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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 March 2005 (10.03.2005)

PCT

(10) International Publication Number
WO 2005/021705 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/IL2004/000795

(22) International Filing Date:
2 September 2004 (02.09.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/499,024 2 September 2003 (02.09.2003) US

(71) Applicant *(for all designated States except US)*: **STATE OF ISRAEL, MINISTRY OF AGRICULTURE, AGRICULTURAL RESEARCH ORGANIZATION** [IL/IL]; Volcani Center, P.O. Box 6, IL-50250 Bet Dagan (IL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **EYAL, Yoram** [IL/IL]; 129 Mesilat Zion Street, IL-99770 Shimshon (IL). **LEWINSOHN, Efraim** [IL/IL]; 32 Moran Street, IL-23840 Timrat (IL).

(74) Agent: **WEBB, Cynthia**; Webb & Associates, P.O. Box 2189, IL-76121 Rehovot (IL).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CITRUS SESQUITERPENE SYNTHASE, METHODS OF PRODUCTION AND USES THEREOF

(57) Abstract: The present invention relates to citrus sesquiterpene synthases, key enzymes in the production of valencene, a sesquiterpene aromatic compound. Particularly, the present invention relates to nucleic acid sequences encoding valecene synthases from angiosperm plant species, in particular citrus, vectors containing the sequences, host cells containing said sequences and transgenic plants expressing said sequences. The present invention further relates to methods of producing recombinant valencene synthase and its products, and uses thereof.

WO 2005/021705 A2

CITRUS SESQUITERPENE SYNTHASE, METHODS OF PRODUCTION AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates to citrus sesquiterpene synthase, a key enzyme in the production of valencene, a sesquiterpene aroma compound. Particularly, the present invention relates to nucleic acid sequences encoding valencene synthase from angiosperm plant species, in particular citrus, vectors containing the sequences, host cells containing said sequences and transgenic plants and microorganisms expressing said sequences. The present invention further relates to methods of producing recombinant valencene synthase and its products, and uses thereof.

BACKGROUND OF THE INVENTION

Flavors and aromas, many of which originate from plants, have always had an important role in human culture. Among the most aromatic plants are citrus species. The flavor and aroma of citrus species is composed of complex combinations of soluble compounds, mostly acids, sugars and flavonoids and of volatile compounds. The latter typically consist of mono- and sesqui- terpenes, secondary metabolites obtained from the enzymatic activity of valence synthase, which are the major components of citrus essential oils.

The profile of the volatile terpenoids in various citrus species and their importance as aroma compounds are well known in the art. Efforts were made to increase the level of volatile terpenoids in citrus end products. For example, US Patent No. 4,970,085 discloses a method for making improved citrus aqueous essences by a fractionation process wherein citrus aqueous essence is passed through a solid adsorbent so that part of the essence compounds exit the adsorbent in a first effluent and part remain on the adsorbent, and then the first effluent is recycled through the adsorbent to recover a fraction of the remaining compounds and to produce a second effluent. US Patent No. 4,973,485 discloses aqueous orange stripper essences and orange stripper oils with high ratios of more desirable to less desirable orange flavor

compounds, wherein these essences and oils are obtained by a method comprising the steps of: (1) heating an orange fed juice stream to a temperature of about 37.7-71°C; (2) stripping the heated feed juice with steam at 37.7-71°C and a stripping column pressure of less than 9 inches of Hg, absolute; (3) condensing the stripped volatiles; (4) centrifuging the condensate in a continuous stacked disk hermetic centrifuge to produce two clear phases; and (5) removing the aqueous orange stripper phase. The methods disclosed in US Patent Nos. 4,970,085 and 4,973,485 are directed to the production of desired essences and oils from citrus fruits but do not address the issue of providing *a priori* citrus fruits with higher ratios of the desired oils and essences.

Methods for the synthesis of aromatic compounds were disclosed in US Patent Nos. 5,847,226 and 6,200,786 among others. US Patent No. 5,847,226 discloses a method for preparing nootkatone, nootkatol or mixtures thereof in vitro by oxidizing valencene in a suitable reaction medium and in the presence of an unsaturated fatty acid hydroperoxide. US Patent No. 6,200,786 discloses a process for producing nootkatone in vitro comprising (a) reacting valencene and a composition having laccase activity in the presence of an oxygen source to form valencene hydroperoxide; (b) degrading the hydroperoxide to form nootkatone; and (c) recovering nootkatone. However, no attempts to provide cells or organisms capable of producing any desired levels of nootkatone were disclosed. Moreover, the complete physiological and biochemical pathways as well as the genetic regulations involved in the production of aromatic compounds within plants remain unresolved.

Terpenoids are found in all plant species and have diverse physiological roles such as phytoalexins, pest deterrents and toxins, growth regulators, pollinator attractants, photosynthetic pigments and electron acceptors. US Patent No. 6,258,602 discloses the isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint that produces the aphid alarm pheromone E-beta-farnesene.

Isolation and expression of Cstps1, a sesquiterpene synthase-encoding gene was published after the priority date of the present application by the inventors of the present invention in Sharon-Asa *et al.*, (*The Plant Journal*, 36:664-674, 2003), which is incorporated herein by reference in its entirety. Sharon-Asa *et al.*, showed that the recombinant enzyme encoded by Cstps1 converts farnesyl diphosphate to a single sesquiterpene product identified as valencene.

A Ph.D. thesis by Mr. Bryan T. Greenhagen at the University of Kentucky outlines the cloning and use of a valencene synthase from Ruby Grapefruit. The thesis was made publicly available after the priority date of the present application.

The biochemical and genetic regulation of fruit aroma has only recently achieved increased attention. Very few genes involved in fruit aroma were described, and thus not much is known on the regulation of aroma formation in fruits. This state of research limits the ability of the agricultural, food and cosmetic industries to use natural fruit aromas, which are highly desirable in products of these industries.

Thus, there is a recognized need for, and it would be highly advantageous to identify specific genetic components involved in the regulation pathways directed to the formation of fruit aromas, and more advantageous to provide host cells and organisms capable of producing the desired quantities of aromatic compounds comprising the aromas.

SUMMARY OF THE INVENTION

The present invention relates to key enzymes in the production of the sesquiterpene valencene, an aroma compound found mainly in citrus species.

The present invention provides novel members of the family of sesquiterpene synthases, which are involved in the terpene biosynthetic pathway converting farnesylpyrophosphate (FPP, also known as farnesyl diphosphate or FDP) to sesquiterpenes. According to one aspect, the sesquiterpene synthase is valencene synthase. The present invention also provides polynucleotide sequences encoding the sesquiterpene synthase, including recombinant DNA molecules. The present invention further provides vectors and host cells, including vectors comprising the polynucleotides of the present invention, host cells engineered to contain the polynucleotides of the present invention and host cells engineered to express the polynucleotides of the present invention. Thus, the present invention provides methods for (i) expressing the recombinant sesquiterpene synthase, specifically valencene synthase, to facilitate the production, isolation and purification of significant quantities of recombinant valencene synthase, or of its primary and secondary products for subsequent use; (ii) expressing or enhancing the expression of a sesquiterpene synthase, specifically valencene synthase, in microorganisms or in

plants; and (iii) regulating the expression of a sesquiterpene synthase, specifically valencene synthase, in an environment where such regulation of expression is desired for the production of the enzyme and for producing the enzyme products and derivatives thereof.

The present invention further provides polynucleotide sequences encoding sesquiterpene synthase characterized in that it converts farnesylpyrophosphate (FPP) to valencene, specifically valencene synthase, for use in a variety of methods and techniques known to those skilled in the art of molecular biology, including, but not limited to the use as hybridization probes, oligomers for PCR, chromosome and gene mapping and the like.

The present invention further provides methods for using sesquiterpene synthase enzymatic products, specifically valencene, in industrial applications selected from agriculture, cosmetics and food.

According to one aspect, the present invention provides an isolated polynucleotide comprising a polynucleotide sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene.

According to one embodiment, the isolated polynucleotide comprising a nucleic acid sequence having at least 80% homology, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to another embodiment, the isolated polynucleotide comprising a nucleic acid sequence capable of hybridizing to a nucleic acid sequence having at least 80% homology, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to yet another embodiment, the isolated polynucleotide comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology, preferably at least 90% homology to SEQ ID NO:2. According to some embodiment, the present invention provides an isolated polynucleotide capable of hybridizing to a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology, preferably at least 90% homology to SEQ ID NO:2.

According to yet another embodiment, the isolated polynucleotide comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid

sequence having at least 80% homology to SEQ ID NO:2, the amino acid comprising at least one consensus motif of contiguous amino acids of the sequence DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid. According to some embodiment, the present invention provides an isolated polynucleotide capable of hybridizing to said nucleic acid sequence.

According to another aspect, the present invention provides an isolated polypeptide comprising an amino acid sequence having an activity of a valencene synthase, the activity is characterized by the ability to convert farnesylpyrophosphate to valencene.

According to one embodiment, the isolated polypeptide comprises an amino acid sequence having at least 80% homology, preferably at least 90% homology to SEQ ID NO:2.

According to another embodiment the isolated polypeptide encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2, the amino acid comprising at least one consensus motif of contiguous amino acids sequence of DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid.

It is to be understood explicitly that the scope of the present invention encompasses homologs, analogs, variants and derivatives, including shorter and longer polypeptides, proteins and polynucleotides, as well as polypeptide, protein and polynucleotide analogs with one or more amino acid or nucleic acid substitution, as well as amino acid or nucleic acid derivatives, non-natural amino or nucleic acids and synthetic amino or nucleic acids as are known in the art, with the stipulation that these variants and modifications must preserve the capacity of sesquiterpene synthase to convert farnesylpyrophosphate (FPP) to valencene. Specifically, any active fragments of the active polypeptide or protein as well as extensions, conjugates and mixtures are disclosed according to the principles of the present invention.

According to one embodiment, the valencene synthase is derived from citrus species, preferably oranges.

According to yet another aspect, the present invention provides an expression vector comprising a nucleic acid sequence encoding a valencene synthase the valencene synthase being capable of converting FPP to valencene.

According to one embodiment, the vector is a plasmid or a virus. According to some embodiments, the vector further comprises at least one element selected from the group consisting of: promoter operatively linked to the polynucleotide encoding the valencene synthase, a selection marker, a signal sequence, an origin of replication, an enhancer and a transcription termination sequence.

According to yet another aspect, the present invention provides a host cell comprising an expression vector, wherein the expression vector comprises a nucleic acid sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene.

According to one embodiment the host cell is prokaryotic or eukaryotic. According to another embodiment, the host cell is a prokaryotic cell, wherein a polynucleotide sequence comprising a nucleic acid sequence encoding valencene synthase is stably integrated into its genome. According to a preferred embodiment, the prokaryotic cell a bacterial cell, preferably, an *E. coli*.

According to yet another embodiment, the host cell produces at least one compound selected from the group consisting of: valencene, valencene metabolite other than nootkatone, nootkatone.

According to yet another aspect, the present invention further provides a method for producing recombinant valencene synthase and recombinant valencene, the method comprising:

(a) culturing a host cell comprising an expression vector, wherein the expression vector comprises a nucleic acid sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene, under conditions suitable for the expression of said valencene synthase; and, optionally,

(b) recovering said valencene synthase.

According to an alternative embodiment, step (b) comprises recovering valencene.

The valencene produced within a host cell according to the present invention can serve as a substrate for producing additional compounds by enzymes present in

the host cell active downstream to valencene synthase in the terpene biosynthesis pathway. Such compounds are designated herein as "valencene metabolites".

According to another embodiment, step (b) comprising:

recovering at least one valencene metabolite.

According to a preferred embodiment, the at least one valencene metabolite is nootkatone.

According to yet another embodiment, the host cell is prokaryotic or eukaryotic.

According to yet another embodiment, the host cell is a prokaryotic cell, wherein a polynucleotide sequence comprising a nucleic acid sequence encoding valencene synthase is stably integrated into its genome.

According to yet another embodiment, the host cell produces at least one compound selected from the group consisting of: valencene, valencene metabolite other than nootkatone, nootkatone. According to a preferred embodiment, the prokaryotic cell a bacterial cell, preferably, an *E. coli*.

According to yet another aspect, the present invention provides a plant comprising a polynucleotide sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene.

According to one embodiment, the polynucleotide sequence encoding a valencene synthase being stably integrated into the genome of the plant.

According to another embodiment, the plant produces at least one compound selected from the group consisting of: valencene, a valencene metabolite other than nootkatone, nootkatone.

According to yet another aspect, the present invention provides valencene and valencene metabolites obtained by any one of the methods of the invention.

According to yet another aspect, the present invention provides use of valencene and valencene metabolites obtained by the methods of the present invention in an industrial application selected from the group consisting of: agriculture, cosmetics and food.

Other objects, features and advantages of the present invention will become clear from the following description and drawings.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 exhibits the nucleotide sequence of *Cstps1*, SEQ ID NO:1.

FIG. 2 shows an amino-acid alignment of sesquiterpene synthase *Cstps1* (SEQ IDNO:1), with sesquiterpene synthase obtained from tobacco species - *Nicotiana tabacum epi*-aristolochene synthase (SEQ ID NO:4, Accession #AAG17667) and from cotton - *Gossypium arboreum* cadinene synthase (SEQ ID NO:5, Accession #CAA77191).

FIG. 3 is a schematic representation of the position of valencene synthase on the phylogenetic map of terpene synthases.

FIG. 4 shows a temporal expression of *Cstps1* in Valencia™ orange flavedo during fruit ripening in fruits collected at monthly intervals during the 2003 season (A) or the 2001 season (B).

FIG. 5 describes the effect of ethylene on *Cstps1* expression and valencene accumulation in citrus flavedo following 7 days (A) or 48 hours of treatment with ethylene.

FIG. 6 describes an activity assay for recombinant *Cstps1* as pmole product per-hour of reaction time.

FIG. 7 shows GC-MS identification of recombinant *Cstps1* sesquiterpene product: (A) chromatogram of valencene; (B) chromatogram of products obtained from assay with lysate of bacteria expressing recombinant Cstps1; (C) chromatogram of products obtained from assay with lysate of bacteria harboring control plasmid pTYB; (D) single ion chromatogram (m/z = 107) of valencene; (E) single ion chromatogram (m/z = 107) of product obtained from lysates expressing recombinant Cstps1; (F) single ion chromatogram profile of valencene (Wiley GC-MS library data base).

FIG. 8 shows the accumulation of valencene during citrus fruit development.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to novel types of sesquiterpene synthases, specifically to valencene synthase.

Ancient cultures cultivated and prized plants for their nutritional value as well as for their flavor, aroma and medicinal properties. However, the development of commercial large-scale agriculture in western civilizations resulted in emphasizing commercial and marketing interests in plant production, such as long shelf life, physical appearance and yield. The content of secondary metabolites, i.e. metabolites which do not have a defined metabolic role and presence of which is restricted to specific tissues, were often overlooked, although they significantly affect the nutritional value and aroma. Nowadays, there is growing awareness to healthy and flavorful plant products. Physically appealing but flavorless and aroma-less fruit is perceived as "synthetic", while scented fruit is perceived as more "natural". There is increasing public interest to "return" the natural flavor and aroma to fruits, emphasizing the importance of elucidating the relevant biosynthetic pathways, enzymes, genes and regulatory mechanisms involved. Cosmetic and food industries also seek natural aroma and flavors.

