(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
11 September 2009 (11.09.2009)

(10) International Publication Number
WO 2009/109597 A1

(51) International Patent Classification:
C12N 15/52 (2006.01) *C12P 7/22* (2006.01)
C12N 9/88 (2006.01) *C12P 7/26* (2006.01)

(21) International Application Number: PCT/EP2009/052546

(22) International Filing Date: 4 March 2009 (04.03.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
08102357.4 6 March 2008 (06.03.2008) EP
08103362.3 3 April 2008 (03.04.2008) EP

(71) Applicant *(for all designated States except US)*: **FIRMENICH SA** [CH/CH]; 1, route des Jeunes, P.O. Box 239, CH-1211 Geneva 8 (CH).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **SCHALK, Michel** [FR/FR]; 195, Allée Des Résidences Du Salève, F-74160 Collonges-sous-saleve (FR).

(74) Agent: **SALVATERRA-GARCIA, Maria de Lurdes**; FIRMENICH SA, Rue de la Bergère 7, P.O. Box 148, CH-1217 Meyrin 2 (CH).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

WO 2009/109597 A1

(54) **Title**: METHOD FOR PRODUCING ALPHA- SANTALENE

(57) **Abstract**: The present invention provides a method of producing α-santalene, said method comprising contacting at least one polypeptide with farnesyl phyrophosphate (FPP). In particular, said method may be carried out in vitro or in vivo to produce α-santalene, a very useful compound in the fields of perfumery and flavoring. The present invention also provides the amino acid sequence of a polypeptide useful in the method of the invention. A nucleic acid encoding the polypeptide of the invention and an expression vector containing said nucleic acid are also part of the present invention. A non-human host organism or a cell transformed to be used in the method of producing α-santalene is also an object of the present invention.

METHOD FOR PRODUCING ALPHA-SANTALENE

Technical field

The present invention provides a method of producing α-santalene, said method comprising contacting at least one polypeptide with farnesyl phyrophosphate (FPP). In particular, said method may be carried out in vitro or in vivo to produce α-santalene, a very useful compound in the fields of perfumery and flavoring. The present invention also provides the amino acid sequence of a polypeptide useful in the method of the invention. A nucleic acid encoding the polypeptide of the invention and an expression vector containing said nucleic acid are also part of the present invention. A non-human host organism or a cell transformed to be used in the method of producing α-santalene is also an object of the present invention.

Prior art

Terpenes are found in most organisms (microorganisms, animals and plants). These compounds are made up of five carbon units called isoprene units and are classified by the number of these units present in their structure. Thus monoterpenes, sesquiterpenes and diterpenes are terpenes containing 10, 15 and 20 carbon atoms respectively. Sesquiterpenes, for example, are widely found in the plant kingdom. Many sesquiterpene molecules are known for their flavor and fragrance properties and their cosmetic, medicinal and antimicrobial effects. Over 300 sesquiterpene hydrocarbons and 3000 sesquiterpenoids have been identified and many new structures are identified each year. Plant extracts obtained by different means such as steam distillation or solvent extraction are used as source of terpenes. Terpene molecules are often used as such, but in some cases chemical reactions are used to transform the terpenes into other high value molecules.

Biosynthetic production of terpenes involves enzymes called terpene synthases. There is virtually an infinity of sesquiterpene synthases present in the plant kingdom, all using the same substrate (farnesyl pyrophosphate, FPP) but having different product profiles. Genes and cDNAs encoding sesquiterpene synthases have been cloned and the corresponding recombinant enzymes characterized. The biosynthesis of terpenes in plants and other organisms has been extensively studied and is not further detailed in here, but reference is made to Dewick, *Nat. Prod. Rep.*, 2002, 19, 181-222, which reviews the state

of the art of terpene biosynthetic pathways.

α-santalene is a naturally occurring sesquiterpene molecule. The (+)- isomer can be used as starting material for the chemical synthesis or the biosynthesis of (Z)-(+)-α-santalol, which is an important constituent of sandalwood oil. Sandalwood oil is an
5 important perfumery ingredient obtained by distillation of the heartwood of Santalum species. Sandalwood is also largely used for incenses and traditional medicine. The oil contains 90% of sesquiterpene alcohols. (Z)-(+)-α-santalol and (Z)-(-)-β-santalol represent the major constituents (respectively 45-47% and 20-30%) and are mainly responsible for the typical sweet-woody and balsamic odour of sandalwood oil. Other constituents such as
10 epi-β-santalol and trans-α-bergamotol are also present and may contribute to the sandalwood note.

Generally, the price and availability of plant natural extracts are dependent on the abundance, oil yield and geographical origin of the plants. In addition, the availability and quality of natural extracts is very much dependent on climate and other local conditions
15 leading to variability from year to year, rendering the use of such ingredients in high quality perfumery very difficult or even impossible some years. Due to over-exploitation of the natural resources, difficulties of cultivation, slow growth of the Santalum plants, the availabilities of sandalwood raw material has dramatically decreased during the past decades. Therefore, it would be an advantage to provide a source of (Z)-(+)-α-santalol,
20 which is less subjected to fluctuations in availability and quality. A chemical synthesis of the sandalwood sesquiterpene constituents is so far not available. A biochemical pathway leading to the synthesis of (+)-α-santalene, which could then be used to produce (Z)-(+)-α-santalol, would therefore be of great interest. Given the difficulty to control sesquiterpene production in Santalum species, alternate plant sources were sought.

25 Santalane type sesquiterpene, and particurly sesquiterpenes with the α-santalane skeleton, were identified in several plant species. *Clausena lansium*, a plant from the Rutaceae family has been reported to contain large quantities of santalane sesquiterpenes in the leaves. Zhao and coworkers (Zhao et al, *Z. Naturforsch*, 2004, 59c, 153-156) have analyzed the leaves of *C. lansium* from China and detected the presence of α-santalol and
30 β-santalol. The analysis of the leaves of *C. lansium* from Cuba, has revealed the presence of (Z)-α-santalol, epi-β-santalol, (Z)-β-santalol and (E)-β-santalol (Pino et al., *J. Essent. Oil Res.*, 2006, 18, 139-141). Surprisingly the analysis of different parts of *C. lansium*

from Thailand origin did not show the presence of sesquiterpenes with santalane skeletons (Chokeprasert et al, *Journal of Food Composition and Analysis*, 2007, 20(1), 52-56).

A sesquiterpene synthase capable of synthesizing at least one bi-cyclic and/or tri-cyclic sesquiterpene having a santalane carbon skeleton, the corresponding nucleic acid
5 and a method for producing such compounds having a santalane carbon skeleton are disclosed in the International patent application WO 2006/134523. (+)-epi-β-santalene, (-)-β-santalene, (+)-β-santalene, (+)-α-santalene and (-)-α-santalene are cited as examples of compounds having a santalane carbon skeleton. Nevertheless, the sesquiterpene synthase provided in the examples does not produce α-santalene. Only epi-β-santalene is produced.
10 The properties of this compound are very different from those of α-santalene. In particular, epi-β-santalene is of no interest in the synthesis of (Z)-(+)-α-santalol. Moreover, the sesquiterpene synthase disclosed in WO 2006/134523 shares only 37% identity with the sequence of the invention.

Terpene synthases having a certain percentage of sequence identity with the
15 sequence of the α-santalene synthase of the present invention have also been found in the sequences databases. Nevertheless, the percentage of identity between the known sesquiterpene synthases and the polypeptide of the invention is very low. The closest protein sequence to the (+)-α-santalene synthase of the invention is a (E)-β-farnesene synthase from Citrus junos (NCBI access No. AAK54279; Maruyama et al, *Biol. Pharm.*
20 *Bull.*, 2001, 24(10), 1171-1175) which shares 67 to 68% amino acid sequence identity with the α-santalene synthase of the invention.

In addition to the difference between the sequences themselves, it also has to be pointed out that the structure and the properties of the products synthesized by the above-mentioned enzyme are very different from those of α-santalene. In particular (E)-β-
25 farnesene is not suitable as a starting material for the synthesis of (Z)-(+)-α-santalol, which is a very useful ingredient in the field of perfumery.

An α-santalene synthase is disclosed in WO 2008/142318. This document was not published at the priority date of the present application. It describes an enzyme capable of catalyzing the transformation of Z,Z-farnesyl pyrophosphate to α-santalene. Therefore the
30 reaction catalyzed by the prior art enzyme is different from the one catalyzed by the synthase of the present invention, which starts from E,E-farnesyl pyrophosphate. Moreover, the α-santalene synthase of the invention shares only 23.8% of sequence

identity with the one described in WO 2008/142318.

Despite extensive studies of terpene cyclization, the isolation and characterization of the terpene synthases is still difficult, particularly in plants, due to their low abundance, their often transient expression patterns, and the complexity of purifying them from the
5 mixtures of resins and phenolic compounds in tissues where they are expressed.

It is an objective of the present invention to provide methods for making (+)-α-santalene in an economic way, as indicated above. Accordingly, the present invention has the objective to produce (+)-α-santalene while having little waste, a more energy and resource efficient process and while reducing dependency on fossil fuels. It is a further
10 objective to provide enzymes capable of synthesizing α-santalene, which is useful as perfumery and/or aroma ingredients.

Abbreviations Used

bp	base pair
kb	kilo base
BSA	bovine serum albumin
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
cDNA	complementary DNA
dT	deoxy thymine
dNTP	deoxy nucleotide triphosphate
DTT	dithiothreitol
FPP	farnesyl pyrophosphate
GC	gaseous chromatograph
idi	isopentenyl diphosphate isomerase
IPP	isopentenyl diphosphate
IPTG	isopropyl-D-thiogalacto-pyranoside
LB	lysogeny broth
MOPSO	3-(N-morpholino)-2-hydroxypropanesulfonic acid
MS	mass spectrometer
mvaK1	mevalonate kinase
mvaK2	mevalonate diphosphate kinase

NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
RMCE	recombinase-mediated cassette exchange
3'-/5'-RACE	3' and 5' rapid amplification of cDNA ends
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid

Description of the invention

The present invention provides a method to biosynthetically produce α-santalene in an economic, reliable and reproducible way.

