METHOD FOR PRODUCING SCLAREOL

The present invention provides a method of producing sclareol, said method comprising contacting a particular polypeptide having a sclareol synthase activity with labdenediol diphosphate (LPP). In particular, said method may be carried out in vitro or in vivo to produce sclareol, a very useful compound in the fields of perfumery and flavoring. The present invention also provides the amino acid sequence of the polypeptide used in the method. A nucleic acid derived from Salvia sclarea and encoding the polypeptide of the invention, an expression vector containing said nucleic acid, as well as a non-human organism or a cell transformed to harbor the same nucleic acid, are also part of the present invention.
METHOD FOR PRODUCING SCLAREOL

Technical field

The present invention provides a method of producing sclareol, said method comprising contacting a particular polypeptide having a sclareol synthase activity with labdenediol diphosphate (LPP). In particular, said method may be carried out in vitro or in vivo to produce sclareol, a very useful compound in the fields of perfumery and flavoring. The present invention also provides the amino acid sequence of the polypeptide used in the method. A nucleic acid derived from Salvia sclarea and encoding the polypeptide of the invention, an expression vector containing said nucleic acid, as well as a non-human organism or a cell transformed to harbor the same nucleic acid, are also part of the present invention.

Prior art

Sclareol is one member of the terpenoids or terpenes family, comprising a high number of natural products. 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. Diterpenes, for example, are widely found in the plant kingdom and over 2500 diterpene structures have been described (Connolly and Hill, Dictionary of terpenoids, 1991, Chapman & Hall, London). Terpene molecules have been of interest for thousands of years because of their flavor and fragrance properties and their cosmetic, medicinal and antimicrobial effects. 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. These enzymes convert a precursor in one or more terpene products. Most of the time, the precursor is an acyclic terpene precursor and, in particular, most diterpene synthases catalyze the cyclization of the acyclic precursor geranylgeranyl pyrophosphate. Nevertheless, in some special cases, terpene synthases catalyze the transformation of an already cyclic molecule into one or more terpene products.
Two types of cyclization mechanisms occur in nature and are related to two types of diterpene syntheses which can be classified into class I and class II diterpene synthases (Wendt and Schulz, 1998, Structure. 6(2):127-33). For some diterpenes, the cyclization mechanism is similar to those of monoterpenes and sesquiterpenes as it is initiated by the ionization of the diphosphate ester function of GGPP, followed by the reaction of the resulting carbocation with an internal double bond. The diterpene synthases catalysing this type of cyclization are class I diterpene synthases. The second mode of cyclization in the biosynthesis of diterpenes, catalyzed by class II diterpene synthases, is initiated by the protonation of the terminal double bond of GGPP and leads, after internal rearrangement and proton elimination, to a cyclic diterpene diphosphate intermediate.

Genes and cDNAs encoding diterpene synthases from each of the two classes have been cloned and the recombinant enzymes characterized. The availability of genes encoding different types of diterpene syntheses provides information on the primary structures of the enzymes. Some amino acid motifs are conserved in diterpene syntheses and are related to either the protonation or the ionization dependent cyclization. A DDxxD motif is found in several class I diterpene synthases. Said motif is probably involved in binding and ionization of the diphosphate moiety. In class II synthases, a conserved DxDD motif is found, in which the second aspartate residue is involved as proton donor.

Sclareol is a naturally occurring diterpene molecule extensively used as starting material for the synthesis of fragrance molecules with ambergris notes. These syntheses were developed to provide an alternative to ambergris, a waxy substance secreted by the intestines of sperm whale. Ambergris is highly appreciated for its pleasant odor and has been historically used as a perfume ingredient. Due to its high price and the increasing demand for ambergris, and particularly due to the protection of the whale species, chemical synthesis of ambergris constituents and molecules with ambergris character have been developed. Amongst these molecules, Ambrox® (registered trademark of Firmenich SA, Switzerland) is the most largely appreciated substitute for Ambergris. The most widely used starting material for the synthesis of Ambrox® is the diterpene-diol sclareol.

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.
leading to variability from year to year, rendering the use of such ingredients in high quality perfumery very difficult or even impossible some years. Therefore, it would be an advantage to provide a source of sclareol, which is less subjected to fluctuations in availability and quality. Chemical synthesis would seem to be an evident option for the preparation of sclareol. However, given its highly complex structure, an economic synthetic process for the preparation of sclareol is still difficult. A biochemical pathway leading to the synthesis of sclareol would therefore be of great interest.

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, \textit{Nat. Prod. Rep.}, 2002, 19, 181-222, which reviews the state of the art of terpene biosynthetic pathways.

Several diterpene synthases have already been identified. In particular, US 7,238,514 discloses a number of diterpene synthases, the nucleic acids encoding them, as well as unicellular organisms transformed to express each of these synthases together with a GGPP synthase, thus producing diterpenes in vivo. Nevertheless, no method for the biosynthetic production of sclareol using a polypeptide having a sclareol synthase activity as provided herein is specifically disclosed in that patent. The amino acid and nucleotide sequences disclosed in it are very different from the sequences of the present invention. Among the diterpene synthases described in that document, the closest to the polypeptides of the present invention are a \textit{Cucumis sativus} mRNA for an ent-kaurene synthase designated by SEQ ID NO:389 in US 7,238,514 and a \textit{Cucurbita maxima} mRNA for an ent-kaurene synthase B designated by SEQ ID NO:395 in US 7,238,514 and by the accession number AAB39482.1. These polypeptides and the one of the invention only share 32\% identity. Moreover, there is no suggestion in this prior art document that the described diterpene synthases are useful for the production of sclareol.

Terpene synthases having a certain percentage of sequence identity with the sequences of the present invention have also been found in the sequences databases. Nevertheless, the percentage of identity between the known diterpene synthases and the polypeptides of the invention is very low. The closest synthases to the ones of the invention are a terpenoid cyclase of undefined function (Accession number NCBI AAS98912) having 36\% identity with the polypeptide of the invention, an ent-kaurene synthase of \textit{Cucumis sativus} (accession number BAB19275) having 32\% identity with the polypeptide of the invention, an ent-cassadiene synthase from \textit{Oryza sativa} (accession
number ABH10734 and published in Xu, Wilderman, Morrone, Xu, Roy, Margis-Pinheiro, Upadhyaya, Coates and Peters, Functional characterization of the rice kaurene synthase-like gene family, Phytochemistry, 68(3), 2007, 312-326) having 32% identity with the polypeptide of the invention and an ent-kaurene synthase from Oryza sativa (accession number AAQ72559 and published in Margis-Pinheiro, Zhou, Zhu, Dennis and Upadhyaya, Isolation and characterization of a DS-tagged rice (Oryza sativa L.) GA-responsive dwarf mutant defective in an early step of the gibberellins biosynthesis pathway, Plant Cell Rep., 23(12), 2005, 819-833) having 32% identity with the polypeptide of the invention. The potential ability of any of these sequences to catalyze the production of sclareol is never mentioned in the prior art.

In addition to the difference between the sequences themselves, it also has to be pointed out that the structure and the properties of ent-kaurene and ent-cassadiene are very different from those of sclareol. In particular, ent-kaurene is a tricyclic diterpene which does not contain any alcohol functional groups, unlike sclareol, which is a bicyclic diol. Moreover, ent-kaurene, which is a precursor of a plant hormone regulating growth, is of no use in the field of perfumery and flavoring, whereas sclareol is of high interest in these technical fields, as explained above.

One document of the prior art relates specifically to a sclareol synthase (Banthorpe, Brown and Morris, Partial purification of farnesyl pyrophosphate: Drimenol cyclase and geranylgeranyl pyrophosphate: Sclareol cyclase, using cell culture as a source of material, Phytochemistry 31, 1992, 3391-3395). In this reference, a partially purified protein from Nicotiana glutinosa is identified as a sclareol synthase, but no indication is given regarding the amino acid sequence of that protein, the nucleotide sequence of the nucleic acid encoding it or the use of that protein in a method for the biosynthesis of sclareol in vitro or in vivo.

Despite extensive studies of terpene cyclization, the isolation and characterization of the enzymes is still difficult, particularly in plants, due to their low abundance, their often transient expression patterns, and the complexity of purifying them from the 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 sclareol in an economic way, as indicated above. Accordingly, the present invention has the objective to produce diterpenes while having little waste, a more energy and resource efficient process and while reducing dependency on fossil fuels. It is a further objective to
provide enzymes capable of synthesizing sclareol, which is useful as perfumery and/or aroma ingredients.