Among the most aromatic plants are citrus species. Citrus are second only to grapes in planting and cultivation worldwide and are an important source of secondary metabolites (e.g. Terpenoids, Flavonoids and other polyphenols) for nutrition, health and industrial applications. The flavor and aroma of the citrus is composed of complex combinations of soluble compounds, mostly acids, sugars and flavonoids and of volatile compounds. The latter typically consist of mono- and sesqui- terpenes, which are the major components of citrus essential oils, accumulating in specific oil glands in the flavedo (external part of the peel) and oil bodies in the juice sacs.

Although the monoterpene limonene normally accounts for over 90% of the content of essential oils obtained from citrus species, several unique sesquiterpene compounds, which are present in very small quantities, have a profound effect on the flavor and aroma of the citrus species. For example, the sesquiterpenes valencene, α- and β- sinensal, that are present in minor quantities in oranges, have an important role in the overall flavor and aroma of orange fruit. Nootkatone, an oxygenated

sesquiterpene that is a putative-derivative of valencene, occupies a small fraction of the essential oil but has a dominant role in the flavor and aroma of grapefruit.

Mono- and sesqui- terpenes are among the most important secondary metabolites obtained from the enzymatic activity of valencene synthase, which are involved in fruit and flower aromas. The backbones of the biosynthetic pathways leading to production of mono- and sesqui-terpenes are ubiquitous to all plant species, however the composition of terpenes often differs dramatically between species or even varieties leading to the diversity of flavors between citrus cultivars. This diversity seems to stem mainly from the specific composition and expression of the key-enzymes in the biosynthetic pathway, the terpene synthases, and additional downstream modification enzymes. Homology analysis reveals that although sequence conservation is not high among terpene synthases of different plant species, discrete conserved domains are present suggesting significant structural and functional similarity. These conserved domains have been the basis for isolation of a number of terpene synthases encoding genes from a variety of plant species using degenerate-primer based RT-PCR.

The study of fruit development and ripening has a long history, and a majority of the processes, especially in climacteric fruit, have been elucidated during the years. However, the biochemical and genetic regulation of fruit aroma has only recently achieved increased attention; very few genes involved in fruit aroma have been described, and as a result not much is known on the regulation of aroma formation in fruits. This state of research limits the ability of agriculture, food and cosmetics industries to use natural fruit aromas, which are highly desired products.

The present invention discloses the isolation and characterization of a key gene and the corresponding enzyme encoded by said key gene, valencene synthase, said gene/enzyme is a key element formation of aroma in citrus fruits. Recombinant enzyme activity *in-vitro* shows a single sesquiterpene product identified as valencene. Studies on the accumulation of valencene and the pattern of valencene synthase gene expression show that valencene production in citrus fruit is regulated at the transcript level: (1) valencene synthase expression is developmentally regulated and occurs only at the final stage of fruit maturation, in close correlation with valencene accumulation; and (2) valencene synthase expression as well as valencene accumulation are responsive to ethylene application.

Definitions

The terms “citrus” or " citrus species" are interchangeably used herein to define any plant or fruit of the genus *Citrus*, a genus of often thorny trees and shrubs of the rue family (Rutaceae) grown in warm regions. Most citrus fruits are edible, as the orange, lemon, grapefruit, lime, kumquat, mandarin and shaddock, and often have firm, thick and pulpy flesh.

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L-α-amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations.

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C"), thymine ("T") and inosine ("I"). The four RNA bases are A,G,C and uracil ("U"). The nucleotide sequences described herein comprise a line array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

The terms “homology” or "percent identity" are interchangeably used herein to define the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences are aligned side by side.

The term "percent similarity" is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon encoding the other member of the pair. Calculations are made after a best fit alignment of the two sequences has been made empirically by iterative comparison of all possible alignments. (Henikoff *et al.*, Proc. Nat'l. Acad. Sci. USA 89:10915-10919, 1992).

"Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified on polyacrylamide gels.

The term "sesquiterpene synthase" is used herein to mean an enzyme capable of catalyzing the production of sesquiterpene from FPP. "Valencene synthase" is used herein to mean an enzyme that catalyzes the production of the sesquiterpene valencene from FPP.

The terms "derivative", "analog" and "variant" refer to valencene synthase molecules or polynucleotides encoding same, having some differences in their sequences as compared to the citrus valencene synthase having the amino acid sequence set forth in SEQ ID NO:2 or encoded by the polynucleotide set forth in SEQ ID NO:1, respectively. Ordinarily, the variants will possess at least about 80% homology, preferably at least about 90% homology with the above defined valencene synthase or the polynucleotides encoding same. The sequence variants of valencene synthase falling within this invention possess "alterations", namely, substitutions, deletions, and/or insertions at certain positions. Sequence variants of valencene synthase may be used to attain desired enhanced enzymatic activity or altered substrate utilization or product distribution.

Valencene synthase variants encompassing "substitutions" are those that have at least one amino acid residue in the valencene synthase sequence set forth in SEQ ID NO:2 removed and a different amino acid inserted in its place at the same position. The substitutions may be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule. Substantial changes in the activity of the valencene synthase molecules of the present invention may be obtained by substituting an amino acid with a side chain that is significantly different in charge and/or structure from that of the native amino acid. This type of substitution would be expected to affect the structure of the polypeptide backbone and/or the charge or hydrophobicity of the molecule in the area of the substitution.

Moderate changes in the activity of the valencene synthase molecules of the present invention would be expected by substituting an amino acid with a side chain that is similar in charge and/or structure to that of the native molecule. This type of

substitution, referred to as a conservative substitution, would not be expected to substantially alter either the structure of the polypeptide backbone or the charge or hydrophobicity of the molecule in the area of the substitution.

Valencene synthase variants encompassing "insertions" are those with one or more amino acids inserted immediately adjacent to an amino acid at a particular position in the amino acid sequence of valencene synthase set forth in SEQ ID NO;2. Immediately adjacent to an amino acid means connected to either the α-carboxy or α-amino functional group of the amino acid. The insertion may be one or more amino acids. Ordinarily, the insertion will consist of one or two conservative amino acids. Amino acids similar in charge and/or structure to the amino acids adjacent to the site of insertion are defined as conservative. Alternatively, this invention includes insertion of an amino acid with a charge and/or structure that is substantially different from the amino acids adjacent to the site of insertion.

Valencene synthase variants encompassing "deletions" are those where one or more amino acids in the amino acid sequence of valencene synthase set forth in SEQ ID NO:2 have been removed. Ordinarily, deletion variants will have one or two amino acids deleted in a particular region of the valencene synthase molecule.

The term "biological activity", "biologically active", "activity" and "active" refer to the ability of the sesquiterpene synthase to convert farnesylpyrophosphate (FPP) to a group of sesquiterpenes, of which valencene is the principle and characteristic sesquiterpene synthesized by valencene synthase.

The terms "DNA sequence encoding", "DNA encoding", "nucleic acid sequence encoding" or "polynucleotide sequence encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the translated polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The terms "stringent conditions" or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. Numerous equivalent conditions

comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C to about 25°C below the melting temperature of the probe). One or more factors may be varied to generate conditions of either low or high stringency.

The terms "replicable expression vector" and "expression vector" refer to a piece of DNA, usually double-stranded, which may have inserted into it a piece of foreign DNA. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidental with the host chromosomal DNA, and several copies of the vector and its inserted (foreign) DNA may be generated. In addition, the vector contains the necessary elements that permit translating the foreign DNA into a polypeptide. Many molecules of the polypeptide encoded by the foreign DNA can thus be rapidly synthesized.

The terms "transformed host cell," "transformed" and "transformation" refer to the introduction of DNA into a cell. The cell is termed a "host cell", and it may be a prokaryotic or a eukaryotic cell. Typical prokaryotic host cells include various strains of *E.* coli. Typical eukaryotic host cells are plant cells, yeast cells, insect cells or animal cells. The introduced DNA is usually in the form of a vector containing an inserted piece of DNA. The introduced DNA sequence may be from the same species as the host cell or from a different species from the host cell, or it may be a hybrid DNA sequence, containing some foreign DNA and some DNA derived from the host species.

Novel sesquiterpene synthase

According to one aspect, the present invention relates to polynucleotides encoding for sesquiterpene synthase, specifically valencene synthase.

Great diversity is found for terpenoid aroma compounds in citrus. This diversity limits the efficacy of the reverse genetics approach for the isolation all of genes

responsible for biosynthesis of all such compounds. Therefore, isolation of terpene synthase encoding genes was approached by an exhaustive screen of mRNA isolated from enriched target tissue. Since some important citrus sesquiterpene flavor compounds, such as valencene and nootkatone, were noted to accumulate towards fruit ripening and mainly in the flavedo (outer-peel), the enriched target tissue chosen was Valencia™ orange (*Citrus sinensis* cv. Valencia) flavedo from oranges picked late in development towards ripening. Degenerate primers based on short conserved sequence elements present in most mono- and sesqui- terpene synthases were employed to isolate partial cDNA fragments from mRNA of Valencia™ orange flavedo, using RT-PCR. PCR fragments of the expected size were cloned and sequenced. Relevant clones containing internal conserved sequence elements characteristic of terpene synthases were selected for complete cDNA cloning. A complete cDNA whose sequence resembled that of plant sesquiterpene synthases was designated as *Cstps1*.

According to one aspect, the present invention provides an isolated polynucleotide comprising a polynucleotide sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene.

According to one embodiment, the isolated polynucleotide comprising a nucleic acid sequence having at least 80%, preferably at least 90%, homology to SEQ ID NO:1 or the complement thereof.

According to another embodiment, the isolated polynucleotide comprising a nucleic acid sequence capable of hybridizing to a nucleic acid sequence having at least 80%, preferably at least 90%, homology to SEQ ID NO:1 or the complement thereof.

According to yet another embodiment, the isolated polynucleotide comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:2.

According to yet another embodiment, the isolated polynucleotide comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2, the amino acid comprising at least one consensus motif of contiguous amino acids of the sequence DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid.

The isolation of cDNA encoding valencene synthase permits the development of efficient expression systems for this functional enzyme; provides useful tools for examining the developmental regulation of valencene biosynthesis; permits investigation of the reaction mechanism(s) of this unique enzyme and permits the transformation of a wide range of organisms in order to introduce valencene biosynthesis de novo, or to modify endogenous valencene biosynthesis.

The present invention further relates to polypeptides having sesquiterpene synthase activity, specifically valencene synthase activity, i.e., the polypeptides of the invention are capable of converting FPP to sesquiterpene, specifically valencene.

According to another aspect, the present invention provides an isolated polypeptide having valencene synthase activity, said activity being characterized by converting farnesylpyrophosphate (FPP) to valencene.

According to one embodiment, the isolated polypeptide comprises an amino acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:2.

According to another embodiment the isolated polypeptide encodes a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2, the amino acid comprising at least one consensus motif of contiguous amino acids sequence of DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid.

It is to be understood explicitly that the scope of the present invention encompasses homologs, analogs, variants and derivatives, including shorter and longer polypeptides, proteins and polynucleotides, as well as polypeptide, protein and polynucleotide analogs with one or more amino acid or nucleic acid substitution, as well as amino acid or nucleic acid derivatives, non-natural amino or nucleic acids and synthetic amino or nucleic acids as are known in the art, with the stipulation that these variants and modifications must preserve the capacity of sesquiterpene synthase to convert farnesylpyrophosphate (FPP) to valencene. Specifically, any active fragments of the active polypeptide or protein as well as extensions, conjugates and mixtures are disclosed according to the principles of the present invention.

According to one embodiment, the valencene synthase is derived from citrus species, preferably oranges.

According to yet another aspect, the present invention provides an expression vector comprising a nucleic acid sequence encoding a valencene synthase the valencene synthase being capable of converting FPP to valencene.

Vectors of various types may be used in the practice of the present invention. A specific vector type is used according to the host cell in which expression is desired, as is known to a person with ordinary skill in the art, and as described herein below. The vector usually has a replication site, marker genes that provide phenotypic selection in transformed cells, one or more promoters, and a polylinker region containing several restriction sites for insertion of foreign DNA. For example, plasmids typically used for transformation of *E. coli* include pBR322, pUC18, pUC19, pUCI18, pUC119, and Bluescript M13, all of which are described in sections 1.12-1.20 of Sambrook et al., (Molecular Cloning: A Laboratory Manual. 3rd edn., 2001, Cold Spring Harbor Laboratory Press). These vectors contain genes coding for ampicillin and/or tetracycline resistance which enables cells transformed with these vectors to grow in the presence of these antibiotics. However, many other suitable vectors, harboring different genes encoding for selection markers are available as well. The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes and the valencene synthase DNA of interest are prepared using standard recombinant DNA procedures. Isolated plasmids and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well known in the art (see, for example, Sambrook et al., *ibid*).

Vector components generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. The expression elements of these vectors vary in their strength and specificities. Depending on the host and the vector system utilized, any one of a number of suitable transcription and translation elements may be used. For example, when cloning in prokaryotic cell systems, promoters isolated from the genome of prokaryotic cells, (e.g., the bacterial tryptophan promoter) may be used. Promoters produced by recombinant DNA or synthetic techniques may also be used to provide for transcription of the inserted sequences.

"A signal sequence" may be a component of the vector, or it may be a part of

the nucleic acid, encoding the proteins of the invention, that is inserted into the vector. The signal sequence may be the naturally occurring sequence or a non-naturally occurring sequence. The signal sequence should be one that is recognized and processed by the host cell.

"An origin of replication" refers to the unique site of initiation of replication of a host organism.

It is desirable for cloning and expression vectors to comprise a selection gene, also termed a "selectable marker" or a "selection marker". This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, e.g. ampicillin; complement auxotrophic deficiencies; or supply critical nutrients not available from complex media. One example of a selection scheme utilizes a drug to arrest growth of a host cell. Those cells that are successfully transformed with a heterologous gene express a protein conferring drug resistance and thus survive the selection regimen. Expression vectors used in prokaryotic host cells will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA.

Construction of suitable vectors containing one or more of the above listed components and including the desired coding and control sequences employs standard ligation techniques. Isolated plasmids or nucleic acid fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required.

According to one embodiment, the expression vector comprises a nucleic acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to another embodiment, the expression vector comprises a nucleic acid sequence capable of hybridizing to a nucleic acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to yet another embodiment, the expression vector comprises a nucleic acid sequence a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:2.

According to yet another embodiment, the expression vector comprises a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2, the amino acid comprising at least one consensus motif of contiguous amino acids of the sequence DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid.

The present invention provides methods for the production, isolation and purification of the valencene synthases according to the present invention, as well as of the products of its enzymatic activity.

According to yet another aspect, the present invention further provides a method for producing recombinant valencene synthase, the method comprising:

(a) culturing a host cell comprising an expression vector, wherein the expression vector comprises a nucleic acid sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene, under conditions suitable for the expression of said valencene synthase; and, optionally,

(b) recovering said valencene synthase.

According to an alternative embodiment, step (b) comprises:

recovering significant amounts of said valencene synthase.

According to another alternative embodiment, step (b) comprises:

recovering valencene.

The valencene produced within a host cell according to the present invention can serve as a substrate for producing additional compounds by enzymes present in the host cell active downstream to valencene synthase in the terpene biosynthesis pathway. Such compounds are designated herein as "valencene metabolites".