A "sesquiterpene synthase" or a "polypeptide having a sesquiterpene synthase activity", is intended here as a polypeptide capable of catalyzing the synthesis of a sesquiterpene molecule or of a mixture of sesquiterpene molecules from the acyclic terpene precursor FPP.

As an "α-santalene synthase" or as a "polypeptide having an α-santalene synthase activity", we mean here a polypeptide capable of catalyzing the synthesis of α-santalene, in the form of any of its stereoisomers or a mixture thereof, starting from FPP. α-Santalene may be the only product or may be part of a mixture of sesquiterpenes.

As a "(+)-α-santalene synthase" or as a "polypeptide having a (+)-α-santalene synthase activity", we mean here a polypeptide capable of catalyzing the synthesis of (+)-α-santalene starting from FPP. (+)-α-santalene may be the only product or may be part of a mixture of sesquiterpenes. The (+)-α-santalene synthase is a particular example of α-santalene synthase.

The ability of a polypeptide to catalyze the synthesis of a particular sesquiterpene (for example (+)-α-santalene) can be simply confirmed by performing the enzyme assay as detailed in Example 4.

According to a preferred embodiment of the invention, FPP is in the form of (2E,6E)-FPP.

According to the present invention, polypeptides are also meant to include truncated polypeptides provided that they keep their sesquiterpene synthase activity as defined in any of the above embodiments and that they share at least the defined percentage of identity with the corresponding fragment of SEQ ID NO:1.

As intended herein below, "a nucleotide sequence obtained by modifying SEQ ID NO:2" encompasses any sequence that has been obtained by changing the sequence of SEQ ID NO:2 using any method known in the art, for example by introducing any type of mutations such as deletion, insertion or substitution mutations. Examples of such methods 5 are cited in the part of the description relative to the variant polypeptides and the methods to prepare them.

The percentage of identity between two peptidic or nucleotidic sequences is a function of the number of amino acids or nucleotide residues that are identical in the two sequences when an alignment of these two sequences has been generated. Identical 10 residues are defined as residues that are the same in the two sequences in a given position of the alignment. The percentage of sequence identity, as used herein, is calculated from the optimal alignment by taking the number of residues identical between two sequences dividing it by the total number of residues in the shortest sequence and multiplying by 100. The optimal alignment is the alignment in which the percentage of identity is the highest 15 possible. Gaps may be introduced into one or both sequences in one or more positions of the alignment to obtain the optimal alignment. These gaps are then taken into account as non-identical residues for the calculation of the percentage of sequence identity.

Alignment for the purpose of determining the percentage of amino acid or nucleic acid sequence identity can be achieved in various ways using computer programs and for 20 instance publicly available computer programs available on the world wide web. Preferably, the BLAST program (Tatiana et al, *FEMS Microbiol Lett.*, 1999, 174:247-250, 1999) set to the default parameters, available from the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi, can be used to obtain an optimal 25 alignment of peptidic or nucleotidic sequences and to calculate the percentage of sequence identity.

One object of the present invention is therefore a method for producing α-santalene comprising

30 a) contacting FPP with at least one polypeptide having an α-santalene synthase activity and comprising an amino acid sequence at least 50% identical to SEQ ID NO:1;
b) optionally, isolating the α-santalene produced in step a).

According to a preferred embodiment, the method is a method for producing α-santalene as a major product. According to an even more preferred embodiment, α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the product produced by the method of the invention.

5 According to a more preferred embodiment, the method is a method for producing (+)-α-santalene and the polypeptide having an α-santalene synthase activity has a (+)-α-santalene synthase activity.

According to an even more preferred embodiment, the method is a method for producing (+)-α-santalene as a major product. According to a most preferred embodiment, 10 (+)-α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the products produced by the method of the invention.

The method can be carried out in vitro as well as in vivo, as will be explained in details further on.

The polypeptide to be contacted with FPP in vitro can be obtained by extraction 15 from any organism expressing it, using standard protein or enzyme extraction technologies. If the host organism is a unicellular organism or cell releasing the polypeptide of the invention into the culture medium, the polypeptide may simply be collected from the culture medium, for example by centrifugation, optionally followed by washing steps and re-suspension in suitable buffer solutions. If the organism or cell accumulates the 20 polypeptide within its cells, the polypeptide may be obtained by disruption or lysis of the cells and further extraction of the polypeptide from the cell lysate.

The polypeptide having an α-santalene synthase activity, either in an isolated form or together with other proteins, for example in a crude protein extract obtained from cultured cells or microorganisms, may then be suspended in a buffer solution at optimal 25 pH. If adequate, salts, BSA and other kinds of enzymatic co-factors, may be added in order to optimize enzyme activity. Appropriate conditions are described in more details in the Examples further on.

The precursor FPP may then be added to the suspension or solution, which is then incubated at optimal temperature, for example between 15 and 40°C, preferably between 30 25 and 35°C, more preferably at 30°C. After incubation, the α-santalene produced may be isolated from the incubated solution by standard isolation procedures, such as solvent extraction and distillation, optionally after removal of polypeptides from the solution.

According to another preferred embodiment, the method of any of the above-described embodiments is carried out in vivo. In this case, step a) comprises cultivating a non-human host organism or cell capable of producing FPP and transformed to express at least one polypeptide comprising an amino acid sequence at least 50% identical to SEQ ID NO:1 and having an α-santalene synthase activity, under conditions conducive to the production of α-santalene.

According to a more preferred embodiment, the method further comprises, prior to step a), transforming a non human organism or cell capable of producing FPP with at least one nucleic acid encoding a polypeptide comprising an amino acid sequence at least 50% identical to SEQ ID NO:1 and having an α-santalene synthase activity, so that said organism expresses said polypeptide.

These embodiments of the invention are particularly advantageous since it is possible to carry out the method in vivo without previously isolating the polypeptide. The reaction occurs directly within the organism or cell transformed to express said polypeptide.

According to a particular embodiment of the invention, the at least one nucleic acid encoding the α-santalene synthase comprises a nucleotide sequence at least 50%, preferably at least 55%, preferably at least 60%, preferably at least 65%, preferably at least 70%, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 98% identical to SEQ ID NO:2 or the complement thereof. According to a more preferred embodiment, said nucleic acid comprises the nucleotide sequence SEQ ID NO:2 or the complement thereof. In an even more preferred embodiment, said nucleic acid consists of SEQ ID NO:2 or the complement thereof.

According to a more preferred embodiment the at least one nucleic acid used in any of the above embodiments comprises a nucleotide sequence that has been obtained by modifying SEQ ID NO:2. According to an even more preferred embodiment, said at least one nucleic acid consists of a nucleotide sequence that has been obtained by modifying SEQ ID NO:2.

According to another embodiment, the at least one nucleic acid is isolated from *Clausena lansium*.

The organism or cell is meant to "express" a polypeptide, provided that the

organism or cell is transformed to harbor a nucleic acid encoding said polypeptide, this nucleic acid is transcribed to mRNA and the polypeptide is found in the host organism or cell. The term "express" encompasses "heterologously express" and "over-express", the latter referring to levels of mRNA, polypeptide and/or enzyme activity over and above
5 what is measured in a non-transformed organism or cell. A more detailed description of suitable methods to transform a non-human host organism or cell will be described later on in the part of the specification that is dedicated to such transformed non-human host organisms or cells as specific objects of the present invention and in the examples.

A particular organism or cell is meant to be "capable of producing FPP" when it
10 produces FPP naturally or when it does not produce FPP naturally but is transformed to produce FPP, either prior to the transformation with a nucleic acid as described herein or together with said nucleic acid. Organisms or cells transformed to produce a higher amount of FPP than the naturally occurring organism or cell are also encompassed by the "organisms or cells capable of producing FPP". Methods to transform organisms, for
15 example microorganisms, so that they produce FPP are already known in the art. Such methods can for example be found in the literature, for example in the following publications Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. Nat Biotechnol., 2003, 21(7), 796-802 (transformation of E. coli); Wu, S., Schalk, M., Clark, A., Miles, R.B., Coates, R., and Chappell, J., *Nat Biotechnol.*, 2006, 24(11), 1441-
20 1447 (transformation of plants); Takahashi, S., Yeo, Y., Greenhagen, B. T., McMullin, T., Song, L., Maurina-Brunker, J., Rosson, R., Noel, J., Chappell, J, *Biotechnology and Bioengineering*, 2007, 97(1), 170-181 (transformation of yeast). .

To carry out the invention in vivo, the host organism or cell is cultivated under conditions conducive to the production of α-santalene. Accordingly, if the host is a
25 transgenic plant, optimal growth conditions are provided, such as optimal light, water and nutrient conditions, for example. If the host is a unicellular organism, conditions conducive to the production of α-santalene may comprise addition of suitable cofactors to the culture medium of the host. In addition, a culture medium may be selected, so as to maximize α-santalene synthesis. Optimal culture conditions are described in a more detailed manner in
30 the following Examples.

Non-human host organisms suitable to carry out the method of the invention in vivo may be any non-human multicellular or unicellular organisms. In a preferred

embodiment, the non-human host organism used to carry out the invention in vivo is a plant, a prokaryote or a fungus. Any plant, prokaryote or fungus can be used. Particularly useful plants are those that naturally produce high amounts of terpenes. In a more preferred embodiment, the plant is selected from the family of *Solanaceae, Poaceae,* 5 *Brassicaceae, Fabaceae, Malvaceae, Asteraceae or Lamiaceae*. For example, the plant is selected from the genera *Nicotiana, Solanum, Sorghum, Arabidopsis, Brassica (rape), Medicago* (alfalfa), *Gossypium* (cotton), *Artemisia, Salvia and Mentha*. Preferably, the plant belongs to the species of *Nicotiana tabacum*.

In a more preferred embodiment the non-human host organism used to carry out 10 the method of the invention in vivo is a microorganism. Any microorganism can be used but according to an even more preferred embodiment said microorganism is a bacteria or yeast. Most preferably, said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*.