**Abbreviations Used**

- **bp**: base pair
- **kb**: kilo base
- **BSA**: bovine serum albumine
- **DNA**: deoxyribonucleic acid
- **cDNA**: complementary DNA
- **dT**: deoxy thymine
- **dNTP**: deoxy nucleotide triphosphate
- **DTT**: dithiothreitol
- **GC**: gaseous chromatograph
- **GGPP**: Geranylgeranyl pyrophosphate
- **IPTG**: isopropyl-D-thiogalacto-pyranoside
- **LB**: lysogeny broth
- **LPP**: labdenediol diphosphate
- **MOPSO**: 3-(N-morpholino)-2-hydroxypropanesulfonic acid
- **MS**: mass spectrometer
- **ORF**: open reading frame
- **PCR**: polymerase chain reaction
- **RMCE**: recombinase-mediated cassette exchange
- **RT-PCR**: reverse transcription – polymerase chain reaction
- **3’-5’-RACE**: 3’ and 5’ rapid amplification of cDNA ends
- **RNA**: ribonucleic acid
- **mRNA**: messenger ribonucleic acid
- **nt**: nucleotide
- **RNase**: ribonuclease
- **RuBisCO**: ribulose-1,5-bisphosphate carboxylase
- **SDS-PAGE**: SDS-polyacrylamid gel electrophoresis
- **SsLPPs**: Salvia sclarea labdenediol diphosphate synthase
- **UTR**: Untranslated Region
**Description of the invention**

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

One object of the present invention is therefore a method for producing sclareol comprising:

a) contacting labdenediol diphosphate (LPP) with at least one polypeptide having a sclareol synthase activity and comprising an amino acid sequence at least 50\% identical to SEQ ID NO:1; and

b) optionally, isolating the sclareol produced in step a).

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

Sclareol and LPP are defined by the way of their formulae as represented in Figure 1.

As a “sclareol synthase” or as a “polypeptide having a sclareol synthase activity”, we mean here a polypeptide capable of catalyzing the synthesis of sclareol starting from (LPP). The ability of a polypeptide to catalyze the synthesis of sclareol can be confirmed by performing the enzyme assay as detailed in the Examples.

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

According to a preferred embodiment, the method for producing sclareol comprises contacting LPP with a polypeptide having a sclareol synthase activity and comprising 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, 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.

According to a preferred embodiment, the sclareol synthase is a truncated polypeptide comprising an amino acid sequence at least 50\% identical to SEQ ID NO:102. Preferably 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:96. According to another preferred embodiment, the polypeptide comprises the amino acid sequence SEQ ID NO:96. According to a more preferred embodiment, the polypeptide consists of SEQ ID NO:96.

The percentage of identity between two peptidic or nucleotidic sequences is a function of the number of amino acids or nucleic acids residues that are identical in the two sequences when an alignment of these two sequences has been generated. Identical 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 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 purposes of determining the percentage of amino acid or nucleic acid sequence identity can be achieved in various ways using computer programs and for 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 alignment of peptidic or nucleotidic sequences and to calculate the percentage of sequence identity.

The polypeptide to be contacted with LPP in vitro can be obtained by extraction 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 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 polypeptides, 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 pH. If adequate, salts, DTT, 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.

LPP 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 25 and 35°C, more preferably at 30°C. After incubation, the sclareol 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.

LPP can be obtained by contacting GGPP with an isolated LPP synthase. Examples 1 to 3 below show a method to isolate a LPP synthase encoding cDNA from *Salvia sclarea*, a method for the heterologous expression of said cDNA in E. coli, a method for the purification of the LPP synthase so produced and a method for the in vitro production of LPP using the isolated LPP synthase.

According to another preferred embodiment, the method for producing sclareol is carried out in vivo. In this case, step a) of the above-described method comprises cultivating a non-human organism or cell capable of producing LPP and transformed to express a polypeptide having a sclareol synthase activity and comprising an amino acid sequence at least 70% identical to SEQ ID NO:1 under conditions conducive to the production of sclareol.

According to a more preferred embodiment, the method further comprises, prior to step a), transforming a non human organism or cell capable of producing LPP with at least one nucleic acid encoding a polypeptide having a sclareol synthase activity and comprising an amino acid sequence at least 70% identical to SEQ ID NO:1, so that said organism expresses said polypeptide.

According to a preferred embodiment, the nucleic acid used to transform the host organism or cell 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 another preferred embodiment, the nucleic acid comprises the nucleotide sequence SEQ ID NO:2 or the complement thereof. According to a more preferred embodiment, the nucleic acid consists of SEQ ID NO:2 or the complement thereof.

According to a further preferred embodiment, the nucleic acid used to transform the host organism or cell is a truncated nucleic acid comprising a nucleotide sequence at least 50%, preferably at least 55%, preferably at least 60%, 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:93 or the complement thereof. According to another preferred embodiment, the nucleic acid comprises the nucleotide sequence SEQ ID NO:93 or the complement thereof. According to a more preferred embodiment, the nucleic acid consists of SEQ ID NO:93 or the complement thereof.

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.

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 what is measured in a non-transformed organism or cell. A more detailed description of suitable methods to transform a non-human organism or cell will be described later on in the part of the specification that is dedicated to such transformed non-human 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 LPP” when it produces LPP naturally or when it does not produce LPP naturally but produces GGPP (or is so transformed) and is transformed to express a LPP synthase, either prior to the transformation with a nucleic acid encoding a sclareol synthase or together with said nucleic acid. Organisms or cells transformed to produce a higher amount of LPP than the naturally occurring organism or cell are also encompassed by the “organisms or cells
capable of producing LPP”. According to a preferred embodiment, the organism accumulates LPP naturally or is transformed to accumulate LPP.

Methods for transforming organisms so that they express a LPP synthase, can be any method known in the art to transform a host organism. Such methods are exposed in more details later on and a specific example of the expression of a LPP synthase in E. coli is given in Example 2. Methods for transforming an organism to produce GGPP are also known in the art. Such methods can for example be found in Huang, Roessner, Croteau and Scott, Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol, Bioorg Med Chem., 9(9), 2001, 2237-2242.

To carry out the invention in vivo, the host organism or cell is cultivated under conditions conducive to the production of sclareol. Accordingly, if the host is a 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 sclareol 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 sclareol synthesis. Optimal culture conditions are described in a more detailed manner in the following Examples.

Non-human 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 organism used to carry out the invention in vivo is a plant, a prokaryote or a fungus.

Any plant, prokaryote or fungus may be used to carry out the method of the invention in vivo. 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 organism is a microorganism. According to an even more preferred embodiment said microorganism is a bacteria or a fungus, preferably said fungus is yeast. Most preferably, said bacteria is E. coli and said yeast is Saccharomyces cerevisiae.

Most of these organisms do not produce LPP naturally. To be suitable to carry out
the method of the invention, these organisms have to be transformed to produce said precursor. They can be so transformed either before the modification with the nucleic acid encoding the polypeptide having a sclareol synthase activity 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 non-human cell, but are preferably plant cells.

According to another preferred embodiment, the polypeptide or the nucleic acid used in the method of any of the embodiments above is derived from *Salvia sclarea*.

An important tool to carry out the method of the invention is the polypeptide itself. A polypeptide having a sclareol 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 sclareol synthase 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 another preferred embodiment, the polypeptide comprises the amino acid sequence SEQ ID NO:1. According to a more preferred embodiment, the polypeptide consists of SEQ ID NO:1.

According to another preferred embodiment of the invention, the polypeptide is derived from *Salvia sclarea*.

As used herein, the terms “sclareol synthase” or “polypeptide having a sclareol synthase activity” refers to a genus of polypeptides or peptide fragments that encompasses the amino acid sequences identified herein, as well as truncated or variant polypeptides, provided that they keep their sclareol synthase activity as defined above and that they share at least the defined percentage of identity with the corresponding fragment of SEQ ID NO:1.

According to a preferred embodiment, the sclareol synthase comprises an amino acid sequence at least 50% identical to SEQ ID NO:96. Preferably the sclareol synthase 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:96. According to another preferred embodiment, the polypeptide comprises the amino acid sequence SEQ ID NO:96. According to a more preferred embodiment, the polypeptide consists of SEQ ID NO:96.

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 mutation of a nucleic acid of the invention, as described thereafter, are also encompassed by the invention.

The nucleic acid encoding the polypeptide having a sclareol synthase activity, as defined above, is a necessary tool to modify non-human organisms or cells intended to be used when the method is carried out in vivo. A nucleic acid encoding a polypeptide as defined in any of the above embodiments is therefore another object of the invention.

According to a preferred embodiment, the nucleic acid comprises a nucleotide sequence at least 50% identical to SEQ ID NO:2 or the complement thereof. According to a more preferred embodiment, said nucleic acid comprises a nucleotide sequence at least 55%, preferably at least 60%, preferably at least 65%, preferably at least 70%, preferably at least 75%, preferably at least 80%, more preferably at least 85%, more 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, the nucleic acid comprises a nucleotide sequence identical to 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 preferred embodiment of the invention, the nucleic acid is derived from Salvia sclarea.

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 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. Particularly useful truncated nucleic acids are the nucleic acids at least 70% identical to SEQ ID NO:93 or the complement thereof.

The nucleic acids obtained by mutations of SEQ ID NO:2 or of the complement thereof are also encompassed by the invention, provided that they share at least the defined percentage of identity with the corresponding fragment of SEQ ID NO:2 and that they encode polypeptides having a sclareol synthase activity, as defined above. Mutations may be any kind of mutations of these nucleic acids, such as point mutations, deletion mutations, insertion mutations and/or frame shift mutations. Variant nucleic acids 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.

According to a further preferred embodiment, the nucleic acid is a truncated nucleic acid comprising a nucleotide sequence at least 50%, preferably at least 55%, preferably at least 60%, 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:93 or the complement thereof. According to another preferred embodiment, the nucleic acid comprises the nucleotide sequence SEQ ID NO:93 or the complement thereof. According to a more preferred embodiment, the nucleic acid consists of SEQ ID NO:93 or the complement thereof.

