol cuando el aceite se extrae a temperatura ambiente.

15 **EJEMPLO 5.** *Mutagénesis y caracterización bioquímica de las versiones modificadas del polipéptido CsTPS3 de naranja*

Una de las características más interesantes de los enzimas sesquiterpeno sintasa vegetales es su capacidad de formar múltiples productos con diferente estructura (esqueleto de átomos de carbón), aun cuando comparten una estructura tridimensional (3D) común. La actividad catalítica y diversidad química de sus productos radica en gran medida en el plegamiento diferencial del sustrato FPP en el interior de la cavidad catalítica (Starks y col., [1997] *Science* **277**, 1815-1820; Bouwmeester y col., [2002] *Plant Physiol.* **129**, 134-144; Deligeorgopoulou y Alleman [2003] *Biochemistry* **42**, 7741-7747). La localización espacial de las cadenas laterales de algunos aminoácidos en el sitio activo de la proteína, distingue catalíticamente una sesquiterpeno sintasa del resto. Además se ha sugerido que el carácter multiproducto de las sesquiterpenos sintasas vegetales puede ser también una consecuencia de la presencia de dos compartimentos independientes en el interior del sitio activo del enzima. Así, el análisis de mutagénesis de la terpeno sintasa multirproducto ZmTPS4 de maíz, demostró que las primeras etapas en la vía de transformación del FPP hasta la formación del intermediario monocíclico bisabolil catión, se catalizan en el compartimento I; mientras que las etapas secundarias de ciclización hasta la formación de los sesquiterpenos bicíclicos tiene lugar en la compartimento II (Köllner y col., [2006] *Arch Biochem Biophys.* **448**:83-92). Las semejanzas estructurales de las diferentes sesquiterpeno sintasas y sus productos de reacción, sugieren que pequeños cambios en la geometría del centro activo del enzima, debido a las diferencias de identidad y localización espacial de las

5 cadenas laterales de ciertos aminoácidos claves, son suficientes para alterar la especificidad de sus mecanismos de reacción. Los residuos de aminoácidos localizados en las capas que rodean el sitio activo pueden también influenciar la conformación y la especificidad de producto de las sesquiterpenos sintasas. En este sentido, se ha demostrado que la biotecnología a través del uso de técnicas del DNA recombinante e ingeniería genética constituye una poderosa estrategia para manipular la actividad catalítica de las sesquiterpeno sintasas, permitiendo incrementar el rendimiento y/o especificidad de ciertos sesquiterpenos de interés.

10 *Materiales y Métodos:* Las versiones modificadas de la δ -cadineno sintasa (CsTPS3, SEQ ID NO:2) objeto de la invención, se generaron *in vitro* mediante mutagénesis dirigida. Para ello se introdujeron de forma específica mutaciones puntuales o combinación de múltiples mutaciones puntuales en la secuencia nucleotídica del cDNA del gen *Cstps3* (SEQ ID NO:2). Los cDNAs modificados objetos de la invención codifican polipéptidos modificados que contienen sustituciones específicas en su secuencia de aminoácidos, con respecto a la
15 secuencia natural del enzima δ -cadineno sintasa de naranjo (CsTPS3, SEQ ID NO:4). Para generar sustituciones individuales o múltiples de pares de bases, se utilizaron los sistemas comerciales QuickChange® II Site-direct y QuickChange® Multi Site-Direct Mutagenesis kits (Stratagene), respectivamente. Ambos tipos de sistemas permitieron la introducción de mutaciones específicas (puntuales o múltiples) en un único ensayo, en la secuencia
20 nucleotídica del cDNA que codifica el polipéptido natural CsTPS3 (SEQ ID NO:2), el cual había sido previamente clonado en el vector de expresión pET45b(+). Ambos métodos requieren la síntesis de oligonucleotidos modificados de 25 – 45 bases de longitud, que llevan la mutación o mutaciones de interés y que se utilizaran como cebadores en una única
25 reacción de PCR. La sustitución de las bases deseadas debe localizarse en el centro del oligonucleotido y con una secuencia correcta a ambos lados de 10-15 nucleótidos. Para el diseño de los cebadores se utilizó el programa informático QuickChange (<http://www.stratagene.com/qcprimerdesign>), según las instrucciones del fabricante. Los cebadores utilizados para la mutagénesis del cDNA del gen *Cstps3* se muestran en la **Tabla**
30 **7.**

TABLA 7

Oligonucleótidos usados como cebadores para la mutagénesis del cDNA del gen *Cstps3* (SEQ ID NO:1) que codifica la δ -cadineno sintasa de naranjo (CsTPS3, SEQ ID NO:2).

SEQ ID NO	Oligonucleótido	Sequence (5'→3')
13	CsTPS3_S ⁴¹⁸ H	GATGTGGTTGACAACATTGGGTGGCCAGATGTA AATGCAAC
14	CsTPS3_C ⁴⁵⁵ G	ATCCATGAGTCGGCCAACATAGAAAGAAGCCCTA AAATCT
15	CsTPS3_S ⁴¹⁶ T, G ⁴¹⁷ T, S ⁴¹⁸ Y_Forward	GATGAGTACATGACGGTTGCATTAGTTACA TATCCAATGTTGTCAACCACATCC
16	CsTPS3_S ⁴¹⁶ T, G ⁴¹⁷ T, S ⁴¹⁸ Y_Reverse	GGATGTGGTTGACAACATTGGATAGGTAGTTGTA TAATGCAACCGTCATGTAATCATC

Resultados: En la Tabla 2 se muestran las sustituciones simples y múltiples de aminoácidos que fueron introducidas en la secuencia de aminoácidos del polipéptido natural CsTPS3 (SEQ ID NO:2), como consecuencia de las sustituciones de pares de bases introducidas en la secuencia nucleotídica natural del cDNA del gen *Cstps3* (SEQ ID NO:1) (**Fig. 1A**). La secuencia nucleotídica de las moléculas de ácidos nucleicos mutados objeto de la presente invención, que codifican variantes modificadas de la δ -cadineno sintasa de naranjo (CsTPS3, SEQ ID NO:2) corresponden a las versiones VB2 (SEQ ID NO:3; **Fig. 1B**), VB5 (SEQ ID NO:5; **Fig. 1C**), VC1 (SEQ ID NO: 7; **Fig. 1D**) y VC2 (SEQ ID NO: 9; **Fig. 1E**). Las moléculas mutadas de cDNA codifican polipéptidos modificados que contienen la sustitución no conservada S⁴¹⁸H o C⁴⁵⁵G, o la sustitución conservada S⁴¹⁶T en combinación con las sustituciones no conservadas G⁴¹⁷T y S⁴¹⁸Y (**Fig. 3**). El alto grado de similitud que se ha encontrado entre cientos de sesquiterpeno sintasas vegetales sugiere que la actividad catalítica de estos enzimas es consecuencia directa del plegamiento de la proteína. La comparación de la secuencia de las sesquiterpeno sintasas vegetales ha demostrado que sus motivos altamente conservados controlan reacciones de ciclización comunes a todas ellas, mientras que las regiones variables contribuyen a las diferencias en el plegamiento del sitio activo. Los cambios en la geometría del sitio activo pueden forzar diferentes conformaciones del sustrato FPP y de los intermediarios macrocíclicos que controlan la catálisis selectiva y la racionalización de sus diferentes productos. La identificación de los residuos de aminoácidos implicados en los mecanismos de la reacción de diferentes TPS y la generación de variantes con propiedades catalíticas alteradas, ha sido posible gracias a la comparación y análisis de secuencias en combinación con el modelaje *in silico* usando como molde la estructura 3D de algunas sesquiterpeno sintasas

(Chang y col., [2005] *Biotechnol Lett.* **27**, 285-288; Deligeorgopoulo y Alleman, [2003] *Biochemistry* **42**, 7741-7747; Kampranis y col., [2007] *Plant Cell* **19**, 1994-2005; Yoshikuni y col., [2006], *Chem Biol.* **13**, 91-98). El alineamiento de la secuencia de aminoácidos del polipéptido natural CsTPS3 de naranjo con sesquiterpeno sintasas filogenéticamente relacionadas, reveló algunos cambios de aminoácidos en su región C-terminal, en 5 posiciones análogas a aquellos residuos que se han implicado en la formación de varios sesquiterpenos bicíclicos (**Fig. 5**). El modelaje *in silico* del polipéptido CsTPS3 realizado en base a la homología con la estructura cristalizada de la 5-*epi*-aristolóqueno sintasa de tabaco (TEAS, PBD: 5EAT; Starks y col., [1997] *Science* **277**, 1815-1820), demostró que el 10 enzima CsTPS3 tiene en efecto el plegamiento típico de enzimas sesquiterpeno sintasas (**Fig. 10**). El modelaje molecular de la δ -cadineno sintasa de naranjo (CsTPS3) reveló además que la mayoría de los aminoácidos involucrados en la catálisis del sustrato FPP están altamente conservados, y que tienen una disposición espacial casi idéntica a la correspondiente a la proteína modelo TEAS (**Fig. 11**). Sin embargo, la identidad y 15 localización espacial de algunos aminoácidos que están en contacto directo con el sustrato FPP en la cavidad catalítica de la CsTPS3, difieren con respecto a la observada en la proteína modelo TEAS (**Fig. 10**). Por ejemplo, los aminoácidos de la tétada ⁴⁰¹TTTY⁴⁰⁴, involucrados en la estabilización del intermediario carbocatiónico mediante cargas parciales distribuidas localmente en la cavidad catalítica de la TEAS (Starks y col., [1997] *Science* 20 **277**, 1815-1820), corresponden a la secuencia ⁴¹⁵TSGS⁴¹⁸ en el polipéptido CsTPS3 (**Figs. 5 y 11**). Para analizar el impacto de estos residuos en la racionalización de los sesquiterpenos producidos por la ciclase CsTPS3, los residuos S⁴¹⁶ (T⁴⁰² in TEAS), G⁴¹⁷ (T⁴⁰³ in TEAS), y S⁴¹⁸ (T⁴⁰⁴ in TEAS) se modificaron mediante mutagénesis dirigida (**Tabla 2**). La mutagénesis de los aminoácidos seleccionados (**Tabla 2**), de forma individual o múltiple, resultó en 25 proteínas enzimáticamente competentes pero con una actividad catalítica alterada. Las mutaciones introducidas causaron incremento/ y/o supresión del rendimiento de los sesquiterpenos específicos, en comparación con los producidos por el enzima natural CsTPS3 de naranjo. Para investigar la posible contribución del residuo C⁴⁵⁵ (ubicado cerca del motivo NSE/DTE; **Fig. 5**) a la especificidad de producto del polipéptido natural CsTPS3, su cadena lateral nucleofílica se reemplazó por la cadena lateral menos reactiva del 30 aminoácido Gly (G). El reemplazamiento C⁴⁵⁵G causó la conversión de la actividad δ -cadineno sintasa del polipéptido natural CsTPS3 (SEQ ID NO:2), en la actividad elemol sintasa exhibida por el polipéptido modificado CsTPS3-VB2 (**Fig. 7B y 8B**). El polipéptido CsTPS3-VB2 produjo mayoritariamente elemol (77,49% [**Tabla 5**], 66,70% [**Tabla 6**]), además de guaiol (22,51% [**Tabla 5**]), o una cantidad discreta bulnesol (15,48% [**Tabla 6**]) 35 junto con una pequeña cantidad de δ -cadineno (3,19% [**Tabla 6**]). Adicionalmente, la

sustitución C⁴⁵⁵G causó una drástica reducción de la racionalización de todos los productos colaterales minoritarios formados por el enzima natural δ -cadineno sintasa (CsTPS3) (**Fig. 7 y 8, Tablas 5 y 6**). Es probable que la sustitución C⁴⁵⁵G haya afectado la isomerización del *Z,E*-germacradienil catión (**Fig. 9A**), a la vez que favorecería el cierre del *E,E*-farnesil catión entre los carbonos C10 – C1. Esta etapa de ciclización daría lugar a la formación del hedycariol, el precursor del elemol (**Fig. 9B**). La sustitución múltiple de C⁴⁵⁵G y S⁴¹⁸H en la secuencia natural del polipéptido CsTPS3 (SEQ ID NO:2), tuvo un efecto aditivo sobre la producción de elemol (**Fig. 7C y 8C**). Asimismo, la combinación de las mutaciones C⁴⁵⁵G y S⁴¹⁸H tuvo además de un efecto negativo sobre la producción de otros productos colaterales minoritarios (**Fig. 7C y 8C, Tablas 5 y 6**). Así el polipéptido modificado CsTPS3-VB5 (SEQ ID NO:6), el cual lleva las sustituciones C⁴⁵⁵G y S⁴¹⁸H, produjo elemol de una forma casi exclusiva (87,78% [**Tabla 5**], 83,86% [**Tabla 6**]). El polipéptido modificado CsTPS3-VB5 produjo además una cantidad discreta de guaiol (12,22%, [**Tabla 5**]), o bulnesol (11,4%, [**Tabla 6**]) junto con una pequeña cantidad de δ -cadineno (1,15% [**Tabla 6**]), y solo trazas de otros sesquiterpenos colaterales. El efecto sinérgico de las mutaciones C⁴⁵⁵G y S⁴¹⁸H sobre la producción de elemol se estimó entre un 13 – 29%, dependiendo del método enzimático utilizado (**Tablas 5 y 6**). En el escenario mecanístico, parece probable que la sustitución del grupo hidroxilo de la Ser en la posición 418 (S⁴¹⁸) por la cadena lateral más larga y voluminosa de la His (S⁴¹⁸H), establezca el *E,E*-germacradienil catión en la parte inferior hidrofóbica de la cavidad catalítica del polipéptido modificado CsTPS3-VB5 (SEQ ID NO:6). El modelaje *in silico* del polipéptido CsTPS3-VB5 realizado sobre la estructura de la proteína TEAS, mostró que las cadenas laterales de los aminoácidos mutados alteran la geometría de la cavidad catalítica del sitio activo del enzima, en comparación con la correspondiente al polipéptido natural CsTPS3 (SEQ ID NO:2) (**Fig. 12**).

25

En un intento de transmutar la actividad δ -cadineno sintasa del polipéptido natural CsTPS3 (SEQ ID NO:4), en una germacreno D sintasa, la triada ⁴¹⁶SGS⁴¹⁸ se reemplazó por la secuencia ⁴¹⁶TTY⁴¹⁸ (**Figs. 1D y 3**), ya que la mayoría de estos residuos se encuentran bien conservados en casi todas las germacreno D sintasas filogenéticamente relacionadas con la CsTPS3 de naranja, incluyendo la proteína modelo 5-*epi*-aristolobeno sintasa (**Fig. 5 y 6**). Efectivamente, la sustitución de la triada ⁴¹⁶SGS⁴¹⁸ por ⁴¹⁶TTY⁴¹⁸ transformó la actividad δ -cadineno sintasa del polipéptido CsTPS3 (SEQ ID NO:2), en una germacreno D sintasa altamente eficiente correspondiente al polipéptido modificado CsTPS3-VC1 (SEQ ID NO:8) (**Figs. 7D y 8D; Tabla 5 y 6**). Así, el polipéptido CsTPS3-VC1 produjo casi exclusivamente germacreno D (77,61% [**Tabla 5**], 84,46% [**Tabla 6**]). El modelaje *in silico* del polipéptido modificado CsTPS2-VC1 sobre la estructura de la proteína TEAS, mostró que las cadenas

35

laterales de los aminoácidos ⁴¹⁶TTY⁴¹⁸ alteran la geometría de la cavidad del sitio activo, en comparación con el correspondiente a la del enzima natural CsTPS3 (**Fig. 12**). La sustitución del grupo hidroxilo del residuo S⁴²² por la corta cadena lateral alifática, de la Ala, cuya posición se encuentra conservada en todos los enzimas germacreno D sintasas
5 relacionados filogenéticamente (**Fig. 5**), tuvo poco efecto sobre la actividad de la variante polipeptídica CsTPS-VC2 (SEQ ID NO:10), ya que esta mutación no alteró la producción de germacreno D (**Fig. 7E y 8E, Tablas 5 y 6**).