Some of these organisms do not produce FPP naturally. To be suitable to carry out the method of the invention, these organisms have to be transformed to produce said 15 precursor. They can be so transformed either before the modification with the nucleic acid described according to any of the above embodiments or simultaneously, as explained above.

Isolated higher eukaryotic cells can also be used, instead of complete organisms, as hosts to carry out the method of the invention in vivo. Suitable eukaryotic cells may be any 20 non-human cell, but are preferably plant or fungal cells.

According to a preferred embodiment, the at least one polypeptide having an α-santalene synthase activity used in any of the above-described embodiments or encoded by the nucleic acid used in any of the above-described embodiments comprises an amino acid sequence at least 55%, preferably at least 60%, preferably at least 65%, preferably at least 25 70%, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 98% identical to SEQ ID NO:1. According to a more preferred embodiment, said polypeptide comprises the amino acid sequence SEQ ID NO:1. In an even more preferred embodiment, said polypeptide consists of SEQ ID NO:1.

30 According to another preferred embodiment, the at least one polypeptide having an α-santalene synthase activity used in any of the above-described embodiments or encoded by the nucleic acid used in any of the above-described embodiments comprises an amino

acid sequence that is a variant of SEQ ID NO:1 obtained by genetic engineering. In other terms, said polypeptide comprises an amino acid sequence encoded by a nucleotide sequence that has been obtained by modifying SEQ ID NO:2. According to a more preferred embodiment, the at least one polypeptide having an α-santalene synthase activity used in any of the above-described embodiments or encoded by the nucleic acid used in any of the above-described embodiments consists of an amino acid sequence that is a variant of SEQ ID NO:1 obtained by genetic engineering, i.e. an amino acid sequence encoded by a nucleotide sequence that has been obtained by modifying SEQ ID NO:2.

As used herein, the polypeptide is intended as a polypeptide or peptide fragment that encompasses the amino acid sequences identified herein, as well as truncated or variant polypeptides, provided that they keep their activity as defined above and that they share at least the defined percentage of identity with the corresponding fragment of SEQ ID NO:1.

Examples of variant polypeptides are naturally occurring proteins that result from alternate mRNA splicing events or form proteolytic cleavage of the polypeptides described herein. Variations attributable to proteolysis include, for example, differences in the N- or C- termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids from the polypeptides of the invention. Polypeptides encoded by a nucleic acid obtained by natural or artificial mutation of a nucleic acid of the invention, as described thereafter, are also encompassed by the invention.

Polypeptide variants resulting from a fusion of additional peptide sequences at the amino and carboxyl terminal ends can also be used in the methods of the invention. In particular such a fusion can enhance expression of the polypeptides, be useful in the purification of the protein or improve the enzymatic activity of the polypeptide in a desired environment or expression system. Such additional peptide sequences may be signal peptides, for example. Accordingly, the present invention encompasses methods using variant polypeptides, such as those obtained by fusion with other oligo- or polypeptides and/or those which are linked to signal peptides. Polypeptides resulting from a fusion with another functional protein, such as another protein from the terpene biosynthesis pathway, can also be advantageously be used in the methods of the invention.

According to another embodiment, the at least one polypeptide having an α-santalene synthase activity used in any of the above-described embodiments or encoded

by the nucleic acid used in any of the above-described embodiments is isolated from *Clausena lansium*.

An important tool to carry out the method of the invention is the polypeptide itself.
5 A polypeptide having an α-santalene synthase activity and comprising an amino acid sequence at least 50% identical to SEQ ID NO:1 is therefore another object of the present invention.

According to a preferred embodiment, the polypeptide is capable of producing α-santalene as a major product. According to an even more preferred embodiment, it is
10 capable of producing a mixture of sesquiterpenes wherein α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the sesquiterpenes produced.

According to a more preferred embodiment, the polypeptide has a (+)-α-santalene synthase activity.

15 According to an even more preferred embodiment, the polypeptide is capable of producing (+)-α-santalene as a major product. According to an even more preferred embodiment, it is capable of producing a mixture of sesquiterpenes wherein (+)-α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the sesquiterpenes produced.

20 According to a preferred embodiment, the polypeptide comprises an amino acid sequence at least 55%, preferably at least 60%, preferably at least 65%, preferably at least 70%, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 98% identical to SEQ ID NO:1. According to a more preferred embodiment, the polypeptide comprises
25 the amino acid sequence SEQ ID NO:1. According to an even more preferred embodiment, the polypeptide consists of SEQ ID NO:1.

According to another preferred embodiment, the at least one polypeptide comprises an amino acid sequence that is a variant of SEQ ID NO:1 obtained by genetic engineering. In other terms, said polypeptide comprises an amino acid sequence encoded by a
30 nucleotide sequence that has been obtained by modifying SEQ ID NO:2. According to a more preferred embodiment, the at least one polypeptide having an α-santalene synthase activity consists of an amino acid sequence that is a variant of SEQ ID NO:1 obtained by

genetic engineering, i.e. an amino acid sequence encoded by a nucleotide sequence that has been obtained by modifying SEQ ID NO:2.

According to another embodiment, the polypeptide is isolated form *Clausena lansium*.

5 As used herein, the polypeptide is intended as a polypeptide or peptide fragment that encompasses the amino acid sequences identified herein, as well as truncated or variant polypeptides, provided that they keep their activity as defined above and that they share at least the defined percentage of identity with the corresponding fragment of SEQ ID NO:1.

10 Examples of variant polypeptides are naturally occurring proteins that result from alternate mRNA splicing events or form proteolytic cleavage of the polypeptides described herein. Variations attributable to proteolysis include, for example, differences in the N- or C- termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids from the polypeptides of the invention. Polypeptides 15 encoded by a nucleic acid obtained by natural or artificial mutation of a nucleic acid of the invention, as described thereafter, are also encompassed by the invention.

Polypeptide variants resulting from a fusion of additional peptide sequences at the amino and carboxyl terminal ends are also encompassed by the polypeptides of the invention. In particular such a fusion can enhance expression of the polypeptides, be useful 20 in the purification of the protein or improve the enzymatic activity of the polypeptide in a desired environment or expression system. Such additional peptide sequences may be signal peptides, for example. Accordingly, the present invention encompasses variants of the polypeptides of the invention, such as those obtained by fusion with other oligo- or polypeptides and/or those which are linked to signal peptides. Polypeptides resulting from 25 a fusion with another functional protein, such as another protein from the terpene biosynthesis pathway, are also encompassed by the polypeptides of the invention.

As mentioned above, the nucleic acid encoding the polypeptide of the invention is a useful tool to modify non-human host organisms or cells intended to be used when the 30 method is carried out in vivo.

A nucleic acid encoding a polypeptide according to any of the above-described embodiments is therefore also an object of the present invention.

According to a preferred embodiment, the nucleic acid comprises a nucleotide sequence at least 50%, preferably at least 55%, preferably at least 60%, preferably at least 65%, preferably at least 70%, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, more preferably at least 95% and even more
5 preferably at least 98% identical to SEQ ID NO:2 or the complement thereof. According to a more preferred embodiment, the nucleic acid comprises the nucleotide sequence SEQ ID NO:2 or the complement thereof. According to an even more preferred embodiment, the nucleic acid consists of SEQ ID NO:2 or the complement thereof

According to another embodiment, the nucleic acid is isolated from *Clausena*
10 *lansium*.

The nucleic acid of the invention can be defined as including deoxyribonucleotide or ribonucleotide polymers in either single- or double-stranded form (DNA and/or RNA). The terms "nucleotide sequence" should also be understood as comprising a polynucleotide molecule or an oligonucleotide molecule in the form of a separate fragment
15 or as a component of a larger nucleic acid. Nucleic acids of the invention also encompass certain isolated nucleotide sequences including those that are substantially free from contaminating endogenous material. The nucleic acid of the invention may be truncated, provided that it encodes a polypeptide encompassed by the present invention, as described above.

20 According to a more preferred embodiment, the at least one nucleic acid according to any of the above embodiments comprises a nucleotide sequence that has been obtained by modifying SEQ ID NO:2. Preferably said nucleic acid consists of a nucleotide sequence that has been obtained by modifying SEQ ID NO:2.

The nucleic acids comprising a sequence obtained by mutation of SEQ ID NO:2 or
25 the complement thereof are encompassed by the invention, provided that the sequences they comprise share at least the defined percentage of identity with the corresponding fragments of SEQ ID NO:2 or with the complement thereof and provided that they encode a polypeptide having an α-santalene synthase activity, as defined in any of the above embodiments. Mutations may be any kind of mutations of these nucleic acids, such as point
30 mutations, deletion mutations, insertion mutations and/or frame shift mutations. A variant nucleic acid may be prepared in order to adapt its nucleotide sequence to a specific expression system. For example, bacterial expression systems are known to more

efficiently express polypeptides if amino acids are encoded by a preferred codon. Due to the degeneracy of the genetic code, wherein more than one codon can encode the same amino acid, multiple DNA sequences can code for the same polypeptide, all these DNA sequences being encompassed by the invention.

5

Another important tool for transforming host organisms or cells suitable to carry out the method of the invention in vivo is an expression vector comprising a nucleic acid according to any embodiment of the invention. Such a vector is therefore also an object of the present invention.

10 An “expression vector” as used herein includes any linear or circular recombinant vector including but not limited to viral vectors, bacteriophages and plasmids. The skilled person is capable of selecting a suitable vector according to the expression system. In one embodiment, the expression vector includes the nucleic acid of the invention operably linked to at least one regulatory sequence, which controls transcription, translation, 15 initiation and termination, such as a transcriptional promoter, operator or enhancer, or an mRNA ribosomal binding site and, optionally, including at least one selection marker. Nucleotide sequences are "operably linked" when the regulatory sequence functionally relates to the nucleic acid of the invention.

The expression vectors of the present invention may be used in the methods for 20 preparing a genetically transformed host organism and/or cell, in host organisms and/or cells harboring the nucleic acids of the invention and in the methods for producing or making polypeptides having an α-santalene synthase activity, as disclosed further below.