According to another embodiment, step (b) comprising:

recovering at least one valencene metabolite.

According to a preferred embodiment, the at least one valencene metabolite is nootkatone.

According to yet another embodiment, the host cell is prokaryotic or eukaryotic.

According to yet another embodiment, the host cell is a prokaryotic cell, wherein a polynucleotide sequence comprising a nucleic acid sequence encoding valencene synthase is stably integrated into its genome. According to a preferred embodiment, the prokaryotic cell a bacterial cell, preferably, an *E. coli*. According to yet another embodiment, the host cell produces at least one compound selected from the group consisting of: valencene, valencene metabolite other than nootkatone, nootkatone.

According to yet another embodiment, the polynucleotide sequence comprising a nucleic acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to another embodiment, the polynucleotide sequence comprising a nucleic acid sequence capable of hybridizing to a nucleic acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:1 or the complement thereof.

According to yet another embodiment, the polynucleotide sequence comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80%, preferably at least 90% homology to SEQ ID NO:2.

According to yet another embodiment, the polynucleotide sequence comprising a nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2, the amino acid sequence comprising at least one consensus motif of contiguous amino acids of the sequence DDXXD (SEQ ID NO:3), wherein X corresponds to any amino acid.

The complete deduced amino acid sequence of the *Cstps1* cDNA clone (SEQ ID NO:2, Fig. 2) showed the characteristic conserved elements of mono- and sesqui-terpene synthase including the most highly conserved metal ion binding motif DDXX D (SEQ ID NO:3). Conserved sequence elements are framed and the mono- and sesqui-terpene synthase universally conserved motif DDxxD (SEQ ID NO:3) is underlined with stars.

Cstps1 was found to be most similar to sesquiterpene synthases of angiosperms, with the highest level of similarity (61%) to β-farnesene synthase from citrus (Maruyama *et al.*, *Biol Pharm Bull.*, 2001, 24(10):1171-5) followed by δ-cadinene synthase (50% similarity) from cotton (Chen *et al.*, Arch. Biochem. Biophys., 1995, 324: 255-266).

The deduced amino-acid sequence of the citrus valencene synthase *Cstps1* groups phylogenetically to the angiosperm sesquiterpene synthase group (Fig. 3), one of five distinct groups obtained by a non-rooted analysis of sequences of angiosperm and gymnosperm mono-, sesqui- and diterpene synthases (BioEdit and Treeview software, Hall, *Nucl. Acids. Symp. Ser.*, 1999, 41:95-98; and Page, *Comput Appl Biosci.* 1996, 12(4):357-8, respectively). Within the angiosperm sesquiterpene group, valencene synthase (*Cstps1*), citrus farnesene synthase (Maruyama *et al.*, *ibid*) and an additional sesquiterpene synthase gene (*Cmtps1*) isolated from citrus flowers disclosed by an inventor of the present invention form a distinct citrus subgroup.

In addition to the native valencene synthase amino acid sequence, sequence variants produced by deletions, substitutions, mutations and/or insertions are intended to be within the scope of the invention. The valencene synthase amino acid sequence variants of this invention may be constructed by mutating the DNA sequence that encodes the wild-type synthase, such as by using techniques commonly referred to as site-directed mutagenesis. Nucleic acid molecules encoding the valencene synthases of the present invention can be mutated by a variety of PCR techniques well known to one of ordinary skill in the art. See, e.g., "PCR Strategies", M. A. Innis et al. eds., 1995, Academic Press, San Diego, Calif. (Chapter 14); "PCR Protocols: A Guide to Methods and Applications", M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White, eds., Academic Press, NY (1990).

By way of non-limiting example, the two-primer system utilized in the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for introducing site-directed mutants into the valencene synthase gene of the present invention. Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried out, tightly linking these two mutations, and the resulting plasmids are transformed into a mutS strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subjoining or generation of single-stranded phagemids. The tight linkage of the two mutations and

the subsequent linearization of unmutated plasmids result in high mutation efficiency and allow minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed to confirm that no other alterations in the sequence have occurred (e.g., by band shift comparison to the unmutagenized control).

In the design of a particular site directed mutagenesis, it is generally desirable to first make a non-conservative substitution (e.g., Ala for Cys, His or Glu) and determining if activity is greatly impaired as a consequence. The properties of the mutagenized protein are then examined with particular attention to the kinetic parameters of K_m and k_{cat} as sensitive indicators of altered function, from which changes in binding and/or catalysis per se may be deduced by comparison to the native enzyme. If the residue is demonstrated to be important by activity impairment, or knockout, then conservative substitutions can be made, such as Asp for Glu to alter side chain length, Ser for Cys, or Arg for His. For hydrophobic segments, it is commonly size that is usefully altered, although aromatics can also be substituted for alkyl side chains. Changes in the normal product distribution can indicate which step(s) of the reaction sequence have been altered by the mutation. Modification of the hydrophobic pocket can be employed to change binding conformations for substrates.

Other site directed mutagenesis techniques may also be employed with the nucleotide sequences of the invention. For example, restriction endonuclease digestion of DNA followed by ligation may be used to generate deletion variants of valencene synthase, as described in section 15.3 of Sambrook et al. (*ibid*). A similar strategy may be used to construct insertion variants, as described in section 15.3 of Sambrook et al. (*ibid*).

Oligonucleotide-directed mutagenesis may also be employed for preparing substitution variants of this invention. It may also be used to conveniently prepare the deletion and insertion variants of this invention. This technique is well known in the

art as described, for example, by Adelman et al. (DNA 2:183 1983); Sambrook et al., (*ibid*).

Generally, oligonucleotides of at least 25 nucleotides in length are used to insert, delete or substitute two or more nucleotides in the nucleic acid molecules encoding valencene synthase of the invention. An optimal oligonucleotide will have 12 to 15 perfectly matched nucleotides on either side of the nucleotides coding for the mutation. To mutagenize nucleic acids encoding the native valencene synthases of the invention, the oligonucleotide is annealed to the single-stranded DNA template molecule under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of *E. coli* DNA polymerase I, is then added. This enzyme uses the oligonucleotide as a primer to complete the synthesis of the mutation-bearing strand of DNA. Thus, a heteroduplex molecule is formed such that one strand of DNA encodes the native synthase inserted in the vector, and the second strand of DNA encodes the mutated form of the synthase inserted into the same vector. This heteroduplex molecule is then transformed into a suitable host cell.

Mutants substituted with more than one amino acid may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located in some distance from each other (e.g., separated by more than ten amino acids) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each substituted amino acid. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second DNA strand synthesized from the template will encode all of the desired amino acid substitutions. An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: native valencene synthase DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid

substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and the second rounds of mutagenesis. The mutagenized DNA can then be used as a template in a third round of mutagenesis, and so on.

According to one currently preferred embodiment, the valencene synthase of the present invention is derived from citrus species, particularly an orange, optionally the Valencia™ orange.

Valencia™ orange (*Citrus sinensis* cv. Valencia) is a late ripening Spanish variety of oranges that is commercially harvested in Israel starting in April. The accumulation of valencene during the development and ripening of Valencia™ oranges was analyzed (Fig. 4). For this purpose Valencia™ orange fruit was collected at monthly intervals during the 2003 season (Fig. 4A) or the 2001 season (Fig. 4B). In 2003, total RNA was extracted from the flavedo of the fruits and was subjected to quantitative RT-PCR analysis using *Cstps1* specific primers (Fig. 4A). Results are presented with reference to the amplification of the 18S rRNA reference gene.

In 2001, total RNA was extracted from the flavedo of the fruits and was subjected to RNA-blot analysis using a *Cstps1* probe. Ribosomal RNA is visualized by ethidium bromide fluorescence served as a loading reference for the RNA-blot analysis (Fig. 4B).

A minor peak of valencene was observed in fruit picked during October. However, significant levels of valencene were detected starting January (approximately 1-2 months after fruit color-break) and continued to accumulate until fruit was fully mature (May). The level of *Cstps1* transcript in Valencia™ orange throughout fruit development and maturation was measured using quantitative RT-PCR as well as RNA blot. *Cstps1* transcript was detected (in both detection systems) starting December (Fig. 4A-B) and continued to progressively accumulate towards fruit maturation, thus corresponding well with the timing of valencene accumulation.

While citrus are classified as non-climacteric fruits, ethylene has been implicated in various aspects of citrus fruit ripening. The accumulation of valencene in Valencia™ orange fruit picked during March and treated with ethylene for 7 days was therefore monitored. Fruit treated with ethylene accumulated over 20% more valencene than the level measured for fruit treated with air (Figure 5A). Correlatively,

quantitative RT-PCR analysis showed that *Cstps1* expression was enhanced over 8 fold in fruit treated with ethylene for 48 hours compared to fruit treated with air (Figure 5B).

A gene encoding valencene synthase may be incorporated into any organism capable of synthesizing terpenes, or cell culture derived thereof.

The valencene-encoding gene may be incorporated into the organism for a variety of purposes, including but not limited to production of valencene synthase; production of valencene, production of products downstream to valencene and production or modification of flavor and aroma compounds.

According to yet another aspect, the present invention provides a host cell comprising an expression vector comprising a nucleic acid sequence encoding a valencene synthase the valencene synthase being capable of converting FPP to valencene.

Host cells are transfected and preferably transformed with the above-described expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

The host cell may be transformed with the expression vector according to the present invention by using any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The transformation process results in the expression of the inserted DNA such as to change the recipient cell into a transformed, genetically modified or transgenic cell.

Prokaryotic as well as eukaryotic expression systems may be utilized for the production of valencene synthase and its product valencene, and valencene metabolites downstream in the terpene biosynthesis pathway. Both systems comprise the necessary elements for posttranslational modification enabling the proper activity of the enzyme, as well as the necessary substrates for the synthesis of valencene and the enzymes for the synthesis of downstream valencene metabolites.

As is known to a person skilled in the art, many bacterial strains are suitable as host cells for the over-expression of sesquiterpene synthases according to the present invention, including *E. coli* strains and many other species and genera of prokaryotes including bacilli such *as Bacillus subtilis*, other enterobacteriaceae such as *Salmonella*

typhimurium or *Serratia marcesans*, and various *Pseudomonas* species. Prokaryotic host cells or other host cells with rigid cell walls are preferably transformed using the calcium chloride method as described in section 1.82 of Sambrook et al., (*ibid.* Alternatively, electroporation may be used for transformation of these cells. Prokaryote transformation techniques are known in the art, e.g. Dower, W. J., in Genetic Engineering, Principles and Methods, 12:275-296, Plenum Publishing Corp., 1990; Hanahan et al., Meth. Enzymol., 204:63 1991.

Since the reading frame encoded by *Cstps1* mostly resembles plant sesquiterpene synthases, which are cytosolic, and apparently no sequences coding for transit peptides are present (as predicted by Target P software: Emanuelsson *et al.*, *J. Mol. Biol.*, 2000, 300:1005-1016), the entire reading frame of *Cstps1* was cloned into expression vectors in order to obtain a catalytically active recombinant enzyme in bacteria.

Two systems for recombinant enzyme production were employed to enhance the prospect of obtaining soluble and functional enzyme; (1) Production of a nearly native recombinant protein containing an amino terminal histidine tag using the expression vector pRSETb (Invitrogen, The Netherlands) resulted in a recombinant product of 64 KDa. (2) Production as part of a large fusion protein (with a 55 kDa intein-CBD) using the expression vector pTYB2 (New England Biolabs, MA, USA) resulted in a recombinant product of 120 kDa. Conditions were worked out to obtain substantial amounts of soluble and enzymatically active gene products. Lysates prepared from bacteria over expressing recombinant *Cstps1* (from both constructs) supplemented with Mg^{2+} were found to convert radio-labeled FPP into an hexane soluble product as exemplified herein below (Fig. 2). No such product was obtained in bacterial lysates derived from bacteria harboring the expression plasmids without the *Cstps1* insert.

The present invention thus discloses the conversion of radiolabeled FPP into a putative sesquiterpene olephin by the activity of *Cstps1* gene product. Similar preparations were unable to convert geranyl diphosphate (GPP) into monoterpene olephins (Fig. 6).

As detailed herein below, up scaled *in-vitro* assays using non-radiolabeled FPP substrate were used to produce sufficient product of the recombinant enzyme activity, to enable the chemical analysis of such product. A unique sesquiterpene peak was

detected in assays conducted with extracts of bacteria producing recombinant Cstps1 relative to control assays conducted with extracts of bacteria lacking the *Cstps1* gene (Fig. 7A and 7C). This peak was analyzed by mass-spectrometry and was identified as valencene by: (1) comparison of its retention time to the retention time of valencene (Fig. 7A-B); (2) comparison of the mass spectrum obtained to the spectrum of valencene from the internal mass spectra library and to the spectrum of valencene obtained by the assay of the invention (Fig. 7D-F). Thus, valencene is the only product produced by recombinant Cstps1 from FPP. Accordingly, the present invention discloses that the enzyme encoded by *Cstps1* is valencene synthase.

Valencene exerts an odor characterized as orange/woody/citrus and is one of the important albeit low concentration products of citrus essential oils. It is of interest to the food and cosmetic industry on its own, and as a substrate for synthesis of nootkatone, which exerts a dominant grapefruit/citrus aroma. The oxygenation of valencene to nootkatone is a straightforward chemical conversion in the industry and has been suggested to occur in citrus fruit via hydroxylation followed by dehydrogenation (del Rio et al., *J. Agric. Food Chem.*, 1992, 40: 1488-1490), but no direct proof of this pathway *in-planta* is available. The observation that in Valencia oranges valencene accumulates continuously throughout fruit maturation (Fig. 8), while its accumulation in grapefruit stops concomitant with the appearance of putative downstream compounds (2-hydroxy-valencene followed by nootkatone; del Rio et al., 1992, *ibid*) is consistent with valencene being an end product in oranges but not in grapefruit.

According to yet another aspect, the present invention provides a plant comprising a polynucleotide sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene.

According to one embodiment, the polynucleotide sequence encoding a valencene synthase being stably integrated into the genome of the plant.

According to another embodiment, the plant produces at least one compound selected from the group consisting of: valencene, valencene metabolite other than nootkatone, nootkatone.

Transgenic plants comprising the valencene synthase according to the present invention may be obtained, for example, by transferring a plasmid that encode

valencene synthase comprising the necessary regulatory elements for expression in plants, and a selectable marker gene, e.g., the *kan* gene encoding resistance to kanamycin, into *Agrobacterium tumifaciens* containing a helper Ti plasmid. This transformation can be carried out using procedures known in the art as disclosed for example in European Patent No. 116718.

Alternatively, any type of vector can be used to transform the plant cell, applying methods such as direct gene transfer (e.g., by microinjection or electroporation), pollen-mediated transformation (as described, for example, in EP270356, WO085/01856 and U.S. Patent No. 4,684,611), plant RNA virus-mediated transformation (as described, for example, in EP067553 and U.S. Patent No. 4,407,956), liposome-mediated transformation (as described, for example, in U.S. Patent No. 4,536,475), and the like.

Other methods, such as microprojectile bombardment are suitable as well. Cells of monocotyledonous plants, such as the major cereals, can also be transformed using wounded and/or enzyme-degraded compact embryogenic tissue capable of forming compact embryogenic callus, or wounded and/or degraded immature embryos as described in International Patent Application WO 92/09696. The resulting transformed plant cell can then be used to regenerate a transformed plant in a conventional manner.