REIVINDICACIONES

1. Un polipéptido caracterizado porque presenta actividad sesquiterpeno sintasa y su secuencia comprende al menos un 90% de identidad con SEQ ID NO:14.
5
2. Un polipéptido según la reivindicación 1, donde la actividad sesquiterpeno sintasa es al menos una seleccionada del grupo que consiste en: actividad δ -cadineno sintasa, actividad elemol sintasa y actividad germacreno D sintasa.
- 10 3. Un polipéptido según una cualquiera de las reivindicaciones 1 ó 2, en el que su secuencia comprende al menos un 90% de identidad con SEQ ID NO:2, o con la SEQ ID NO:2 modificada con al menos una sustitución en un aminoácido que altera la actividad sesquiterpeno sintasa y/o especificidad de producto, y donde dicha sustitución se selecciona entre al menos una del grupo compuesto por:
15 a. una sustitución de la cisteína en posición 455 por glicina (C⁴⁵⁵G), y
 b. una triple sustitución de la serina en posición 416 por treonina, de la glicina en posición 417 por treonina, y de la serina en posición 418 por tirosina (S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y).
- 20 4. Un polipéptido según la reivindicación 3, donde la SEQ ID NO:2 modificada comprende además otra sustitución en un aminoácido de su secuencia, y que:
 a. cuando la SEQ ID NO:2 comprende la sustitución C⁴⁵⁵G, es una sustitución de la serina en posición 418 por histidina (S⁴¹⁸H), y
 b. cuando la SEQ ID NO:2 comprende la triple sustitución S⁴¹⁶T/G⁴¹⁷T/S⁴¹⁸Y, es una
25 sustitución de la serina en posición 422 por alanina (S⁴²²A).
5. Un polipéptido según una cualquiera de las reivindicaciones 1 a 4 que comprende una secuencia seleccionada del grupo compuesto por: SEQ ID NO:2, SEQ ID NO:16 y SEQ ID NO:18.
30
6. Un polipéptido según la reivindicación 5, en el que:
 a. cuando su secuencia comprende SEQ ID NO:2, presenta actividad δ -cadineno sintasa,
 b. cuando su secuencia comprende SEQ ID NO:16, presenta actividad elemol sintasa, o
35 c. cuando su secuencia comprende SEQ ID NO:18, presenta actividad germacreno D sintasa.

7. Un polipéptido según una cualquiera de las reivindicaciones 1 a 5, que es una sesquiterpeno sintasa seleccionada del grupo compuesto por: la δ -cadineno sintasa CsTPS3 que consiste en SEQ ID NO:2, la elemol sintasa CsTPS3-VB2 que consiste en SEQ ID NO:4, la elemol sintasa CsTPS3-VB5 que consiste en SEQ ID NO:6, la germacreno D sintasa CsTPS3-VC1 que consiste en SEQ ID NO:8 y la germacreno D sintasa CsTPS3-VC2 que consiste en SEQ ID NO:10 .

10

8. Un ácido nucleico caracterizado porque comprende una secuencia que codifica un polipéptido definido en una cualquiera de las reivindicaciones 1 a 7.

9. Un ácido nucleico según la reivindicación 8, que cuando codifica un polipéptido comprendiendo SEQ ID NO:2, consiste en un polinucleótido que comprende SEQ ID NO:1.

15

10. Un ácido nucleico según la reivindicación 8 que, cuando codifica un polipéptido que comprende la secuencia SEQ ID NO:2 modificada, consiste en un polinucleótido que comprende la secuencia SEQ ID NO:1 con al menos una mutación seleccionada del grupo compuesto por:

20

- a. una transversión T→G en la posición +1363 con respecto al codón de inicio (ATG) de SEQ ID NO:1; y
- b. un grupo de sustituciones de los nucleótidos T→A/G→A/AGC→TAT en las posiciones +1246/1249/1252-1254 con respecto al codón de inicio (ATG) de SEQ ID NO:1.

25

11. Un ácido nucleico según la reivindicación 10 que consiste en un polinucleótido en el que, adicionalmente a la mutación en el polinucleótido de secuencia SEQ ID NO:1, comprende además otra mutación en SEQ ID NO:1 que se selecciona entre:

30

- a. una transversión/transición AG→CA en las posiciones +1252/1253 con respecto al codón de inicio, cuando SEQ ID NO:1 comprende la transversión T→G; y
- b. una transversión T→A en la posición +1264 con respecto al codón de inicio de SEQ ID NO:1, cuando SEQ ID NO:1 comprende el grupo de sustituciones T→A/G→A/AGC→TAT.

35

12. Un ácido nucleico según la reivindicación 8 que comprende una secuencia polinucleotídica seleccionada del grupo que consiste en: SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 y SEQ ID NO:9.

13. Una construcción génica caracterizada porque comprende al menos un ácido nucleico definido en una cualquiera de las reivindicaciones 8 a 12.

5 14. Un vector de expresión caracterizado porque comprende una construcción génica definida en la reivindicación 13.

15. Una cepa bacteriana caracterizada porque comprende un vector de expresión definido en la reivindicación 14.

10 16. Un método para obtener un polipéptido con actividad sesquiterpeno sintasa, caracterizado porque comprende cultivar una célula bacteriana transformada con un ácido nucleico definido en una cualquiera de las reivindicaciones 8 a 12, o con un vector recombinante que lo comprende, en condiciones adecuadas para favorecer o inducir la expresión de dicho ácido nucleico.

15

17. Uso de al menos un polipéptido definido en una cualquiera de las reivindicaciones 1 a 7 para la síntesis sesquiterpenos.

18. Uso según la reivindicación 17 caracterizado porque dicha síntesis se lleva a cabo *in vitro*.

20

19. Uso según una cualquiera de las reivindicaciones 17 ó 18 caracterizado porque el sesquiterpeno sintetizado es al menos uno seleccionado del grupo que consiste en: α -cubebeno, α -copaeno, β -cubebeno, β -elemeno, α -gurjuneno, *E*- β -cariofileno, α -humuleno, germacreno D, biciclogermacreno, α -muuroleno, α -bulneseno, cubebol, δ -cadineno, elemol, germacreno D-4-ol, bulnesol y guaiol, o una combinación cualquiera de dos o más de los anteriores.

25

20. Uso según una cualquiera de las reivindicaciones 17 a 19 caracterizado porque el sesquiterpeno sintetizado es al menos uno seleccionado del grupo que consiste en: δ -cadineno, elemol y germacreno D, o una combinación cualquiera de dos o más de los anteriores.

30

21. Un procedimiento para sintetizar al menos un sesquiterpeno caracterizado porque comprende:

a. incubar el substrato farnesil-pirofosfato con al menos un polipéptido definido en una cualquiera de las reivindicaciones 1 a 6,

35

b. dejar reaccionar hasta obtener un producto volátil que comprende el sesquiterpeno.

(A)

```

1 ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT GCT GGG AAA GAT TCT
1 M S L E V S A S P A K V I Q N A G K D S
61 ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT
21 T R R S A N Y H P S I W G D H F L Q Y T
121 TGT GAC ACC CAG GAA ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA
41 C D T Q E T D D G S N V K H L E L K K E
181 ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT CAA TTG ATT GAT GCA
61 I R R M L K A D N K P S R T L Q L I D A
241 ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG
81 I Q R L G V S Y H F E S E I D E I L G K
301 ATG CAT AAG GCT TCC CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT
101 M H K A S Q D S D L C D N E N D E L Y Y
361 ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT TCC GCT GAT GTG TTC
121 I S L H F R L L R Q N G Y K I S A D V F
421 AAA AAG TTC AAA GAC ACG GAT GGG AAC TTT AAA ACA TCT CTT GCG AAA GAT GTT CGA GGA
141 K K F K D T D G N F K T S L A K D V R G
481 ATG TTA AGC TTG TAT GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA
161 M L S L Y E A T H L G V H E E D I L D E
541 GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT CAA ATC AGG TCT CCA
181 A L A F T S H L E S I A T H Q I R S P
601 CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT CAG CCT ATC CAC AGG GGC TTA CAA AGG CTT
201 L V E Q V K H A L V Q P I H R G L Q R L
661 GAG GCA AGA CAG TAC ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA
221 E A R Q Y I P I Y I Y Q E E S P H N E A L L
721 ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CTT CAC CAG AAG GAA CTC GGT GAT
241 T F A K L D F N K L Q K L H Q K E L G D
781 ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA
261 I S R W K E L D F A H K L P F I R D R
841 GTT GCA GAG TGC TAC TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA
281 V A E C Y F W I L G V Y F E P Q Y S F A
901 AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT GAT ATC TAT GAT GTG
301 R R I L T K V I S M T S V I D D V
961 TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT
321 Y G K I E E L E L F T S A I E R W D I S
1021 GCC ATA GAT CAA CTT CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT
341 A I D Q L P E Y M K L C Y R A L L D V F
1081 AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC CTC TAT TAT GCA AAA
361 S E A E K D L A P Q G K S Y R L Y Y A K
1141 GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT
381 E A M K N M V K N Y F Y E A K W C L Q N
1201 TAT GTA CCT ACA GTG GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA TCT GGC TCC CCA ATG
401 Y V P T V D E Y M T V A L V T S G S P M
1261 TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA GAA TCT TTT GAG TGG
421 L S T T S F V G M G D I V T K E S F E W
1321 TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT TCT TCT ATA GTT TGC CGA CTC ATG GAT GAC
441 L F S N P R F I R A S S I V C R L M D D
1381 ATA GTG TCA CAC AAG TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC
461 I V S H K F E Q S R G H V A S S V E C Y
1441 ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT CGG AAA CAA GTT TCA
481 M K Q H G A T E E E A C N E F R K Q V S
1501 AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT
501 N A W K D I N E D C L R P T V V P M P L
1561 CTG ATG CGA ATT CTC AAT CTT ACA CGC GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC
521 L M R I L N L T R V I D V I Y K Y E D G
1621 TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG TTT ATT AAT CCT GTG
541 Y T H S A V V L K D F V A S L F I N P V
1681 CCG ATA TGT GTA TGA
561 P I C V *

```

FIGURA 1

(B)

```

1 ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT GCT GGG AAA GAT TCT
1 M S L E V S A S P A K V I Q N A G K D S
61 ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT
21 T R R S A N Y H P S I W G D H F L Q Y T
121 TGT GAC ACC CAG GAA ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA
41 C D T Q E T D D G S N V K H L E L K K E
181 ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT CAA TTG ATT GAT GCA
61 I R R M L K A D N K P S R T L Q L I D A
241 ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG
81 I Q R L G V S Y H F E S E I D E I L G K
301 ATG CAT AAG GCT TCC CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT
101 M H K A S Q D S D L C D N E N D E L Y Y
361 ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT TCC GCT GAT GTG TTC
121 I S L S H F R L L R Q N G Y K I S A D V F
421 AAA AAG TTC AAA GAC ACG GAT GAG AAC TTT AAA ACA TCT CTT GCG AAA GAT DTT CGA
141 K K F K D T D G N F K T S L A K D V R G
481 ATG TTA AGC TTG TAT GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA
161 M L R S L Y E A T H L G V H E E D I L D E
541 GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT CAA ATC AGG TCT CCA
181 A L A F T T S H L E S I A T H Q I R S P
601 CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT CAG CCT ATC CAC AGG GGC TTA CAA AGG CTT
201 L V E V K H A L V Q P I H R G L
661 GAG GCA AGA CAG TAC ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA
221 E A R Q Y I P I Y Q E E S P H N E A L L
721 ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CTT CAC CAG AAG GAA CTC GGT GAT
241 T F A K L D F N K L Q K L H Q K E L G D
781 ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA
261 I S R W W K E L D F A H K L P F I R D R
841 GTT GCA GAG TGC TAC TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA
281 V A E C Y F W I L G V Y F E P Q Y S F A
901 AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT GAT ATC TAT GAT GTG
301 R R I L T K V I S M T S V I D D I Y D V
961 TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT
321 Y G K I E E L E L F T S A I E R W D I S
1021 GCC ATA GAT CAA CTT CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT
341 A I D Q L P E Y M K L C Y R A L L D V F
1081 AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC CTC TAT TAT GCA AAA
361 S E A E K D L A P Q G K S Y R L Y Y A K
1141 GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT
381 E A M K N M V K N Y F Y E A K W C L Q N
1201 TAT GTA CCT ACA GTG GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA TCT GGC TCC CCA ATG
401 Y V P T V D E Y M T V A L V T S G S P M
1261 TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA GAA TCT TTT GAG TGG
421 L S T T S F V G M G D I V T K E S F E W
1321 TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT TCT TCT ATA GTT GGC CGA CTC ATG GAT GAC
441 L F S N P R F I R A S S I V G R L M D D
1381 ATA GTG TCA CAC AAG TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC
461 I V S H K F E Q S R G H V A S S V E C Y
1441 ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT CGG AAA CAA GTT TCA
481 M K Q H G A T E E E A C N E F R K Q V S
1501 AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT
501 N A W K D I N E D C L R P T V V P M P L
1561 CTG ATG CGA ATT CTC AAT CTT ACA GCG GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC
521 L M R I L N L T R V I D V I Y K Y E D G
1621 TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG TTT ATT AAT CCT GTG
541 Y T H S A V V L K D F V A S L F I N P V
1681 CCG ATA TGT GTA TGA
561 P I C V *

```

FIGURA 1

(C)

```

1 ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT GCT GGG AAA GAT TCT
1 M S L E V S A S P A K V I Q N A G K D S
61 ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT
21 T R R S A N Y H P S I W G D H F L Q Y T
121 TGT GAC ACC CAG GAA ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA
41 C D T Q E T D D G S N V K H L E L K K E
181 ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT CAA TTG ATT GAT GCA
61 I R R M L K A D N K P S R T L Q L I D A
241 ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG
81 I Q R L G V S Y H F E S E I D E I L G K
301 ATG CAT AAG GCT TCC CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT
101 M H K A S Q D S D L C D N E N D E L Y Y
361 ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT TCC GCT GAT GTG TTC
121 I S L S H F R L L R Q N G Y K I S A D V F
421 AAA AAG TTC AAA GAC ACG GAT GAG AAC TTT AAA ACA TCT CTT GCG AAA GAT TT CGA
141 K K F K D T D G N F K T S L A K D V R G
481 ATG TTA AGC TTG TAT GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA
161 M L R L Y E A T H L G V H E E D I L D E
541 GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT CAA ATC AGG TCT CCA
181 A L A F T T S H L E S I A T H Q I R S P
601 CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT CAG CCT ATC CAC AGG GGC TTA CAA AGG CTT
201 L V E C Y F W I L G V Q P I H R G L
661 GAG GCA AGA CAG TAC ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA
221 E A R Q Y I P I Y Q E E S P H N E A L L
721 ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CTT CAC CAG AAG GAA CTC GGT GAT
241 T F A K L D F N K L Q K L H Q K E L G D
781 ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA
261 I S R W W K E L D F A H K L P F I R D R
841 GTT GCA GAG TGC TAC TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA
281 V A E C Y F W I L G V Y F E P Q Y S F A
901 AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT GAT ATC TAT GAT GTG
301 R R I L T K V I S M T S V I D D I Y D V
961 TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT
321 Y G K I E E L E L F T S A I E R W D I S
1021 GCC ATA GAT CAA CTT CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT
341 A I D Q L P E Y M K L C Y R A L L D V F
1081 AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC CTC TAT TAT GCA AAA
361 S E A E K D L A P Q G K S Y R L Y Y A K
1141 GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT
381 E A M K N M V K N Y F Y E A K W C L Q N
1201 TAT GTA CCT ACA GTG GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA TCT GGC CAC CCA ATG
401 Y V P T V D E Y M T V A L V T S G H P M
1261 TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA GAA TCT TTT GAG TGG
421 L S T T S F V G M G D I V T K E S F E W
1321 TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT TCT TCT ATA GTT GGC CGA CTC ATG GAT GAC
441 L F S N P R F I R A S S I V G R L M D D
1381 ATA GTG TCA CAC AAG TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC
461 I V S H K F E Q S R G H V A S S V E C Y
1441 ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT CGG AAA CAA GTT TCA
481 M K Q H G A T E E E A C N E F R K Q V S
1501 AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT
501 N A W K D I N E D C L R P T V V P M P L
1561 CTG ATG CGA ATT CTC AAT CTT ACA GCG GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC
521 L M R I L N L T R V I D V I Y K Y E D G
1621 TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG TTT ATT AAT CCT GTG
541 Y T H S A V V L K D F V A S L F I N P V
1681 CCG ATA TGT GTA TGA
561 P I C V *