Generally speaking, the nucleic acid of the invention can be isolated using a massively parallel sequencing approach, which is extensively developed in Examples 5 and 6. The first step of this method is the global sequencing of the cDNA library. The cDNA library is first fragmented by nebulization. The fragments are then amplified by PCR and the sequencing reaction is carried out. Short sequences of 35 bases named “reads” are obtained. “Reads” are reassembled in contiguous sequences (“contigs”) using a software with defined minimum length of overlap and percentage of homology settings.
“Reads” and “contigs” are then searched for sequence identity with known enzymes of the same type. On the basis of these homologies, “reads” and “contigs” are selected and used to synthesize primers in order to carry out the PCR amplification of the full length sclareol synthase.

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 at least one nucleic acid according to any embodiment of the invention. Such a vector is therefore also an object of the present invention.

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 vectors include the nucleic acid of the invention operably linked to at least one regulatory sequence, which controls transcription, translation, 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 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 a sclareol synthase activity, as disclosed further below.

Recombinant non-human organisms and cells transformed to harbor at least 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 organisms and cells are therefore another object of the present invention.

Non-human organisms of the invention may be any non-human multicellular or unicellular organisms. In a preferred embodiment, the non-human organism of the invention is a plant, a prokaryote or a fungus. Said organism may be any plant, prokaryote or fungus. 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 organism is a microorganism. According to an even more preferred embodiment said microorganism is a bacteria or yeast and 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 any of the nucleic acids of the invention. Preferably the term “transformed” relates to hosts heterologously expressing the polypeptides of the invention, as well as over-expressing them. Accordingly, in an embodiment, the present invention provides a transformed organism, in which the polypeptide of the invention is 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 cell cultures of higher eukaryotic organisms. 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, such as those of the present invention, 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 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 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 sclareol in vitro, as exposed herein above, it is very advantageous to provide a method of making at least one polypeptide having a sclareol synthase activity. Therefore, the invention provides a method for producing at least one polypeptide having a sclareol synthase activity comprising

a) culturing a non-human organism or cell transformed with the expression vector of the invention, so that it harbors a nucleic acid according to the invention and expresses or over-expresses a polypeptide encoded by said nucleic acid and having a sclareol synthase activity;

b) isolating the polypeptide having a sclareol synthase activity from the non-human 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 at least one expression vector of the invention, so that it harbors at least one nucleic acid according to the invention and expresses or over-expresses at least one polypeptide encoded by said nucleic acid.

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

A “polypeptide variant” as referred to herein means a polypeptide having a sclareol synthase activity and being substantially homologous to a native polypeptide, 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 physicochemical 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, Addison-Wesley Pub. Co., (1983). The effects of such substitutions can be calculated using substitution score matrices such as PAM-120, PAM-200, and PAM-250 as discussed in Altschul, (J. Mol. Biol. 219:555-65, 1991). 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 desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution. Furthermore, variants may be prepared to have at least one modified property, for example an increased affinity for the substrate, an improved specificity for the production of one or more desired compounds, a different product distribution, a different enzymatic activity, an increase of the velocity of the enzyme reaction, a higher activity or stability in a specific environment (pH, temperature, solvent, etc), or an improved expression level in a desired expression system. A variant or site directed mutant may be made by any method known in the art. As stated above, the invention provides recombinant and non-recombinant, isolated and purified polypeptides, such as from *Salvia sclarea*. 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 native sclareol synthases. 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.

Therefore, in an embodiment, the present invention provides a method for preparing a variant polypeptide having a sclareol synthase activity 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 sclareol synthase activity, repeat the process steps (a) to (d) until a polypeptide with a desired variant sclareol synthase activity is obtained;
(f) optionally, if a polypeptide having a desired variant sclareol synthase activity was identified in step d), isolating the corresponding mutant nucleic acid obtained in step (c).

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, SEQ ID NO:2 or 93 may be recombined with any other diterpene synthase encoding nucleic acids, for example isolated from an organism other than Salvia
Thus, mutant nucleic acids may be obtained and separated, which may be used for transforming a host cells according to standard procedures, for example such as disclosed in the present Examples.

In step (d), the polypeptide obtained in step (e) 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 chances of obtaining a desired variant polypeptide may thus be increased at the discretion of the skilled person.

In an embodiment, the present invention provides a method for preparing a nucleic acid encoding a variant polypeptide having a sclareol synthase activity, the method comprising the steps (a)-(e) disclosed above and further comprising the step of:

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 publications are cited.

**Description of the drawings**

Figure 1: Structures of the diverse compounds cited in the description.

Figure 2: Putative biosynthesis of sclareol from LPP, which is catalyzed by the SsTps1132 (SEQ ID NO:1).

Figure 3: Alignment of amino acid sequences from the class I diterpene synthase-like fragments with the sequence of the stemodene synthase from *Oryza sativa* (Access. No. AAZ76733).
Figure 4: Alignment of the amino acid sequence deduced from SsTps1132 (SEQ ID NO:1) and SsTps1137 (SEQ ID NO:86) with diterpene synthases amino acid sequences selected from the database.

Figure 5: Alignment of the amino acid sequences of SsTps1132 (SEQ ID NO:1) and 1132-2-5 (SEQ ID NO:96) which were heterologously expressed in E. coli.

Figure 6: GC analysis of the products obtained after incubation of the different 1132 recombinant proteins with LPP. Crude protein extracts from E. coli expressing the recombinant SsTps1132 and 1132-2-5 proteins (SEQ ID NO:1 and 96) were incubated with LPP in a final volume of 1 mL 50 mM MOPS pH 7 supplemented with 15 mM MgCl2 and 1 mM DTT.

Figure 7: GC-MS analysis of the products generated from LPP by the recombinant 1132-2-5 protein. (A) Total ion chromatogram of the products obtained from the incubation of LPP with a crude protein extract from E. coli transformed with pET101-1132-2-5 (SEQ ID NO:93). (B) Mass spectrum of the peak at retention time of 14.3. (C) Mass spectrum of an authentic sclareol standard.

Figure 8: GC analysis of the products obtained after co-incubation of SsTps1132 and 1132-2-5 recombinant proteins (SEQ ID NO:1 and 96) with the SsLPPs3 (SEQ ID NO:24) recombinant protein in the presence of GGPP.

**Specific embodiments of the invention or Examples**

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

**Example 1**

Isolation of LPP synthase encoding cDNAs from *Salvia clarea* by a PCR approach.

**A. Plant material and RNA extraction.**

*Salvia clarea* developing flower buds (1.5 to 2 cm length, 1-2 days old) were collected in fields of Bassins (Switzerland) and directly frozen in liquid nitrogen.

Total RNA was extracted using the Concert™ Plant RNA Reagent from Invitrogen (Carlsbad, CA) and the mRNA was purified by oligo(dT)-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 using the Marathon™ cDNA Amplification Kit (Clontech, Mountain View, CA).

**B. Polymerase Chain Reactions for amplification of diterpene synthases cDNAs**

PCR were performed using the forward primer DT3F (5’-GAYRTNGAYGAYACNGCNATGG-3’ (SEQ ID NO:3)) and the reverse primer DT4R (5’-GTYTTNCCNACKCANACRTCRYT-3’ (SEQ ID NO:4)). The PCR mixture contained 0.4 μM of each primer, 300 μM each dNTPs, 5 μL of 10X HotStartTaq® DNA polymerase buffer (Qiagen), 2 μL of 100 fold diluted cDNA, 0.5 μL of HotStartTaq® DNA polymerase in a final volume of 50 μL. The cycling conditions were: 35 cycles of 45 sec at 94°C, 45 sec at 50°C and 2 min at 72°C; and 10 min at 72°C. The sizes of the PCR products were evaluated on a 1% agarose gel. The bands corresponding to the expected size were excised from the gel, purified using the QIAquick® Gel Extraction Kit (Qiagen) and cloned in the pCR®2.1-TOPO vector using the TOPO TA cloning Kit (Invitrogen, Carlsbad, CA). Inserted cDNAs fragments were then subject to DNA sequencing and the sequence was compared against the GenBank non-redundant protein database (NCBI) using the BLASTX algorithm (Altschul et al., *J. Mol. Biol.* 215, 403-410, 1990). A 354 bp sequence named FN23 (SEQ ID NO:5) was obtained. This DNA fragment possessed the expected size and showed sequence homology to diterpene synthases.

**C. Full length cDNA isolation by rapid amplification of cDNA ends (RACE).**

Oligonucleotides specific for the FN23 sequence (SEQ ID NO:5) were designed: FN23-F1 (3’-GCACGGATACGACGTCGATCCAAATGTAC-5’ (SEQ ID NO:6)), FN23-F2 (3’-GGGTGTATCCGACCACCTTATTTGATGAG-5’ (SEQ ID NO:7)) and FN23-F3 (5’-GGGTGATATCCGACCACCTTATTTGATGAG-5’ (SEQ ID NO:8)). These primers were used in RT-PCR in combination with oligodT primers extended with an adaptor sequence (5’-AATTCGGTACCCGGGATCC(T)_{17}-3’) (SEQ ID NO:9). The composition of the RT-PCR reaction mixture was the following: 10 μl 5X Qiagen OneStep RT-PCR buffer, 400 μM each dNTP, 400 nM each primer, 2 μl Qiagen OneStep RT-PCR Enzyme Mix, 1 μl RNasin® Ribonuclease Inhibitor (Promega Co., Madison, WI) and 1250 ng total RNA in a final volume of 50 μl. The thermal cycler conditions were: 30 min at 50°C
(reverse transcription); 15 min at 95°C (DNA polymerase activation); 35 cycles of 45 sec at 94°C, 45 sec at 50°C and 90 sec at 72°C; and 10 min at 72°C. A second round of PCR was performed using the RT-PCR products as template with the adapterP primer (5’-AATTCGGTACCCGGGATCC-3’ (SEQ ID NO:10)) in combination with the same or nested FN23-specific primers. This PCR approach provided a 1271 bp cDNA fragment (FN30 (SEQ ID NO:11)) having a 192 bp perfect overlap with the FN23 fragment (SEQ ID NO:5) and containing the 3’end including the stop codon and the 3’ non-coding sequence of the corresponding cDNA.