The obtained transformed plant can be used in a conventional breeding scheme to produce more transformed plants with the same characteristics or to introduce the valencene synthase in other varieties of the same or related plant species, or in hybrid plants. Seeds obtained from the transformed plants contain the valencene synthase as a stable genomic insert. Seeds, fruits, roots, and other organs or isolated organs thereof obtained from the transformed plants contain the chimeric genes of the invention as a stable genomic insert, and are also encompassed by the present invention.

Although grapevine flowers have been reported to emit valencene and the compound has been detected in low amounts in celery (*Apium graveolens*), mango (*Mangifera indica*), olives (*Olea europea*) and even coral, valencene is mostly associated with citrus fruit. No other genes encoding valencene synthases from any other genus are known.

The present invention discloses for the first time a gene encoding valence synthase. This novel valencene synthase evolved, most likely, within the genus *Citrus*. As described herein above, the two additional sesquiterpene synthase-encoding genes from citrus were found to be more similar to valencene synthase than to any of the other published sesquiterpene synthases. Therefore, one can speculate that much of the diversity of sesquiterpene synthases developed in citrus after speciation, as previously noted for other genera.

According to yet another aspect, the present invention provides valencene and velencene metabolite produced downstream in the sesquiterpene biosynthesis pathway obtained by the methods of the present invention for industrial uses.

According to one embodiment, the present invention provides valencene and at least one valencene metabolite, obtained by the methods of the present invention, for use in an industrial applications selected from the agriculture, cosmetics and food.

According to another embodiment, the at least one valencene metabolite is nootkatone.

The principles of the invention, disclosing a novel sesquiterpene synthase, polynucleotides encoding same, methods of productions and methods for use may be better understood with reference to the following non-limiting examples.

EXAMPLES

Plant material

Citrus sinensis cv. Valencia fruit was harvested from the Hebrew University Faculty of Agriculture grove in Rehovot courtesy of Prof. Eliezer Goldschmidt. Fruits for analysis of developmental accumulation of valencene and gene expression were collected at monthly intervals. The flavedo tissue was isolated and frozen at -80°C until use. Fruits used for determining the response to ethylene were treated post-harvest with either ethylene (14 µl/ l) or air for 48 h or 7 days; the flavedo tissue was isolated and frozen at -80°C until use.

RNA extraction and analysis

Total RNA was extracted from orange flavedo as previously described (Jacob-Wilk *et al.*, Plant J. 1999, 20(6):653-61). Poly-A RNA was purified for RT-PCR applications by PolyAtract mRNA Isolation system III (Promega, WI, USA). Northern analysis was performed using total RNA (20 μg per-lane) according to standard procedures (Sambrook *et al.*, *ibid*). Ribosomal RNA served as a loading reference and was visualized by ethidium bromide staining.

Quantitative PCR was performed using an ABI Prism 7000 sequence detection system and a SyberGreen kit (Applied Biosystems, Warrington, UK) according to the manufacturer's instructions. 5 μg total RNA from each sample was reverse transcribed using MMLV reverse transcriptase (Gibco-BRL, MD, USA) at a reaction volume of 20 μl according to the manufacturer's instructions. The Quantitative PCR reaction consisted of 10 μl master mix (Applied Biosystems, Warrington, UK), 3 μl of amplicon primers (2 μM) and 1 μl of the reverse-transcription reaction (for 18S rRNA reference gene the reverse transcription reaction was diluted 10^5 fold) in a final volume of 20 μl.

The amplicon for *Cstps1* consisted of 150 bp with the following primers:

"Sense" primer- 5' CCCAGGCGTTGTACTTCATCA (SEQ ID NO:6).

"Antisense" primer- 5' CGACACGAGGCACTGAAAGA (SEQ ID NO:7). The amplicon for the 18S rRNA reference gene consisted of 150 bp with the following primers:

5' GCGACGCATCATTCAAATTTC (SEQ ID NO:8).

5' TCCGGAATCGAACCCTAATTC (SEQ ID NO:9).

Each sample was analyzed in triplicate, and the results represent normalized mean values and standard deviation.

Example 1: Isolation of the cDNA Cstps1

Poly-A RNA (500ng) from Valencia™ Orange flavedo was reverse transcribed using Superscript II reverse transcriptase (BRL, Life Technologies, UK) and 10 pmoles of modified oligo-(dT) primer (5'CGGCTAGCATGCTTTTTTTTTTTTTT,

SEQ ID NO:10). PCR was performed using 2 degenerate primers matching conserved sequence elements between various mono and sesquiterpene synthases:

5' GAYGAYIIIITWYGAYGYITWYGG (SEQ ID NO:11).

5' YTKCATRTAITCNGG (SEQ ID NO:12).

Conditions were worked out to enhance the expected band of 110 bp (40 amplification cycles of 94°C 20 seconds; 44°C 20 seconds; 72°C 20 seconds), which was purified and cloned into a pGEM-T plasmid (Promega, WI, USA). Relevant clones were selected based on the presence of an internal terpene synthase conserved element in the sequence.

One clone was selected to isolate a complete cDNA by 5' and 3' by rapid amplification of cDNA ends (RACE). Sequence information from this clone was used to design 2 gene-specific primers, one for each direction of RACE:

5' CAGTAAAGAGGCTGAGTTCTTC for 5' RACE (SEQ ID NO:13).

5' GAAGAACTCAGCCTCTTTACTG for 3' RACE (SEQ ID NO:14).

Amplification of 5' and 3' segments of the cDNA was performed with DNA extracted from a cDNA library of near-ripe Valencia orange flavedo (Jacob-Wilk *et al.*, *ibid*) using one gene-specific primer (above) and one universal primer (T3 for 5' RACE and T7 for 3' RACE). The complete cDNA, designated as *Cstps1* comprises 1647 nucleotides, as described in Fig. 1.

Example 2: Cloning of recombinant *Cstps1* in *E.coli*

The entire *Cstps1* reading frame was amplified from the cDNA library by PCR using EXTAQ polymerase (Takara, Japan), digested at the relevant restriction sites and cloned into the expression vectors pTYB2 (New England Biolabs, MA, USA) and pRSET (Invitrogen, The Netherlands). Cloning into pTYB2 at the restriction sites NdeI and SmaI involved PCR amplification using the primers:

5' GAGAGTCCATATGTCGTCTGGAGAAACATTTCG (SEQ ID NO:15).

5' AAATGGAACGTGGTCTCCTAGC (SEQ ID NO:16).

Cloning into pRSET at the restriction sites NheI and SalI involved PCR amplification using the primers:

5' CGATGCTAGCTCGTCTGGAGAAACATTTC (SEQ ID NO:17).

5' CGTAGTCGACTCAAAATGGAACGTGGTCTCCTAG (SEQ ID NO:18).

The pTYB2 construct was expressed in *E.coli* ER2566 according to the manufacturers instructions (New England Biolabs, MA, USA) for 6 hours post-induction at 25°C. The pRSET construct was expressed in *E.coli* BL21 (DE3) LysE according to the manufacturers instructions (Invitrogen, The Netherlands) for 6 hours post-induction at 25°C. Bacterial cells were either lysed in sample buffer and subjected to SDS-PAGE (Laemmli, *Nature*, 1970, 227(259):680-5) or stored frozen at –20°C until further use.

Example 3: Identification and analysis of products resulted from Cstps1 activity

Growth of bacteria and induction for *Cstsp1* expression

Recombinant BL21 (D3) p*LysS*- bacteria were plated in LB-agar containing 50 μg/ml ampicillin and 34 μg/ ml chloramphenicol. The recombinant bacteria contained either an empty plasmid (control), or plasmid comprising the *Cstsp1* gene (assay). Individual colonies were grown in 2 ml LB liquid medium containing 50 μg/ml ampicillin over night, to be used as starter cultures. Five hundred μl of bacterial-cell-suspensions were transferred into 50 ml LB liquid medium containing ampicillin and grown at 37°C with shaking (200 rpm) until OD_{600} reached 0.6. IPTG was then added to a final concentration of 0.3 mM and the cultures grown for another 5 hours at room temperature. 1.5 ml aliquots were transferred to 2ml polypropylene tubes. Cells were harvested by centrifugation at 20,000 *g* for 10 min at 4°C and frozen at –20°C until use.

Preparation of bacterial lysates

Individual bacterial pellets were suspended in 500 μl of a 50 mM bis-tris buffer pH 6.9 containing 10 % v/v glycerol, 10 mM DTT and 5 mM sodium metabisulfite (Reaction Buffer). Ten μg/ml chicken egg white lysozyme chloride (grade VI, Sigma, 60,000 u/mg protein) were added. The samples were vigorously mixed and incubated in ice water (4°C) for 15 min. After the cells lysed, the suspensions were centrifuged (20,000 g, 10 min at 4°C). The supernatants were used fresh for characterization of the enzymatic activity by examining the gene products.

Radioactive (micro) Assay

Lysate aliquots (10 to 40 µl) containing the recombinant enzyme were mixed with 10 mM $MgCl_2$, 0.015 µCi of either ^{3}H-FPP (specific activity 20.5 Ci/mmole, Sigma) to test for sesquiterpene synthase activity or ^{3}H-GPP (specific activity 15 Ci/mmole, ARC) to test for monoterpene synthase activity, and reaction buffer to a total volume of 100 µl. The reactions were overlaid with 1 ml hexane, briefly vortexed and spun, and then incubated for 1 to 2 h at 30°C. The samples were then vortexed and spun and 850 µl of the upper hexane layers transferred to a new tube containing 10 to 50 mg Silica Gel 60 (Merck, Darmstaad, FRG). The tubes vortexed and spun for 1 min at 10,000 g at room temperature, and 600 µl were transferred to 5 ml scintillation tubes containing 3 ml scintillation liquid [2,5 phenyloxazol (PPO, 4 g/l), 2,2-p-phenylene-bis 5-phenyloxazol (POPOP, 0.05 g/l), and 30% (v/v) Triton in toluene]. The radioactivity was quantified using a liquid scintillation counter (Kontron model 810). Enzyme activity was calculated based on the specific activity of the substrate and using appropriate correction factors for the counting efficiency of the scintillation machine (Shalit *et al.*, *J Agric Food Chem.* 2001, 49(2):794-9). Lysates obtained from recombinant bacteria expressing the *Cstps1* gene were found to convert FDP into a hexane-soluble product, which did not bind to silica-gel under the experimental conditions used (Figure 2). No such product was obtained from control lysates derived from bacteria harboring the expression plasmids without the *Cstps1* insert (not shown). This result indicated the conversion of the radiolabeled FPP into a putative sesquiterpene olephin.

Up-scaled production of Cstsp1 activity products

The radioactive assay was scaled-up in aluminum-foil capped glass tubes. One µM of non-radioactive farnesyl diphosphate (Sigma, MO, USA), 10 mM $MgCl_2$ and 400 µl bacterial lysate were mixed in a total volume of 2 ml reaction buffer, overlaid with 2 ml of hexane and incubated overnight at 30°C. Each tube was then shaken and extracted repeatedly (three to five times) with 2 ml of hexane. The hexane layers containing the *in vitro* formed sesquiterpenes were pooled, passed through a small Pasteur pipette filled with Silica Gel, dried with sodium sulfate and concentrated by a Turbo Vac II (Zymark, MA, USA) to a final volume of 400 µl.

GC-MS analysis: Volatile compounds obtained as described above were

analyzed on an HP-GCD apparatus equipped with an HP-5 (30 m × 0.25 mm) fused-silica capillary column. Helium (1 ml/min) was used as a carrier gas. The injector temperature was 250°C, set for splitless injection. The oven was set to 70°C for 2 min, the temperature was increased to 200°C at a rate of 4°C/min and set on hold for 6 min. The detector temperature was 280°C. The mass range was recorded from 45 to 450 *m/z*, with electron energy of 70 eV. Identification of valencene was done by comparison of mass spectra and retention time data known for valencene with those of citrus samples (extracted as described for tomato, Lewinsohn *et al.*, *Plant Physiol.*, 2001, 127:1256-1265), valencene (Frutarom, Israel) and supplemented with those stored in a Wiley GC-MS library data base. As shown in Fig. 7, only one peak was observed in the extract obtained from the lysate of the recombinant bacteria, indicating the production of only one hexane-soluble, detectable product. Comparison of this peak to the peak obtained for a commercially available purified valencene, and to the library database identified the product of the Cstps1 activity as valencene. Therefore, the Cstps1 polypeptide of the present invention was designated as valencene synthase.

Example 4: Sequence analysis and phylogenetic tree software tools

General sequence data manipulations and homology searches were conducted using CuraTools integrated bioinformatics tools (Curagen, CT, USA). Transit peptide analysis was conducted by Target P (Emanuelsson *et al.*, 2000, *ibid*). Phylogenetic trees were obtained using the following software: BioEdit (Hall, 1999, *ibid*) and Treeview (Page, *ibid*). Accession numbers of the sequences used are as follows: (1) AAB95209; (2) AAG17667; (3) AAD02223; (4) AAA86339; (5) CAC36896; (6) CAA77191; (7) AAF80333; (8) AAC39432; (9) AAK15697; (10) Naaman and Eyal, unpublished; (11) AF441124 (SEQ ID NOS:1-2); (12) AAK54279; (13) AAK58723; (14) AAG09310; (15) AAM53946 (16) AAC26016; (17) AAD50304; (18) AAC26017; (19) AAC26018; (20) AAK5990; (21) AAD34295; (22) AAC39443; (23) AAB58822; (24) AAK83563; (25) AAK83561; (26) AAC05728; (27) AAB70707; (28) 024475; (29) 024474; (30) 022340. As presented in Fig. 3, citrus sesquiterpene synthases (numbers 10, 11 and 12 on the phylogenetic tree) form a distinct subgroup of known angiosperm sesquiterpene synthases. Alignment and phylogenetic tree programs, BioEdit (Hall, 1999, *ibid*) and TreeView (Page, *ibid*)

were used to obtain the phylogenetic tree.

Example 5: Valencene production in squash using Cstps1 gene

The cDNA *Cstps1* encoding citrus valencene synthase (Sharon-Asa et al., *ibid*) was cloned into a ZYMV-based viral expression vector in frame with the viral polyprotein sequence. Viral vector containing either no insert (AG) or the *Cstps1* insert (AG-*Cstps1*) were used to infect squash cotyledons. Leaves (3rd, 4th and 5th true leaves of each plant) of viral-vector infected plants were harvested following the appearance of typical infection symptoms and confirmation, by ELISA analysis, was conducted approximately 18 days post-infection. Leaves of wild-type plants were harvested at the corresponding age. Volatiles were extracted from all leaves as previously described for other plant tissues (Sharon-Asa et al., *ibid*) and the presence of valencene was analyzed using GC-MS.

Non-infected squash plants and plants infected with the "empty" viral vector (AG) did not accumulate the sequiterpene valencene, which is common in citrus, but has never been documented in cucurbits. Only plants infected with the viral vector containing citrus valencene synthase (AG-*Cstps1*) were found to accumulate the sesquiterpene valencene (Table 1).

Table 1: Valencene product in squash plants.