```

FIGURA 1

(D)

```

1 ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT GCT GGG AAA GAT TCT
1 M S L E V S A S P A K V I Q N A G K D S
61 ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT
21 T R R S A N Y H P S I W G D H F L Q Y T
121 TGT GAC ACC CAG GAA ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA
41 C D T Q E T D D G S N V K H L E L K K E
181 ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT CAA TTG ATT GAT GCA
61 I R R M L K A D N K P S R T L Q L I D A
241 ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG
81 I Q R L G V S Y H F E S E I D E I L G K
301 ATG CAT AAG GCT TCC CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT
101 M H K A S Q D S D L C D N E N D E L Y Y
361 ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT TCC GCT GAT GTG TTC
121 I S L S L F R L L R Q N G Y K I S A D V F
421 AAA AAG TTC AAA GAC ACG GAT GAG AAC TTT AAA ACA TCT CTT GCG AAA GAT TT CGA
141 K K F K D T D G N F K T S L A K D V R G
481 ATG TTA AGC TTG TAT GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA
161 M L R L Y E A T H L G V H E E D I L D E
541 GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT CAA ATC AGG TCT CCA
181 A L A F T T S H L E S I A T H Q I R S P
601 CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT CAG CCT ATC CAC AGG GGC TTA CAA AGG CTT
201 L V E V K H A L V Q P I H R G L
661 GAG GCA AGA CAG TAC ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA
221 E A R Q Y I P I Y Q E E S P H N E A L L
721 ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CTT CAC CAG AAG GAA CTC GGT GAT
241 T F A K L D F N K L Q K L H Q K E L G D
781 ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA
261 I S R W W K E L D F A H K L P F I R D R
841 GTT GCA GAG TGC TAC TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA
281 V A E C Y F W I L G V Y F E P Q Y F A
901 AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT GAT ATC TAT GAT GTG
301 R R I L T K V I S M T S V I D D I Y D V
961 TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT
321 Y G K I E E L E L F T S A I E R W D I S
1021 GCC ATA GAT CAA CTT CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT
341 A I D Q L P E Y M K L C Y R A L L D V F
1081 AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC CTC TAT TAT GCA AAA
361 S E A E K D L A P Q G K S Y R L Y Y A K
1141 GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT
381 E A M K N M V K N Y F Y E A K W C L Q N
1201 TAT GTA CCT ACA GTG GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA ACT ACC TAT CCA ATG
401 Y V P T V D E Y M T V A L V T T T Y P M
1261 TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA GAA TCT TTT GAG TGG
421 L S T T S F V G M G D I V T K E S F E W
1321 TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT TCT TCT ATA GTT TGC CGA CTC ATG GAT GAC
441 L F S N P R F I R A S S I V C R L M D D
1381 ATA GTG TCA CAC AAG TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC
461 I V S H K F E Q S R G H V A S S V E C Y
1441 ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT CGG AAA CAA GTT TCA
481 M K Q H G A T E E E A C N E F R K Q V S
1501 AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT
501 N A W K D I N E D C L R P T V V P M P L
1561 CTG ATG CGA ATT CTC AAT CTT ACA CGC GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC
521 L M R I L N L T R V I D V I Y K Y E D G
1621 TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG TTT ATT AAT CCT GTG
541 Y T H S A V V L K D F V A S L F I N P V
1681 CCG ATA TGT GTA TGA
561 P I C V *

```

FIGURA 1

(E)

1 ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT GCT GGG AAA GAT TCT
 1 M S L E V S A S P A K V I Q N A G K D S
 61 ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT
 21 T R R S A N Y H P S I W G D H F L Q Y T
 121 TGT GAC ACC CAG GAA ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA
 41 C D T Q E T D D G S N V K H L E L K K E
 181 ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT CAA TTG ATT GAT GCA
 61 I R R M L K A D N K P S R T L Q L I D A
 241 ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG
 81 I Q R L G V S Y H F E S E I D E I L G K
 301 ATG CAT AAG GCT TCC CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT
 101 M H K A S Q D S D L C D N E N D E L Y Y
 361 ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT TCC GCT GAT GTG TTC
 121 I S L H F R L L R Q N G Y K I S A D V F
 421 AAA AAG TTC AAA GAC ACG GAT GGC AAC TTT AAA ACA TCT CTT GCG AAA GAT TTT CGA
 141 K K F K D T D G N F K T S L A K D V R G
 481 ATG TTA AGC TTG TAT GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA
 161 M L R L Y E A T H L G V H E E D I L D E
 541 GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT CAA ATC AGG TCT CCA
 181 A L A F T T S H L E S I A T H Q I R S P
 601 CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT CAG CCT ATC CAC AGG GGC TTA CAA AGG CTT
 201 L V E K H A L V Q P I H R G L
 661 GAG GCA AGA CAG TAC ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA
 221 E A R Q Y I P I Y Q E E S P H N E A L L
 721 ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CTT CAC CAG AAG GAA CTC GGT GAT
 241 T F A K L D F N K L Q K L H Q K E L G D
 781 ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA
 261 I S R W W K E L D F A H K L P F I R D R
 841 GTT GCA GAG TGC TAC TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA
 281 V A E C Y F W I L G V Y F E P Q Y S F A
 901 AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT GAT ATC TAT GAT GTG
 301 R R I L T K V I S M T S V I D D I Y D V
 961 TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT
 321 Y G K I E E L E L F T S A I E R W D I S
 1021 GCC ATA GAT CAA CTT CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT
 341 A I D Q L P E Y M K L C Y R A L L D V F
 1081 AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC CTC TAT TAT GCA AAA
 361 S E A E K D L A P Q G K S Y R L Y Y A K
 1141 GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT
 381 E A M K N M V K N Y F Y E A K W C L Q N
 1201 TAT GTA CCT ACA GTG GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA **ACT ACC TAT** CCA ATG
 401 Y V P T V D E Y M T V A L V T **T T Y** P M
 1261 TTG **GCA** ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA GAA TCT TTT GAG TGG
 421 L **A** T T S F V G M G D I V T K E S F E W
 1321 TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT TCT TCT ATA GTT TGC CGA CTC ATG GAT GAC
 441 L F S N P R F I R A S S I V C R L M D D
 1381 ATA GTG TCA CAC AAG TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC
 461 I V S H K F E Q S R G H V A S S V E C Y
 1441 ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT CGG AAA CAA GTT TCA
 481 M K Q H G A T E E E A C N E F R K Q V S
 1501 AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT
 501 N A W K D I N E D C L R P T V V P M P L
 1561 CTG ATG CGA ATT CTC AAT CTT ACA GCG GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC
 521 L M R I L N L T R V I D V I Y K Y E D G
 1621 TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG TTT ATT AAT CCT GTG
 541 Y T H S A V V L K D F V A S L F I N P V
 1681 CCG ATA TGT GTA TGA
 561 P I C V *

FIGURA 1

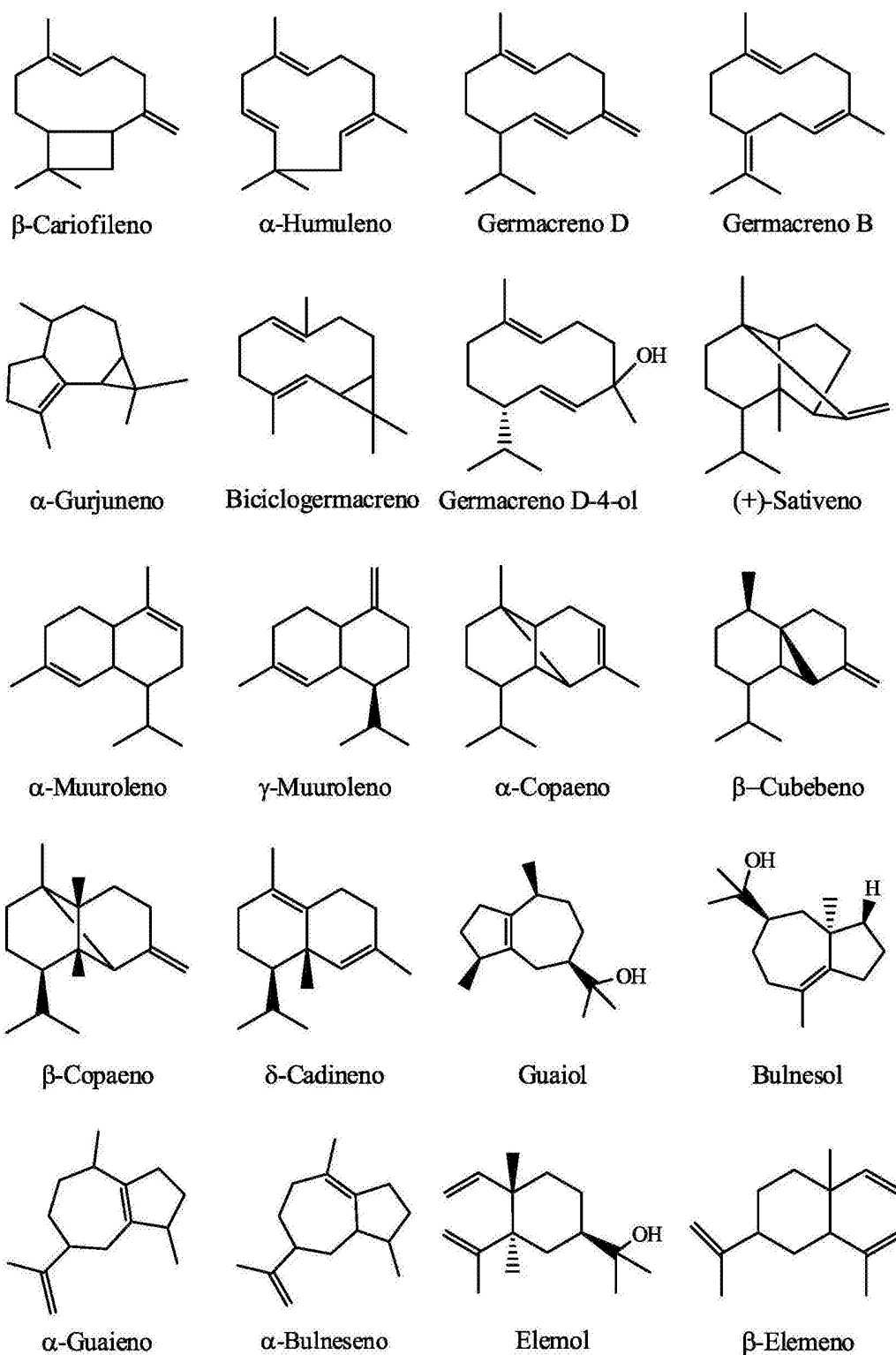


FIGURA 2

ES 2 540 791 B1

	* 20 * 40 * 60	
CsTPS3	MSLEVSASPAKVIQNAGKDSRRSANYHPSIWGDHFLQYTCDTQETDDGSNVKHLELKE	60
CsTPS3_VB2	60
CsTPS3_VB5	60
CsTPS3_VC1	60
CsTPS3_VC2	60
	* 80 * 100 * 120	
CsTPS3	IRRMLKADNKPSTLQQLIDATQRLGVSYHFSEIDEILGKMHKASQSDSLCDNENDELYY	120
CsTPS3_VB2	120
CsTPS3_VB5	120
CsTPS3_VC1	120
CsTPS3_VC2	120
	* 140 * 160 * 180	
CsTPS3	ISLHFRLLRQNGYKISADVFKKEKDTDGNFKTSLAKDVRGMLSLYEATHLGVHEEDILDE	180
CsTPS3_VB2	180
CsTPS3_VB5	180
CsTPS3_VC1	180
CsTPS3_VC2	180
	* 200 * 220 * 240	
CsTPS3	ALAFTTSHLESIAHQIRSPLEVEQVKHALVQPIHRGLQRLEARQYIPIYQEEsphneall	240
CsTPS3_VB2	240
CsTPS3_VB5	240
CsTPS3_VC1	240
CsTPS3_VC2	240
	* 260 * 280 * 300	
CsTPS3	TFAKLDNFNKLQKLHQKELGDISRWWKELDFAHKLPFIRDRVAECYFWILGVYFEPQYSFA	300
CsTPS3_VB2	300
CsTPS3_VB5	300
CsTPS3_VC1	300
CsTPS3_VC2	300
	* 320 * 340 * 360	
CsTPS3	RRILTKVISMTSVIDDIVVYVKIELELFTSAIERWDISAIDQLPEYMKLCYRALLDVF	360
CsTPS3_VB2	360
CsTPS3_VB5	360
CsTPS3_VC1	360
CsTPS3_VC2	360
	* 380 * 400 * 420	
CsTPS3	SEAEKDLAPQGKSYRLYYAKEAMKNMVKNFYEAkWCLQNYVPTVDEYMTVALVTSGSPM	420
CsTPS3_VB2	420
CsTPS3_VB5	420
CsTPS3_VC1	420
CsTPS3_VC2	420
	* 440 * 460 * 480	
CsTPS3	LSTTSFVGMGDIIVTKESFEWLFNSNPRFIRASSIVCRLMDDIVSHKFEQSRGHVASSVECY	480
CsTPS3_VB2	480
CsTPS3_VB5	480
CsTPS3_VC1	480
CsTPS3_VC2	480
	* 500 * 520 * 540	
CsTPS3	MKQHGATEEBEACNEFRKQVSNWAKDINEDCLRPTVVPMLLMRILNLTRVLDVIYKYEDG	540
CsTPS3_VB2	540
CsTPS3_VB5	540
CsTPS3_VC1	540
CsTPS3_VC2	540
	* 560	
CsTPS3	YTHSAVVLKDFVASLFINPVICV	564
CsTPS3_VB2	564
CsTPS3_VB5	564
CsTPS3_VC1	564
CsTPS3_VC2	564