For amplification of the 5’ end of the cDNA, anti-sense oligonucleotides specific for FN23 were designed: FN23-R1 (5’-CATGGCATCTTCAACCCACGTCTTATCTCATC-3’ (SEQ ID NO:12)), FN23-R2 (5’-GTGGTCCGATATCACCTTTATTGCTGCGG-3’ (SEQ ID NO:13)), FN23-R3 (5’-CATTGGAGATGCAGACTCGACCGATCGTCTGACC-3’ (SEQ ID NO:14)). These primers were used for 5’RACE using the S. sclarea cDNA library following the Marathon™ cDNA Amplification Kit protocol (Clontech, Mountain View, CA). The thermal Cycling conditions were as follows: 1 min at 94°C, 5 cycles of 30 sec at 94°C and 4 min at 72°C, 5 cycles of 30 sec at 94°C and 4 min at 70°C, 20 cycles of 30 sec at 94°C and 4 min at 68°C. This 5’RACE provided a 1449 bp cDNA fragment (FN40 (SEQ ID NO:15) having a 227 bp perfect overlap with FN23 (SEQ ID NO:5). Comparison with known diterpene synthases sequences revealed that the FN40 fragment (SEQ ID NO:15) contained the translation initiation codon and a 87 bp non-coding region. The assembling of the three cDNA fragments (FN23, FN30 and FN40 (SEQ ID NO:5, 11 and 15) provided a full length cDNA sequence (SaTps1) of 2655 bp (SEQ ID NO:16) with an open reading frame of 2355 bp coding for a 785 residues protein (SEQ ID NO:17).

Example 2

Heterologous expression of the S. sclarea LPP synthase in E. coli.

The pETDuet-1 (Novagen, Madison, WI), designed for expression under the control of a T7 promoter, was used for expression in E. coli cells. To construct the expression plasmid, the open reading frame of SaTps1 (SEQ ID NO:16) was amplified by PCR from the cDNA library with the forward and reverse primers SaTps-Nde
(3’- TACTGACATATGACTTCTGTAAATTTGAGCAGAGCACC-5’ (SEQ ID NO:18)) and SaTps-Kpn (3’- TTGGTACCTCATAAACCAGGTCGAAGAGTACTTTG-5’ (SEQ ID NO:19)) designed to introduce an NdeI site immediately before the start codon and a KpnI site after the stop codon. Since the open reading frame contains an NdeI site at position of 1614 of the open reading frame, this amplification was performed in two steps by overlap extension PCR (Horton et al, Gene 77, 61-68, 1989), using the primers SaTps-Nde (SEQ ID NO:18) and SaTps-Kpn (SEQ ID NO:19) in combination with the primers Satps-mut1f (5’- GTTGGAGTGGATCCACATGCAGGAATGGTAC-3’ (SEQ ID NO:20)) and Satps-mut1r (3’- GTACCATTCCTGCATCTGGATCCACTCCAAC-5’ (SEQ ID NO:21)), designed to remove the NdeI site without altering the amino acid sequence. The resulting cDNA were first ligated in the PCR2.1-Topo plasmid using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) and the sequences of the inserts were verified prior to sub-cloning as NdeI-KpnI fragment into the pETDuet-1 vector.

Analysis of the sequence of several clones obtained by amplification from the cDNA library with the SaTps1 specific primers showed some variability in several positions of the cDNA sequence. Seven positions were identified, in which two different amino acids can be found. One position was found were insertion of a serine residue occurred in some of the clones. These positions are listed in the table below.

<table>
<thead>
<tr>
<th>Positions (relative to the aminoacid sequence)</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Ile or Thr</td>
</tr>
<tr>
<td>40</td>
<td>Phe or Leu</td>
</tr>
<tr>
<td>174</td>
<td>Gln or His</td>
</tr>
<tr>
<td>222</td>
<td>Gly or Asp</td>
</tr>
<tr>
<td>538</td>
<td>Gln or His</td>
</tr>
<tr>
<td>560</td>
<td>Arg or Leu</td>
</tr>
<tr>
<td>596</td>
<td>Asn or Lys</td>
</tr>
<tr>
<td>612</td>
<td>Insertion of a Ser</td>
</tr>
</tbody>
</table>

These variations seemed to occur in a random manner in eleven different clones sequenced, suggesting that at least two very closely related isoforms of a diterpene
synthase are present in the *S. sclarea* genome and that the PCR amplification approach leaded to shuffling of the sequences. Two clones, SsLPPs3 (SEQ ID NO:22) and SsLPPs9 (SEQ ID NO:23) representative of the sequences variability, were selected for the heterologous expression and enzyme characterization experiments. The plasmids pETDuet-SsLPPs3 and pETDuet-SsLPPs9 were transferred into BL21(DE3) *E. Coli* cells (Novagen, Madison, WI). Single colonies of transformed cells were used to inoculate 5 ml LB medium. After 5 to 6 hours incubation at 37°C, the cultures were 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, resuspended in 0.1 volume of 50 mM MOPS pH 7, 10% glycerol and lyzed by sonication. The extracts were cleared by centrifugation (30 min at 20,000 g), and the supernatants containing the soluble proteins were used for further experiments. The crude protein extracts from pETDuet-SsLPPs3 and pETDuet-SsLPPs9 transformed cells were analyzed by SDS-PAGE and compared to protein extracts obtained from cells transformed with the empty pETDuet plasmid. The recombinant SsLPPs3 and SsLPPs9 proteins (SEQ ID NO:24 and 25) were clearly detected and the apparent molecular weight estimated at 90 KDa, a value in concordance with the calculated molecular weight of 83 KDa.

**Example 3**

Purification of the LPP synthase from *Salvia sclarea* and enzymatic activities

The PCR2.1-Topo plasmids containing the SsLPPs3 and SsLPPs9 cDNA (SEQ ID NO:22 and 23) (Example 2) were digested with NdeI and SacI and the inserts were ligated into the pET28a(+) plasmid (Novagen). The resulting expression plasmids (pET28-SsLPPs3 and pET28-SsLPPs9) contain the cDNAs with a 5'-end modification designed to express the proteins with an N-terminal hexa-histidine tag (His-tag). Purification was performed under native conditions using the ProBond™ Purification System (Invitrogen) following the manufacturer protocol excepted that, for the elution, imidazole was replaced by L-histidine to minimize inhibition of the enzyme. Using this approach, the SsLPPs3 and SsLPPs9 “His-tag” recombinant enzymes (SEQ ID NO:97 and 98) could be purified to
The affinity purified enzymes were incubated 12 hours at 30°C with 200 μM GGPP and 1 mM DTT in MOPS buffer pH 7, 10% glycerol, 1 mM DTT. No diterpene product was observed by extracting the incubation medium with pentane and analysis of the extract by GC or GC-MS. Treatment of the same extract by alkaline phosphatase (Sigma, 6 units/ml), followed by extraction with pentane and GC analysis, showed the formation of labdenediol and demonstrated the enzymatic formation of labdenediol-diphosphate (LPP) as the unique product from GGPP by the recombinant diterpene synthase.

The GC analysis was performed on an Agilent 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 100°C (1 min hold) followed by a gradient of 10°C/min to 300°C. The GC-MS analysis was performed in the same conditions and the spectra were recorded on an Agilent 5975 mass detector.

**Example 4**

**PCR approach for the homology cloning of class I diterpene synthases (sclareol synthase) from S. sclarea.**

The cloning and characterization of SsLPPs3 (SEQ ID NO:24) and SsLPPs9 (SEQ ID NO:25), in Examples 1 to 3, suggest that the biosynthesis of sclareol in S. sclarea involves two proteins, the SsLPPs and a class I diterpene synthase, the sclareol synthase, catalyzing the conversion of LPP to sclareol.