Plants and viral vectors	**Valencene product (μg/g fresh weight)**
AG infected plant 1	0.000
AG infected plant 2	0.000
Non-infected plant 1	0.000
Non-infected plant 2	0.000
AG-*Cstps1* infected plant 1	14.717
AG-*Cstps1* infected plant 2	13.375

The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying current knowledge, readily modify and/or adapt for various applications such specific embodiments without undue experimentation and without departing from the generic concept, and, therefore, such adaptations and modifications should and are intended to be comprehended within the meaning and range of equivalents of the disclosed embodiments. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation. The means, materials, and steps for carrying out various disclosed chemical structures and functions may take a variety of alternative forms without departing from the invention.

CLAIMS

1. An isolated polynucleotide comprising a nucleic acid sequence encoding valencene synthase, said valencene synthase being capable of converting farnesylpyrophosphate to valencene.

2. The isolated polynucleotide of claim 1, wherein the valence synthase is derived from citrus species.

3. The isolated polynucleotide of claim 2, wherein the valence synthase is derived from oranges.

4. The isolated polynucleotide of claim 1, comprising a nucleic acid having at least 80% homology to SEQ ID NO:1 or the complement thereof.

5. The isolated polynucleotide of claim 4, comprising a nucleic acid having at least 90% homology to SEQ ID NO:1 or the complement thereof.

6. The isolated polynucleotide of claim 1, comprising a nucleic acid encoding a valencene synthase comprising an amino acid sequence having (a) at least 70% homology to SEQ ID NO:2; and (b) at least one consensus motif of contiguous amino acids as set forth in SEQ ID NO:3.

7. The isolated polynucleotide of claim 1, comprising a nucleic acid encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2.

8. The isolated polynucleotide of claim 7, comprising a nucleic acid encoding a valencene synthase comprising an amino acid sequence having at least 90% homology to SEQ ID NO:2.

9. An isolated polynucleotide comprising a nucleic acid capable of hybridizing to the polynucleotide of any one of claims 1 to 8.

10. An isolated polypeptide comprising an amino acid sequence having an activity of a valencene synthase, the activity is characterized by the ability to convert farnesylpyrophosphate to valencene.

11. The isolated polypeptide of claim 10, wherein the valencene synthase is derived from citrus species.

12. The isolated polypeptide of claim 11, wherein the valencene synthase is derived from oranges.

13. The isolated polypeptide of claim 10, comprising an amino acid sequence having (a) at least 70% identity to SEQ ID NO:2; and (b) at least one consensus motif of contiguous amino acid sequence as set forth in SEQ ID NO:3.

14. The isolated polypeptide of claim 10, comprising an amino acid sequence having at least 80% identity to SEQ ID NO:2 and fragments, derivatives and analogs thereof.

15. The isolated polypeptide of claim 14, comprising an amino acid sequence having at least 90% identity to SEQ ID NO:2 and fragments, derivatives and analogs thereof.

16. An expression vector comprising a nucleic acid sequence encoding valencene synthase, said valencene synthase being capable of converting farnesylpyrophosphate to valencene.

17. The expression vector of claim 16, wherein the valence synthase is derived from citrus species.

18. The isolated polynucleotide of claim 17, wherein the valence synthase is derived from oranges.

19. The expression vector of claim 16, wherein the nucleic acid sequence having at least 80% homology to SEQ ID NO:1 or the complement thereof.

20. The expression vector of claim 19, wherein the nucleic acid sequence having at least 90% homology to SEQ ID NO:1 or the complement thereof.

21. The expression vector of claim 16, wherein the nucleic acid sequence

encoding a valencene synthase comprising an amino acid sequence having (a) at least 70% homology to SEQ ID NO:2; and (b) at least one consensus motif of contiguous amino acids as set forth in SEQ ID NO:3.

22. The expression vector of claim 16, the nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2.

23. The expression vector of claim 22, comprising a nucleic acid encoding a valencene synthase comprising an amino acid sequence having at least 90% homology to SEQ ID NO:2.

24. The expression vector of any one of claims 16 to 23, selected from a plasmid or a virus.

25. The expression vector of any one of claims 16 to 23, further comprising at least one element selected from the group consisting of: promoter operatively linked to the polynucleotide encoding the valencene synthase, a selection marker, a signal sequence, an origin of replication, an enhancer and a transcription termination sequence.

26. A host cell comprising an expression vector, the expression vector comprising a nucleic acid sequence encoding valencene synthase, said valencene synthase being capable of converting farnesylpyrophosphate to valencene.

27. The host cell of claim 26, wherein the valence synthase is derived from citrus species.

28. The host cell of claim 27, wherein the valence synthase is derived from oranges.

29. The host cell of claim 26, the nucleic acid sequence having at least 80% homology to SEQ ID NO:1 or the complement thereof.

30. The host cell of claim 29, the nucleic acid sequence having at least 90% homology to SEQ ID NO:1 or the complement thereof.

31. The host cell of claim 26, wherein the nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having (a) at least 70% homology to SEQ ID NO:2; and (b) at least one consensus motif of contiguous amino acids as set forth in SEQ ID NO:3.

32. The host cell of claim 26, the nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2.

33. The host cell of claim 32, comprising a nucleic acid encoding a valencene synthase comprising an amino acid sequence having at least 90% homology to SEQ ID NO:2.

34. The host cell of any one of claims 26 to 33, being prokaryotic or eukaryotic.

35. The host cell of claim 34, said host cell produces at least one compound selected from the group consisting of: valencene, a valencene metabolite other than nootkatone, nootkatone.

36. The host cell of claim 35, the host cell is a bacterial cell.

37. The host cell of claim 36, the host cell is E. Coli.

38. The host cell of claim 34, the host cell is prokaryotic wherein the nucleic acid sequence encoding valencene synthase is stably integrated into the genome of said cell.

39. A method for producing recombinant valencene synthase, the method comprising:

culturing a host cell comprising an expression vector, wherein the expression vector comprises a nucleic acid sequence encoding a valencene synthase, the valencene synthase being capable of converting FPP to valencene, under conditions suitable for the expression of said valencene synthase.

40. The method of claim 39, further comprising:

recovering said valencene synthase.

41. The method of claim 39, further comprising:

recovering valencene.

42. The method of claim 39, further comprising:

recovering at least one valencene metabolite.

43. The method of claim 42, wherein the at least one valence metabolite is nootkatone.

44. The method of claim 39, wherein the valence synthase is derived from citrus species.

45. The method of claim 44, wherein the valence synthase is derived from oranges.

46. The method of claim 39, wherein the nucleic acid having at least 80% homology to SEQ ID NO:1 or the complement thereof.

47. The method of claim 46, wherein the nucleic acid having at least 90% homology to SEQ ID NO:1 or the complement thereof.

48. The method of claim 39, wherein the nucleic acid encodes a valencene synthase comprising an amino acid sequence having (a) at least 70% homology to SEQ ID NO:2; and (b) an amino acid sequence comprising at least one consensus motif of contiguous amino acids as set forth in SEQ ID NO:3

49. The method of claim 39, wherein the nucleic acid encodes a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2.

50. The method of claim 49, wherein the nucleic acid encodes a valencene synthase comprising an amino acid sequence having at least 90% homology to SEQ ID NO:2.

51. The method of claim 39, wherein the host cell is selected from

prokaryotic or eukaryotic.

52. The method of claim 51, wherein the host cell is prokaryotic and wherein the nucleic acid sequence encoding valencene synthase is stably integrated into the genome of said cell.

53. The method of claim 51, wherein the host cell is a bacterial cell.

54. The method of claim 53, wherein the host cell is E. Coli.

55. The method of claim 39, wherein the expression vector is selected from a plasmid or a virus.

56. The method of claim 55, wherein the expression vector further comprises at least one element selected from the group consisting of: promoter operatively linked to the polynucleotide encoding the valencene synthase, a selection marker, a signal sequence, an origin of replication, an enhancer and a transcription termination sequence.

57. Use of valencene obtained by the methods according to any one of claims 39 to 56, in an industrial application selected from the group consisting of: agriculture, cosmetics and food.

58. A plant comprising a nucleic acid sequence encoding valencene synthase, said valencene synthase being capable of converting farnesylpyrophosphate to valencene.

59. The plant of claim 58, wherein the valence synthase is derived from citrus species.

60. The plant of claim 59, wherein the valence synthase is derived from oranges.

61. The plant of claim 58, the nucleic acid sequence having at least 80% homology to SEQ ID NO:1 or the complement thereof.

62. The plant of claim 61, the nucleic acid sequence having at least 90% homology to SEQ ID NO:1 or the complement thereof.

63. The plant of claim 58, the nucleic acid sequence encoding a valencene synthase comprising an amino acid sequence having (a) at least 70% homology to SEQ ID NO:2 and (b) at least one consensus motif of contiguous amino acids as set forth in SEQ ID NO:3.

64. The plant of claim 58, the nucleic acid sequence encodes a valencene synthase comprising an amino acid sequence having at least 80% homology to SEQ ID NO:2.

65. The plant of claim 64, comprising a nucleic acid encodes a valencene synthase comprising an amino acid sequence having at least 90% homology to SEQ ID NO:2.

66. The plant of any one of claims 58 to 65, said plant produces at least one compound selected from the group consisting of: valencene, a valencene metabolite other than nootkatone, nootkatone.

67. The plant of claim 58, wherein the nucleic acid sequence encoding valencene synthase is stably integrated into the genome of said plant.

```
1       ATGTCGTCTG GAGAAACATT TCGTCCTACT GCAGATTTCC ATCCTAGTTT
51      ATGGAGAAAC CATTTCCTCA AAGGTGCTTC TGATTTCAAG ACAGTTGATC
101     ATACTGCAAC TCAAGAACGA CACGAGGCAC TGAAAGAAGA GGTAAGGAGA
151     ATGATAACAG ATGCTGAAGA TAAGCCTGTT CAGAAGTTAC GCTTGATTGA
201     TGAAGTACAA CGCCTGGGGG TGGCTTATCA CTTTGAGAAA GAAATAGGAG
251     ATGCAATACA AAAATTATGT CCAATCTATA TTGACAGTAA TAGAGCTGAT
301     CTCCACACCG TTTCCCTTCA TTTTCGGTTG CTTAGGCAGC AAGGAATCAA
351     GATTTCATGT GATGTGTTTG AGAAGTTCAA AGATGATGAG GGTAGATTCA
401     AGTCATCGTT GATAAACGAT GTTCAAGGGA TGTTAAGTTT GTACGAGGCA
451     GCATACATGG CAGTTCGCGG AGAACATATA TTAGATGAAG CCATTGCTTT
501     CACTACCACT CACCTGAAGT CATTGGTAGC TCAGGATCAT GTAACCCCTA
551     AGCTTGCGGA ACAGATAAAT CATGCTTTAT ACCGTCCTCT TCGTAAAACC
601     CTACCAAGAT TAGAGGCGAG GTATTTTATG TCCATGATCA ATTCAACAAG
651     TGATCATTTA TGCAATAAAA CTCTGCTGAA TTTTGCAAAG TTAGATTTTA
701     ACATATTGCT AGAGCTGCAC AAGGAGGAAC TCAATGAATT AACAAAGTGG
751     TGGAAAGATT TAGACTTCAC TACAAAACTA CCTTATGCAA GAGACAGATT
801     AGTGGAGTTA TATTTTTGGG ATTTAGGGAC ATACTTCGAG CCTCAATATG
851     CATTTGGGAG AAAGATAATG ACCCAATTAA ATTACATATT ATCCATCATA
901     GATGATACTT ATGATGCGTA TGGTACACTT GAAGAACTCA GCCTCTTTAC
951     TGAAGCAGTT CAAAGATGGA ATATTGAGGC CGTAGATATG CTTCCAGAAT
1001    ACATGAAATT GATTTACAGG ACACTCTTAG ATGCTTTTAA TGAAATTGAG
1051    GAAGATATGG CCAAGCAAGG AAGATCACAC TGCGTACGTT ATGCAAAAGA
1101    GGAGAATCAA AAAGTAATTG GAGCATACTC TGTTCAAGCC AAATGGTTCA
1151    GTGAAGGTTA CGTTCCAACA ATTGAGGAGT ATATGCCTAT TGCACTAACA
1201    AGTTGTGCTT ACACATTCGT CATAACAAAT TCCTTCCTTG GCATGGGTGA
1251    TTTTGCAACT AAAGAGGTTT TTGAATGGAT CTCCAATAAC CCTAAGGTTG
1301    TAAAAGCAGC ATCAGTTATC TGCAGACTCA TGGATGACAT GCAAGGTCAT
1351    GAGTTTGAGC AGAAGAGAGG ACATGTTGCG TCAGCTATTG AATGTTACAC
1401    GAAGCAGCAT GGTGTCTCTA AGGAAGAGGC AATTAAAATG TTTGAAGAAG
1451    AAGTTGCAAA TGCATGGAAA GATATTAACG AGGAGTTGAT GATGAAGCCA
1501    ACCGTCGTTG CCCGACCACT GCTCGGGACG ATTCTTAATC TTGCTCGTGC
1551    AATTGATTTT ATTTACAAAG AGGACGACGG CTATACGCAT TCTTACCTAA
1601    TTAAAGATCA AATTGCTTCT GTGCTAGGAG ACCACGTTCC ATTTTGA
```

FIGURE 1

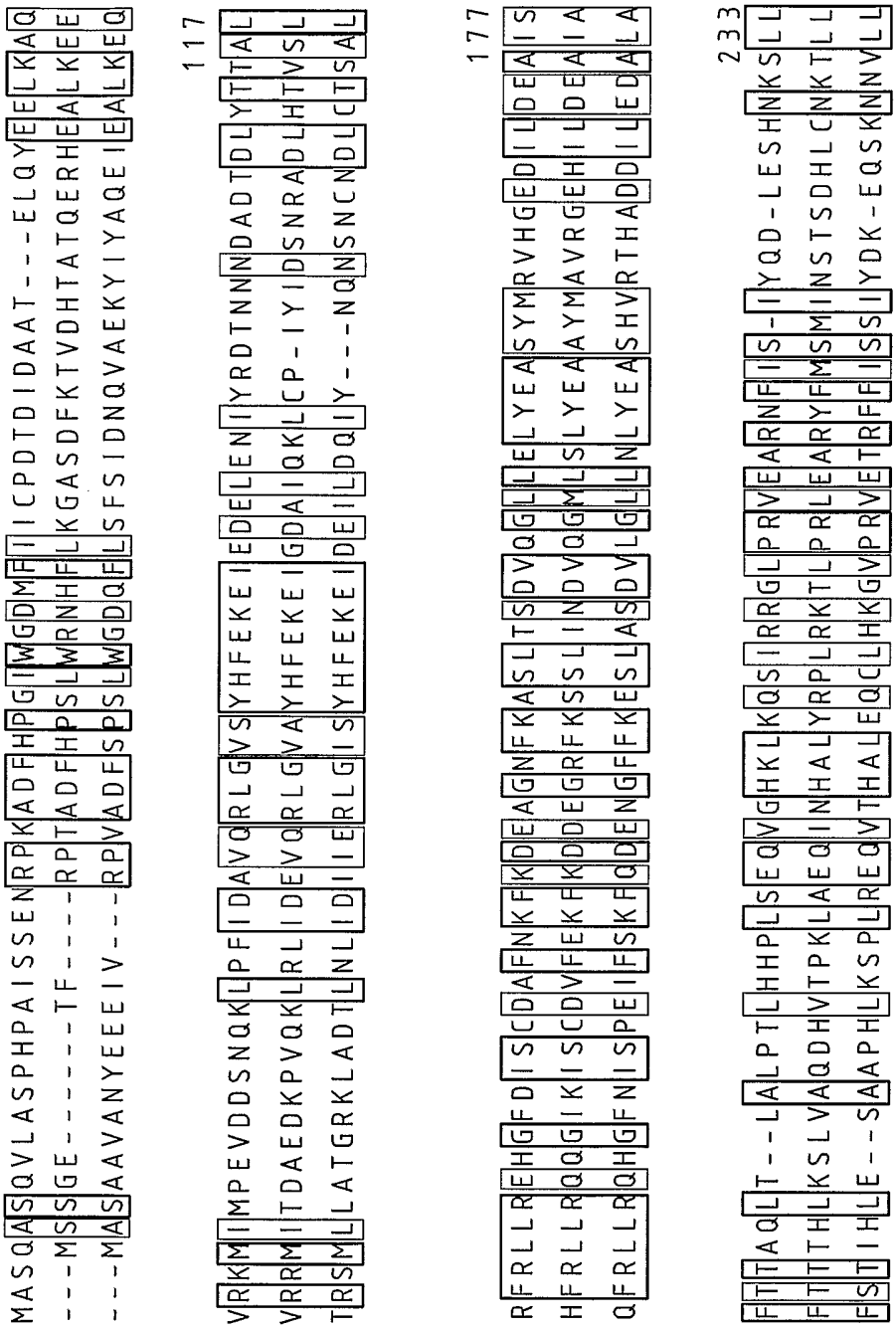
57
117
177
233
cdainene syn
cstps1
epi-aristolochene syn
delta-cdainene syn
cstps1
epi-aristolochene syn
delta-cdainene syn
cstps1
epi-aristolochene syn
delta-cdainene syn
cstps1
epi-aristolochene syn