FIGURA 3

ES 2 540 791 B1

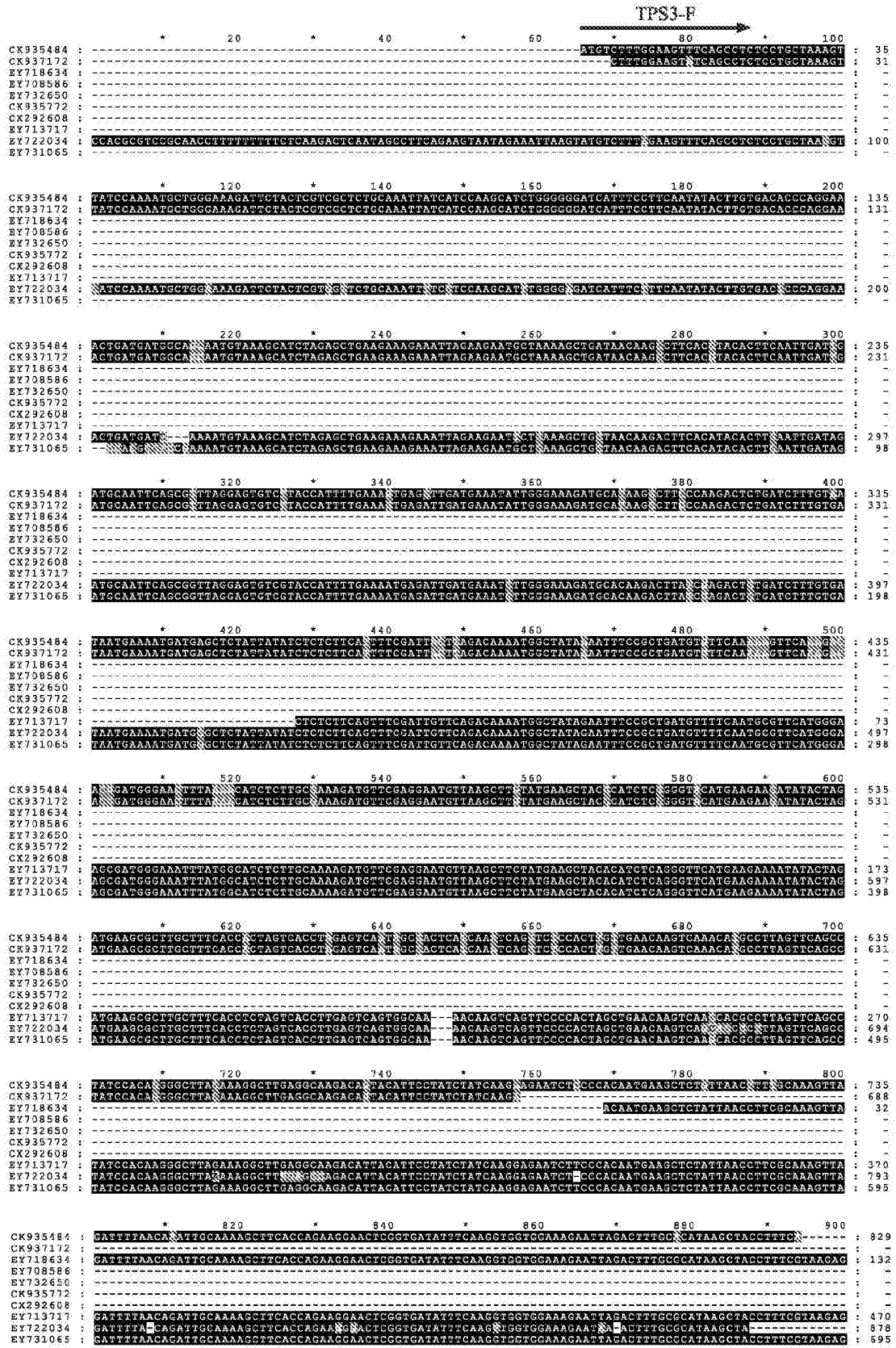


FIGURA 4

ES 2 540 791 B1

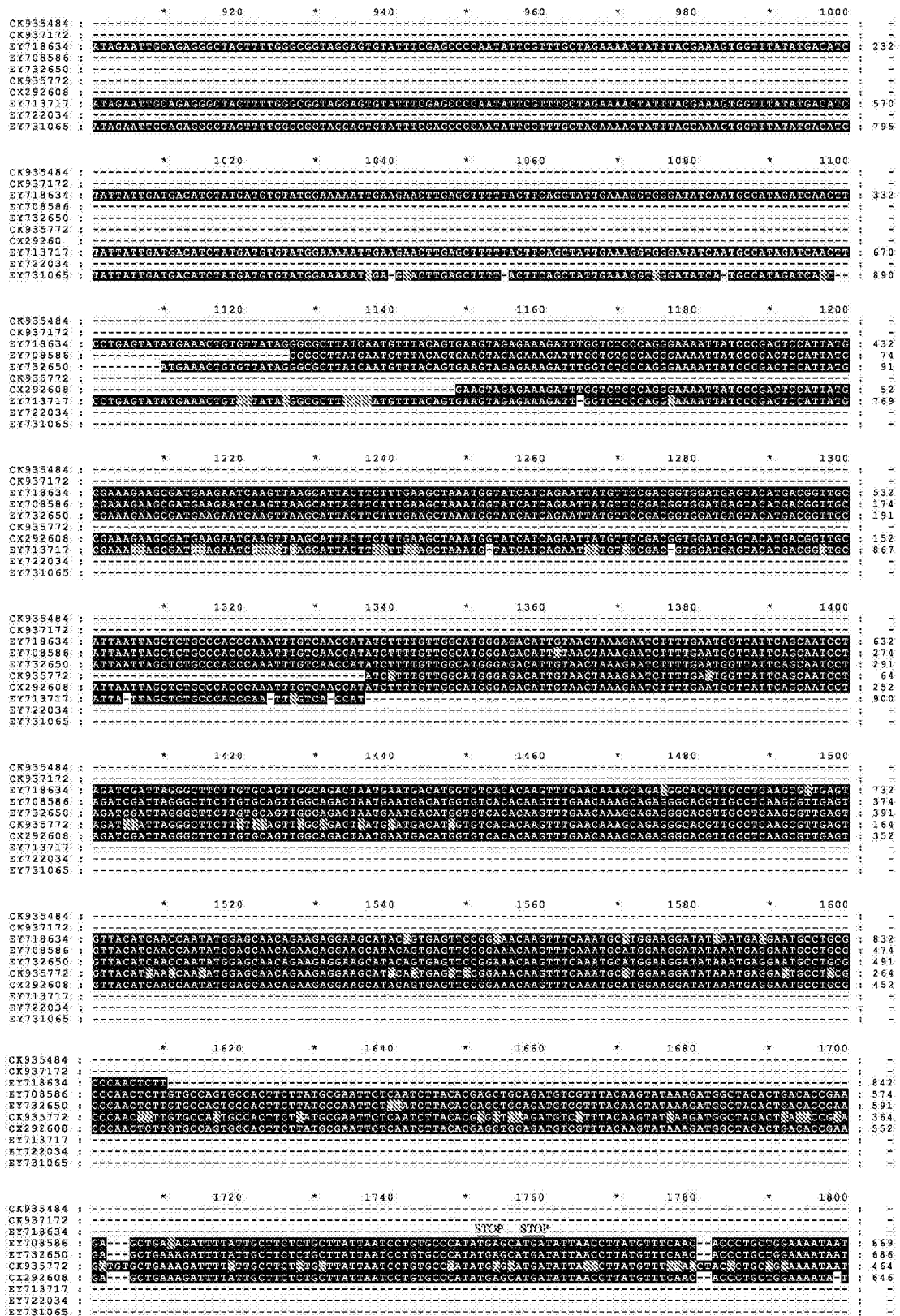


FIGURA 4. (Cont.)

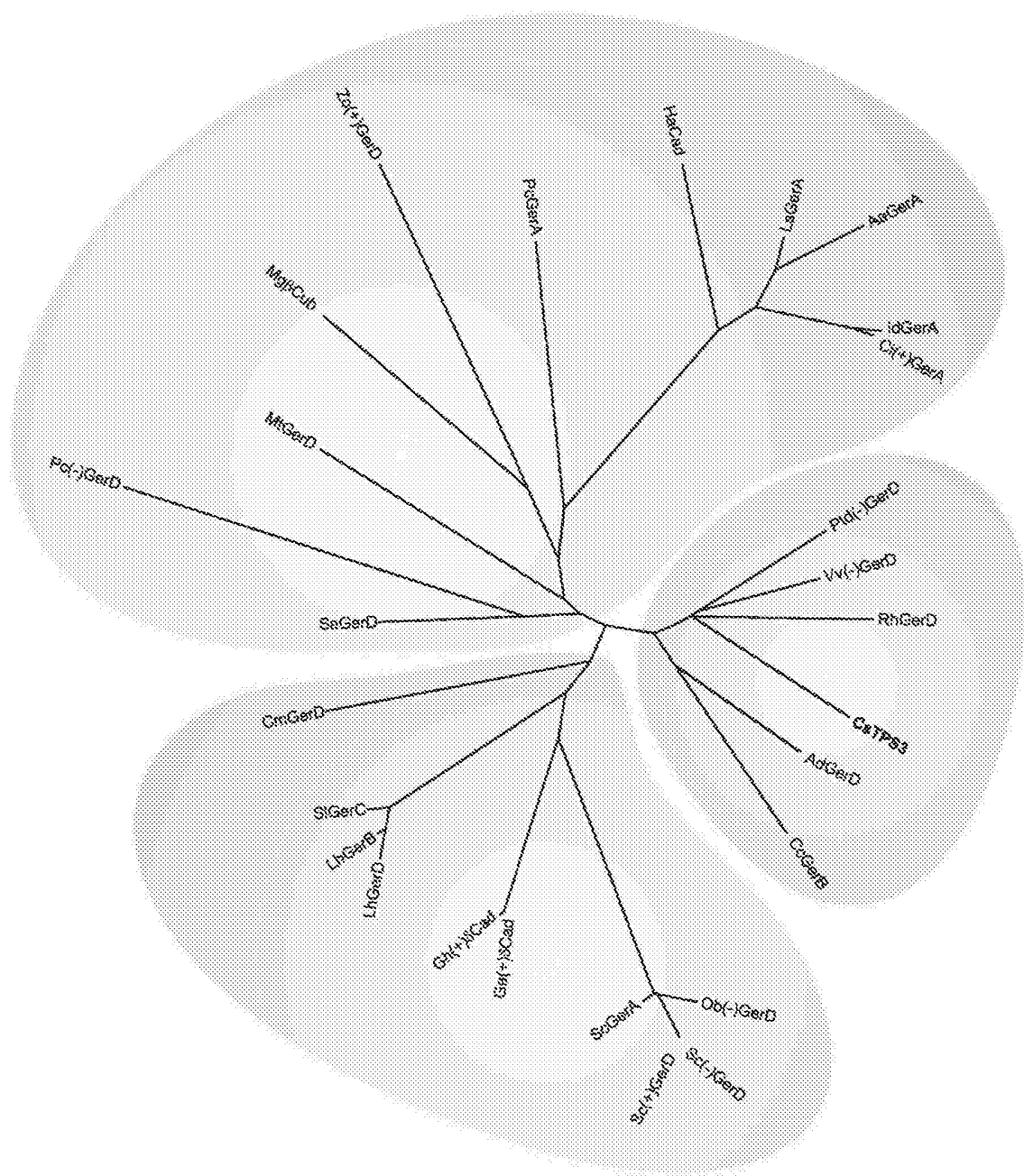


FIGURA 6

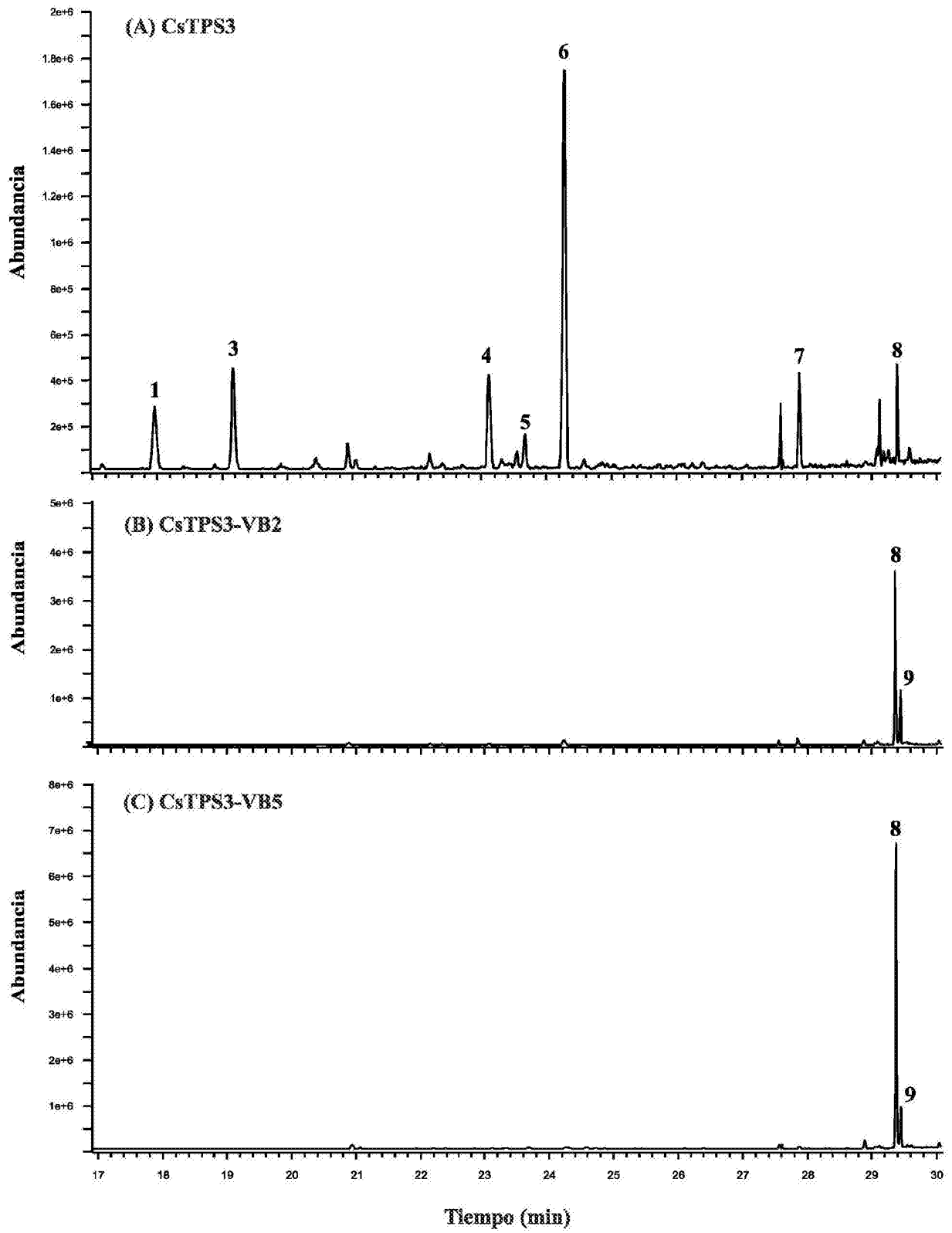


FIGURA 7

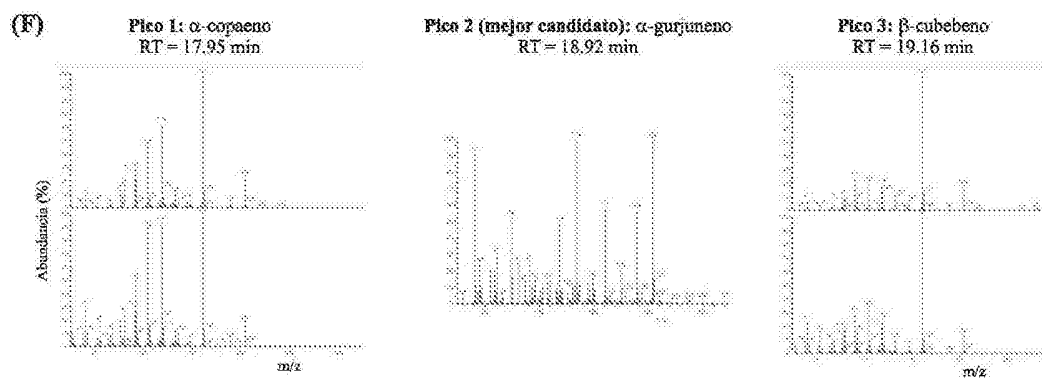
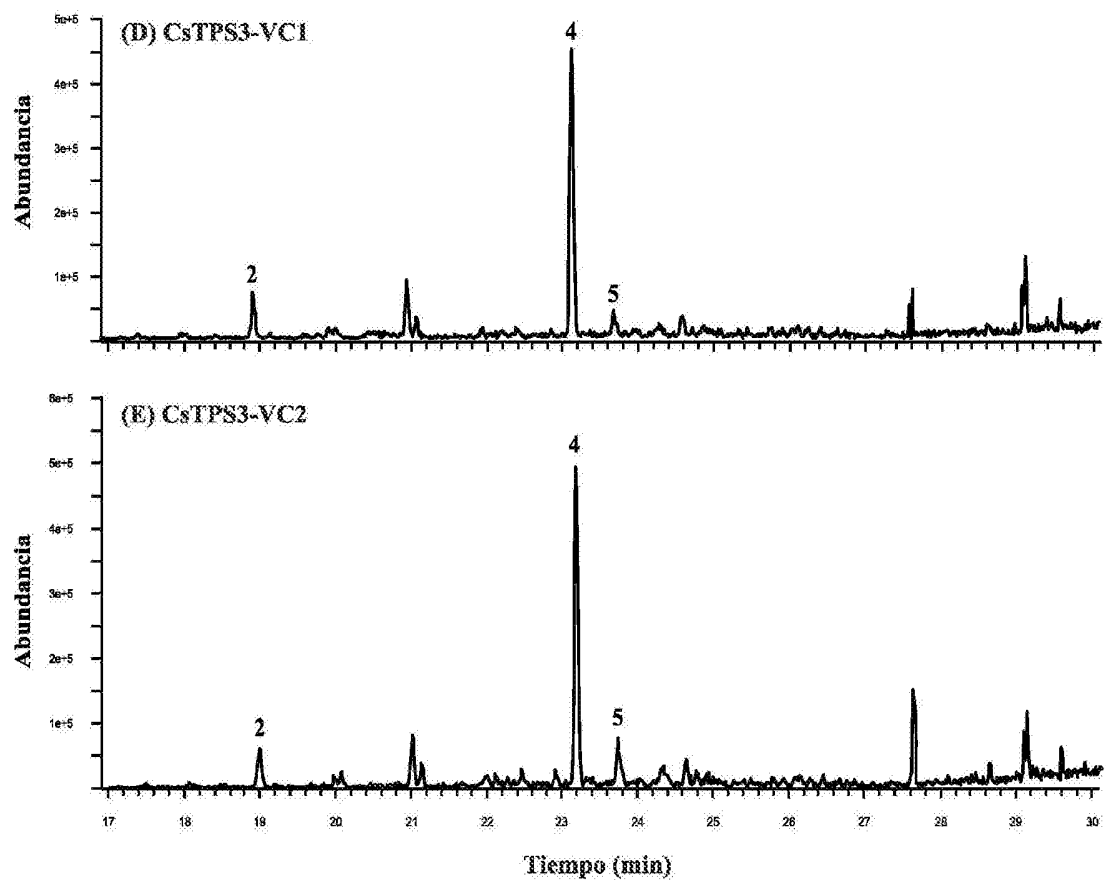


FIGURA 7. (Cont.)