A PCR approach was used in a first attempt for the isolation of class I diterpene synthases cDNA sequences. Oligonucleotides were designed based on conserved sequences in plant diterpene synthases and especially in diterpene synthases catalyzing the cyclization of C20-diphosphate esters via an ionization mechanism. The sequences with accession numbers BAB19275, AAB39482, AAD30231, AAD34295, CAE05201, BAB12441, AAT49066, CAE05199, AAU05906, BAD17672, AAQ72565, AAL09965, AAK83563, AAS47691, AAS47690 and AAR13860, were selected from the public sequence databases (http://www.ncbi.nlm.nih.gov). All these protein sequences correspond to class I diterpene synthases and contain the DDxxD motif (wherein x represents any amino acid).
characteristic of ionization-dependent cyclization mechanism in terpene synthases. From the alignment of these sequences, two conserved motifs were first selected in the N-terminal region and used for the design of sense oligonucleotides: YDT(A/S)WVA and (D/N)GSWG. In the amino acid sequence of the SsLPPs (SEQ ID NO:24 and 25, Examples 1 to 3) these two motifs were also conserved, though with some differences for the first motif (YDTAVIA). Thus the sequence of SsLPPs was also taken into account for the design of the sense oligonucleotides. From the first motif, three oligonucleotides were design to cover all the sequences variations: DiTpsTB_F1, 5’-TATGATACNGCNGTNATDGC-3’ (SEQ ID NO:26); DiTpsTB_F2, 5’-TATGACACGGCAGTGATCGC-3’ (SEQ ID NO:27); DiTpsTB_F3, 3’-TATGACACGGCACKKRNGC-5’ (SEQ ID NO:28). From the second motif, two oligonucleotides were designed: DiTpsTB_F4, 5’-CAACTGGCTGATGGNTCNTGGGG-3’ (SEQ ID NO:29); DiTpsTB_F5, 5’-CAACTGGCTGATGGCTCATGGGG-3’ (SEQ ID NO:30). The DDxxD motif, located in the C-terminal region of the proteins and involved in the binding of the diphosphate moiety in the active site, was used to design two anti-sense oligonucleotides: DiTpsTB_R1, 5’-GATCCTCCAACRTCRWARARRTCRTC-3’ (SEQ ID NO:31); DiTpsTB_R2, 5’-GATCCTCCACGTCGWAGAAGTCGTC-3’ (SEQ ID NO:32).

These primers were used for PCR amplification from a Salvia sclarea cDNA library (prepared as described in Example 1). The PCRs were performed using the Advantage® 2 Polymerase Mix (Clontech). Each PCR mixture contained, in total volume of 50 μL, 5 μL of Advantage® 2 PCR Buffer, 200 μM dNTPs, 200 nM each oligonucleotide primer, 5 μL of 200 fold diluted cDNA, 1 μL of Advantage® 2 Polymerase Mix. The following conditions were used for the amplifications: 3 minutes of denaturation at 94°C; 15 cycles of 1 minutes denaturation at 94°C, 1 min of annealing at 65°C for the first cycle and minus one degree for each following cycle, and 2 minutes extension at 72°C; 20 cycles of 1 minutes denaturation at 94°C, 1 min of annealing at 58°C and 2 minutes extension at 72°C; and finally 10 minutes extension at 72°C. Different PCR were performed with the possible combinations of sense and anti-sense oligonucleotides. The amplicons were screened for the expected sizes and for sequence homology to diterpene synthases. Unfortunately, using this PCR approach, no diterpene-related sequence could be obtained.
Example 5

Massively parallel sequencing of a *S. sclarea* flower cDNA library.

Since the classical homology-based cloning approach did not succeed in the cloning of class I diterpene synthase from *S. sclarea*, we undertook to use an approach based on the global sequencing of the 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 all the transcripts (transcriptome) present in the Salvia *sclarea* flowers. 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 technology and equipment was set up at Fasteris SA (Geneva, Switzerland) and the preparation of the DNA samples and the sequencing were performed by Fasteris SA. An aliquot (1 μg) of the cDNA library generated from *S. sclarea* developing flowers and using the Marathon™ cDNA Amplification Kit (Clontech, Mountain View, CA) (Example 1), was treated using the Genomic Sample Prep Kit (Illumina). Briefly, the DNA is fragmented by nebulization, the ends are repaired to 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 35 bases were obtained.

The Edena software (Dr David Hernandez, Genomic Research Laboratory, University of Geneva Hospitals, Geneva, Switzerland) 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 sequencing procedure. The parameters of the software were set such as to allow 15 bases minimum length for the overlaps with strict (100%) identity. The contigs (contiguous sequences) with a length of at least 50 bases were retained. In these conditions, 2054 contigs of 50 to 1330 bases in length could be reconstituted.
To evaluate the quality of the assembling, the contigs were searched for sequence identity with the DNA sequence of SsLPPs, the class II diterpene synthases first isolated from the *S. sclarea* cDNA library (SsLPPs3 (SEQ ID NO:22), Example 2). This search was performed using the BLASTn method (Altschul et al, *J. Mol. Biol.* 215, 403-410, 1990).

Surprisingly, only 3 contigs of lengths of 81, 73 and 52 bases were found and only forty reads had been used by the Eland software to generate these contigs. Alignment with the SsLPPs3 reference sequence showed that the 3 contigs (SEQ ID NO:33 to 35 covered only 8.7% of the full-length sequence although with an identity of 99%).

Very limited sequence information has been reported in the public databases for *Salvia sclarea*. The only gene sequence available from the NCBI database (http://www.ncbi.nlm.nih.gov) was the sequence of the large subunit of the ribulose-1,5-bisphosphate carboxylase (RuBisCO) from salvia sclarea (NCBI access No. Z37450). Search of the contigs for DNA identity with this *S. sclarea* RuBisCO DNA sequence (BLASTn Search) provided two contigs of 870 and 547 bases respectively (SEQ ID NO:36 and 37). Alignment of the two contigs with the RuBisCO sequence showed coverage of 98%: only 27 bases (between position 858 and 884) out of 1420 bases were not present in the contigs. In addition to this almost complete coverage, the identity between the reference sequence and the contigs was 99.5%, representing a difference of only 7 nucleotides.

All reads (non-assembled data) were then searched for sequence identity with the SsLPPs3 sequence (SEQ ID NO:22). The Eland software (Illumina) was used to perform this search allowing a maximum of 2 mismatches with the reference sequence. A total of 616 reads where recovered. Alignment of the selected fragments with the reference sequence revealed that the SsLPPs3 sequence (SEQ ID NO:22) was covered on the whole length with a slightly higher coverage (more reads) towards the 3’end. The same manipulation with the RuBisCO sequence showed that 1650 reads were obtained for this sequence. The coverage of the reference sequence with the reads was much higher for the RuBisCo than for SsLPPs3 (SEQ ID NO:22). For SsLPPs3 (SEQ ID NO:22), several small regions with no coverage and regions with sequence ambiguity between reads were found. This incomplete coverage prevents the complete re-assembling and is certainly the reason for the generation of only a few very small contigs.
The number of reads obtained for a given cDNA is proportional to the abundance of this cDNA. Thus, relative abundances can be estimated by dividing the number of reads obtained for given cDNAs by their total lengths. Performing this calculation for the RuBisCO and SsLPPs3 (SEQ ID NO:22) gave values of 1160 and 260 reads/Kb respectively, reflecting a 4.5 higher abundance of the RuBisCO cDNA relative to the SsLPPs cDNAs. The RubisCo is an enzyme involved in the primary metabolism of plants and catalyzing the fixation of carbon in the Calvin cycle. The higher relative abundances of the RuBisCO reflects a high representation of genes involved in primary metabolisms compared to gene involved in secondary metabolism such as diterpene synthesis. BLAST search analysis with the contigs showed that other enzymes from the Calvin cycles (e.g. phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, triosephosphate isomerase) and primary metabolism were also abundantly represented in the cDNA library used herein. Thus, the cDNA coding for the enzymes involved in secondary metabolism and particularly in diterpene biosynthesis were in too low abundance to obtain a sufficient coverage and complete reassembling.

Example 6

Extraction of class I diterpene synthases-like sequences from the sequencing data.

The Blast algorithm (Altschul et al, J. Mol. Biol. 215, 403-410, 1990) was used to search for homology of the deduced amino acid sequences with class I diterpene synthases sequences.

A Blastx search against a protein database was first performed with the 2054 contigs. This search provided only one contig (contig1610, SEQ ID NO:38) presenting sequence homology with class I diterpene synthases. The amino acid sequence deduced from this contig contained the DDxxD motif characteristic of ionization-initiated cyclization of prenyl-diphosphates.

A fraction of the row data, representing approximately $3 \times 10^5$ reads was then search for homology with class I diterpene synthases. The reads were search using the tBlastn algorithm with five selected class I diterpene synthase amino acid sequences (NCBI accession numbers AAC39443, BAB19275, BAB12441, AAD34295, AAS98912). This search selected 462 reads, which were then processes using the CAP program (Huang,
Genomics 14(1), 18-25, 1992) to identify overlapping sequences. A small portion of the reads could be assembled in short contigs of maximum length of 111 bases. These contigs as well as the remaining isolated reads were used for Blastx search against a protein database to confirm their identity with class I diterpene synthases. Finally, 5 DNA fragments were retained (SEQ ID NO:39 to 43).

The amino acid sequences were deduced from the selected fragments (SEQ ID NO:44 to 48) and were aligned with references diterpene synthases sequences, allowing their relative positioning. Figure 3 shows an alignment of these sequences with a full-length diterpene synthase sequence, the stemodene synthase from *Oriza sativa* (Morrone et al., 2006; NCBI access No. AAZ76733) taken as reference.