Recombinant non-human host organisms and cells transformed to harbor at least 25 one nucleic acid of the invention so that it heterologously expresses or over-expresses at least one polypeptide of the invention are also very useful tools to carry out the method of the invention. Such non-human host organisms and cells are therefore another object of the present invention.

A nucleic acid according to any of the above-described embodiments can be used 30 to transform the non-human host organisms and cells and the expressed polypeptide can be any of the above-described polypeptides.

Non-human host organisms of the invention may be any non-human multicellular or

unicellular organisms. In a preferred embodiment, the non-human host organism is a plant, a prokaryote or a fungus. Any plant, prokaryote or fungus is suitable to be transformed according to the present invention. Particularly useful plants are those that naturally produce high amounts of terpenes. In a more preferred embodiment, the plant is selected from the family of *Solanaceae, Poaceae, Brassicaceae, Fabaceae, Malvaceae, Asteraceae or Lamiaceae*. For example, the plant is selected from the genera *Nicotiana, Solanum, Sorghum, Arabidopsis, Brassica (rape), Medicago* (alfalfa), *Gossypium* (cotton), *Artemisia, Salvia and Mentha*. Preferably, the plant belongs to the species of *Nicotiana tabacum*.

In a more preferred embodiment the non-human host organism is a microorganism. Any microorganism is suitable for the present invention, but according to an even more preferred embodiment said microorganism is a bacteria or yeast. Most preferably, said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*.

Isolated higher eukaryotic cells can also be transformed, instead of complete organisms. As higher eukaryotic cells, we mean here any non-human eukaryotic cell except yeast cells. Preferred higher eukaryotic cells are plant cells or fungal cells.

The term “transformed” refers to the fact that the host was subjected to genetic engineering to comprise one, two or more copies of each of the nucleic acids required in any of the above-described embodiment. Preferably the term “transformed” relates to hosts heterologously expressing the polypeptides encoded by the nucleic acid with which they are transformed, as well as over-expressing said polypeptides. Accordingly, in an embodiment, the present invention provides a transformed organism, in which the polypeptides are expressed in higher quantity than in the same organism not so transformed.

There are several methods known in the art for the creation of transgenic host organisms or cells such as plants, fungi, prokaryotes, or cultures of higher eukaryotic cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, plant and mammalian cellular hosts are described, for example, in Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, Elsevier, New York and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd edition, 1989, Cold Spring Harbor Laboratory Press. Cloning and expression vectors for higher plants and/or plant cells in particular are available to the skilled person. See for example Schardl et al. *Gene 61: 1-11*, 1987.

Methods for transforming host organisms or cells to harbor transgenic nucleic acids are familiar to the skilled person. For the creation of transgenic plants, for example, current methods include: electroporation of plant protoplasts, liposome-mediated transformation, agrobacterium-mediated transformation, polyethylene-glycol-mediated 5 transformation, particle bombardement, microinjection of plant cells, and transformation using viruses.

In one embodiment, transformed DNA is integrated into a chromosome of a non-human host organism and/or cell such that a stable recombinant system results. Any chromosomal integration method known in the art may be used in the practice of the 10 invention, including but not limited to recombinase-mediated cassette exchange (RMCE), viral site-specific chromosomal insertion, adenovirus and pronuclear injection.

In order to carry out the method for producing α-santalene in vitro, as exposed herein above, it is very advantageous to provide a method of making at least one 15 polypeptide having an α-santalene synthase activity as described in any embodiment of the invention. Therefore, the invention provides a method for producing at least one polypeptide according to any embodiment of the invention comprising

a) culturing a non-human host organism or cell transformed with the expression vector of the invention, so that it harbors a nucleic acid according to the invention and 20 expresses or over-expresses a polypeptide of the invention;
b) isolating the polypeptide from the non-human host organism or cell cultured in step a).

According to a preferred embodiment, said method further comprises, prior to step a), transforming a non-human host organism or cell with the expression vector of the invention, so that it harbors a nucleic acid according to the invention and expresses or 25 over-expresses the polypeptide of the invention.

A nucleic acid according to any of the above-described embodiments can be used.

Transforming and culturing of the non-human host organism or cell can be carried out as described above for the method of producing α-santalene in vivo. Step b) may be performed using any technique well known in the art to isolate a particular polypeptide 30 from an organism or cell.

A “polypeptide variant" as referred to herein means a polypeptide having an α-santalene synthase activity and being substantially homologous to the polypeptide according to any of the above embodiments, but having an amino acid sequence different from that encoded by any of the nucleic acid sequences of the invention because of one or more deletions, insertions or substitutions.

Variants can comprise conservatively substituted sequences, meaning that a given amino acid residue is replaced by a residue having similar physiochemical characteristics. Examples of conservative substitutions include substitution of one aliphatic residue for another, such as Ile, Val, Leu, or Ala for one another, or substitutions of one polar residue for another, such as between Lys and Arg; Glu and Asp; or Gln and Asn. See Zubay, *Biochemistry*, 1983, Addison-Wesley Pub. Co. The effects of such substitutions can be calculated using substitution score matrices such a PAM-120, PAM-200, and PAM-250 as discussed in Altschul, J. *Mol. Biol.*, 1991, 219, 555-565. Other such conservative substitutions, for example substitutions of entire regions having similar hydrophobicity characteristics, are well known.

Naturally occurring peptide variants are also encompassed by the invention. Examples of such variants are proteins that result from alternate mRNA splicing events or from proteolytic cleavage of the polypeptides described herein. Variations attributable to proteolysis include, for example, differences in the N- or C-termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids from the polypeptides encoded by the sequences of the invention.

Variants of the polypeptides of the invention may be used to attain for example desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution, increased affinity for the substrate, improved specificity for the production of one or more desired compounds, increased velocity of the enzyme reaction, higher activity or stability in a specific environment (pH, temperature, solvent, etc), or improved expression level in a desired expression system. A variant or site directed mutant may be made by any method known in the art. Variants and derivatives of native polypeptides can be obtained by isolating naturally-occurring variants, or the nucleotide sequence of variants, of other or same plant lines or species, or by artificially programming mutations of nucleotide

sequences coding for the polypeptides of the invention. Alterations of the native amino acid sequence can be accomplished by any of a number of conventional methods.

Polypeptide variants resulting from a fusion of additional peptide sequences at the amino and carboxyl terminal ends of the polypeptides of the invention can be used to enhance expression of the polypeptides, be useful in the purification of the protein or improve the enzymatic activity of the polypeptide in a desired environment or expression system. Such additional peptide sequences may be signal peptides, for example. Accordingly, the present invention encompasses variants of the polypeptides of the invention, such as those obtained by fusion with other oligo- or polypeptides and/or those which are linked to signal peptides. Fusion polypeptide encompassed by the invention also comprise fusion polypeptides resulting from a fusion of other functional proteins, such as other proteins from the terpene biosynthesis pathway.

Therefore, in an embodiment, the present invention provides a method for preparing a variant polypeptide having an α-santalene synthase activity, as described in any of the above embodiments, and comprising the steps of:

(a) selecting a nucleic acid according to any of the embodiments exposed above;
(b) modifying the selected nucleic acid to obtain at least one mutant nucleic acid;
(c) transforming host cells or unicellular organisms with the mutant nucleic acid sequence to express a polypeptide encoded by the mutant nucleic acid sequence;
(d) screening the polypeptide for at least one modified property; and,
(e) optionally, if the polypeptide has no desired variant α-santalene synthase activity, repeating the process steps (a) to (d) until a polypeptide with a desired variant α-santalene synthase activity is obtained;
(f) optionally, if a polypeptide having a desired variant α-santalene synthase activity was identified in step d), isolating the corresponding mutant nucleic acid obtained in step (c).

According to a preferred embodiment, the variant polypeptide prepared is capable of producing α-santalene as a major product. According to an even more preferred embodiment, it is capable of producing a mixture of sesquiterpenes wherein α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the sesquiterpenes produced.

According to a more preferred embodiment, the variant polypeptide prepared has a

(+)-α-santalene synthase activity.

According to an even more preferred embodiment, the variant polypeptide prepared is capable of producing (+)-α-santalene as a major product. According to an even more preferred embodiment, it is capable of producing a mixture of sesquiterpenes wherein (+)-α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, preferably at least 92% of the sesquiterpenes produced.

In step (b), a large number of mutant nucleic acid sequences may be created, for example by random mutagenesis, site-specific mutagenesis, or DNA shuffling. The detailed procedures of gene shuffling are found in Stemmer, DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A., 1994, 91(22): 10747–1075. In short, DNA shuffling refers to a process of random recombination of known sequences in vitro, involving at least two nucleic acids selected for recombination. For example mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion, substitution, or deletion. Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered gene wherein predetermined codons can be altered by substitution, deletion or insertion.

Accordingly, the polypeptide comprising SEQ ID NO:1 may be recombined with any other sesquiterpene synthase encoding nucleic acids, for example isolated from an organism other than *Clausena lansium*. Thus, mutant nucleic acids may be obtained and separated, which may be used for transforming a host cell according to standard procedures, for example such as disclosed in the present examples.

In step (d), the polypeptide obtained in step (c) is screened for at least one modified property, for example a desired modified enzymatic activity. Examples of desired enzymatic activities, for which an expressed polypeptide may be screened, include enhanced or reduced enzymatic activity, as measured by K_M or V_{max} value, modified regio-chemistry or stereochemistry and altered substrate utilization or product distribution. The screening of enzymatic activity can be performed according to procedures familiar to the skilled person and those disclosed in the present examples.

Step (e) provides for repetition of process steps (a)-(d), which may preferably be performed in parallel. Accordingly, by creating a significant number of mutant nucleic acids, many host cells may be transformed with different mutant nucleic acids at the same time, allowing for the subsequent screening of an elevated number of polypeptides. The
5 chances of obtaining a desired variant polypeptide may thus be increased at the discretion of the skilled person.

All the publications mentioned in this application are incorporated by reference to disclose and describe the methods and/or materials in connection with which the
10 publications are cited.