(F) Continuación

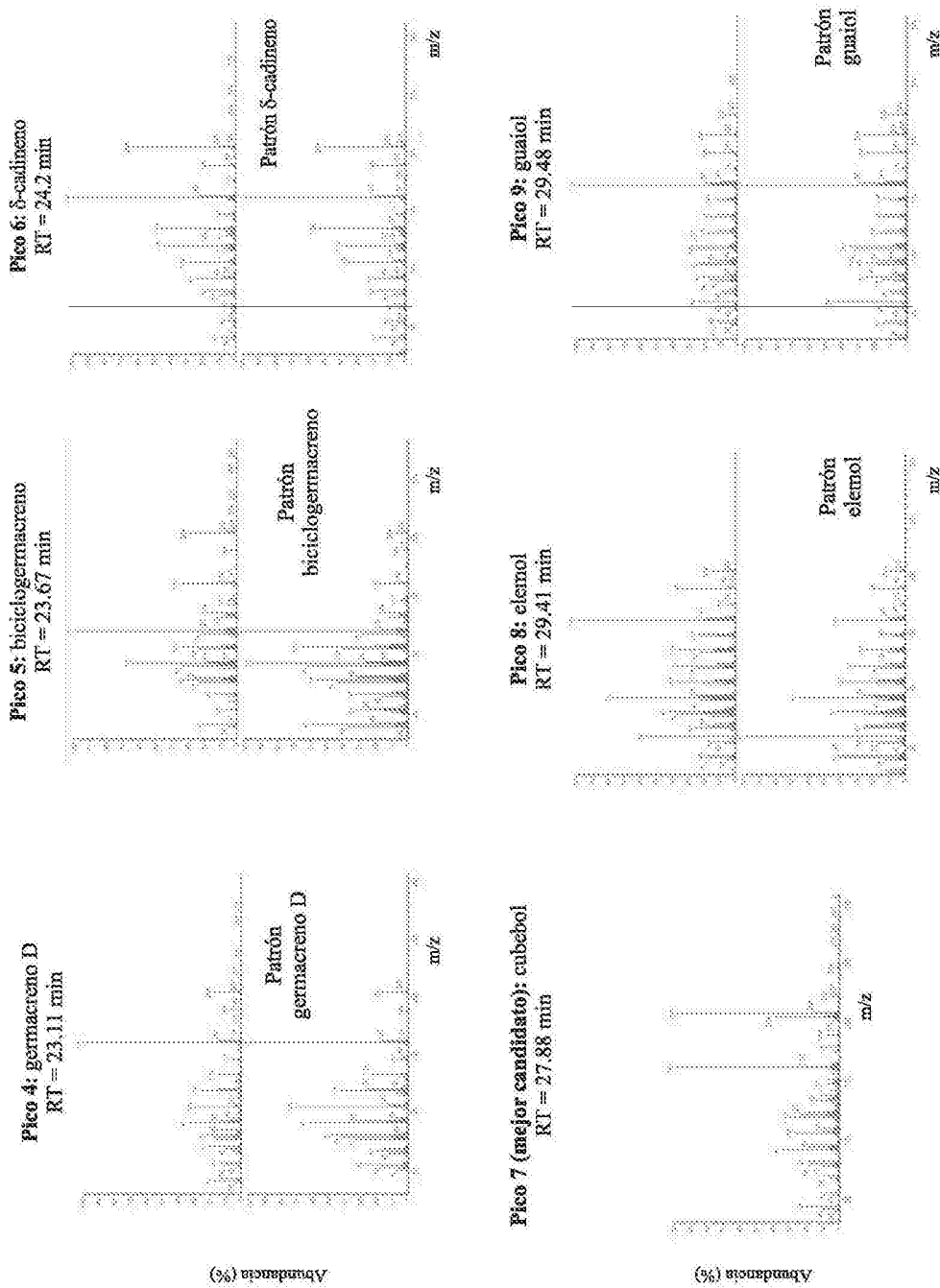


FIGURA 7. (Cont.)

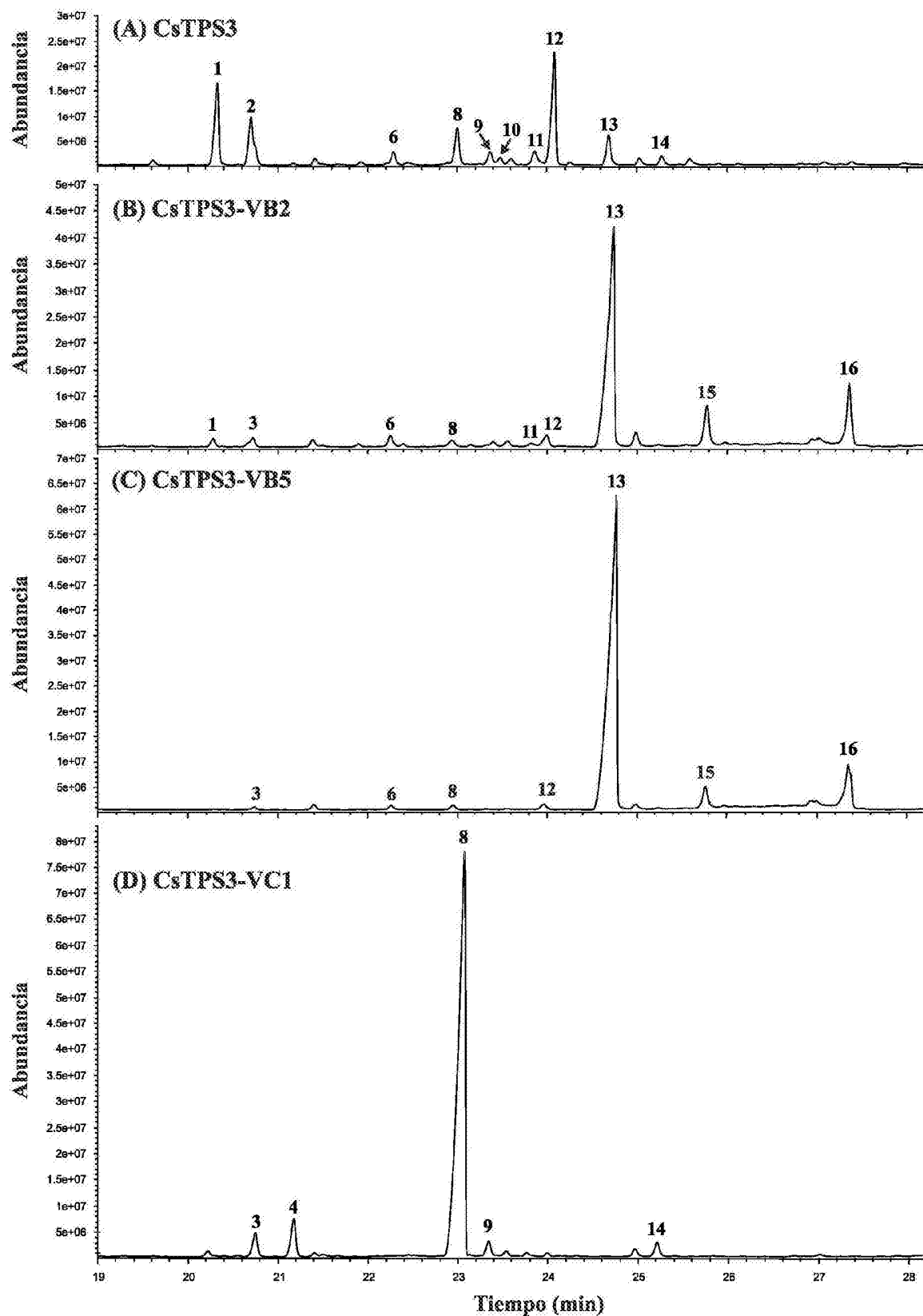


FIGURA 8.

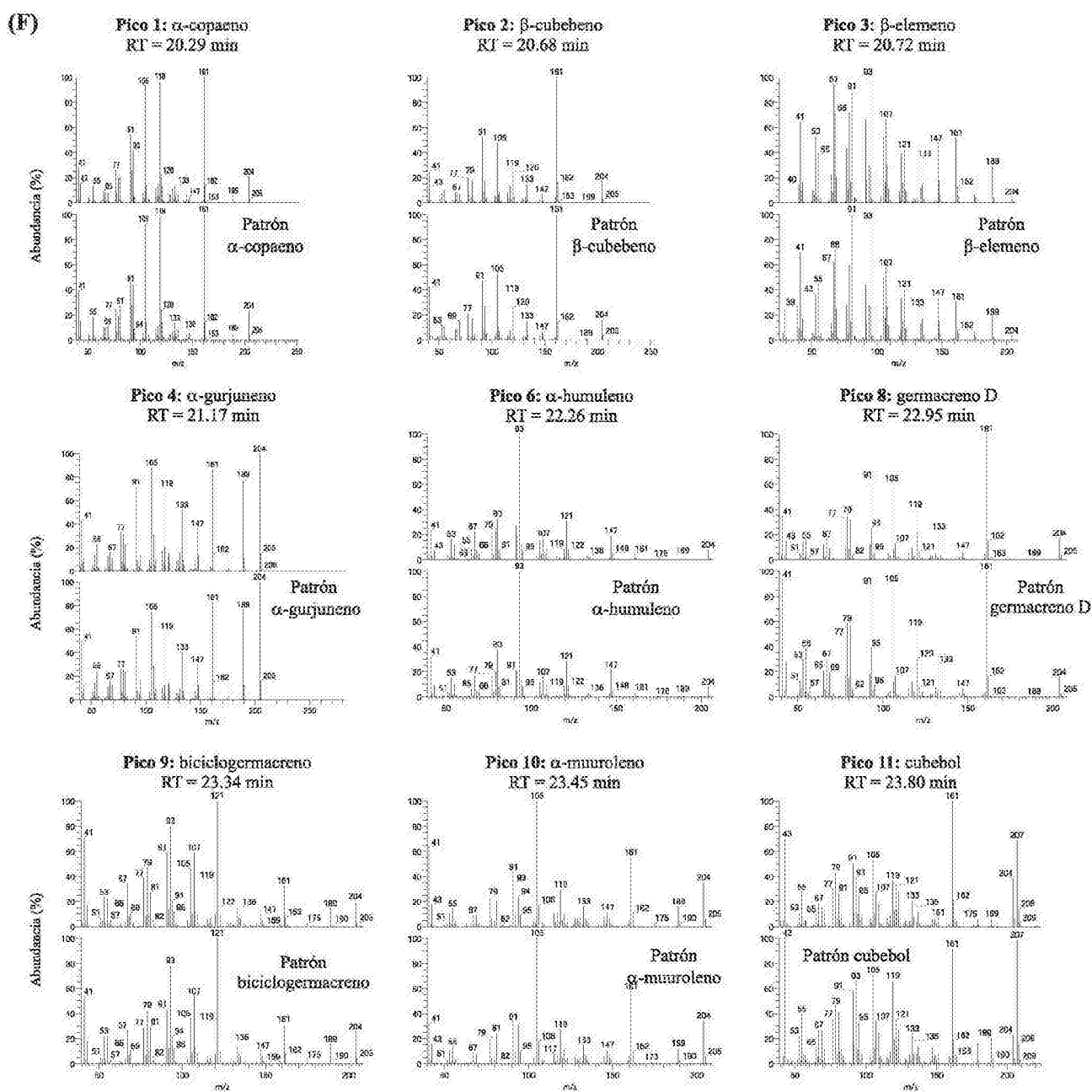
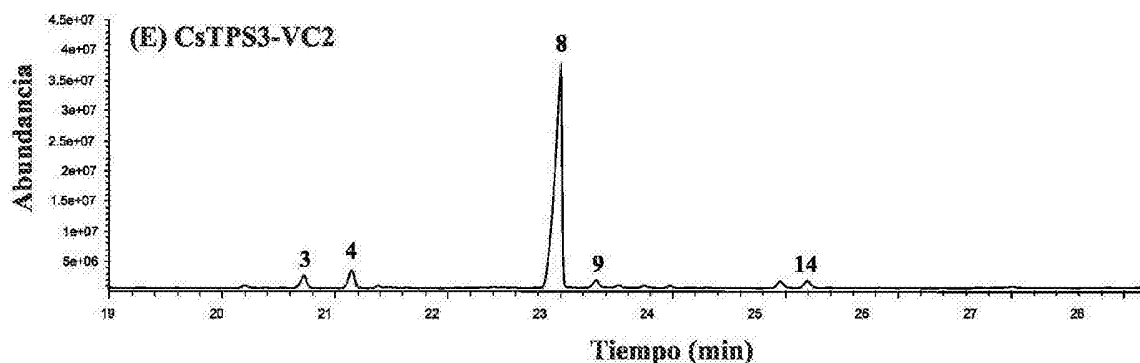


FIGURA 8. (Cont.)

(F) Continuación

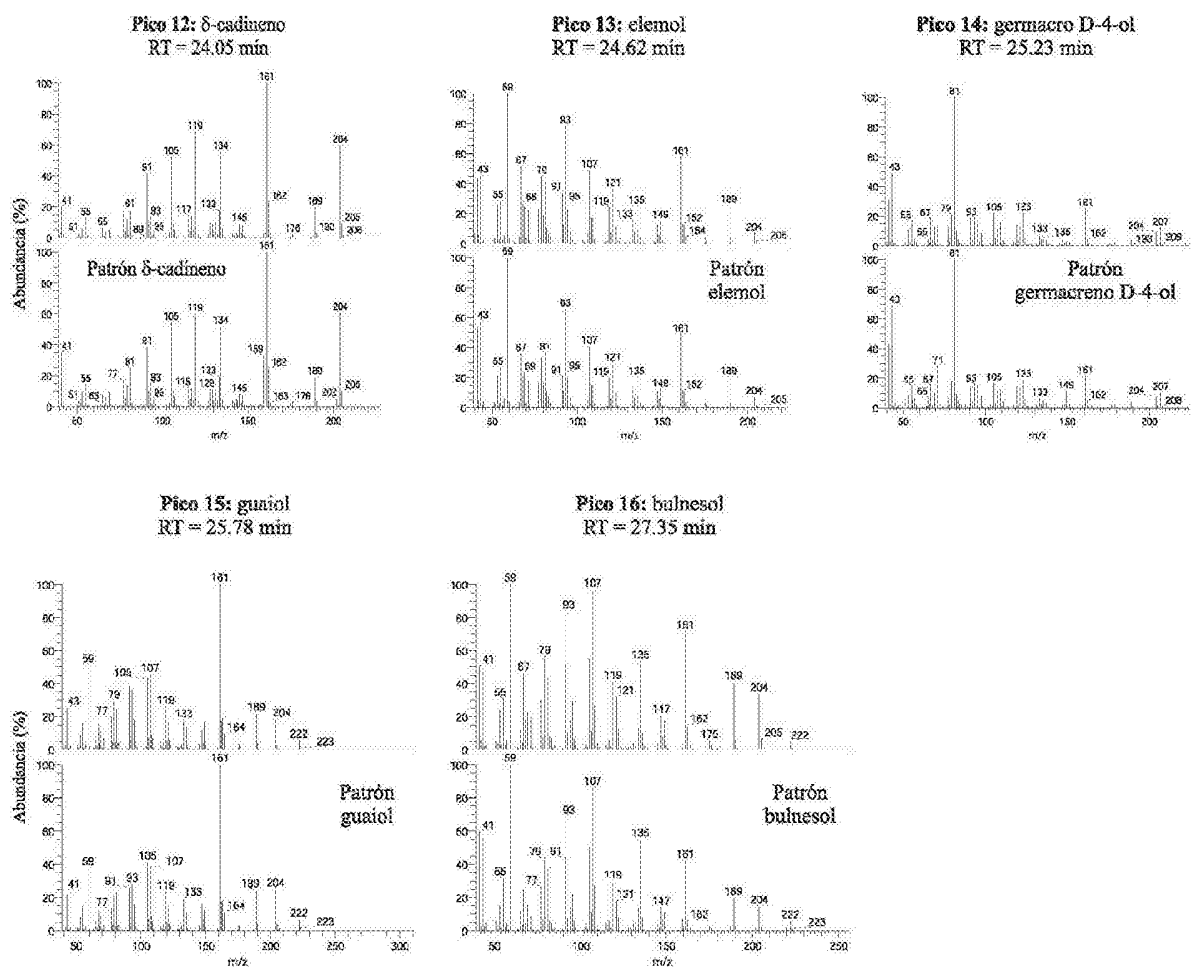


FIGURA 8. (Cont.)