FIG.2

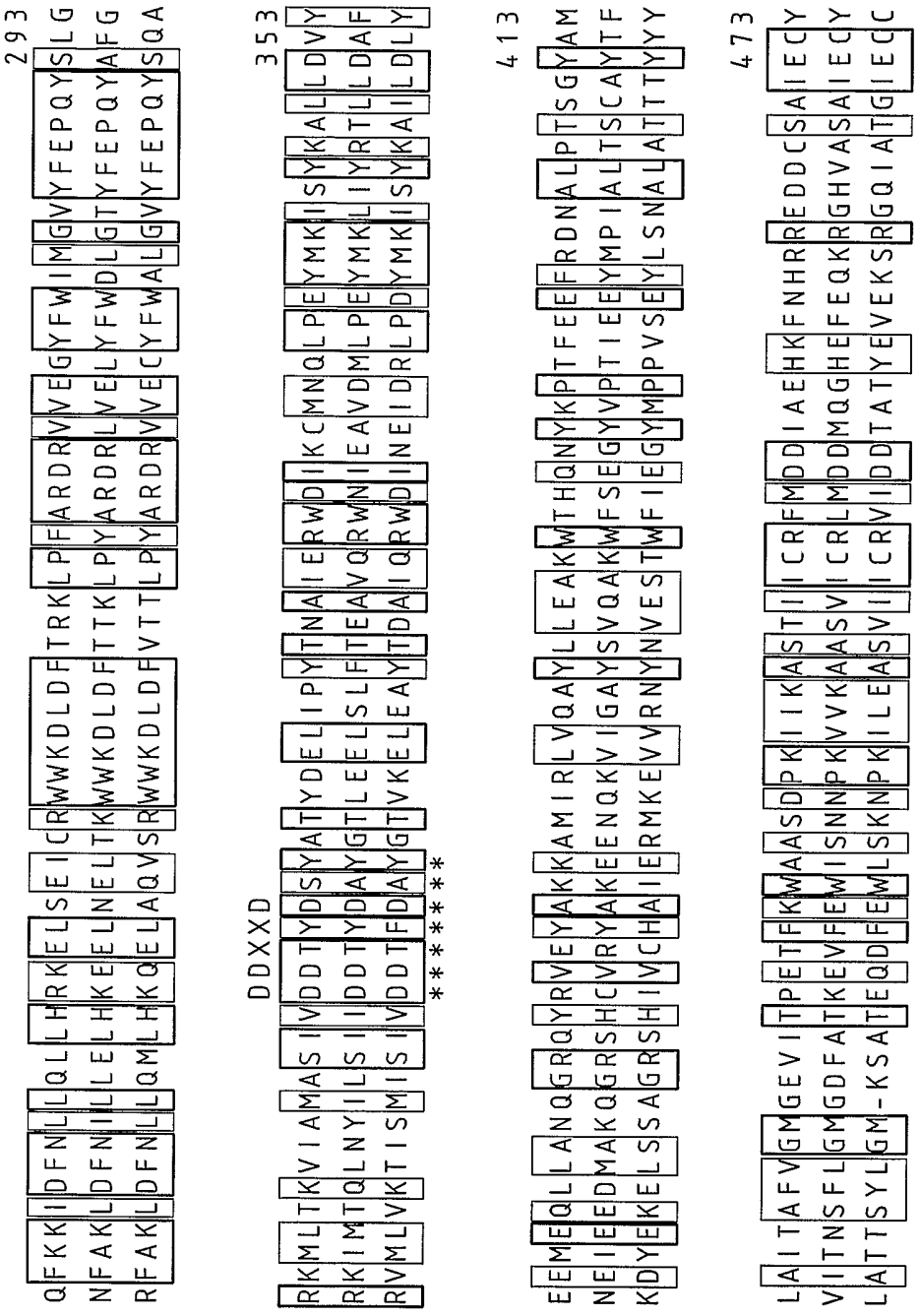
293
cdainene syn
cstps1
epi-aristolochene syn
353
DDXXD
413
473

FIG.2 cont. (1)

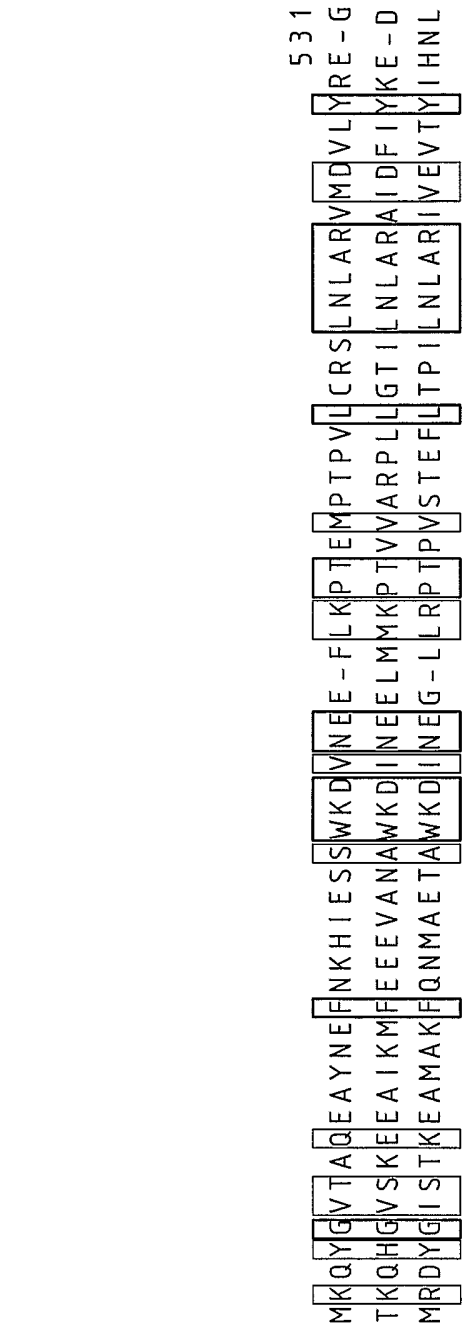
531
cdainene syn
cstps1
epi-aristolochene syn
555
cdainene syn
cstps1
epi-aristolochene syn

FIG.2 cont. (2)

Sesquiterpene synthases (angiosperms)
1. Farnesene synthase (*Mentha piperita*)
2. epi-Aristolochene synthase (*Nicotiana tabacum*)
3. Vetispiradiene synthase (*Solanum tuberosum*)
4. Vetispiradiene synthase (*Hyoscyamus muticus*)
5. Germacrene A synthase (*Solidago canadensis*)
6. delta-Cadinene synthase (*Gossypium arboreum*)
7. 8-Epicedrol synthase (*Artemisia annua*)
8. Germacrene C synthase (*Lycopersicon esculentum*)
9. Amorpha 4,11 diene synthase (*Artemisia annua*)
10. Cmtps1 (*Citrus grandis*)
11. Farnesene synthase (*Citrus junos*)
12. Valencene synthase, Cstps1 (*Citrus sinensis*)
Monoterpene synthases (gymnosperms)
27. Camphene synthase (*Abies grandis*)
28. Pinene synthase (*Abies grandis*)
29. Myrcene synthase (*Abies grandis*)
30. Limonene synthase (*Abies grandis*)
Sesquiterpene synthases (gymnosperms)
25. delta-Selinene synthase (*Abies grandis*)
26. Humulene synthase (*Abies grandis*)
Diterpene synthases
20. Kaurene synthase (*Zea mays*)
21. Kaurene synthase (*Stevia rebaudiana*)
22. ent-Kaurene synthase (*Arabidopsis thaliana*)
23. ent-Kaurene synthase (*Pisum sativum*)
24. Abietadiene synthase (*Abies grandis*)
Monoterpene synthases (angiosperms)
13. beta-Pinene synthase (*Artemisia annua*)
14. Myrcene-Ocimene synthase (*Arabidopsis thaliana*)
15. Limonene synthase (*Citrus limon*)
16. 8-Cineole synthase (*Salvia officinalis*)
17. Limonene synthase (*Mentha longifolia*)
18. Bornyl diphosphate synthase (*Salvia officinalis*)
19. Sabinene synthase (*Salvia officinalis*)