Description of the drawings

Figure 1: Amino acid sequences deduced from the fragments of sesquiterpene synthases obtained from the sequencing of the C. lansium library and aligned with the
15 amino acid sequence of sesquiterpene synthase with the NCBI access No. AAK54279.

Figure 2: Comparison of the product profiles obtained from E,E-FPP with the Cont2-1, Cont2B_22, Cont2B_26 and Cont2B_29 recombinant proteins. The analysis were made by GC-MS and the total ion chromatograms are shown.

Figure 3: Identification α-santalene by comparison of the mass spectrum from the
20 peak at retention time of 12.63 minutes and the mass spectrum of an α-santalene authentic standard.

Specific embodiments of the invention or Examples

25 The invention will now be described in further detail by way of the following Examples.

Example 1

Plant material and cDNA library construction

Seeds of *Clausena lansium* (wampee) were obtained from farmers located in the Hainan province in China and particularly in the town of FuShan (ChengMai County) and the town of Yongxing (Haikou City). The seeds were germinated and the plants cultivated in a greenhouse.

Young leaves (1 to 2 cm long) were collected and used for the construction of a cDNA library. Total RNA was extracted from the leaves using the Concert™ Plant RNA Reagent from Invitrogen (Carlsbad, CA) and the mRNA were purified by oligodT-cellulose affinity chromatography using the FastTrack® 2.0 mRNA isolation Kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. A cDNA library was constructed from this mRNA and using the Marathon™ cDNA Amplification Kit (Clontech, Mountain View, CA).

Example 2

Massively parallel sequencing of the C. lansium leaf cDNA library

We used the technology of massive parallel sequencing of small DNA fragments developed by Illumina (San Diego, California) to obtain sequence information of the whole cDNA library made from wampee small leaves. This sequencing technique uses a reversible terminator-based sequencing chemistry and the Cluster Station and Genome Sequencer apparatuses developed by Solexa and Illumina (www.illumina.com).

The cDNA library (1 μg) was first loaded on an agarose gel and the bands corresponding to a size between 1.5 and 3 Kb were excised, eluted and used for the sequencing. This size enrichment avoids the dilution of the library by some cDNAs encoding for proteins involved in primary metabolism (such as for example the ribulose-1,5-bisphosphate carboxylase) which often are present in high proportion in library made from plant tissues and specially green tissues. The target cDNAs, encoding for sesquiterpene synthases, typically have a size between 1.8 and 2.5 Kb and are thus included in the size enriched library.

The Ilumina technology and equipment was set up at Fasteris SA (Geneva, Switzerland) and the preparation of the DNA sample and the sequencing were performed by Fasteris SA. The cDNA library was treated using the Genomic Sample Prep Kit (Illumina). Briefly, the DNA is fragmented by nebulization, the ends are repaired to
5 generate blunt ends, adapters are ligated to the ends of the DNA fragments and the adapter-modified DNA fragments are amplified by PCR. After controlling the quality of the library by gel electrophoresis, the generation of the DNA clusters on the flow cell and the sequencing reaction is performed on the Cluster Station and Genome Sequencer equipments. Using this technology, 1.9 millions of short sequences (reads) of at least 35
10 bases were obtained.

The Edena software (Dr David Hernandez, Genomic Research Laboratory, University of Geneva Hospitals, Geneva, Switzerland, unpublished result) was used to reassemble contiguous sequences. The five last bases were first removed from each read because of possible miss-incorporations due to the lower fidelity in the last cycles of the
15 sequencing procedure. Several sets of contigs (contiguous sequences) were generated. For each set, the contigs of minimum length of 50 bases were retained. First the software parameters were set to allow assembly with 25 bases minimum overlap and either strict (100%) or non-strict (2 bases miss-match) identity. Two sets of 3634 and 3756 contigs respectively were thus generated. Another set of 4540 contigs was generated by allowing
20 assemble with a minimum of 18 bases and non-strict overlap. The sequences of the contigs were used to search for homology with terpene synthases in publicly available protein databases using the Blastx algorithm (Altschul et al, J. Mol. Biol. 215, 403-410, 1990; http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). From the three set of contigs, 14, 15 and 14 contigs were selected. Throughout the analysis of the sequences obtained from the
25 *Clausena lansium* cDNA library, strong sequence homology was observed with sequences from citrus species, an observation consistent with the phylogenic relationship of *Clausena lansium* and Citrus species (both belonging to the Rutaceae family). Thus, the Eland software (Illumina) was used to search the non-assembled reads for DNA sequence identity with sequiterpene synthases from citrus (NCBI Accession No. CQ813507,
30 CQ813505, CQ813508, CQ813506). From this analysis, 117 reads were selected.

The selected contigs and reads were then processes using the CAP program (Huang, Genomics 14(1), 18-25, 1992) and new contigs were generated. After

confirmation of sequence homology with sesquiterpene synthases, 17 contigs of length from 30 to 436 bases were retained (see SEQ ID NO:3 to 19). The deduced sequences were aligned with a citrus sesquiterpene synthase (the *C. junos* beta-farnesene synthase, NCBI access No. AAK54279) sequence in order to map their relative position along a full-length sesquiterpene synthase sequence and evaluate the number of different sesquiterpene cDNA present (Figure 1). A set of specific oligonucleotides were designed from 6 of the 19 contigs presumably arising from distinct sesquiterpene synthases cDNAs.

Example 3

Amplification of full-length sesquiterpene synthases cDNAs

The sesquiterpene synthases-specific primers deduced from the massively parallel sequencing (Example 2) were used in combination with cDNA adaptor primers in 3'/5'RACE type PCR amplifications. The amplifications were performed using the *C. lansium* cDNA library, prepared as described above in Example 1, and the Advantage® 2 Polymerase Mix (Clontech) following the Marathon™ cDNA Amplification Kit protocol (Clontech, Mountain View, CA). The thermal Cycling conditions were as follows: 1 min at 94°C, 32 cycles of 1 min at 94°C and 3 min at 68°C, and 3 min at 68°C.

Using the FS2_cont2_F1 primer (SEQ ID NO:20), a 1049 bp DNA sequence was obtained. Analysis of the sequences of several clones obtained from this amplification showed that two sequence variants were present (Cont2_RACE_F1 (SEQ ID NO:23) and Cont2_RACE_F2 (SEQ ID NO:25)) with 96% sequence identity. Each of the two sequences corresponded to the 3'end of a sesquiterpene synthase cDNA and contained a 735 bp coding region. The two deduced amino acid sequences (SEQ ID NO:24 and 26) had 92% sequence identity to each other. With the primer FS2_cont2_R1 (SEQ ID NO:21), a 1101 bp fragment (Cont2_RACE_R, SEQ ID NO:27)) was amplified containing the start codon and encoding for the 349 N-terminal amino acids of the sesquiterpene corresponding to the contig2. Alignment of the two sequences from the 3'RACE (Cont2_RACE_F1 and Cont2_RACE_F2, SEQ ID NO:23 and 25) with the sequence from the 5'RACE (cont2_RACE_R, SEQ ID NO:27) showed an overlap of 132 bases. In this overlapping region, the Cont2_RACE_F2 and Cont2_RACE_R sequences (SEQ ID NO:25 and 27) were nearly identical (one single base difference) whereas 9 bases

differences were observed between the Cont2_RACE_F1 and Cont2_RACE_R sequences (SEQ ID NO:23 and 27). Thus the sequences Cont2_RACE_F2 (SEQ ID NO:25) and Cont2_RACE_R (SEQ ID NO:27) were used to reconstitute a full-length cDNA sequence (Cont2_RACE_1, SEQ ID NO:28) encoding for a 551 amino acids protein (SEQ ID
5 NO:29).

With the FS2_Cont10_F primer (SEQ ID NO:22) two 1342 bp sequences (Cont10_RACE_Fa and Cont10_RACE_Fb, SEQ ID NO: 30 and 31) were obtained showing significant differences (67 bp, representing 95% DNA sequence identity) and suggesting the presence of two closely related sesquiterpene synthase cDNAs. The two
10 sequences contained a 1135 bp coding region. Interestingly the sequence of Cont10_RACE_Fa (SEQ ID NO:30) was 99.9% identical to the sequence of Cont2_RACE_F2 (SEQ ID NO:25, only 1 bases difference on the 1 Kb alignment) and the sequence of Cont10_RACE_Fb (SEQ ID NO:31) was 99% identical to the sequence of Cont2_RACE_F1 (SEQ ID NO:23, only 8 bases difference on the 1 Kb alignment), thus
15 suggesting that the DNA fragments amplified with the Cont2 and Cont10 primers allowed amplifications from two related sequences with no real discrimination. Two primers (Cont2_start (SEQ ID NO:32) and Cont2_stop (SEQ ID NO:33)), which are specific to the regions of the start and the stop codons of the sequences from the 5'RACE and the 3'RACE of the cont2 and cont10 fragments, were designed in order to amplify
20 simultaneously the two or more corresponding full-length cDNAs. The primer Cont2_start (SEQ ID NO:32) was extended with the CACC sequence to allow direct insertion into the pET101/D-TOPO plasmid (Invitrogen). The amplification was first performed using the Advantage® 2 Polymerase Mix (Clontech). Each PCR mixture contained, in a total volume of 50 µL, 5 µL of Advantage® 2 PCR Buffer, 200 µM dNTPs, 200 nM each
25 oligonucleotide primer, 5 µL of 100 fold diluted cDNA and 1 µL of Advantage® 2 Polymerase Mix. The thermal cycling conditions were as follows: 2 min at 95°C; 35 cycles of 30 sec at 95°C, 30 sec at 60°C and 4 min at 72°C; and 10 min at 72°C. A second round of amplification was then performed using 5 µl of the purified PCR product from the first round of amplification and using the *Pfu* DNA polymerase (Promega), in a final volume of
30 50 µl containing 5µl of *Pfu* DNA polymerase 10X buffer, 200 µM each dNTP, 0.4 µM each forward and reverse primer, 2.9 units *Pfu* DNA polymerase. The thermal cycling

conditions were identical to the conditions used in the first round. The purified PCR products were ligated in the pET1001/D-TOPO vector following the manufacturer's instructions (Invitrogen). Several clones were selected and after sequencing of the insert, some variations in the sequences were observed. The following clones were selected:
5 Cont2-1 (SEQ ID NO:2), Cont2B_22 (SEQ ID NO:38), Cont2B_26 (SEQ ID NO:39) and Cont2B_29 (SEQ ID NO:40). The sequences of the proteins encoded by these clones are provided in SEQ ID NO:1 and 41 to 43, respectively.