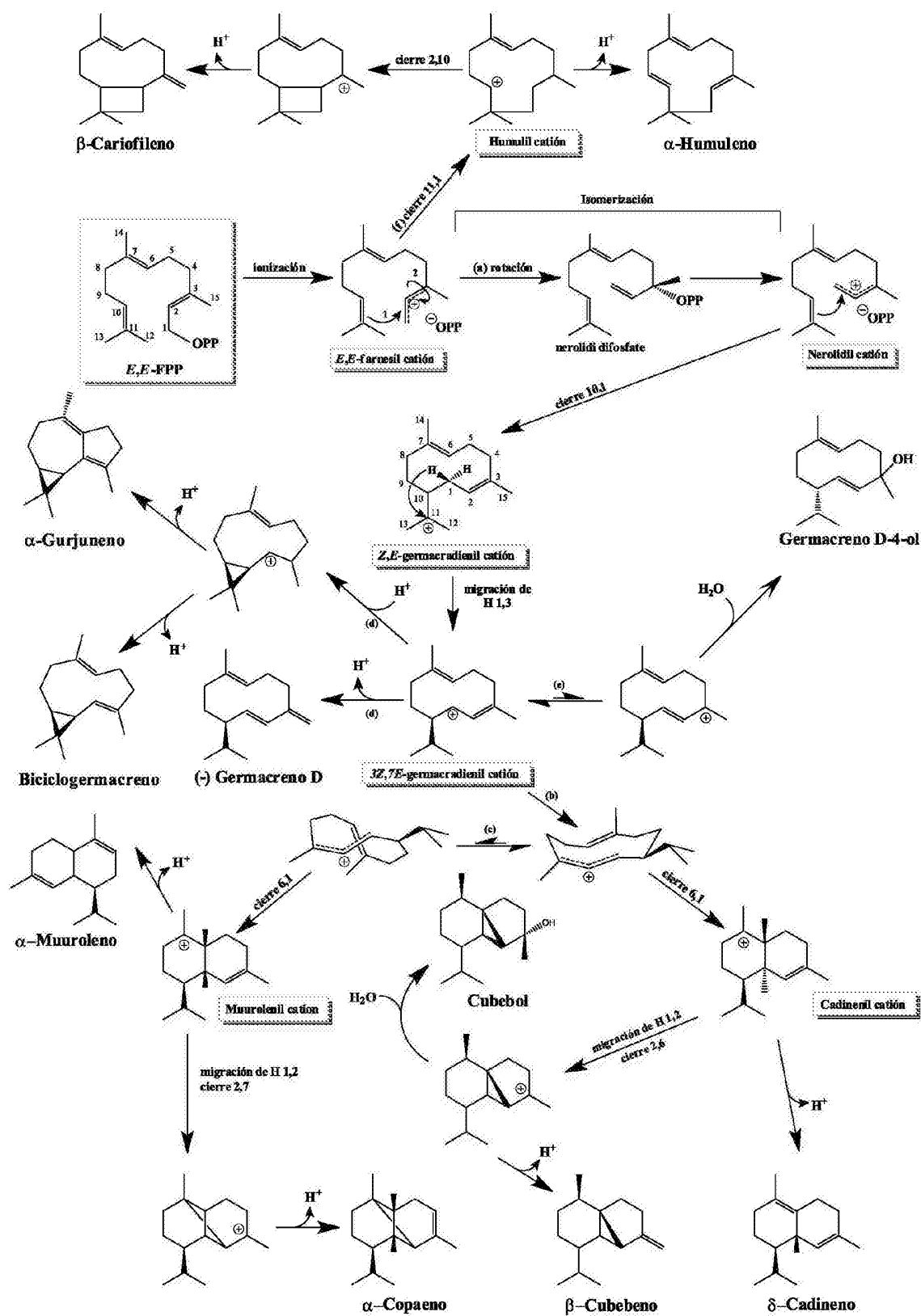


FIGURA 9

(B)

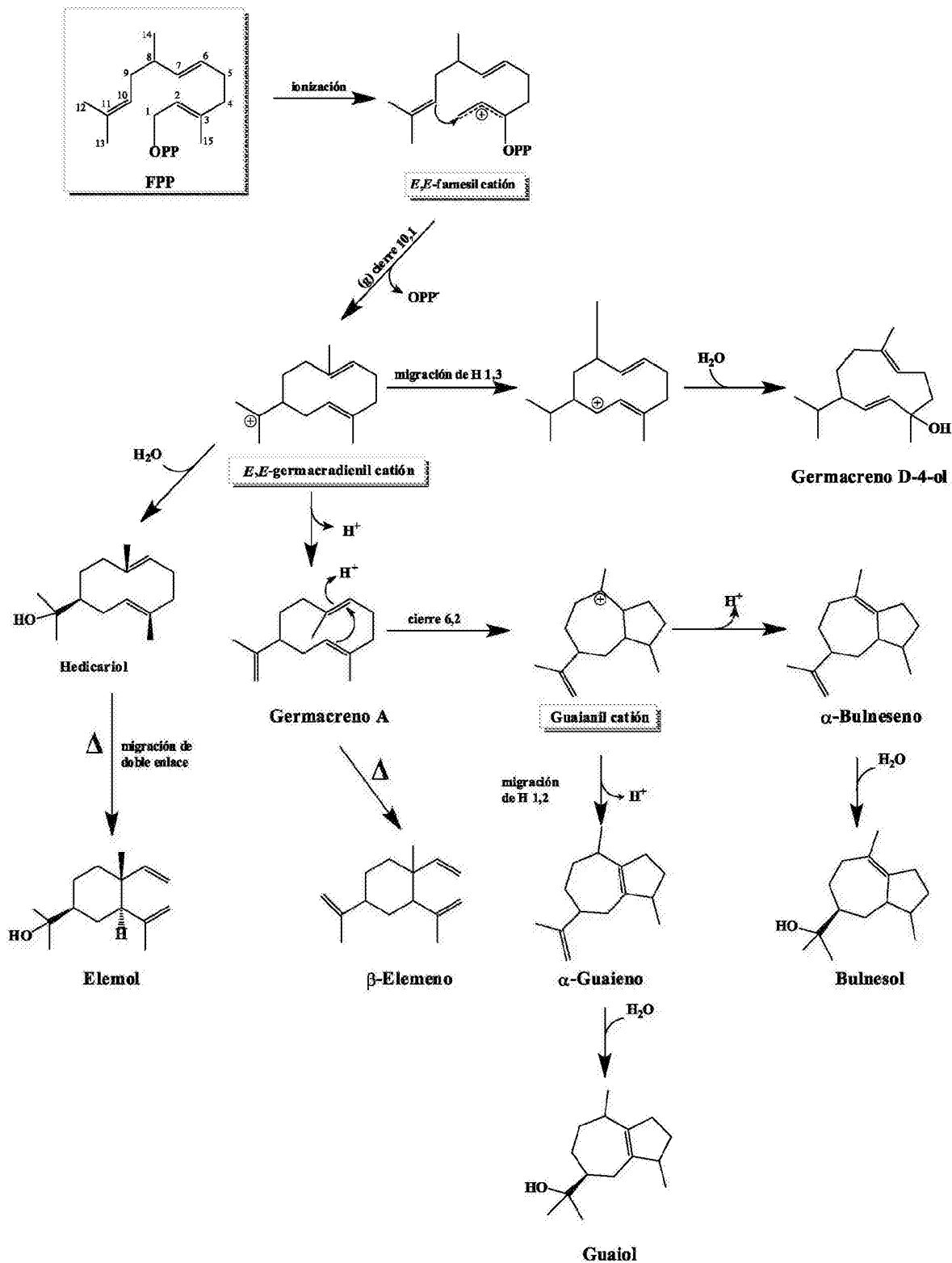


FIGURA 9. (Cont.)

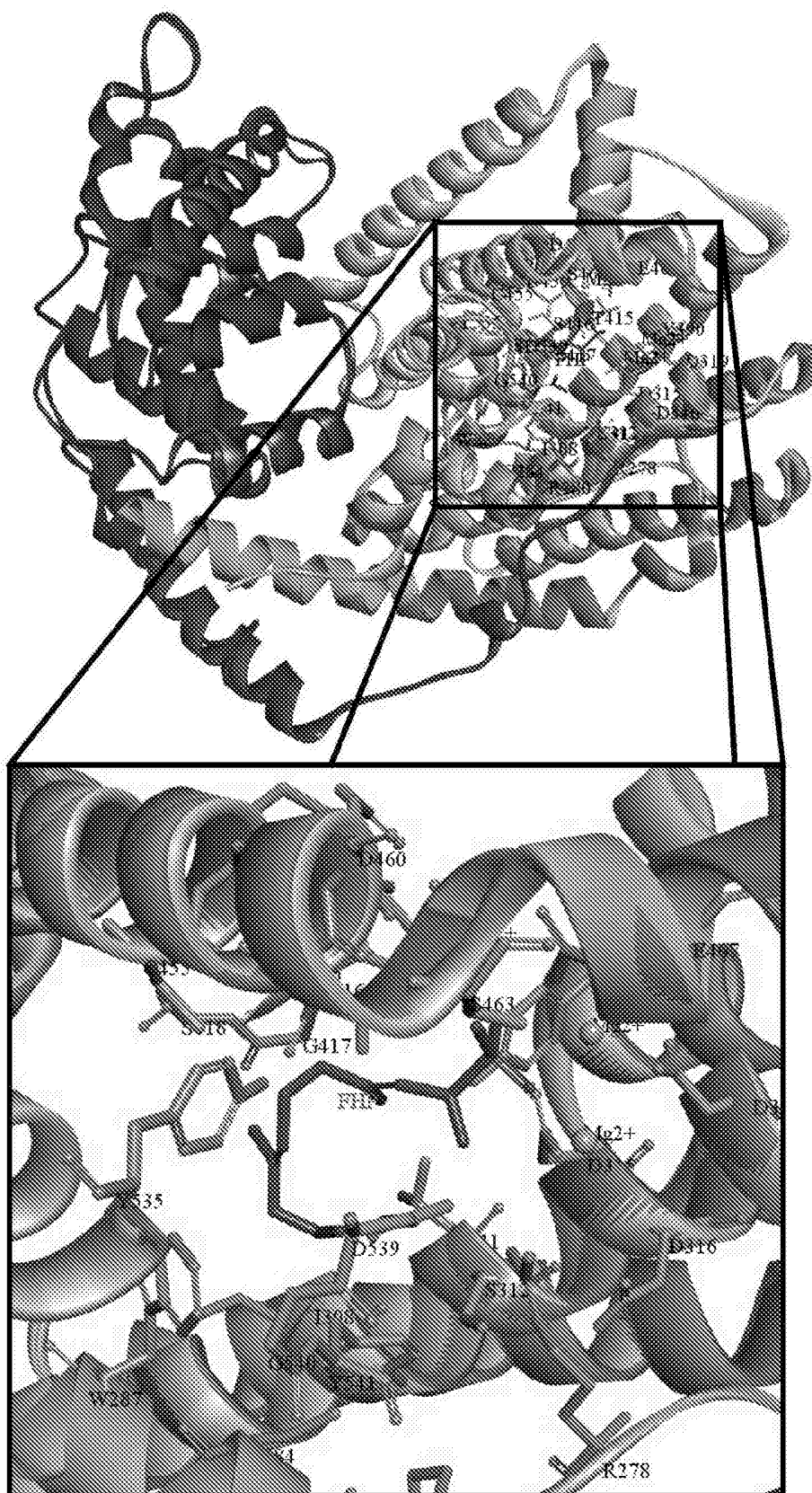


FIGURA 10

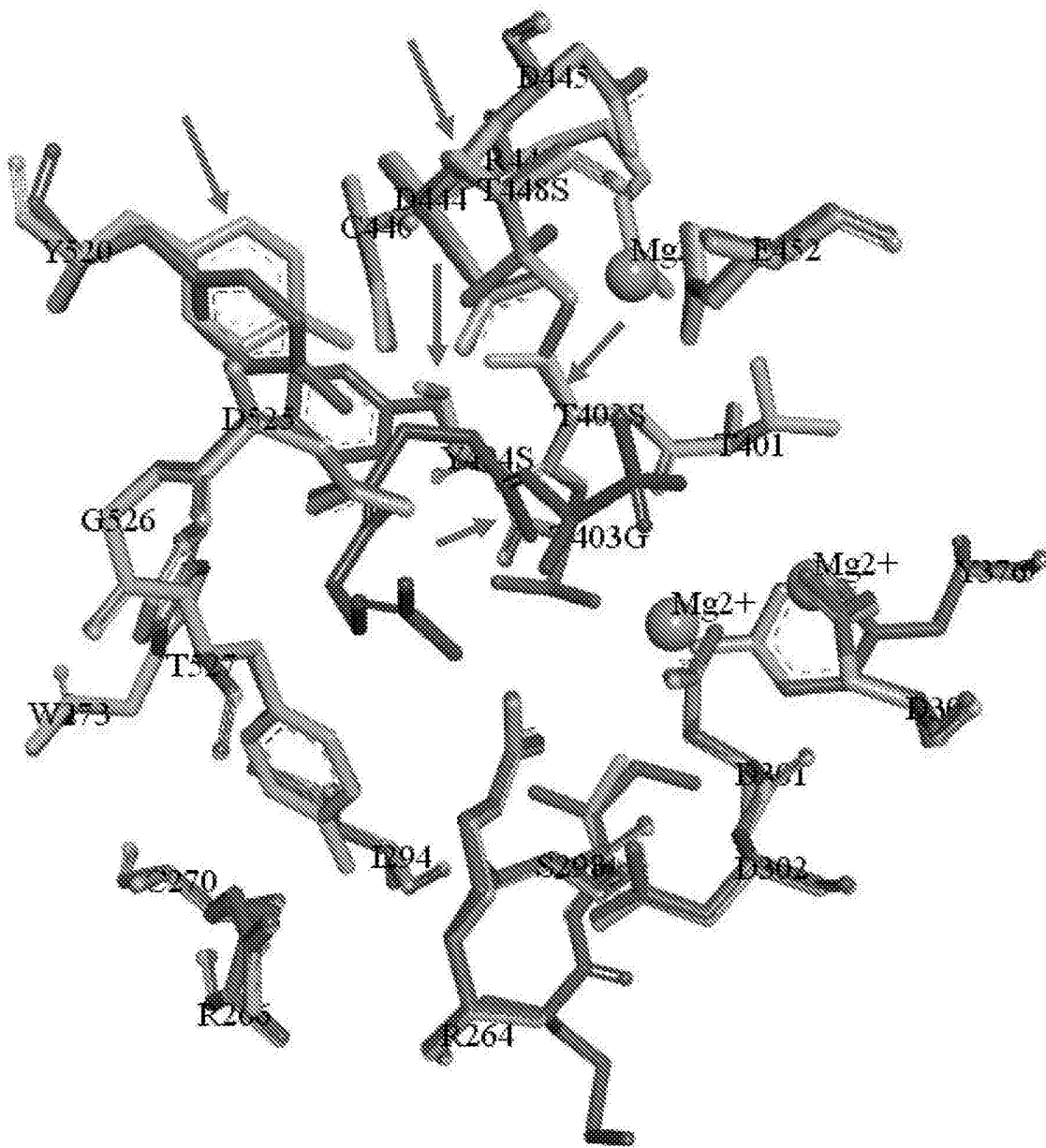


FIGURA 11

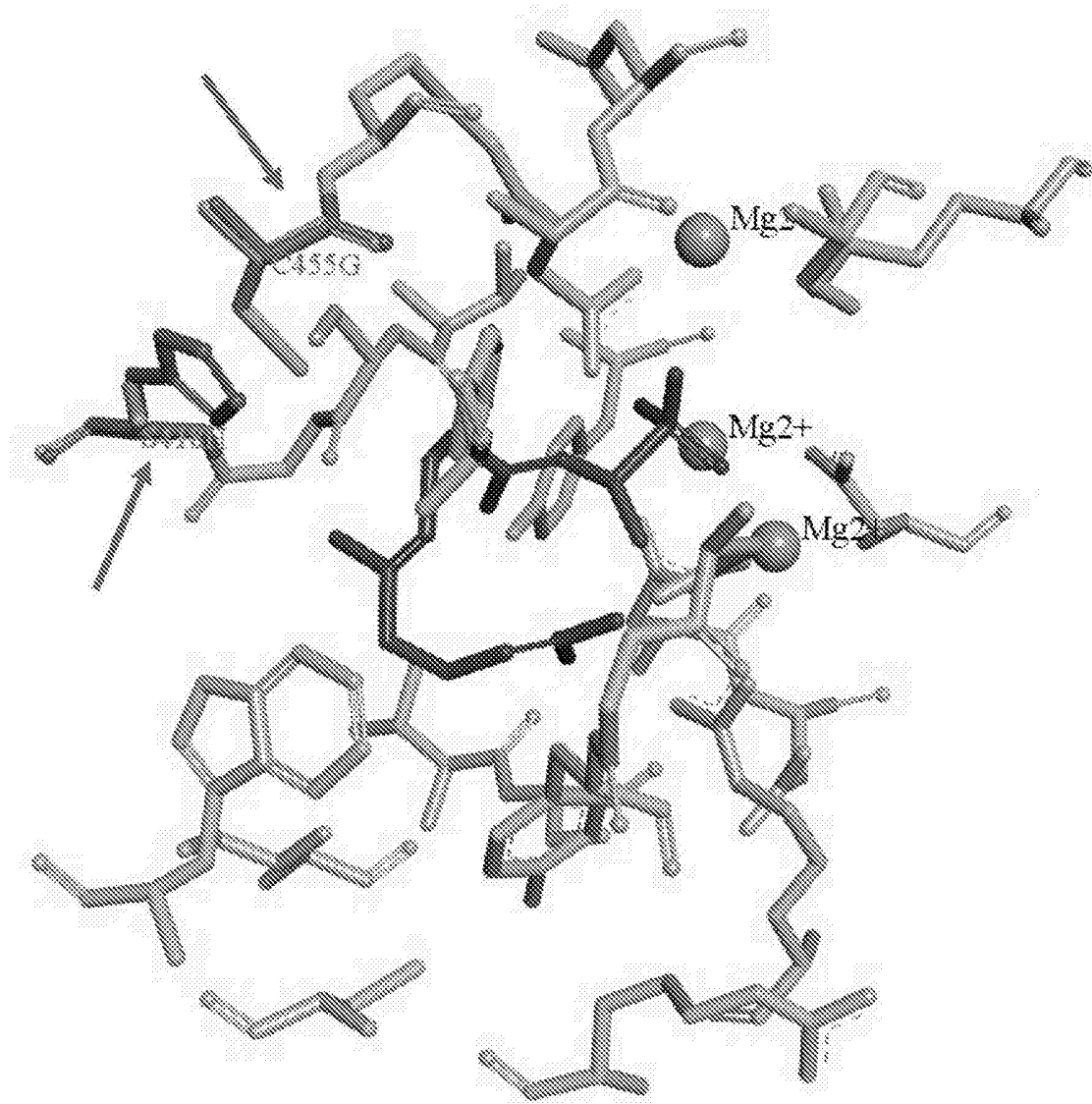


FIGURA 12

ES 2 540 791 B1

LISTA DE SECUENCIAS

<110> Consejo Superior de Investigaciones Científicas
 Instituto Valenciano de Investigaciones Agrarias
 <120> NUEVAS SESQUITERPENO SINTASAS AISLADAS DE LA PIEL DE NARANJA
 <130> 2013-0265
 <160> 18
 <170> BISSAP, 1.2
 <210> 1
 <211> 1695
 <212> DNA
 <213> Citrus sinensis var. Washington Navel
 <220>
 <221> source
 <222> 1..1695
 <223> /organism="Citrus sinensis var. Washington Navel"
 /note="Secuencia polinucleotídica de cDNA correspondiente a la
 región codificante del gen Cstps3 de naranjo"
 /mol_type="unassigned DNA"