**Example 7**

**PCR amplification of full-length class I diterpene synthases cDNAs.**

A set of forward and reverse oligonucleotides was deduced from the diterpene synthases-related DNA sequences selected from the sequencing of the *S. sclarea* cDNA library (Example 6). These primers were used in combination with cDNA adaptor primers in 3’/5’RACE type PCR amplifications. The amplifications were performed using the *S. sclarea* cDNA library, prepared as described above in Example 1, following the Marathon™ cDNA Amplification Kit protocol (Clontech, Mountain View, CA). The thermal Cycling conditions were as follows: 1 min at 94°C, 5 cycles of 30 sec at 94°C and 4 min at 72°C, 5 cycles of 30 sec at 94°C and 4 min at 70°C, 20 cycles of 30 sec at 94°C and 4 min at 68°C.

Using the Cont250-Fwd primer (SEQ ID NO:49) a 547 bp DNA sequence (1130Cont250, SEQ ID NO:81) was obtained. Analysis of the sequence revealed that it corresponded to the 5’end of a diterpene synthase cDNA and contained 348 bp of the coding region. With the primer Cont147_fw1 (SEQ ID NO:51) and Cont147-fw2 (SEQ ID NO:52) we obtained a 1473 bp sequence (1132Cont147, SEQ ID NO:82) containing the 3’end and 1293 bp of the coding region of a diterpene synthase cDNA. The Cont224_fw primer (SEQ ID NO:57) provided a 207 bp DNA fragment (1137Cont224, SEQ ID NO:83) encoding for the 43 C-terminal amino acids of a diterpene synthases with a sequence
distinct from 1132Cont147 (SEQ ID NO:82). The Cont147_rev1 (SEQ ID NO:53) and Cont147_rev2 (SEQ ID NO:54) primers allowed the amplification of a 464 bp DNA fragment (1134Cont147, SEQ ID NO:84). The deduced amino acid showed homology with diterpene synthases but alignment with other diterpene synthases sequences suggested that 200 to 300 codons where still missing to reach the 5’end. All the sequences obtained by this series of amplification differed significantly from the sequences of SsLPPs previously isolated (SEQ ID NO:22 and 23). PCR with the other primers deduced from the diterpene synthases-related DNA sequences (primers cont224-rev (SEQ ID NO:58), cont250-rev (SEQ ID NO:50), cont33-fw1 (SEQ ID NO:55) and cont33-rev (SEQ ID NO:56)) did not provide diterpene synthase related sequences.

From the only sequence containing an obvious translation initiation region of a diterpene synthase (1130Cont250, SEQ ID NO:81), sense oligonucleotides were deduced from the 5’ untranslated region (UTR) (1130-fw1 (SEQ ID NO:59) and 1130-fw2 (SEQ ID NO:60) and from the 5’end of the open reading frame (ORF) (1130-fw3, SEQ ID NO:61). From the two sequences containing the stop codon region of two distinct diterpene synthases (1132Cont147 (SEQ ID NO:82) and 1137Cont224 (SEQ ID NO:83)), reverse-sense primers were deduced either from the 3’ UTR (1132-rev1 (SEQ ID NO:65) and 1137-rev1 (SEQ ID NO:62)) or from the 3’ end of the open reading frame (1132-rev2 (SEQ ID NO:64) and 1137-rev2 (SEQ ID NO:63)). PCR were performed with different combinations of these forward and reverse primers. The combination of primers deduced from the 1130Cont250 (SEQ ID NO:81) sequence with the primers deduced from the 1137Cont224 (SEQ ID NO:83) sequence produced a fragment of 2388 bp (SEQ ID NO:85) coding for a protein of 795 amino acids (SsTps1137, SEQ ID NO:86)). Comparison with published sequences showed homologies with class I diterpene synthases and particularly ent-kaurene synthases B. Highest homology was with an uncharacterized protein from Vitis vinifera (NCBI access No. CA064942, 59% identity), an ent-kauren synthase from Cucumis sativus (NCBI access No. BAB19275, 54% identity) and an ent-kauren synthase from Lactuca sativa (NCBI access No. BAB12441, 54% identity). The SsTps1137 (SEQ ID NO:86) amino acid sequence contained a DDFFD motif typical of ionization-dependent (class I) terpene synthases and did not contain the characteristic class II motif.
The combination of the same forward primers with the reverse primers deduced from the 1132Cont147 (SEQ ID NO:82) did not allow the amplification of any fragment, confirming that these two sequences did not arise from the same cDNA. A 5' RACE approach was then used to identify the 5’ end of the ORF corresponding to the 1132Cont147 sequence (SEQ ID NO:82). Using the primers 1132_race1 (SEQ ID NO:67) and 1132_race2 (SEQ ID NO:68), a 536 bp sequence (1132RACE, SEQ ID NO:87) was obtained which had 41 bases overlap with the 1132Cont147 fragment (SEQ ID NO:82). This RACE product was identical to the previously obtained 1134Cont147 sequence (SEQ ID NO:84) and no extension at the 5’ end was observed. As observed previously, this sequence had homology with diterpene synthases but seemed shorter by at least 200 codons than all other published diterpene synthases sequences. 5’ RACE experiments were performed, in order to try to extend the sequence toward the 5’ end of the 1132Cont147 sequence (SEQ ID NO:82) and to identify the true translation initiation codon. Several sets of oligonucleotides (1132_race3 to 1132_race9, SEQ ID NO:69 to 75) were designed but no additional sequence information was obtained. This led us to suppose that one of the ATG codon in the 1134Cont147 sequence (SEQ ID NO:84) was actually the initiation codon of the corresponding diterpene synthase gene. The nucleotidic sequence of this putative diterpene synthase (named SsTps1132, SEQ ID NO:2) was reconstituted from the 1132Cont147 (SEQ ID NO:82) and 1132RACE (SEQ ID NO:87) sequences. Taking the first ATG, the 1728 bp ORF of SsTps1132 (SEQ ID NO:2) encoded for a 575 amino acid protein (SEQ ID NO:1). This protein contained the ionization-dependent modif (DDFFD) and shared homology, but relatively low, with published diterpene synthases; the closest sequence being a terpene synthase from *Nicotiana tabacum* (NCBI acces No. AAS98912), with 37% identity.

Surprisingly, the identity between the SsTps1137 (SEQ ID NO:86) and SsTps1132 (SEQ ID NO:1) proteins was only 30% and these sequences shared only 21 to 23% identity with the class II SsLPPs first isolated from *S. sclarea* (SEQ ID NO:24 and 25, Examples 1-3). An alignment of these two proteins with selected diterpene synthases sequences is presented in figure 4. The alignment shows that SsTps1132 (SEQ ID NO:1) is truncated at the N-terminal end by 150 to 240 amino acids compared to the other diterpene synthases. The ChloroP method (Emanuelsson et al, *Protein Science* 8, 978-984, 1999; http://www.cbs.dtu.dk/services/ChloroP/) was used to predict the presence of a
chloroplast transit peptide in each protein sequence. For SsTps1137 (SEQ ID NO:86) and SsTps1132 (SEQ ID NO:1) chloroplast transit peptides of 22 and 51 amino acids respectively were predicted, arguing for a chloroplast localization of both proteins.

Search of all reads for sequences identical to the SsTps1137 (SEQ ID NO:85) and SsTps1132 (SEQ ID NO:2) DNA sequences, provided only 24 reads for SsTps1137 and 425 reads for SsTps1132. This difference in the number of reads generated from each transcript reflects a significant difference in the expression levels. Based on the relative number of reads obtained for each transcript, it can be estimated that the expression level of SsTps1132 (220 reads/Kb) was similar to the expression level of SsLPPs (260 reads/Kb) and that SsTps1137 was expressed at a much lower level (10 reads/Kb). With the assumption that enzymes catalyzing steps in the same metabolic pathway are generally expressed at a similar level, it can be speculated that SsTps1132 (SEQ ID NO:1) rather than SsTps1137 (SEQ ID NO:86) is involved in the same metabolic pathway as SsLPPs.

The contigs generated with the Edena software (Example 5) were searched for DNA sequences identical to the sequences of these two new putative class I diterpene synthases. For SsTps1137 (SEQ ID NO:85) no contig was found in accordance with the presumed low expression level of this enzyme. For SsTps1132 (SEQ ID NO:2), 4 contigs were found. The previously identified contig1610 (SEQ ID NO:38) and three additional contigs (of length of 53 to 96 bp) (SEQ ID NO:88 to 90) not previously identified as fragment of a diterpene synthase. Blastx search with these three sequences did not show homology with known protein sequences. The failure in finding homology for these contigs is due to the short lengths of these fragments and to the low homology of SsTps1132 (SEQ ID NO:1) with the diterpene synthases present in the databases.

The observation of an N-terminal deletion of SsTps1132 (SEQ ID NO:1) compared to the other diterpene synthases also explains afterwards why the PCR approach first employed did not succeed. Indeed, the forward primers were designed from conserved regions present in the first 150 amino acids of diterpene synthases, a region absent in SsTps1132 (SEQ ID NO:1). The SsTps1137 sequence (SEQ ID NO:86) contains the conserved motifs used to design the primers and the corresponding DNA sequences are complementary to the primer sequences. Presumably, the amplification of SsTps1137
(SEQ ID NO:85) did not succeed in the PCR approach because of the low abundance of this transcript.

Example 8

Heterologous expression of the S. Sclarea class I diterpene synthases in *E coli*.

To assign an enzymatic activity to SsTps1137 (SEQ ID NO:86) and SsTps1132 (SEQ ID NO:1), the recombinant proteins were expressed in *E coli*. The full-length cDNAs were inserted into the pet101/D-TOPO vector using the Champion pET101 Directional TOPO Expression Kit.