Example 4

10 Heterologous expression and enzymatic activities of the recombinant sesquiterpene synthases.

The plasmids pET101 with Cont2_1 (SEQ ID NO:2), Cont2B_22 (SEQ ID NO:38), Cont2B_26 (SEQ ID NO:39) and Cont2B_29 (SEQ ID NO:40) prepared as
15 described in Example 3 were transformed into Bl21(DE3) *E. Coli* cells. Single colonies of transformed cells were used to inoculate 5 ml LB medium. After 5 to 6 hours incubation at 37°C, the culture was transferred to a 20°C incubator and left 1 hour for equilibration. Expression of the protein was then induced by the addition of 1 mM IPTG and the culture was incubated over-night at 20°C. The next day, the cells were collected by centrifugation,
20 re-suspended in 0.1 volume of 50 mM MOPSO pH 7, 10% glycerol, 1 mM DTT and lysed by sonication. The extract was cleared by centrifugation (30 min at 20,000 g), and the supernatant containing the soluble protein was used for further experiments.

The crude protein extract was used to evaluate the enzymatic activity. The enzymatic assay was performed in a Teflon sealed glass tube using 50 to 100 μl of protein
25 extract in a final volume of 1 mL of 50 mM MOPSO pH 7, 10% glycerol supplemented with 1 mM DTT, 20 mM $MgCl_2$ and 50 to 200 μM purified E,E-farnesyl diphosphate (FPP) (prepared as described by Keller and Thompson, J. Chromatogr 645(1), 161-167, 1993). The tube was incubated 18 to 24 hours at 30°C and the enzyme products were extracted twice with one volume of pentane. After concentration under a nitrogen flux, the
30 extract was analyzed by GC and the identity of the products was confirmed by GC-MS based on the concordance of the retention indices and mass spectra of authentic standards. The GC-MS analysis was performed on a Hewlett-Packard 6890 Series GC system

equipped with a flame ionization detector using a 0.25 mm inner diameter by 30 m SPB-1 capillary column (Supelco, Bellefonte, PA). The carrier gas was He at a constant flow of 1.5 mL/min. The initial oven temperature was 80°C followed by a gradient of 10°C/min to 280°C. The spectra were recorded at 70eV with an electron multiplier voltage of 2200V.

5 The assay revealed the formation of (+)-α-santalene as a major product (92.7% of the total sesquiterpenes produced) and traces amounts of five additional sesquiterpenes accounting for 4.8 to 0.95% of the enzyme products. (+)-α-santalene was identified with GC-MS analysis by coincidence of the mass spectrum and of the retention index with published values (Joulain, D., and König, W.A. The Atlas of Spectral Data of 10 Sesquiterpene Hydrocarbons, EB Verlag, Hamburg, 1998). The The identification of (+)-α-santalene was further confirmed by ^{1}H NMR, ^{13}C NMR and by measurement of the optical rotation. To produce sufficient quantities for these measurements, the enzymatic assay described above was scaled up to 1 L. The enzyme products were extracted with an equal volume of pentane, concentrated and the sesquiterpene hydrocarbons fraction 15 (5.5 mg) purified by filtration on a short silica column. spectral data obtained with Cont2_1 is provided in Figure 2.

The NMR spectrum was recorded on a Bruker-Avance-500 spectrometer. The NMR data is the following :

20 ^{1}H NMR (500.13 MHz, $CDCl_3$): δ 0.82 (s, 2H), 0.83 (s, 3H), 0.99 (s, 3H), 1.00-1.08 (m, 2H), 1.08-1.26 (m, 2H), 1.57-1.63 (m, 6H), 1.68 (s, 3H), 5.12 (t × q, $J = 7.2, 1.4$ Hz, 1H)

^{13}C NMR (125.76 MHz, $CDCl_3$): δ 10.7 (q), 17.5 (q), 19.6 (d), 23.3 (t), 25.7 (q), 27.4 (s), 31.0 (t), 31.5 (t), 34.6 (t), 38.2 (d), 45.9 (s), 125.5 (d), 130.8 (s);

25

The fact that the (+)-α-santalene stereoisomer was produced has been evidenced by measuring the optical rotation (as measured on a Perkin-elmer 241 polarimeter): $[\alpha]_D^{20} = +12.0$ (C = 0.3, $CHCl_3$).

Example 5

In-vivo production of (+)-α-santalene in E coli

The use of the *C. lansium* santalene synthase for the *in-vivo* production of
5 sesquiterpenes in *E coli* cells was evaluated by co-expressing the sesquiterpene synthase with a FPP synthase and the enzymes of a four step biosynthetic pathway allowing the conversion of mevalonate to FPP. The mevalonate pathway genes were organized in a single operon and encoded for a mevalonate kinase (mvaK1), a phosphomevalonate kinase (mvaK2), a mevalonate diphosphate decarboxylase (MvaD) and an isopentenyl
10 diphosphate isomerase (idi), all the enzymes converting exogenous mevalonate to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), the two substrates of the FPP synthase. The co-expression of this partial mevalonate pathway was used to increase the amount of intracellular FPP available for the sesquiterpene synthase and thus the quantities of sesquiterpene produced.

15 The yeast FPP synthase gene (Accession number J05091) was amplified from *S. cerevisiae* genomic DNA using the primers FPPy_NcoI (SEQ ID NO:34) and FPPy-Eco (SEQ ID NO:35). The genomic DNA was isolated from *S. cerevisiae* using the Qiagen RNA/DNA Maxi Kit (Qiagen AG, Basel, Switzerland). The PCR was performed with the *Pfu* DNA polymerase (Promega AG, Dubendorf, Switzerland) in a final volume of
20 50 μl containing 0.4 μl of each primer, 200 μM dNTPs, 0.5 μl DNA polymerase 5 μl *S. cerevisiae* genomic DNA. The PCR cycling condition were as follows: 90 sec at 95°C; 28 cycles of 45 sec at 95°C, 30 sec at 54°C and 4 min at 72°C; 10 min at 72°C. The amplified DNA was ligated as NdeI-EcoRI fragment in the first multi cloning site (MCS1) of the pACYCDuet-1 plasmid (Novagen, Madison, WI) providing the plasmid
25 pACYCDuet-FPPs harbouring the FPPs gene under the control of a T7 promoter.

An operon containing the genes encoding for mvaK1, mvaK2, MvaD and idi was amplified from genomic DNA of *Streptococcus pneumoniae* (ATCC BAA-334, LGC Standards, Molsheim, France) with the primers MVA-up1-start (SEQ ID NO:36) and MVA-up2-stop (SEQ ID NO:37). The PCR was performed using the PfuUltra™ II Fusion
30 HS DNA polymerase (Stratagene, Agilent Technologies Inc., Santa Clara, CA, USA). The composition of the PCR mix was according to the manufacturer's instructions. The thermal cycling conditions were 2 min at 95°C; 30 cycles of 20 sec at 95°C, 20 sec at 58°C

and 90 sec at 72°C; and 3 min at 72°C. The 3.8 Kb fragment was purified on an agarose gel and ligated using the In-Fusion™ Dry-Down PCR Cloning Kit (Clontech Laboratories) into the second MCS of the pACYCDuet-FPPs plasmid digested with *Nde*I and *Xho*I providing the plasmid pACYCDuet-4506. The sequences of the two inserts were fully
5 sequenced to exclude any mutation.

BL21 Star™(DE3) *E. coli* cells (Invitrogen, Carlsbad, CA) were transformed with the plasmids pET101-cont2_1 (SEQ ID NO:2) prepared as described in Example 3 and with the plasmid pACYCDuet-4506. Transformed cells were selected on carbenicillin (50 μg/ml) and chloramphenicol (34 μg/ml) LB-agarose plates. Single colonies were used
10 to inoculate 5 mL liquid LB medium supplemented with the same antibiotics. The culture was incubated overnight at 37°C. The next day 2 mL of TB medium supplemented with the same antibiotics were inoculated with 0.2 mL of the overnight culture. After 6 hours incubation at 37°C, the culture was cooled down to 28°C and 1 mM IPTG, 2 mg/mL mevalonate (prepared by dissolving mevalonolactone (Sigma) in 0.5N NaOH at a
15 concentration of 1 g/mL and incubating the solution for 30 min at 37°C) and 0.2 ml decane were added to each tube. The cultures were incubated for 48 hours at 28°C. The cultures were then extracted twice with 2 volumes of ethyl-acetate, the organic phase was concentrated to 500 μL and analyzed by GC-MS as described above in Example 4. In these conditions the cells produced *(+)*-α-santalene at 250 mg/L culture in 48 hours.

20 This example shows that an E. coli cell transformed with an α-santalene synthase, as defined in the present invention, is capable of producing α-santalene. The other enzymes with which the E. coli cell is transformed are not essential for the production of α-santalene. Indeed α-santalene is also produced when an E. coli cell is transformed with the α-santalene synthase only, but in lower amounts. The other enzymes with which the E. coli
25 cell is transformed are added for the only purpose of increasing the amount of precursor available to the α-santalene synthase.

Claims

1. A method for producing α-santalene comprising

a) contacting FPP with at least one polypeptide having an α-santalene synthase activity and comprising an amino acid sequence at least 50% identical to SEQ ID NO:1;

b) optionally, isolating the α-santalene produced in step a).

2. The method of claim 1, wherein step a) comprises cultivating a non-human host organism or cell capable of producing FPP and transformed to express at least one polypeptide comprising an amino acid sequence at least 50% identical to SEQ ID NO:1 and having an α-santalene synthase activity, under conditions conducive to the production of α-santalene.