<400> 1
 atgtctttg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct 60
 actcgtcgcct ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120
 tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180
 attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240
 attcagcggt taggagtgtc ttaccatctt gaaagtgaga ttgatgaaat attgggaaag 300
 atgcataagg ctcccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat 360
 atctctcttc attttcgatt acttagacia aatggctata aaatttccgc tgatgtgttc 420
 aaaaagttca aagacacgga tgggaacttt aaaacatctc ttgcgaaaga tgttcgagga 480
 atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540
 gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat cagggtctcca 600
 cttgttgaac aagtcaaaca tgccttagtt cagcctatcc acaggggctt acaaaggctt 660
 gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720
 acttttgcaa agtttagatt taacaaattg caaaagcttc accagaagga actcggtgat 780
 atttcaaggt ggtggaaaga attagacttt gcacataagc tacctttcat aagagataga 840
 gttgcagagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca 900
 agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960
 tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagagggtg ggatatcagt 1020
 gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt 1080
 agtgaagcag agaaggattt ggcccccaa ggaaaatcat accgcctcta ttatgcaaaa 1140
 gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200
 tatgtacctc cagtggatga gtacatgacg gttgcattag ttacatctgg ctccccaatg 1260
 ttgtcaacca catcctttgt tggcatggga gacattgtaa ctaaagaatc ttttgagtgg 1320

ES 2 540 791 B1

ttattcagca atcctagatt tattagggct tcttctatag tttgccgact catggatgac 1380
 atagtgtcac acaagtttga acaaagcaga gggcacgttg cctcaagcgt tgagtgttac 1440
 atgaaacaac atggagcaac agaagaggaa gcatgcaatg agtttcggaa acaagtttca 1500
 aatgcctgga aggatataaa tgaggactgc ctacgcccaa cggttgtgcc aatgccactt 1560
 ctgatgcaaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc 1620
 tacactcatt ccgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
 ccgatatgtg tatga 1695

<210> 2
 <211> 564
 <212> PRT
 <213> Citrus sinensis var. Washington Navel

<220>
 <223> Secuencia de aminoácidos de la proteína sesquiterpeno sintasa 3 de naranjo (CSTPS3), codificada por el gen Cstps3 (SEQ ID NO:1)

<220>
 <221> SITE
 <222> 22..559
 <223> Terpeno ciclase vegetal, Clase 1; NCBI CDD: cd00684; PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222> join(22..32,466..467,469..472,474..477,538..541)
 <223> Sitio activo (active site lid residues); PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222> join(278,287,308,310..312,315,319,390,456..457,459..460,463,467,535,539,541)
)
 <223> sitio activo: Cavidad de enlace del sustrato (substrate binding pocket); PSSM-Id: 173832

<220>
 <221> SITE
 <222> join(315,319,459,463,467)
 <223> Enlace del sustrato-Mg2+ (substrate-Mg2+ binding site); PSSM-Id:173832

<220>
 <221> SITE
 <222> 314..319
 <223> Motivo DDX2D (aspartate-rich región 1); PSSM-Id:173832

<220>
 <221> SITE
 <222> 457..467
 <223> Motivo NSE/DTE (aspartate-rich región 2); PSSM-d:173832

<400> 2
 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
 2

ES 2 540 791 B1

65	Ile	Gln	Arg	Leu	Gly	Val	Ser	Tyr	His	Phe	Glu	Ser	Glu	Ile	Asp	Glu	80
					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	
					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	Leu	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	Leu	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			
	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					295				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					
	Ala	Val	Val	Leu	Lys	Asp	Phe	Val	Ala	Ser	Leu	Phe	Ile	Asn	Pro	Val	
						550					555					560	
	Pro	Ile	Cys	Val													

<210> 3
 <211> 1695
 <212> DNA
 <213> Artificial sequence
 <220>

ES 2 540 791 B1

```

<221> source
<222> 1..1695
<223> /organism="Artificial Sequence"
      /note="Secuencia polinucleotídica de cDNA Cstps3-VB2"
      /mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1..1695
<223> /note="Versión mutante derivada de la secuencia natural Cstps3
      (SEQ ID NO:1), en la cual se ha substituido el nucleótido de
      timina (t) en la posición 1363 (t1363)."
```

```

<220>
<221> misc_feature
<222> 1363
<223> /note="Substitución t1363g"
```

```

<400> 3
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
attcagcgtt taggagtgtc ttaccattht gaaagtgaga ttgatgaaat attgggaaag      300
atgcataagg cttcccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat      360
atctctcttc attttcgatt acttagacaa aatggctata aaatttccgc tgatgtgttc      420
aaaaagttca 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 acaaaggctt      660
gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta      720
acttttgcaa agttagatth taacaaattg caaagcttc accagaagga actcgggtgat      780
atthcaaggt ggtggaaaga attagactth gcacataagc tacctthcat aagagataga      840
gttgacagag tctactthtg gatattagga gtgtatthcg agccccaata thcatttgca      900
agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg      960
tatggcaaaa ttgaagaact tgagctthtt acttcagcta ttgagaggtg ggatatcagt     1020
gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgtthtt     1080
agtgaagcag agaaggatth ggcccccaa ggaaaatcat accgcctcta ttatgcaaaa     1140
gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat     1200
tatgtaccta cagtggatga gtacatgacg gttgcattag ttacatctgg ctccccaatg     1260
ttgtcaacca catcctthtg tggcatggga gacattgtaa ctaaagaatc thttgagtgg     1320
ttattcagca atcctagatt tattagggtc tcttctatag ttggccgact catggatgac     1380
atagtgtcac acaagthtg acaaagcaga gggcacgttg cctcaagcgt tgagtgttac     1440
atgaaacaac atggagcaac agaagaggaa gcatgcaatg agthtcggaa acaagthtca     1500
aatgcctgga aggatataaa tgaggactgc ctacgccccaa cggttgtgcc aatgccactt     1560
```

ES 2 540 791 B1

ctgatgcgaa ttctcaatct tacacgcggt atagatgtca ttacaagta tgaagatggc 1620
 tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
 ccgatatgtg tatga 1695

<210> 4
 <211> 564
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Secuencia de aminoácidos del polipéptido CstPS3-VB2, codificado por la secuencia nucleotídica artificial Cstps3-VB2 (SEQ ID NO:3)

<220>
 <221> SITE
 <222> 22..559
 <223> Terpeno ciclasa vegetal, Clase 1; NCBI CDD: cd00684; PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222> join(22..32,466..467,469..472,474..477,538..541)
 <223> Sitio activo (active site lid residues); PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222>
 join(278,287,308,310..312,315,319,390,456..457,459..460,463,467,535,539,541)
)
 <223> sitio activo: Cavidad de enlace del sustrato (substrate binding pocket); PSSM-Id: 173832

<220>
 <221> SITE
 <222> join(315,319,459,463,467)
 <223> Enlace del sustrato-Mg2+ (substrate-Mg2+ binding site); PSSM-Id:173832

<220>
 <221> SITE
 <222> 314..319
 <223> Motivo DDX2D (aspartate-rich región 1); PSSM-Id:173832

<220>
 <221> SITE
 <222> 457..467
 <223> Motivo NSE/DTE (aspartate-rich región 2); PSSM-d:173832

<220>
 <221> MUTAGEN
 <222> 455
 <223> Mutación: C455G

<400> 4
 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

ES 2 540 791 B1

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 Leu 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 Leu 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
 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
 290 295 300
 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 Gly 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
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val
 545 550 555 560
 Pro Ile Cys Val

<210> 5
 <211> 1695
 <212> DNA
 <213> Artificial Sequence

 <220>
 <221> source
 <222> 1..1695
 <223> /organism="Artificial Sequence"

ES 2 540 791 B1

/note="Secuencia polinucleotidica de cDNA Cstps3-VB5"
/mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1..1695
<223> /note="Versión mutante derivada de la secuencia natural Cstps3 (SEQ ID NO:1), en la cual se han substituido los nucleótidos de timina (t) en las posiciones 1252 y 1363, y el nucleótido de citosina (c) en la posición 1253"

<220>
<221> misc_feature
<222> 1252
<223> /note="Substitución t1252c"

<220>
<221> misc_feature
<222> 1253
<223> /note="Substitución c1253a"

<220>
<221> misc_feature
<222> 1363
<223> /note="Substitución t1363g"

<400> 5
atgtctttgg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct 60
actcgtcgtc ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120
tgtgacacc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180
attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240
attcagcgtt taggagtgtc ttaccatttt gaaagtgaga ttgatgaaat attgggaaag 300
atgcataagg cttcccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat 360
atctctcttc attttcgatt acttagacaa aatggctata aaatttccgc tgatgtgttc 420
aaaaagttca aagacacgga tgggaacttt aaaacatctc ttgcgaaaga tgttcgagga 480
atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540
gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat caggctctca 600
cttgttgaac aagtcaaaca tgccttagtt cagcctatcc acaggggctt acaaggctt 660
gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720
acttttgcaa agttagattt taacaaattg caaaagcttc accagaagga actcgggtgat 780
atntcaaggt ggtggaaaga attagacttt gcacataagc tacctttcat aagagataga 840
gttgacagag gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca 900
agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960
tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagagggtg ggatatcagt 1020
gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt 1080
agtgaagcag agaaggattt ggcccccaa ggaaatcat accgcctcta ttatgcaaaa 1140
gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200
tatgtacctc cagtggatga gtacatgacg gttgcattag ttacatctgg ccaccaatg 1260
ttgtcaacca catcctttgt tggcatggga gacattgtaa ctaaagaatc ttttgagtgg 1320

ES 2 540 791 B1

ttattcagca atcctagatt tattagggct tcttctatag ttggccgact catggatgac 1380
atagtgtcac acaagtttga acaaagcaga gggcacgttg cctcaagcgt tgagtgttac 1440
atgaaacaac atggagcaac agaagaggaa gcatgcaatg agtttcggaa acaagtttca 1500
aatgcctgga aggatataaa tgaggactgc ctacgcccaa cggttgtgcc aatgccactt 1560
ctgatgcgaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc 1620
tacactcatt ccgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
ccgatatgtg tatga 1695

<210> 6
<211> 564
<212> PRT
<213> Artificial Sequence

<220>
<223> Secuencia de aminoácidos del polipéptido CsTPS3-VB5, codificado por la secuencia nucleotídica artificial Cstps3-VB5 (SEQ ID NO:4)

<220>
<221> SITE
<222> 22..559
<223> Terpeno ciclasa vegetal, Clase 1; NCBI CDD: cd00684; PSSM-Id: 173832

<220>
<221> ACT_SITE
<222> join(22..32,466..467,469..472,474..477,538..541)
<223> Sitio activo (active site lid residues); PSSM-Id: 173832

<220>
<221> ACT_SITE
<222> join(278,287,308,310..312,315,319,390,456..457,459..460,463,467,535,539,541)
<223> sitio activo: Cavidad de enlace del sustrato (substrate binding pocket); PSSM-Id: 173832

<220>
<221> SITE
<222> join(315,319,459,463,467)
<223> Enlace del sustrato-Mg²⁺ (substrate-Mg²⁺ binding site); PSSM-Id:173832

<220>
<221> SITE
<222> 314..319
<223> Motivo DDX2D (aspartate-rich región 1); PSSM-Id:173832

<220>
<221> SITE
<222> 457..467
<223> Motivo NSE/DTE (aspartate-rich región 2); PSSM-d:173832

<220>
<221> MUTAGEN
<222> 416
<223> Substitución: S416H

<220>
<221> MUTAGEN
<222> 455
<223> Mutación: C455G

ES 2 540 791 B1

<400> 6
 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
 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
 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 Leu 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 Leu 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
 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
 290 295 300
 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 Gln 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 His
 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 Gly 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
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val

ES 2 540 791 B1

545 Pro Ile Cys Val 550 555 560

<210> 7
 <211> 1695
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <222> 1..1695
 <223> /organism="Artificial Sequence"
 /note="Secuencia polinucleotídica de cDNA Cstps3-VC1"
 /mol_type="unassigned DNA"

<220>
 <221> misc_feature
 <222> 1..1695
 <223> /note="Versión mutante derivada de la secuencia natural Cstps3
 (SEQ ID NO:1), en la cual se han substituido los nucleótidos
 t1246,g1249, g1250, c1253 y c1254"

<220>
 <221> misc_feature
 <222> 1246
 <223> /note="Substitución t1246a"

<220>
 <221> misc_feature
 <222> 1249
 <223> /note="Substitución g1249a"

<220>
 <221> misc_feature
 <222> 1250
 <223> /note="Substitución g1250c"

<220>
 <221> misc_feature
 <222> 1253
 <223> /note="Substitución c1253a"

<220>
 <221> misc_feature
 <222> 1254
 <223> /note="Substitución c1254t"

<400> 7
 atgtctttgg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct 60
 actcgtcgcct ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120
 tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180
 attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240
 attcagcggt taggagtgtc ttaccattht gaaagtgaga ttgatgaaat attgggaaag 300
 atgcataagg cttccaaga ctctgatcct tgtgataatg aaaatgatga gctctattat 360
 atctctcttc attttcgatt acttagacia aatggctata aaatttccgc tgatgtgttc 420
 aaaaagttca aagacacgga tgggaacttt aaaacatctc ttgcgaaaga tgttcgagga 480
 atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540
 gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat caggtctcca 600
 cttgttgaac aagtcaaca tgccttagtt cagcctatcc acaggggctt acaaaggctt 660

ES 2 540 791 B1

gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720
 acttttgcaa agttagattt taacaaattg caaaagcttc accagaagga actcgggtgat 780
 atttcaaggt ggtggaaaga attagacttt gcacataagc tacctttcat aagagataga 840
 gttgcagagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca 900
 agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960
 tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagaggtg ggatatcagt 1020
 gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt 1080
 agtgaagcag agaaggattt ggcccccaa ggaaatcat accgcctcta ttatgcaaaa 1140
 gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200
 tatgtaccta cagtggatga gtacatgacg gttgcattag ttacaactac ctatccaatg 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
 ctgatgcaaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc 1620
 tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
 ccgatatgtg tatga 1695

<210> 8
 <211> 564
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> secuencia de aminoácidos del polipéptido CsTPS3-VC1, codificado por la secuencia nucleotídica artificial Cstps3-VC1 (SEQ ID NO:7)

<220>
 <221> SITE
 <222> 22..559
 <223> Terpeno ciclasa vegetal, Clase 1; NCBI CDD: cd00684; PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222> join(22..32,466..467,469..472,474..477,538..541)
 <223> Sitio activo (active site lid residues); PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222>
 join(278,287,308,310..312,315,319,390,456..457,459..460,463,467,535,539,541)
)
 <223> Sitio activo: Cavidad de enlace del sustrato (substrate binding pocket); PSSM-Id: 173832

<220>
 <221> SITE
 <222> 314..319
 <223> Motivo DDX2D (aspartate-rich región 1); PSSM-Id:173832

ES 2 540 791 B1

<220>
 <221> SITE
 <222> join(315,319,459,463,467)
 <223> Enlace del substrato-Mg2+ (substrate-Mg2+ binding site);
 PSSM-Id:173832

<220>
 <221> SITE
 <222> 457..467
 <223> Motivo NSE/DTE (aspartate-rich región 2); PSSM-d:173832

<220>
 <221> MUTAGEN
 <222> 416
 <223> Substitución: S416T

<220>
 <221> MUTAGEN
 <222> 417
 <223> Mutación: G417T

<220>
 <221> MUTAGEN
 <222> 418
 <223> Substitución S418Y

<400> 8
 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 Leu 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 Leu 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
 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
 290 295 300
 Thr Lys Val Ile Ser Met Thr Ser Val Ile 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

ES 2 540 791 B1

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 Thr
 405 410 415
 Thr Tyr 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
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val
 545 550 555 560
 Pro Ile Cys Val

<210> 9
 <211> 1695
 <212> DNA
 <213> Artificial sequence

<220>
 <221> source
 <222> 1..1695
 <223> /organism="Artificial sequence"
 /note="Secuencia polinucleotídica de cDNA Cstps3-VC2"
 /mol_type="unassigned DNA"