FIGURE 3

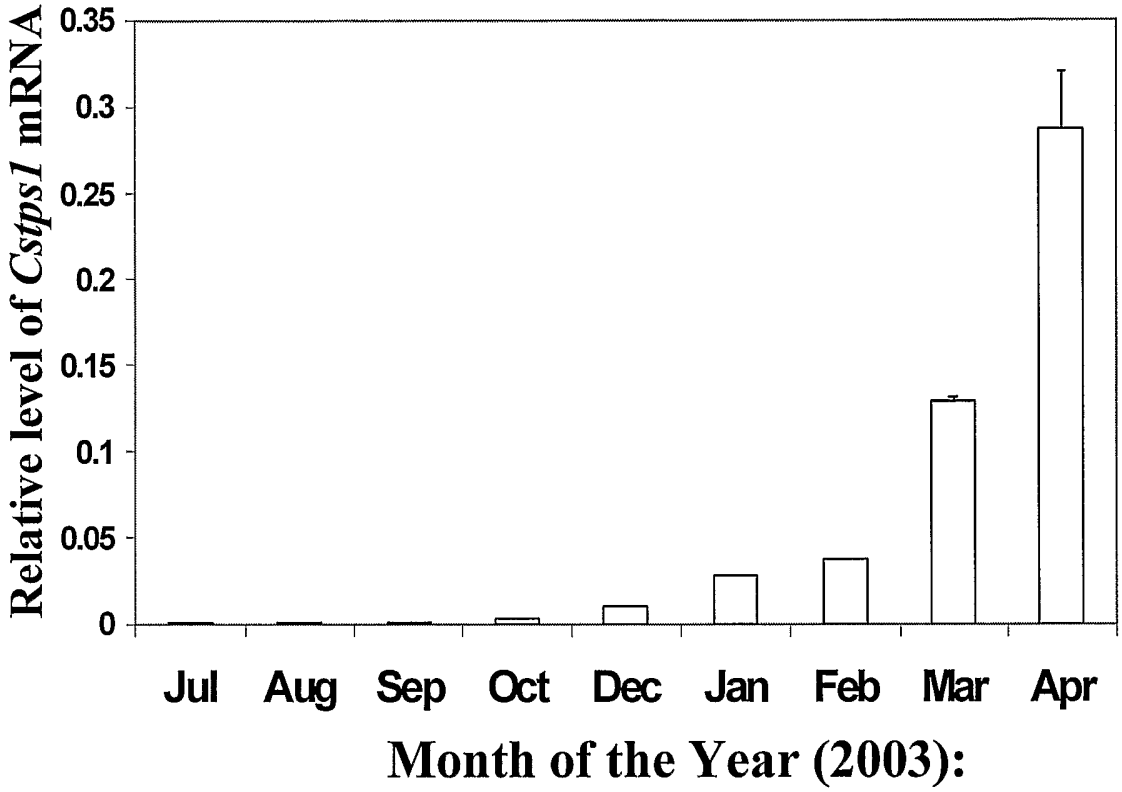

FIGURE 4A

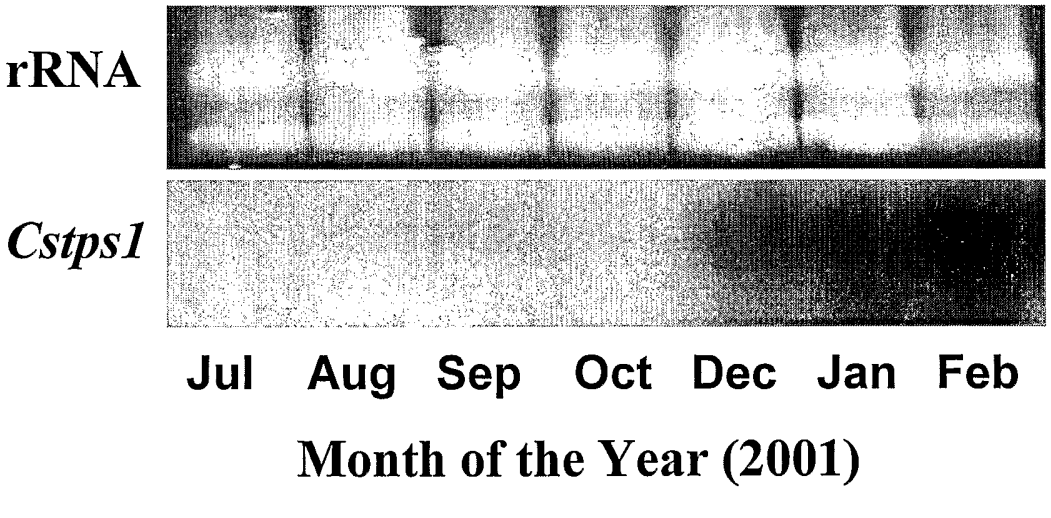

FIGURE 4B

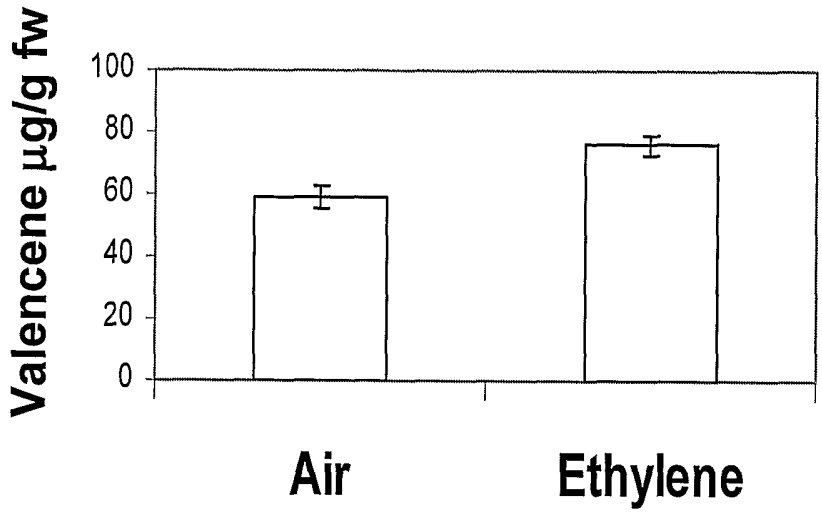

FIGURE 5A

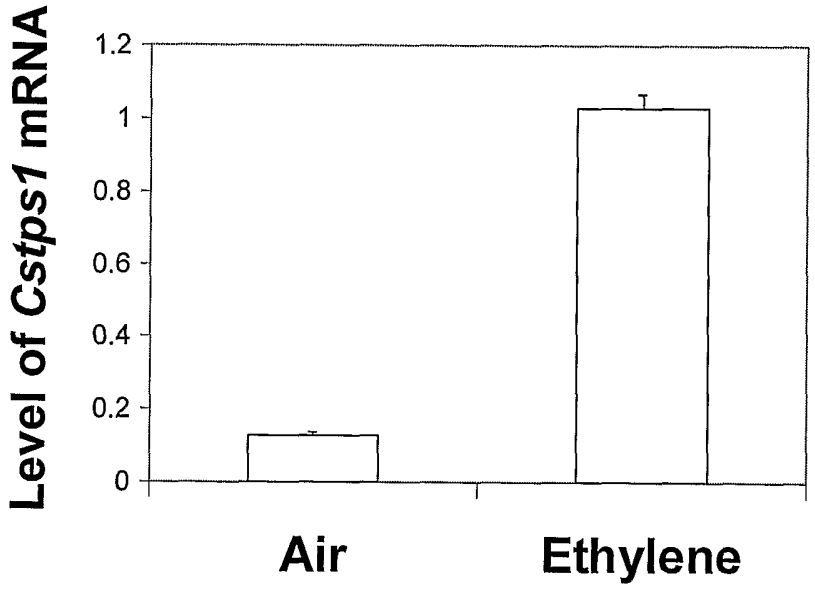

FIGURE 5B

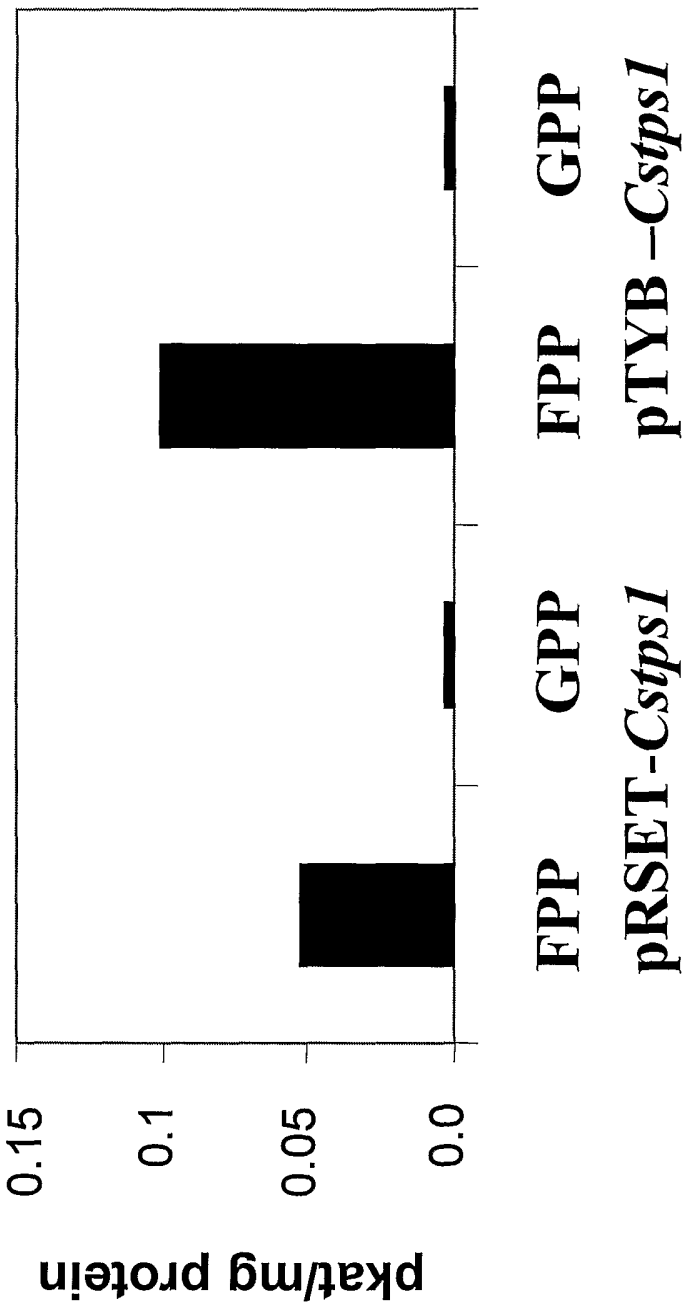
pkat/mg protein
0.15
0.1
0.05
0.0
FPP
GPP
pRSET-Cstps1
FPP
GPP
pTYB –Cstps1

FIGURE 6

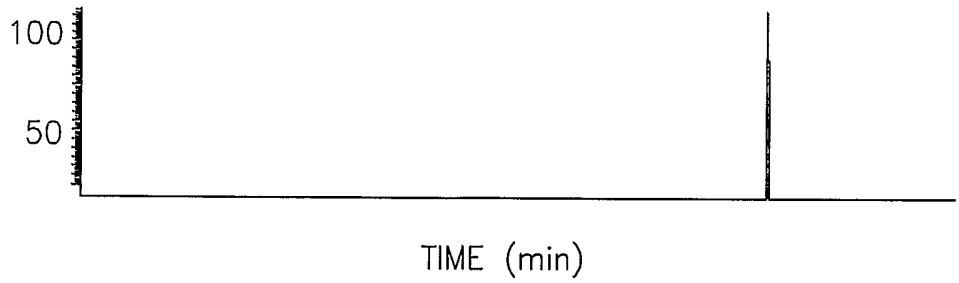

FIG.7A

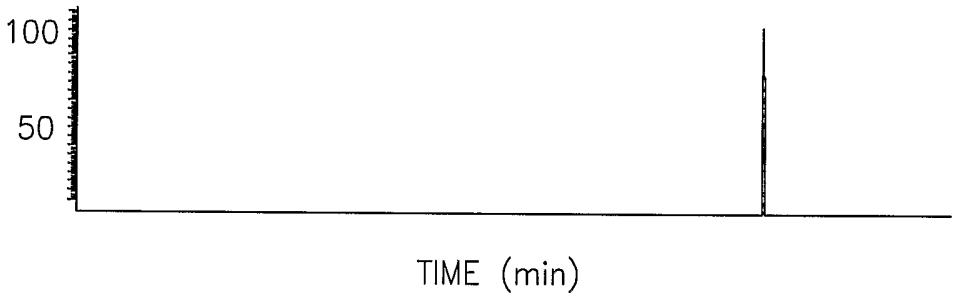

FIG.7B

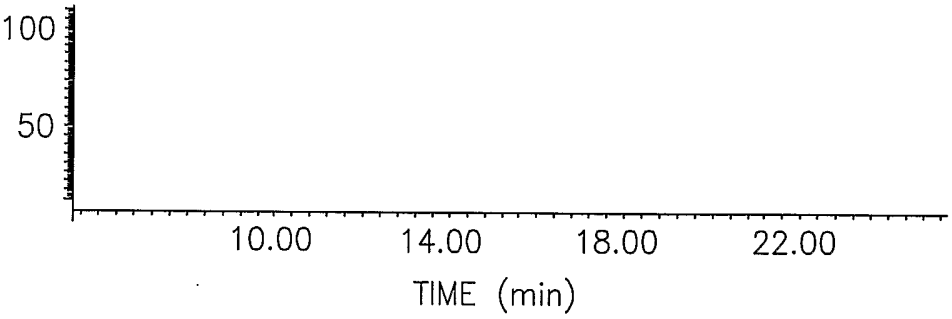

FIG.7C

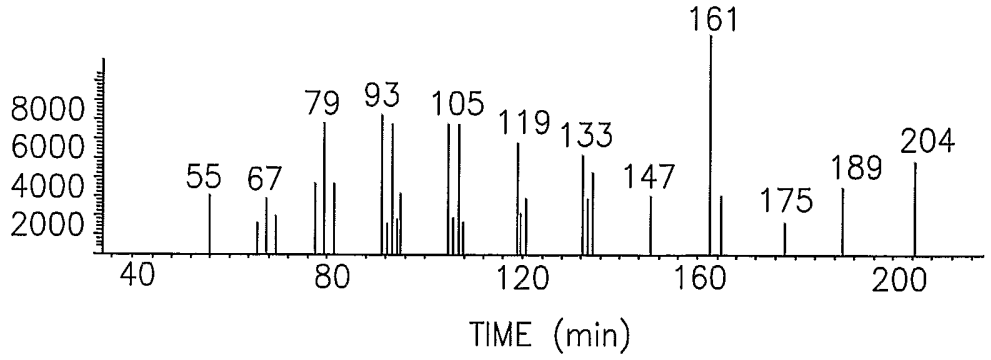

FIG.7D

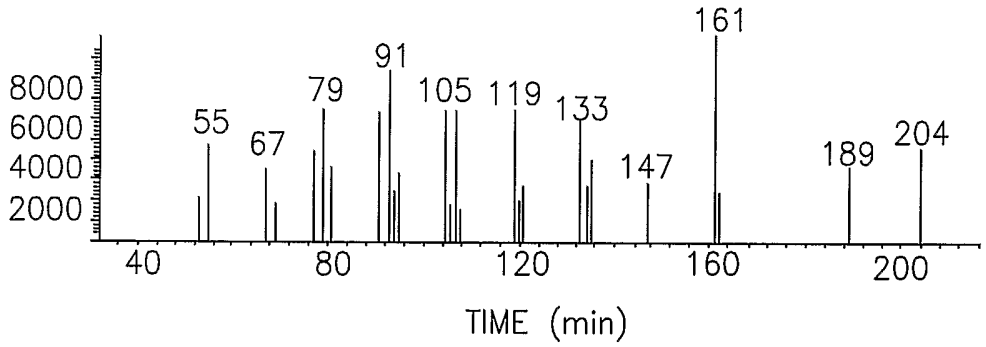

FIG.7E

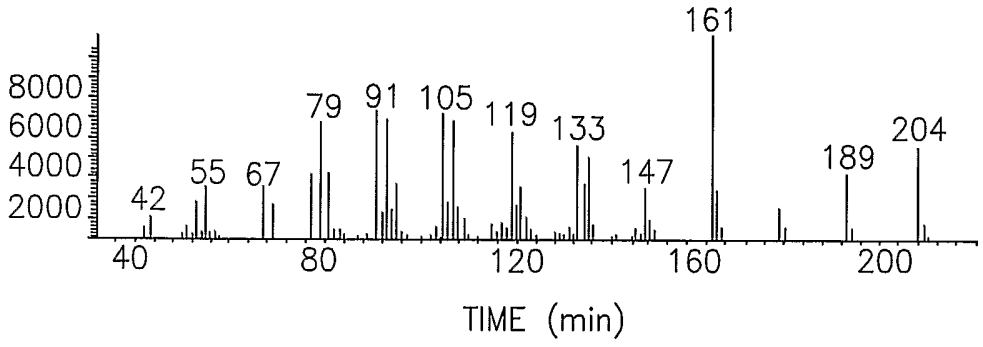

FIG.7F

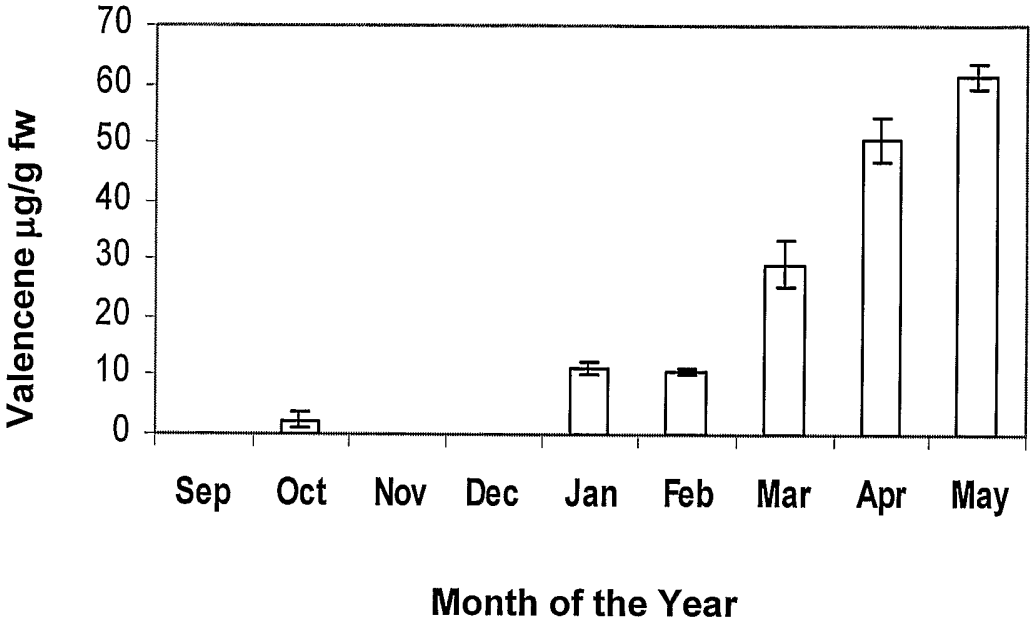
70
60
50
40
30
20
10
0
Valencene µg/g fw
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
Month of the Year