3. The method of claim 2, wherein the method further comprises, prior to step a), transforming a non human host organism or cell capable of producing FPP with at least one nucleic acid encoding a polypeptide comprising an amino acid sequence at least 50% identical to SEQ ID NO:1 and having an α-santalene synthase activity, so that said organism expresses said polypeptide.

4. The method of claim 3, wherein the at least one nucleic acid encoding the α-santalene synthase comprises a nucleotide sequence at least 50%, preferably at least 70%, preferably at least 90% identical to SEQ ID NO:2 or the complement thereof.

5. The method of claim 4, wherein the nucleic acid comprises the nucleotide sequence SEQ ID NO:2 or the complement thereof.

6. The method of claim 5, wherein the nucleic acid consists of SEQ ID NO:2 or the complement thereof.

7. The method of claim 3 or 4, wherein the at least one nucleic acid used in any of the above embodiments comprises a nucleotide sequence that has been obtained by modifying SEQ ID NO:2 or the complement thereof.

8. The method of any of claims 2 to 7, wherein the non-human host organism is a plant, a prokaryote or a fungus.

5 **9.** The method of any of claims 2 to 7, wherein the non-human host organism is a microorganism, preferably a bacteria or a yeast

10. The method of claim 9, wherein said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*

10

11. The method of any of claims 2 to 7, wherein said non-human host cell is a plant or a fungal cell.

12. The method of any of the preceding claims, for producing α-santalene as a 15 major product.

13. The method of claim 12, wherein α-santalene represents at least 60%, preferably at least 80%, preferably at least 90%, of the sesquiterpenes obtained.

20 **14.** The method according to any of the preceding claims, for producing (+)-α-santalene and wherein the polypeptide having an α-santalene synthase activity has a (+)-α-santalene synthase activity.

15. The method of claim 14, wherein (+)-α-santalene is obtained as a major 25 product.

16. The method of claim 15, wherein (+)-α-santalene represents at least 60%, preferably at least 80%, preferably at least 90% of the sesquiterpenes obtained.

30 **17.** The method of any of the preceding claims, wherein the at least one polypeptide comprises an amino acid sequence at least 70%, preferably at least 90% identical to SEQ ID NO:1.

18. The method of claim 17, wherein the at least one polypeptide comprises the amino acid sequence SEQ ID NO:1.

5 **19.** The method of claim 18, wherein the at least one polypeptide consists of SEQ ID NO:1

20. The method of any of claims 1 to 4 or 7 to 17, wherein the at least one polypeptide comprises an amino acid sequence that has been obtained by modifying SEQ 10 ID NO:1.

21. A polypeptide having an α-santalene synthase activity and comprising an amino acid sequence at least 50% identical to SEQ ID NO:1.

15 **22.** The polypeptide of claim 21, wherein said polypeptide is capable of producing α-santalene as a major product.

23. The polypeptide of claim 22, wherein said polypeptide is capable of producing a mixture of sesquiterpenes wherein α-santalene represents at least 60%, 20 preferably at least 80%, preferably at least 90% of the sesquiterpenes produced.

24. The polypeptide of claim 21, wherein said polypeptide has a (+)-α-santalene synthase activity.

25 **25.** The polypeptide of claim 24, wherein said polypeptide is capable of producing (+)-α-santalene as a major product.

26. The polypeptide of claim 25, wherein said polypeptide is capable of producing a mixture of sesquiterpenes wherein (+)-α-santalene represents at least 60%, 30 preferably at least 80%, preferably at least 90% of the sesquiterpenes produced.

27. The polypeptide of any of claims 21 to 26, wherein said polypeptide comprises an amino acid sequence at least 70%, preferably at least 90% identical to SEQ ID NO:1.

5 **28.** The polypeptide of claim 27, wherein said polypeptide comprises the amino acid sequence SEQ ID NO:1.

29. The polypeptide of claim 28, wherein said polypeptide consists of SEQ ID NO:1.

10

30. The polypeptide of any of claims 21 to 27, wherein said polypeptide comprises an amino acid sequence that has been obtained by modifying SEQ ID NO:1.

31. A nucleic acid encoding a polypeptide according to any of claims 21 to 30.

15

32. The nucleic acid of claim 31, comprising a nucleotide sequence at least 50%, preferably at least 70%, preferably at least 90% identical to SEQ ID NO:2 or the complement thereof.

20 **33.** The nucleic acid of claim 32, comprising the nucleotide sequence SEQ ID NO:2 or the complement thereof.

34. The nucleic acid of claim 33, consisting of SEQ ID NO:2 or the complement thereof.

25

35. The nucleic acid of claim 31 or 32, comprising a nucleotide sequence that has been obtained by modifying SEQ ID NO:2 or the complement thereof.

36. An expression vector comprising the nucleic acid of any of claims 31 to 35.

30

37. The expression vector of claim 36, in the form of a viral vector, a bacteriophage or a plasmid.

38. The expression vector of claim 36 or 37, including the nucleic acid of the invention operably linked to at least one regulatory sequence which controls transcription, translation initiation or termination, such as a transcriptional promoter, operator or enhancer or an mRNA ribosomal binding site and, optionally, including at least one selection marker.

39. A non-human host organism or cell transformed to harbor at least one nucleic acid according to any of claims 31 to 35, so that it heterologously expresses or over-expresses at least one polypeptide according to any of claims 21 to 30.

40. The non-human host organism of claim 39, wherein said non-human host organism is a plant, a prokaryote or a fungus.

41. The non-human host organism of claim 39, wherein said non-human host organism is a microorganism, preferably a bacteria or yeast.

42. The non-human host organism of claim 41, wherein said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*.

43. The higher eukaryotic cell of claim 39, wherein said higher eukaryotic cell is a plant cell or a fungal cell.

44. A method for producing at least one polypeptide according to any of claims 21 to 30 comprising

a) culturing a non-human host organism or cell transformed with the expression vector of any of claims 36 to 38, so that it harbors a nucleic acid according to any of claims 31 to 35 and expresses or over-expresses a polypeptide according to any of claims 21 to 30;
b) isolating the polypeptide from the non-human host organism or cell cultured in step a).

45. The method of claim 44, further comprising, prior to step a), transforming a non-human host organism or cell with the expression vector of any of claim 36 to 38, so that it harbors a nucleic acid according to any of claims 31 to 35 and expresses or over-expresses the polypeptide of any of claims 21 to30.

5

46. A method for preparing a variant polypeptide having an α-santalene synthase activity comprising the steps of:

(a) selecting a nucleic acid according to any of the claims 31 to 35;

(b) modifying the selected nucleic acid to obtain at least one mutant nucleic acid;

10 (c) transforming host cells or unicellular organisms with the mutant nucleic acid sequence to express a polypeptide encoded by the mutant nucleic acid sequence;

(d) screening the polypeptide for at least one modified property; and,

(e) optionally, if the polypeptide has no desired variant α-santalene synthase activity, repeating the process steps (a) to (d) until a polypeptide with a desired variant α-

15 santalene synthase activity is obtained;

(f) optionally, if a polypeptide having a desired variant α-santalene synthase activity was identified in step d), isolating the corresponding mutant nucleic acid obtained in step (c).

20 **47.** A method according to claim 46, wherein the variant polypeptide prepared is capable of producing α-santalene as a major product.

48. A method according to claim 47, wherein the variant polypeptide prepared is capable of producing a mixture of sesquiterpenes wherein α-santalene represents at least

25 60%, preferably at least 80%, preferably at least 90% of the sesquiterpenes produced.

49. A method according to claim 46, wherein the variant polypeptide prepared has a (+)-α-santalene synthase activity.

30 **50.** A method according to claim 49, wherein the variant polypeptide prepared is capable of producing (+)-α-santalene as a major product.

51. A method according to claim 50, wherein the variant polypeptide prepared is capable of producing a mixture of sesquiterpenes wherein (+)-α-santalene represents at least 60%, preferably at least 80%, preferably at least 90% of the sesquiterpenes produced.