<220>
 <221> misc_feature
 <222> 1..1695
 <223> /note="Versión mutante derivada de la secuencia natural Cstps3
 (SEQ ID NO:1), en la cual se han substituido los nucleótidos
 t1246,g1249, g1250, c1253, c1254, y t1264"

<220>
 <221> misc_feature
 <222> 1246
 <223> /note="Substitución t1246a"

<220>
 <221> misc_feature
 <222> 1249
 <223> /note="Substitución g1249a"

<220>
 <221> misc_feature
 <222> 1250
 <223> /note="Substitución g1250c"

<220>
 <221> misc_feature
 <222> 1253
 <223> /note="Substitución c1253a"

ES 2 540 791 B1

<220>
 <221> misc_feature
 <222> 1254
 <223> /note="Substitución c1254t"

<220>
 <221> misc_feature
 <222> 1264
 <223> /note="Substitución t1264g"

<400> 9
 atgtctttgg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct 60
 actcgtcgcct ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120
 tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180
 attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240
 attcagcgtt taggagtgtc ttaccattht gaaagtgaga ttgatgaaat attgggaaag 300
 atgcataagg cttcccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat 360
 atctctcttc atthttcgatt acttagacaa aatggctata aaatthccgc tgatgtgttc 420
 aaaaagttca aagacacgga tgggaactth aaaacatctc ttgcgaaaga tgttcgagga 480
 atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540
 gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat caggctctcca 600
 cttgttgaac aagtcaaaca tgccttagtt cagcctatcc acaggggctt acaaaggctt 660
 gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720
 actthttgcaa agtttagatth taacaaattg caaaagcttc accagaagga actcgggtgat 780
 atthcaagggt ggtggaaaga attagactth gcacataagc tacctthtcat aagagataga 840
 gttgcagagt gctactthttg gatattagga gtgtatthtcg agccccaata ttcatttgca 900
 agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960
 tatggcaaaa ttgaagaact tgagctthttt acttcagcta ttgagagggtg ggatatcagt 1020
 gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgtthttt 1080
 agtgaagcag agaaggatth ggcccccaa ggaaaatcat accgcctcta ttatgcaaaa 1140
 gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200
 tatgtaccta cagtggatga gtacatgacg gttgcattag ttacaactac ctatccaatg 1260
 ttggcaacca catcctthtgt tggcatggga gacattgtaa ctaaagaatc thttgagtgg 1320
 ttattcagca atcctagatt tattagggtc tcttctatag thttgccgact catggatgac 1380
 atagtgtcac acaagthttga acaaagcaga gggcacgttg cctcaagcgt tgagtgttac 1440
 atgaaacaac atggagcaac agaagaggaa gcatgcaatg agthttcgaa acaagthttca 1500
 aatgcctgga aggatataaa tgaggactgc ctacgcccaa cggttgtgcc aatgccactt 1560
 ctgatgcgaa thtctaatct tacacgcgtt atagatgtca thtacaagta tgaagatggc 1620
 tacactcatt ccgcagthtgt gctgaaagat thttgttgctt cthttgttht taatcctgtg 1680
 ccgatatgtg tatga 1695

ES 2 540 791 B1

<210> 10
 <211> 564
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Secuencia de aminoácidos del polipéptido CsTPS3-VC2, codificado por la secuencia nucleotídica artificial Cstps3-VC2 (SEQ ID NO:9)

<220>
 <221> SITE
 <222> 22..559
 <223> Terpeno ciclasa vegetal, Clase 1; NCBI CDD: cd00684; PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222> join(22..32,466..467,469..472,474..477,538..541)
 <223> Sitio activo (active site lid residues); PSSM-Id: 173832

<220>
 <221> ACT_SITE
 <222>
 join(278,287,308,310..312,315,319,390,456..457,459..460,463,467,535,539,541)
 <223> sitio activo: Cavidad de enlace del sustrato (substrate binding pocket); PSSM-Id: 173832

<220>
 <221> SITE
 <222> 314..319
 <223> Motivo DDX2D (aspartate-rich región 1); PSSM-Id:173832

<220>
 <221> SITE
 <222> join(315,319,459,463,467)
 <223> Enlace del sustrato-Mg2+ (substrate-Mg2+ binding site); PSSM-Id:173832

<220>
 <221> SITE
 <222> 457..467
 <223> Motivo NSE/DTE (aspartate-rich región 2); PSSM-d:173832

<220>
 <221> MUTAGEN
 <222> 416
 <223> Substitución: S416T

<220>
 <221> MUTAGEN
 <222> 417
 <223> Mutación: G417T

<220>
 <221> MUTAGEN
 <222> 418
 <223> Substitución S418Y

<220>
 <221> MUTAGEN
 <222> 422
 <223> Substitución: S422A

<400> 10
 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
 15

ES 2 540 791 B1

Gly Ser Asn Val Lys His Leu Glu Leu Lys Lys Glu Ile Arg Arg Met
 50 35 55 40 60 45
 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 Leu 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 Leu 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
 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
 290 295 300
 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 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 Thr
 405 410 415
 Thr Tyr Pro Met Leu Ala 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
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val
 545 550 555 560
 Pro Ile Cys Val

<210> 11
<211> 32

ES 2 540 791 B1

```

<212> DNA
<213> Artificial Sequence

<220>
<221> source
<222> 1..32
<223> /organism="Artificial Sequence"
      /note="Secuencia nucleotidica correspondiente al oligonucleótido
      TPS3-F que se utilizó como cebador para la amplificación mediante
      PCR de la región codificante del gen Cstps3 (SEQ ID NO:1)"
      /mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1..32
<223> /note="ssDNA (single strand DNA)"

<220>
<221> misc_feature
<222> 3..8
<223> /note="Sitio de reconocimiento para el enzima de restricción BamH
      I"

<220>
<221> misc_feature
<222> 9..11
<223> /note="El oligonucleótido TPS3 incluye el codon atg"

<400> 11
cgggatccca tgtctttgga agtttcagcc tc                                     32

<210> 12
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<222> 1..29
<223> /organism="Artificial Sequence"
      /note="Secuencia nucleotidica correspondiente al oligonucleótido
      TPS3-R que se utilizó como cebador para la amplificación mediante
      PCR de la región codificante del gen Cstps3 (SEQ ID NO:1)"
      /mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1..29
<223> /note="ssDNA (single strand DNA)"

<220>
<221> misc_feature
<222> 3..8
<223> /note="Sitio de reconocimiento para el enzima de restricción sal
      I"

<220>
<221> misc_feature
<222> 9..11
<223> /note="El oligonucleótido TPS3-R incluye el codón de parada
      (stop) tga"

<400> 12
cggtcgactc atacacatat cggcacagg                                         29

<210> 13
<211> 1695
<212> DNA

```

ES 2 540 791 B1

```

<213> Citrus sinensis var. Washington Navel
<220>
<221> source
<222> 1..1695
<223> /organism="Citrus sinensis var. Washington Navel"
      /note="cDNA de la proteína cSTPS3 aislada de naranjo o uno de sus
      mutantes con actividad sesquiterpeno sintasa"
      /mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1246
<223> /note="n es t o a"

<220>
<221> misc_feature
<222> 1249
<223> /note="n es g o a"

<220>
<221> misc_feature
<222> 1252..1254
<223> /note="nnn es agc, cac o tat"

<220>
<221> misc_feature
<222> 1264
<223> /note="n es t o g"

<220>
<221> misc_feature
<222> 1363
<223> /note="n es t o g"

<400> 13
atgtcttttg aagtttcagc ctctcctgct aaagttatcc aaaatgctgg gaaagattct      60
actcgtcgcct ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact      120
tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa      180
attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca      240
attcagcggt taggagtgtc ttaccatfff gaaagtgaga ttgatgaaat attgggaaag      300
atgcataagg cttccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat      360
atctctcttc attttcgatt acttagacia aatggctata aaatttccgc tgatgtgttc      420
aaaaagttca 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 acaaaggctt      660
gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta      720
acttttgcaa agttagattt taacaaattg caaaagcttc accagaagga actcgggtgat      780
atttcaagggt ggtggaaaga attagacttt gcacataagc tacctttcat aagagataga      840
gttgacagagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca      900
agaagaatat tgacgaaagt gatctcatg acttctgtta ttgatgatat ctatgatgtg      960
tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagagggtg ggatatcagt      1020

```

ES 2 540 791 B1

gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt 1080
 agtgaagcag agaaggattt ggcccccaa ggaaatcat accgcctcta ttatgcaaaa 1140
 gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggg tcttcagaat 1200
 tatgtacctc cagtggatga gtacatgacg gttgcattag ttacancnnc cnnccaatg 1260
 ttgncaacca catcctttgt tggcatggga gacattgtaa ctaaagaatc ttttgagtgg 1320
 ttattcagca atcctagatt tattagggct tcttctatag ttngccgact catggatgac 1380
 atagtgtcac acaagtttga acaagcaga gggcacgttg cctcaagcgt tgagtgttac 1440
 atgaaacaac atggagcaac agaagaggaa gcatgcaatg agtttcggaa acaagtttca 1500
 aatgcctgga aggatataaa tgaggactgc ctacgccccaa cggttggtgcc aatgccactt 1560
 ctgatgcaaa ttctcaatct tacacgcggt atagatgtca tttacaagta tgaagatggc 1620
 tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
 ccgatatgtg tatga 1695

<210> 14
 <211> 564
 <212> PRT
 <213> Citrus sinensis var. Washington Navel

<220>
 <223> Proteína CsTPS3 aislada de naranjo o uno de sus mutantes con actividad sesquiterpeno sintasa

<220>
 <221> SITE
 <222> 416
 <223> Xaa es Thr o Ser

<220>
 <221> SITE
 <222> 417
 <223> Xaa es Ala o Thr

<220>
 <221> SITE
 <222> 418
 <223> Xaa es Ser, Hys o Tyr, y preferentemente es Ser o Tyr

<220>
 <221> SITE
 <222> 422
 <223> Xaa es Ser o Ala

<220>
 <221> SITE
 <222> 455
 <223> Xaa es Cys o Gly

<400> 14
 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

ES 2 540 791 B1

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 Leu 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 Leu 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
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
290 295 300
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 Xaa
405 410 415
Xaa Xaa Pro Met Leu Xaa 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 Xaa 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
Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val
545 550 555 560
Pro Ile Cys Val

<210> 15
<211> 1695
<212> DNA
<213> Citrus sinensis var. Washington Navel

<220>
<221> source

ES 2 540 791 B1

```

<222> 1..1695
<223> /organism="Citrus sinensis var. Washington Navel"
      /note="cDNA de mutantes C455G de la proteína sesquiterpeno
      sintasa CsTPS3 aislada de naranjo"
      /mol_type="unassigned DNA"

<220>
<221> misc_feature
<222> 1252..1253
<223> /note="nn es ag o ca"

<400> 15
atgtctttgg aagtttcagc ctctcctgct aaagtatcc aaaatgctgg gaaagattct      60
actcgtcgcct ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact      120
tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa      180
attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca      240
attcagcgtt taggagtgtc ttaccatttt gaaagtgaga ttgatgaaat attgggaaag      300
atgcataagg cttcccaaga ctctgatcct tgtgataatg aaaatgatga gctctattat      360
atctctcttc attttcgatt acttagacaa aatggctata aaatttccgc tgatgtgttc      420
aaaaagttca 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 acaaggctt      660
gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta      720
acttttgcaa agttagattt taacaaattg caaaagcttc accagaagga actcgggtgat      780
atttcaaggt ggtggaagaa attagacttt gcacataagc tacctttcat aagagataga      840
gttgacagagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca      900
agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg      960
tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagaggty ggatatcagt      1020
gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt      1080
agtgaagcag agaaggattt ggcccccaa ggaaatcat accgcctcta ttatgcaaaa      1140
gaagcgatga agaatatggt taagaattac ttctacgaag ctaaagggtg tcttcagaat      1200
tatgtaccta cagtggatga gtacatgacg gttgcattag ttacatctgc cnncccaatg      1260
ttgtcaacca catcctttgt tggcatggga gacattgtaa ctaaagaatc ttttgagtgg      1320
ttattcagca atcctagatt tattagggct tcttctatag ttggccgact catggatgac      1380
atagtgtcac acaagtttga acaagcaga gggcacgttg cctcaagcgt tgagtgttac      1440
atgaaacaac atggagcaac agaagaggaa gcatgcaatg agtttcggaa acaagtttca      1500
aatgcctgga aggatataaa tgaggactgc ctacgccccaa cggttggtgcc aatgccactt      1560
ctgatgcaaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc      1620
tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg      1680
ccgatatgtg tatga                                             1695

```

ES 2 540 791 B1

<210> 16
 <211> 564
 <212> PRT
 <213> Citrus sinensis var. Washington Navel

<220>
 <223> Mutantes C455G de la proteína sesquiterpeno sintasa CsTPS3
 aislada de naranjo

<220>
 <221> SITE
 <222> 418
 <223> Xaa es Ser o Hys, y preferentemente es Ser

<400> 16
 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
 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 Leu 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 Leu 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
 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
 290 295 300
 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 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
 Ala Xaa Pro Met Leu Ser Thr Thr Ser Phe Val Gly Met Gly Asp Ile
 420 425 430

ES 2 540 791 B1

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

<210> 17
 <211> 1695
 <212> DNA
 <213> Citrus sinensis var. Washington Navel

<220>
 <221> source
 <222> 1..1695
 <223> /organism="Citrus sinensis var. Washington Navel"
 /note="cdna de mutantes S416T/A417T/S418/Y de la proteina
 sesquiterpeno sintasa CstPS3 aislada de naranjo"
 /mol_type="unassigned DNA"

<220>
 <221> misc_feature
 <222> 1264
 <223> /note="n e s t o g"

<400> 17
 atgtctttgg aagtttcagc ctctcctgct aaagtatcc aaaatgctgg gaaagattct 60
 actcgtcgtc ctgcaaatta tcatccaagc atctgggggg atcatttcct tcaatatact 120
 tgtgacaccc aggaaactga tgatggcagc aatgtaaagc atctagagct gaagaaagaa 180
 attagaagaa tgctaaaagc tgataacaag ccttcacgta cacttcaatt gattgatgca 240
 attcagcgtt taggagtgtc ttaccatttt gaaagtgaga ttgatgaaat attgggaaag 300
 atgcataagg cttcccaaga ctctgatctt tgtgataatg aaaatgatga gctctattat 360
 atctctcttc attttcgatt acttagacaa aatggctata aaatttccgc tgatgtgttc 420
 aaaaagttca aagacacgga tgggaacttt aaaacatctc ttgcgaaaga tgttcgagga 480
 atgttaagct tgtatgaagc tacgcatctc ggggtacatg aagaagatat actagatgaa 540
 gcgcttgctt tcaccactag tcacctagag tcaatagcga ctcatcaaat caggctctcca 600
 cttgttgaac aagtcaaca tgccttagtt cagcctatcc acaggggctt acaaggcctt 660
 gaggcaagac agtacattcc tatctatcaa gaagaatctc cccacaatga agctctgtta 720
 acttttgcaa agttagattt taacaaattg caaaagcttc accagaagga actcgggtgat 780
 atttcaaggt ggtggaaaga attagacttt gcacataagc tacctttcat aagagataga 840
 gttgcagagt gctacttttg gatattagga gtgtatttcg agccccaata ttcatttgca 900
 agaagaatat tgacgaaagt gatctccatg acttctgtta ttgatgatat ctatgatgtg 960

ES 2 540 791 B1

tatggcaaaa ttgaagaact tgagcttttt acttcagcta ttgagaggtg ggatatcagt 1020
gccatagatc aacttcctga gtatatgaaa ttgtgttata gggcccttct tgatgttttt 1080
agtgaagcag agaaggattt ggcccccaa ggaaaatcat accgcctcta ttatgcaaaa 1140
gaagcgatga agaatatggt taagaattac ttctacgaag ctaaattggtg tcttcagaat 1200
tatgtacctc cagtggatga gtacatgacg gttgcattag ttacaactac ctatccaatg 1260
ttgncaacca 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
ctgatgcaaa ttctcaatct tacacgcgtt atagatgtca tttacaagta tgaagatggc 1620
tacactcatt cgcagttgt gctgaaagat tttgttgctt ctttgtttat taatcctgtg 1680
ccgatatgtg tatga 1695

<210> 18
<211> 564
<212> PRT
<213> Citrus sinensis var. Washington Navel

<220>
<223> Mutantes S416T/A417T/S418/Y de la proteína sesquiterpeno sintasa cSTPS3 aislada de naranjo

<220>
<221> SITE
<222> 422
<223> xaa es Ala o Ser, y más preferentemente es Ala

<400> 18
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 Leu Gln Arg Leu Glu Ala Arg Gln
210 215 220

ES 2 540 791 B1

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 Leu 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
 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 Thr
 405 410 415
 Thr Tyr Pro Met Leu Xaa 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
 Ala Val Val Leu Lys Asp Phe Val Ala Ser Leu Phe Ile Asn Pro Val
 545 550 555 560
 Pro Ile Cys Val