For each enzyme, two constructs were prepared: one to express the full-length protein and one to express a truncated protein based on the chloroplast transit peptide prediction. The full-length SsTps1137 (SEQ ID NO:85) and SsTps1132 (SEQ ID NO:2) open reading frames were amplified from the cDNA library using the primer pairs 1137-start (SEQ ID NO:78) with 1137-stop (SEQ ID NO:80) and 1132-start1 (Seq ID NO 76) with 1132-stop (SEQ ID NO:66) respectively. The primers 1137 _start2 (SEQ ID NO:79) and 1137 _stop (SEQ ID NO:80) were used to amplify a 72 bp truncated version of SsTps1137 designed to express the protein with 24 amino acids deleted at the N-terminal end. In the same manner, the primers 1132 _start2 (SEQ ID NO:77) and 1132-stop (SEQ ID NO:66) were used to prepare a truncated version of SsTps1132 designed to express the protein with a 50 amino acid N-terminal deletion. All amplifications of cDNA for expression of the expression constructs were performed using the *Pfu* DNA polymerase (Promega), in a final volume of 50 µl containing 5 µl of *Pfu* DNA polymerase 10X buffer, 200 µM each dNTP, 0.4 µM each forward and reverse primer, 2.9 units *Pfu* DNA polymerase and 5 µl of 100-fold diluted cDNA (prepared as described herein in Example 1 using the Marathon™ cDNA Amplification Kit (Clontech)). The thermal cycling conditions were as follows: 1.5 min at 95°C; 30 cycles of 45 sec at 95°C, 30 sec at 58°C and 5 min at 72°C; and 10 min at 72°C.

After the ligation in the pET101 vector, several clones were selected for each construct and were sequenced to ensure that no mutation had been introduced during the PCR amplification. For SsTps1137 the two constructs 1137-B12 (SEQ ID NO:91) and 1137-2-
WO 2009/095366

B12 (SEQ ID NO:92) were selected containing the SsTps1137 cDNA respectively with and without the peptide signal (corresponding polypeptide sequences are SEQ ID NO:94 and 95). For SsTps1132, two constructs were selected: one with the complete sequence of SsTps1132 (SEQ ID NO:2) and a construct without peptide signal (1132-2-5, SEQ ID NO:93). The alignment of the two amino acid sequences (SEQ ID NO:1 and 96) deduced from these constructs is shown in figure 5.

The plasmids pET101-1137-B12, pET101-1137-2-B12, pET101-SsTps1132, and pET101-1132-2-5 were transferred into Bl21(DE3) E. Coli cells (Novagene, Madison, WI). Single colonies of transformed cells were used to inoculate 5 ml LB medium. After 5 to 6 hours incubation at 37°C, the cultures were 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, resuspended in 0.1 volume of 50 mM MOPSO pH 7, 10% glycerol and lyzed by sonication. The extracts were cleared by centrifugation (30 min at 20,000 g), and the supernatants containing the soluble proteins were used for further experiments. The crude protein extracts were analysed by SDS-PAGE and compared to protein extracts obtained from cells transformed with the empty pET101 plasmid.

Example 9

Enzymatic activity of the recombinant S. Sclarea class I diterpene synthases in E coli.

The crude E coli protein extracts containing the recombinant proteins and prepared as described in Example 8 were used for the characterization of the enzymatic activities. The enzymatic assays were performed as described in Example 3. All assays were performed in 50 mM MOPSO pH 7, 10% glycerol, 1 mM DTT.

The enzymatic activities were first evaluated using as substrate either GGPP or LPP, the product of SsLPPs (SEQ ID NO:22) and the presumed intermediate in the biosynthesis of sclareol (Examples 1 to 3). GGPP was synthesized as described by Keller and Thompson (J. Chromatogr 645(1), 1993, 161-167) and LPP was prepared enzymatically as described in Example 3. The assays were performed in the presence of 10 to 100 μM of substrate, 15 mM MgCl₂ and 0.1 to 0.5 mg of crude protein in a total volume of 1 mL. The tubes
were incubated 4 to 12 hours at 30°C and extracted twice with one volume of pentane.
After concentration under a nitrogen flux, the extracts were analysed by GC and GC-MS
(using the conditions described in Example 3) and compared to extracts from assay with
control proteins (obtained from cells transformed with the empty plasmid). With GGPP as
substrate, no activity was observed with any of recombinant proteins (data not shown).
With LPP as substrate, no activity was observed with the proteins extracts containing
SsTps1137 recombinant proteins but with SsTps1132, activity was observed with both
SsTps1132 and 1132-2-5 (SEQ ID NO:1 and 96) (Figure 6). The enzymes were also
active in the absence of MgCl₂ and the same product profiles were observed with an
overall activity roughly the same. The identity of product was confirmed by concordance
of the retention times (Figure 6) and matching of the mass spectrum with the spectrum of
an authentic standard (Figure 7). In all assays, a single peak of sclareol was observed with
no trace of additional product.
Assays were then performed with co-incubation of the class II diterpene synthases
(SsLPPs3, SEQ ID NO:24; Examples 1-3) and the class I diterpene synthases (1132
series, SEQ ID NO:1 and 96). Assays were performed in 50 mM MOPS pH 7, 10%
glycerol, 1 mM DTT, 50 μM GGPP, with 1 mM MgCl₂ and in the presence of 50 μL of
the crude protein extracts from E coli expressing the different constructs. Thus assays in
the presence of 50 μL of crude protein extracts containing the SsLPPs3 (SEQ ID NO:24)
recombinant enzyme and 50 μL of extracts containing SsTps1132 (SEQ ID NO:1) or
1132-2-5 (SEQ ID NO:96) were evaluated for the production of diterpene products.
Figure 8 shows the GC profiles of extracts from such incubations in the presence of
MgCl₂. Sclareol was produced with both 1132 constructs (SEQ ID NO: 1 and 96)
(Figure 8), a result consistent with the assay described above with LPP as substrate. No
significant difference was observed when omitting MgCl₂ from the incubations (data not
shown).
In conclusion the SsTps1132 (SEQ ID NO:2) encodes for the sclareol synthase (SEQ ID
NO:1) and catalyses the conversion of LPP to sclareol.
Claims

1. A method for producing sclareol comprising
   a) contacting labdenediol diphosphate (LPP) with a polypeptide having a sclareol synthase activity and comprising an amino acid sequence at least 70% identical to SEQ ID NO:1; and
   b) optionally, isolating the sclareol produced in step a).

2. The method of claim 1, wherein said polypeptide comprises an amino acid sequence at least 80%, more preferably at least 90% identical to SEQ ID NO:1.

3. The method of claim 2, wherein said polypeptide comprises the amino acid sequence set out in SEQ ID NO:1.

4. The method of claim 3, wherein said polypeptide consists of the amino acid sequence set out in SEQ ID NO:1.

5. The method of any of claims 1 to 4, wherein step a) is carried out by cultivating a non-human organism or cell capable of producing LPP and transformed to express said polypeptide under conditions conducive to the production of sclareol.

6. The method of claim 5, further comprising, prior to step a), transforming the non human organism or cell capable of producing LPP with a nucleic acid encoding said polypeptide, so that said organism expresses said polypeptide.

7. The method of claim 5 or 6, wherein said non-human organism is a plant, a prokaryote or a fungus.

8. The method of claim 5 or 6, wherein said non-human organism is a microorganism.

9. The method of claim 8, wherein said microorganism is a bacteria or yeast.
10. The method of claim 9, wherein said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*.

11. The method of claim 5 or 6, wherein said non-human cell is a higher eukaryotic cell selected from plant cells or fungal cells.

12. An isolated polypeptide having a sclareol synthase activity and comprising an amino acid sequence at least 70% identical to SEQ ID NO:1.

13. The isolated polypeptide of claim 12, comprising an amino acid sequence at least 80%, more preferably at least 90%, identical to SEQ ID NO: 1.

14. The isolated polypeptide of claim 13, comprising the amino acid sequence set out in SEQ ID NO: 1.

15. The isolated polypeptide of claim 14, consisting of the amino acid sequence set out in SEQ ID NO: 1.

16. The isolated polypeptide of any of claims 12 to 15, derived from *Salvia Sclarea*.

17. An isolated nucleic acid encoding the polypeptide of any of claims 12 to 16.

18. The isolated nucleic acid of claim 17, comprising a nucleotide sequence at least 70% identical to SEQ ID NO: 2 or the complement thereof.

19. The isolated nucleic acid of claim 18, comprising a nucleotide sequence at least 80%, preferably at least 90%, identical to SEQ ID NO: 2 or the complement thereof.

20. The isolated nucleic acid of claim 19, comprising a nucleotide sequence identical to SEQ ID NO:2 or the complement thereof.
21. The isolated nucleic acid of claim 20, consisting of the nucleotide sequence set out in SEQ ID NO:2 or in the complement thereof.

22. The isolated nucleic acid of any of claims 17 to 21, derived from *Salvia sclarea*.

23. An expression vector comprising the nucleic acid of any of claims 17 to 22.

24. The expression vector of claim 23, in the form of a viral vector, a bacteriophage or a plasmid.

25. The expression vector of claim 23 or 24, 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.

26. A non-human organism transformed with the expression vector of any of claims 23 to 25, so that it harbors the nucleic acid of any of claims 17 to 22 and heterologously expresses or over-expresses the polypeptide of any of claims 12 to 16.