FIGURE 8

29-8-04-kidum 004 seq.ST25
SEQUENCE LISTING

<110> Kidum

<120> Citrus sesquiterpene synthase, methods of production and uses thereof

<130> Kidum/004 PCT

<160> 18

<170> PatentIn version 3.3

<210> 1
<211> 1647
<212> DNA
<213> Citrus sinensis

<400> 1
atgtcgtctg gagaaacatt tcgtcctact gcagatttcc atcctagttt atggagaaac 60

catttcctca aaggtgcttc tgatttcaag acagttgatc atactgcaac tcaagaacga 120

cacgaggcac tgaaagaaga ggtaaggaga atgataacag atgctgaaga taagcctgtt 180

cagaagttac gcttgattga tgaagtacaa cgcctggggg tggcttatca ctttgagaaa 240

gaaataggag atgcaataca aaaattatgt ccaatctata ttgacagtaa tagagctgat 300

ctccacaccg tttcccttca ttttcggttg cttaggcagc aaggaatcaa gatttcatgt 360

gatgtgtttg agaagttcaa agatgatgag ggtagattca agtcatcgtt gataaacgat 420

gttcaaggga tgttaagttt gtacgaggca gcatacatgg cagttcgcgg agaacatata 480

ttagatgaag ccattgcttt cactaccact cacctgaagt cattggtagc tcaggatcat 540

gtaaccccta agcttgcgga acagataaat catgctttat accgtcctct tcgtaaaacc 600

ctaccaagat tagaggcgag gtattttatg tccatgatca attcaacaag tgatcattta 660

tgcaataaaa ctctgctgaa ttttgcaaag ttagatttta acatattgct agagctgcac 720

aaggaggaac tcaatgaatt aacaaagtgg tggaaagatt tagacttcac tacaaaacta 780

ccttatgcaa gagacagatt agtggagtta tatttttggg atttagggac atacttcgag 840

cctcaatatg catttgggag aaagataatg acccaattaa attacatatt atccatcata 900

gatgatactt atgatgcgta tggtacactt gaagaactca gcctctttac tgaagcagtt 960

caaagatgga atattgaggc cgtagatatg cttccagaat acatgaaatt gatttacagg 1020

acactcttag atgcttttaa tgaaattgag gaagatatgg ccaagcaagg aagatcacac 1080

tgcgtacgtt atgcaaaaga ggagaatcaa aaagtaattg gagcatactc tgttcaagcc 1140

aaatggttca gtgaaggtta cgttccaaca attgaggagt atatgcctat tgcactaaca 1200

agttgtgctt acacattcgt cataacaaat tccttccttg gcatgggtga ttttgcaact 1260

aaagaggttt ttgaatggat ctccaataac cctaaggttg taaaagcagc atcagttatc 1320

tgcagactca tggatgacat gcaaggtcat gagtttgagc agaagagagg acatgttgcg 1380

tcagctattg aatgttacac gaagcagcat ggtgtctcta aggaagaggc aattaaaatg 1440

tttgaagaag aagttgcaaa tgcatggaaa gatattaacg aggagttgat gatgaagcca 1500

29-8-04-kidum 004 seq.ST25

accgtcgttg cccgaccact gctcgggacg attcttaatc ttgctcgtgc aattgatttt 1560

atttacaaag aggacgacgg ctatacgcat tcttacctaa ttaaagatca aattgcttct 1620

gtgctaggag accacgttcc attttga 1647

<210> 2
<211> 548
<212> PRT
<213> Citrus sinensis

<400> 2

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

Leu Trp Arg Asn His Phe Leu Lys Gly Ala Ser Asp Phe Lys Thr Val
20 25 30

Asp His Thr Ala Thr Gln Glu Arg His Glu Ala Leu Lys Glu Glu Val
35 40 45

Arg Arg Met Ile Thr Asp Ala Glu Asp Lys Pro Val Gln Lys Leu Arg
50 55 60

Leu Ile Asp Glu Val Gln Arg Leu Gly Val Ala Tyr His Phe Glu Lys
65 70 75 80

Glu Ile Gly Asp Ala Ile Gln Lys Leu Cys Pro Ile Tyr Ile Asp Ser
85 90 95

Asn Arg Ala Asp Leu His Thr Val Ser Leu His Phe Arg Leu Leu Arg
100 105 110

Gln Gln Gly Ile Lys Ile Ser Cys Asp Val Phe Glu Lys Phe Lys Asp
115 120 125

Asp Glu Gly Arg Phe Lys Ser Ser Leu Ile Asn Asp Val Gln Gly Met
130 135 140

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

Leu Asp Glu Ala Ile Ala Phe Thr Thr Thr His Leu Lys Ser Leu Val
165 170 175

Ala Gln Asp His Val Thr Pro Lys Leu Ala Glu Gln Ile Asn His Ala
180 185 190

Leu Tyr Arg Pro Leu Arg Lys Thr Leu Pro Arg Leu Glu Ala Arg Tyr
195 200 205

Phe Met Ser Met Ile Asn Ser Thr Ser Asp His Leu Cys Asn Lys Thr

210 215 220

Leu Leu Asn Phe Ala Lys Leu Asp Phe Asn Ile Leu Leu Glu Leu His
225 230 235 240

Lys Glu Glu Leu Asn Glu Leu Thr Lys Trp Trp Lys Asp Leu Asp Phe
245 250 255

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

Trp Asp Leu Gly Thr Tyr Phe Glu Pro Gln Tyr Ala Phe Gly Arg Lys
275 280 285

Ile Met Thr Gln Leu Asn Tyr Ile Leu Ser Ile Ile Asp Asp Thr Tyr
290 295 300

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

Gln Arg Trp Asn Ile Glu Ala Val Asp Met Leu Pro Glu Tyr Met Lys
325 330 335

Leu Ile Tyr Arg Thr Leu Leu Asp Ala Phe Asn Glu Ile Glu Glu Asp
340 345 350

Met Ala Lys Gln Gly Arg Ser His Cys Val Arg Tyr Ala Lys Glu Glu
355 360 365

Asn Gln Lys Val Ile Gly Ala Tyr Ser Val Gln Ala Lys Trp Phe Ser
370 375 380

Glu Gly Tyr Val Pro Thr Ile Glu Glu Tyr Met Pro Ile Ala Leu Thr
385 390 395 400

Ser Cys Ala Tyr Thr Phe Val Ile Thr Asn Ser Phe Leu Gly Met Gly
405 410 415

Asp Phe Ala Thr Lys Glu Val Phe Glu Trp Ile Ser Asn Asn Pro Lys
420 425 430

Val Val Lys Ala Ala Ser Val Ile Cys Arg Leu Met Asp Asp Met Gln
435 440 445

Gly His Glu Phe Glu Gln Lys Arg Gly His Val Ala Ser Ala Ile Glu
450 455 460

Cys Tyr Thr Lys Gln His Gly Val Ser Lys Glu Glu Ala Ile Lys Met
465 470 475 480

Phe Glu Glu Glu Val Ala Asn Ala Trp Lys Asp Ile Asn Glu Glu Leu

485 490 495

Met Met Lys Pro Thr Val Val Ala Arg Pro Leu Leu Gly Thr Ile Leu
500 505 510

Asn Leu Ala Arg Ala Ile Asp Phe Ile Tyr Lys Glu Asp Asp Gly Tyr
515 520 525

Thr His Ser Tyr Leu Ile Lys Asp Gln Ile Ala Ser Val Leu Gly Asp
530 535 540

His Val Pro Phe
545

<210> 3
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<223> valencene synthase consensus motif

<220>
<221> MISC_FEATURE
<222> (3)..(4)
<223> any amino acid

<400> 3

Asp Asp Xaa Xaa Asp
1 5

<210> 4
<211> 441
<212> PRT
<213> Nicotiana tabacum

<400> 4

Ser Ala Leu Gln Phe Arg Leu Leu Arg Gln His Gly Phe Asn Ile Ser
1 5 10 15

Pro Glu Ile Phe Ser Lys Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu
20 25 30

Ser Leu Ala Ser Asp Val Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser
35 40 45

His Val Arg Thr His Ala Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe
50 55 60

Ser Thr Ile His Leu Glu Ser Ala Val Pro His Leu Lys Ser Pro Leu
65 70 75 80

Arg Glu Gln Val Thr His Ala Leu Glu Gln Cys Leu His Lys Gly Val
85 90 95

Pro Arg Val Glu Thr Arg Phe Phe Ile Ser Ser Ile Tyr Glu Lys Glu
100 105 110

Gln Ser Lys Asn Asp Val Leu Leu Arg Phe Ala Lys Leu Asp Tyr Asn
115 120 125

Leu Leu Gln Met Leu His Lys Gln Glu Leu Ala Glu Val Ser Arg Trp
130 135 140

Trp Lys Asp Leu Asp Phe Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg
145 150 155 160

Val Val Glu Cys Tyr Phe Trp Ala Leu Gly Val Tyr Phe Glu Pro Gln
165 170 175

Tyr Ser Gln Ala Arg Val Met Leu Val Lys Thr Ile Ser Met Ile Ser
180 185 190

Ile Val Asp Asp Thr Phe Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu
195 200 205

Ala Tyr Thr Asp Ala Ile Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg
210 215 220

Leu Pro Asp Tyr Met Lys Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr
225 230 235 240

Lys Asp Tyr Glu Lys Asp Leu Ser Val Pro Glu Ser His Ile Val Cys
245 250 255

His Ala Ile Glu Arg Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu
260 265 270

Ser Thr Trp Phe Ile Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu
275 280 285

Ser Asn Ala Leu Ala Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser
290 295 300

Tyr Leu Gly Met Lys Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser
305 310 315 320

Lys Asn Pro Lys Ile Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile
325 330 335

Asp Asp Thr Ala Thr Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala
340 345 350

Thr Gly Ile Glu Cys Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu
355 360 365

Ala Met Ala Lys Phe Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile
370 375 380

Asn Glu Gly Leu Leu Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr
385 390 395 400

Pro Ile Leu Asn Leu Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn
405 410 415

Leu Asp Gly Tyr Thr His Pro Glu Lys Val Leu Lys Pro His Ile Ile
420 425 430

Ala Leu Leu Val Asp Ser Ile Asp Ile
435 440

<210> 5
<211> 555
<212> PRT
<213> Gossypium arboreum

<400> 5

Met Ala Ser Gln Ala Ser Gln Val Leu Ala Ser Pro His Pro Ala Ile
1 5 10 15

Ser Ser Glu Asn Arg Pro Lys Ala Asp Phe His Pro Gly Ile Trp Gly
20 25 30

Asp Met Phe Ile Ile Cys Pro Asp Thr Asp Ile Asp Ala Ala Thr Glu
35 40 45

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

Pro Val Asp Asp Ser Asn Gln Lys Leu Pro Phe Ile Asp Ala Val Gln
65 70 75 80

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

Glu Asn Ile Tyr Arg Asp Thr Asn Asn Asn Asp Ala Asp Thr Asp Leu
100 105 110

Tyr Thr Thr Ala Leu Arg Phe Arg Leu Leu Arg Glu His Gly Phe Asp
115 120 125

Ile Ser Cys Asp Ala Phe Asn Lys Phe Lys Asp Glu Ala Gly Asn Phe
130 135 140

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

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

Ser Phe Thr Thr Ala Gln Leu Thr Leu Ala Leu Pro Thr Leu His His
180 185 190

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

Gly Leu Pro Arg Val Glu Ala Arg Asn Phe Ile Ser Ile Tyr Gln Asp
210 215 220

Leu Glu Ser His Asn Lys Ser Leu Leu Gln Phe Ala Lys Ile Asp Phe
225 230 235 240

Asn Leu Leu Gln Leu Leu His Arg Lys Glu Leu Ser Glu Ile Cys Arg
245 250 255

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

Arg Val Val Glu Gly Tyr Phe Trp Ile Met Gly Val Tyr Phe Glu Pro
275 280 285

Gln Tyr Ser Leu Gly Arg Lys Met Leu Thr Lys Val Ile Ala Met Ala
290 295 300

Ser Ile Val Asp Asp Thr Tyr Asp Ser Tyr Ala Thr Tyr Asp Glu Leu
305 310 315 320

Ile Pro Tyr Thr Asn Ala Ile Glu Arg Trp Asp Ile Lys Cys Met Asn
325 330 335

Gln Leu Pro Asn Tyr Met Lys Ile Ser Tyr Lys Ala Leu Leu Asp Val
340 345 350

Tyr Glu Glu Met Glu Gln Leu Leu Ala Asn Gln Gly Arg Gln Tyr Arg
355 360 365

Val Glu Tyr Ala Lys Lys Ala Met Ile Arg Leu Val Gln Ala Tyr Leu
370 375 380

Leu Glu Ala Lys Trp Thr His Gln Asn Tyr Lys Pro Thr Phe Glu Glu
385 390 395 400

Phe Arg Asp Asn Ala Leu Pro Thr Ser Gly Tyr Ala Met Leu Ala Ile
405 410 415

Thr Ala Phe Val Gly Met Gly Glu Val Ile Thr Pro Glu Thr Phe Lys
420 425 430

Trp Ala Ala Ser Asp Pro Lys Ile Ile Lys Ala Ser Thr Ile Ile Cys
435 440 445

Arg Phe Met Asp Asp Ile Ala Glu His Lys Phe Asn His Arg Arg Glu
450 455 460

Asp Asp Cys Ser Ala Ile Glu Cys Tyr Met Lys Gln Tyr Gly Val Thr
465 470 475 480

Ala Gln Glu Ala Tyr Asn Glu Phe Asn Lys His Ile Glu Ser Ser Trp
485 490 495

Lys Asp Val Asn Glu Glu Phe Leu Lys Pro Thr Glu Met Pro Thr Pro
500 505 510

Val Leu Cys Arg Ser Leu Asn Leu Ala Arg Val Met Asp Val Leu Tyr
515 520 525

Arg Glu Gly Asp Gly Tyr Thr His Val Gly Lys Ala Ala Lys Gly Gly
530 535 540

Ile Thr Ser Leu Leu Ile Asp Pro Ile Gln Ile
545 550 555

<210> 6
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> PCR primer

<400> 6
cccaggcgtt gtacttcatc a 21

<210> 7
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 7
cgacacgagg cactgaaaga 20

<210> 8
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> PCR primer

<400> 8
gcgacgcatc attcaaattt c 21

<210> 9
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 9
tccggaatcg aaccctaatt c 21

<210> 10
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 10
cggctagcat gctttttttt ttttttt 27

<210> 11
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<220>
<221> misc_feature
<222> (3)..(3)
<223> y = t/u or c (pyrimidine)

<220>
<221> misc_feature
<222> (6)..(6)
<223> y = t/u or c (pyrimidine)

<220>
<221> modified_base
<222> (7)..(9)
<223> I

<220>
<221> misc_feature
<222> (7)..(9)
<223> inosine

<220>
<221> misc_feature
<222> (11)..(11)
<223> w = a or t/u

<220>
<221> misc_feature
<222> (12)..(12)
<223> y = t/u or c (Pyrimidine)

<220>
<221> misc_feature
<222> (15)..(15)
<223> y = t/u or c (Pyrimidine)

<220>
<221> misc_feature
<222> (17)..(17)
<223> y = t/u or c (Pyrimidine)

<220>
<221> modified_base
<222> (18)..(18)
<223> I

<220>
<221> misc_feature
<222> (18)..(18)
<223> inosine

<220>
<221> misc_feature
<222> (20)..(20)
<223> w = a or t/u

<220>
<221> misc_feature
<222> (21)..(21)
<223> y = t/u or c (pyrimidine)

<400> 11
gaygaynnnt wygaygyntw ygg 23

<210> 12
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<220>
<221> modified_base
<222> (10)..(10)
<223> I

<220>
<221> misc_feature
<222> (10)..(10)
<223> inosine

<220>
<221> misc_feature
<222> (13)..(13)
<223> a or g or c or t/u

<400> 12
ytkcatrtan tcngg 15

<210> 13
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 13
cagtaaagag gctgagttct tc 22

<210> 14
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 14
gaagaactca gcctctttac tg 22

<210> 15
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 15
gagagtccat atgtcgtctg gagaaacatt tcg 33

<210> 16
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 16
aaatggaacg tggtctccta gc 22

<210> 17
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 17
cgatgctagc tcgtctggag aaacatttc 29

<210> 18
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 18
cgtagtcgac tcaaaatgga acgtggtctc ctag 34