5

Figure 1

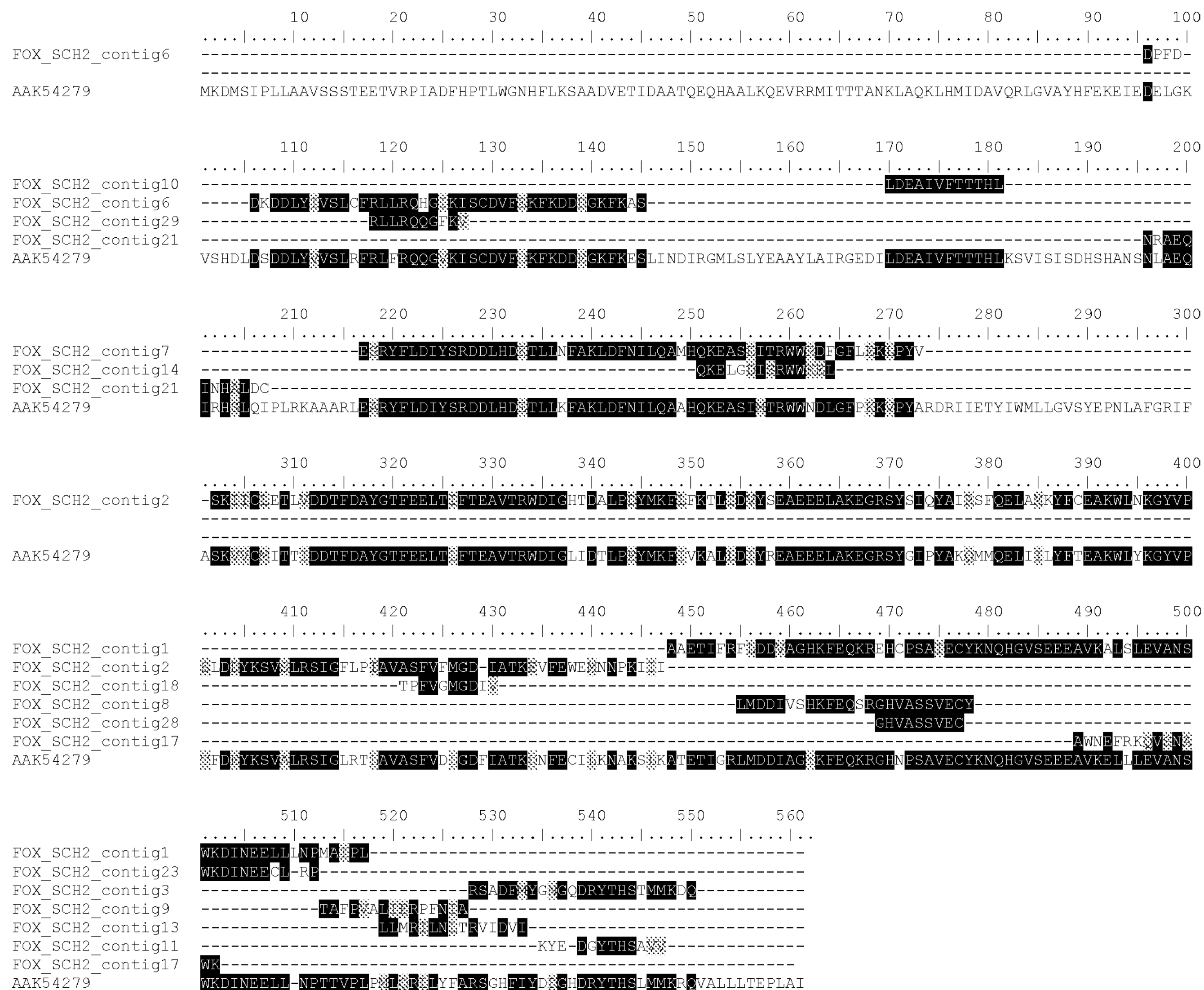


Figure 2

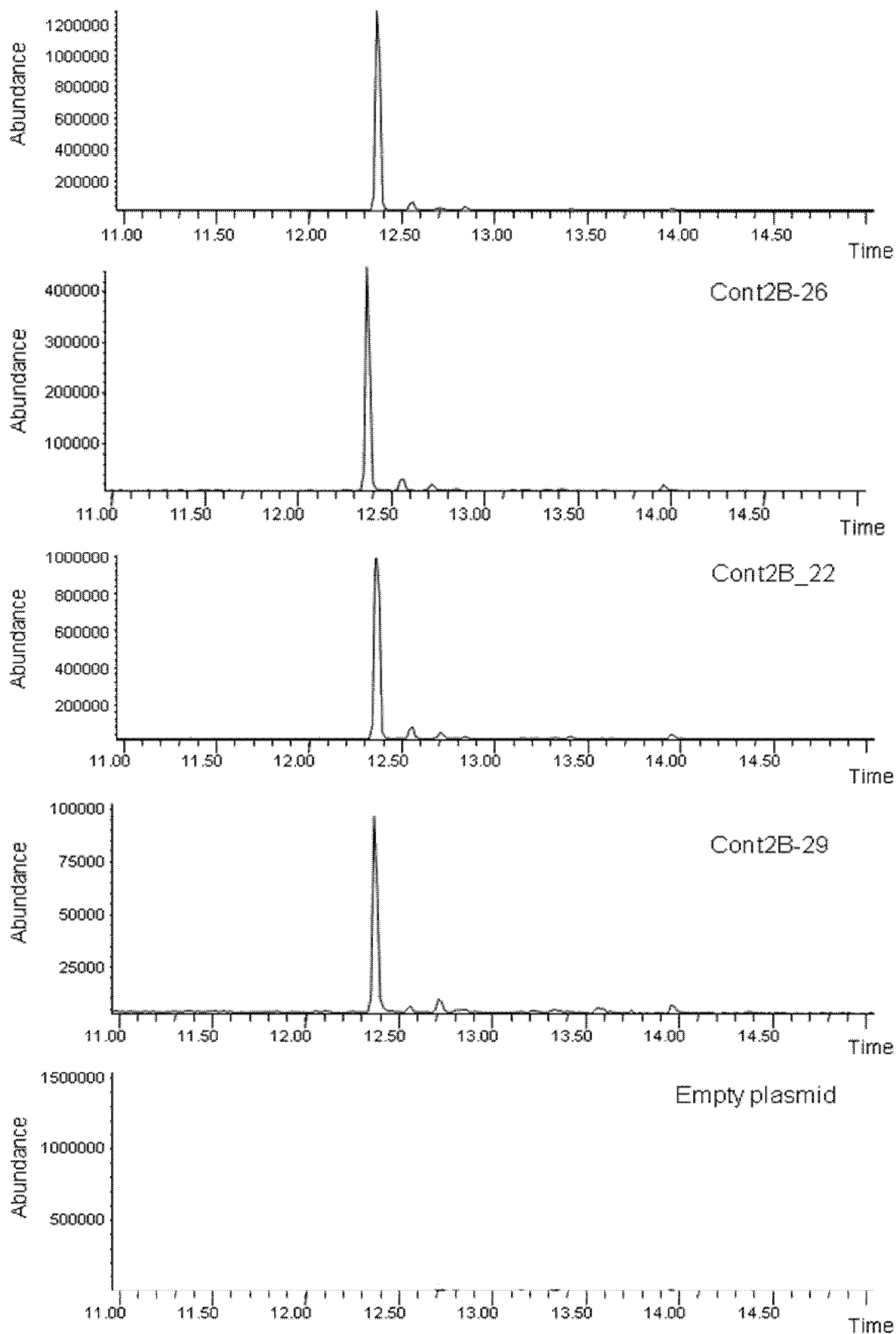
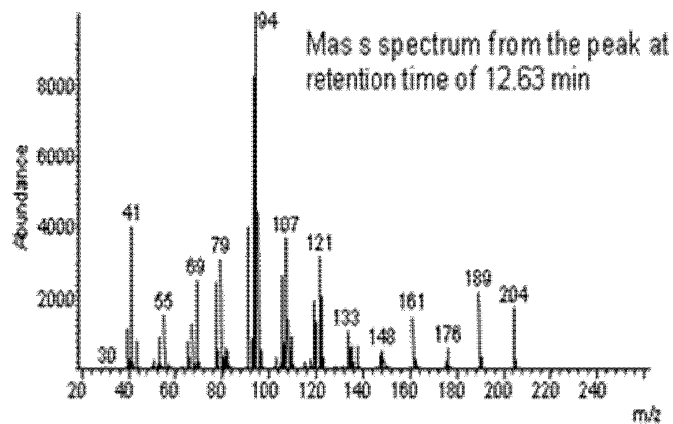
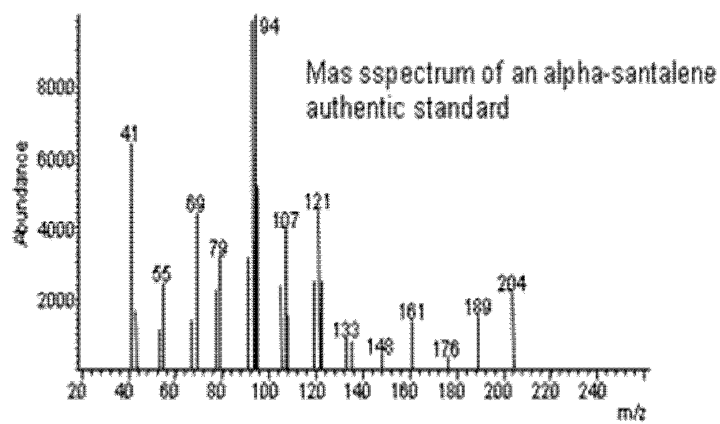


Figure 3





INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/052546

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/52 C12N9/88 C12P7/22 C12P7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, CHEM ABS Data, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARUYAMA T ET AL: "Molecular cloning, functional expression and characterization of (E)-beta-farnesene synthase from Citrus junos" BIOLOGICAL & PHARMACEUTICAL BULLETIN (OF JAPAN), PHARMACEUTICAL SOCIETY OF JAPAN, TOKYO, JP, vol. 24, no. 10, 1 October 2001 (2001-10-01), pages 1171-1175, XP002282504 ISSN: 0918-6158 cited in the application the whole document ----- -/--	21-27, 30-32, 35-45

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 May 2009	15/05/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Espen, Josée

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2005/021705 A (ISRAEL STATE [IL]; EYAL YORAM [IL]; LEWINSOHN EFRAIM [IL]) 10 March 2005 (2005-03-10) sequences 1,2	21,22, 24,25, 31,32, 36-45
X	WO 2004/031376 A (FIRMENICH & CIE [CH]; SCHALK MICHEL [FR]; CLARK ANTHONY [FR]) 15 April 2004 (2004-04-15) sequence 8	21,22, 24,25, 31,32, 36-45
P,X	WO 2008/142318 A (LIBROPHYT [FR]; SALLAUD CHRISTOPHE [FR]; RONTEIN DENIS [FR]; TISSIER A) 27 November 2008 (2008-11-27) cited in the application page 13 - page 30	46-51
A	VAN DER HOEVEN, RUTGER S. ET AL: "Genetic control and evolution of sesquiterpene biosynthesis in Lycopersicon esculentum and L. hirsutum" PLANT CELL , 12(11), 2283-2294 CODEN: PLCEEW; ISSN: 1040-4651, 2000, XP002495621 page 2286, right-hand column - page 2287; figure 4	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2009/052546

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005021705 A	10-03-2005	AU 2004269197 A1	10-03-2005
		CA 2537505 A1	10-03-2005
		EP 1659996 A2	31-05-2006
		JP 2007517496 T	05-07-2007
WO 2004031376 A	15-04-2004	BR 0314979 A	02-08-2005
		CN 1703507 A	30-11-2005
		EP 1554379 A2	20-07-2005
		JP 2006500942 T	12-01-2006
WO 2008142318 A	27-11-2008	FR 2914652 A1	10-10-2008