27. The non-human organism of claim 26, wherein said non-human organism is a plant, a prokaryote or a fungus.

28. The non-human organism of claim 26, wherein said non-human organism is a microorganism.

29. The non-human organism of claim 28, wherein said microorganism is a bacteria or yeast.

30. The non-human organism of claim 29, wherein said bacteria is *E. coli* and said yeast is *Saccharomyces cerevisiae*. 
31. A higher eukaryotic cell transformed with the expression vector of any of claims 23 to 25, so that it harbors the nucleic acid of any of claims 17 to 22 and expresses the polypeptide of any of claims 12 to 16.

32. The higher eukaryotic cell of claim 31, wherein said higher eukaryotic cell is a plant cell or a fungal cell.

33. A method for producing at least one polypeptide having a sclareol synthase activity comprising:
   a) culturing a non-human organism or cell transformed with the expression vector of any of claims 23 to 25, so that it harbors a nucleic acid according to any of claims 17 to 22 and expresses or over-expresses a polypeptide encoded by said nucleic acid and having a sclareol synthase activity;
   b) isolating the polypeptide having a sclareol synthase activity from the non-human organism or cell cultured in step a).

34. The method of claim 33, further comprising, prior to step a), transforming a non-human host organism or cell with the expression vector of any of claims 23 to 25, so that it harbors a nucleic acid according to any of claims 17 to 22 and expresses or over-expresses the polypeptide encoded by said nucleic acid.

35. A method for preparing a variant polypeptide having a sclareol synthase activity comprising the steps of:
   a) selecting a nucleic acid according to any of claims 17 to 23;
   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 sclareol synthase activity, repeat the process steps (a) to (d) until a polypeptide with a desired variant sclareol synthase activity is obtained;
f) optionally, if a polypeptide having a desired variant scclareol synthase activity was identified in step d), isolating the corresponding mutant nucleic acid obtained in step (c).
Figure 1

- Labdenediol diphosphate
- Sclareol (-)-(13R)-14-labdene-8,13-diol
- (-)-Ambrox
- Geranylgeranyl diphosphate (GGPP)
- Ent-kaurene
- Ent-cassa-12,15-diene
Figure 2

[Chemical diagram showing the transformation of Sclareol to LPP through various reactions involving protonation and water addition.]
Figure 3 (2nd part)

<table>
<thead>
<tr>
<th>490</th>
<th>500</th>
<th>510</th>
<th>520</th>
<th>530</th>
<th>540</th>
<th>550</th>
<th>560</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAVDKFHYQSVYQQELRYLESWVAEFGLDELKFarvip1q81LsalvlfpaelsDarF2gCQLJTFLDDFLFSGS</td>
<td>Contig-33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>570</td>
<td>580</td>
<td>590</td>
<td>600</td>
<td>610</td>
<td>620</td>
<td>630</td>
<td>640</td>
</tr>
<tr>
<td>MEEMVNFVALIDEWDNHEIGFCSNNVEIMFNAIYNTTKRNCAAKAVQNRCVMHIAKQWQVMVRAMKTEAEWAASRH</td>
<td>Stemodene synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>650</td>
<td>660</td>
<td>670</td>
<td>680</td>
<td>690</td>
<td>700</td>
<td>710</td>
<td>720</td>
</tr>
<tr>
<td>IPATMEEMYSVGEPSFALGPIVPLSAYLLGEELPEEAVRSPEYQQLRHASAVGRLLNDVMTYKEVLWTIPNSVLLQAL</td>
<td>Stemodene synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>730</td>
<td>740</td>
<td>750</td>
<td>760</td>
<td>770</td>
<td>780</td>
<td>790</td>
<td>800</td>
</tr>
<tr>
<td>AAARGGGEAPTPSFACAAARGEvRRAIQASWRDLHLRVFRDDGSS</td>
<td>Contig-224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>810</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGMANAVILDPLHLQO</td>
<td>Stemodene synthase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4 (3rd part)
Figure 4 (4\textsuperscript{th} part)
Figure 5

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stps1132 MSLAFWGVTVFQSGQRVSGSKKEFVQRVPFTTPNSRLIVCSLTDPMARKNKENFKREDXKFFTTLASIDPSNLCLIDLQQLQVDQFFQYRIENTLMTLPRWQEKHKYGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1132-2-5</td>
<td>MSLAFWGVTVFQSGQRVSGSKKEFVQRVPFTTPNSRLIVCSLTDPMARKNKENFKREDXKFFTTLASIDPSNLCLIDLQQLQVDQFFQYRIENTLMTLPRWQEKHKYGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stps1132 VTTHAMAFRLRLRVGYSBSSEELAPGQRAVESQQTNDLPMIELYEAANERIVIEERSELKILAWTTFLNQVOQNSIPDKKLHKLVEFYLNQSTVILQRLGVDQFFQYEINTLMTLPRWQEKHKYGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1132-2-5</td>
<td>VTTHAMAFRLRLRVGYSBSSEELAPGQRAVESQQTNDLPMIELYEAANERIVIEERSELKILAWTTFLNQVOQNSIPDKKLHKLVEFYLNQSTVILQRLGVDQFFQYEINTLMTLPRWQEKHKYGN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stps1132 KSTNRFDNLCREDFLVFAKQDFDIIEAQONQKLQQLQLRQWDTCRDLDDLNFLQGDDVVIANYLASLIIGDASFDVRLAFAKTSLVFVTDMDFFCDRGGSEQEDKIKIIEVAK WKENPDAEYG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1132-2-5</td>
<td>KSTNRFDNLCREDFLVFAKQDFDIIEAQONQKLQQLQLRQWDTCRDLDDLNFLQGDDVVIANYLASLIIGDASFDVRLAFAKTSLVFVTDMDFFCDRGGSEQEDKIKIIEVAK WKENPDAEYG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stps1132 SEELEILMALYNTVEIAERADEVQCSVKEIFIKLWVEILSAFKIERTTVSNSTQCSFDEYITSSWLSNGSATCLLMQVFVGLSDEMLSEETDILRVCMVGRLNDVCSER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1132-2-5</td>
<td>SEELEILMALYNTVEIAERADEVQCSVKEIFIKLWVEILSAFKIERTTVSNSTQCSFDEYITSSWLSNGSATCLLMQVFVGLSDEMLSEETDILRVCMVGRLNDVCSER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stps1132 EREENIAGKSYLLLATEKGRKVSDEAIARRIMVENEWHWKLQTVKKKESILPRACKDVFLSMAGKTFAYGINDLTSPPQSKSKEDMKSFVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1132-2-5</td>
<td>EREENIAGKSYLLLATEKGRKVSDEAIARRIMVENEWHWKLQTVKKKESILPRACKDVFLSMAGKTFAYGINDLTSPPQSKSKEDMKSFVF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6

Empty plasmid

1132-1-3

1132-2-5

Sclareol standard
Figure 7
Figure 8

[Graph showing chromatograms labeled 1132-1-3, 1132-2-5, and Standard sclareol]
**INTERNATIONAL SEARCH REPORT**

**International application No**
PCT/EP2009/00816

**A. CLASSIFICATION OF SUBJECT MATTER**

| INV. | C12P7/02 | C12N9/14 | C12N15/52 |

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)

| C12P | C12N |

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, Sequence Search, BIOSIS, EMBASE, WPI Data |

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>DATABASE EMBL [Online] 1 February 2005 (2005-02-01), &quot;Nicotiana tabacum terpenoid cyclase mRNA, complete cds.&quot; XP002494697 retrieved from EBI accession no. EMBL:AY528645 Database accession no. AY528645 the whole document</td>
<td>12,17</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

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 document or other special reason (as specified);
  - "O" document referring to an oral disclosure, use, exhibition or other means of disclosure prior to the priority date claimed;
  - "P" document published prior to the international filing date but later than the priority date claimed.

**Date of the actual completion of the International search**

15 May 2009

**Date of mailing of the international search report**

29/05/2009

**Name and mailing address of the ISN**

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk

Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016

**Authorized officer**

Roscoe, Richard
### DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DATABASE EMBL [Online] 16 March 2000 (2000-03-16), &quot;ML458 peppermint glandular trichome Mentha x piperita cDNA, mRNA sequence.&quot; XP002494698 retrieved from EBI accession no. EMBL:AW255436 Database accession no. AW255436 the whole document</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>GUO ZHENHUA ET AL: &quot;Biosynthesis of labdenediol and sclareol in cell-free extracts from trichomes of Nicotiana glutinosa&quot; PLANTA, SPRINGER VERLAG, DE, vol. 197, no. 4, 1 January 1995 (1995-01-01), pages 627-632, XP008089701 ISSN: 0032-0935 the whole document</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>BANTHORPE D V ET AL: &quot;ACCUMULATION OF THE ANTI-FUNGAL DITERPENE SCLAREOL BY CELL CULTURES OF SALVIA SCLAREA AND NICOTIANA GLUTINOSA&quot; PHYTOCHEMISTRY, PERGamon PRESS, GB, vol. 29, no. 7, 1 January 1990 (1990-01-01), pages 2145-2148, XP000671796 ISSN: 0031-9422 the whole document</td>
<td>A</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to claim No.</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
</tbody>
</table>