Cytochrome P450S and Uses Thereof

Joseph Chappell
University of Kentucky, chappell@uky.edu

Lyle F. Ralston

Follow this and additional works at: https://uknowledge.uky.edu/ps_patents

Part of the Pharmacy and Pharmaceutical Sciences Commons

Recommended Citation
https://uknowledge.uky.edu/ps_patents/179

This Patent is brought to you for free and open access by the Pharmaceutical Sciences at UKnowledge. It has been accepted for inclusion in Pharmaceutical Sciences Faculty Patents by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
The invention relates to isolated cytochrome P450 polypeptides and nucleic acid molecules, as well as expression vectors and transgenic plants containing these molecules. In addition, the invention relates to uses of such molecules in methods of increasing the level of resistance against a disease caused by a plant pathogen in a transgenic plant, in methods for producing altered compounds, for example, hydroxylated compounds, and in methods of producing isoprenoid compounds.

15 Claims, 11 Drawing Sheets
(56) References Cited

U.S. PATENT DOCUMENTS

8,642,815 B2 2/2014 Julien
8,722,363 B2 5/2014 Chappell
8,741,651 B2 6/2014 Chappell
8,753,842 B2 6/2014 Julien
8,835,131 B2 9/2014 Chappell
2010/0120110 A1 5/2010 Chappell
2010/0129306 A1 5/2010 Julien
2012/0129235 A1 5/2012 Julien
2012/0196340 A1 8/2012 Chappell
2013/0071877 A1 3/2013 Chappell
2013/0122560 A1 5/2013 Julien
2013/0236943 A1 9/2013 Julien
2013/0330793 A1 12/2013 Chappell
2014/0212939 A1 7/2014 Chappell
2014/0242660 A1 8/2014 Chappell

OTHER PUBLICATIONS

Danielle et al. Genome Biology (2000); online review.*

References Cited

OTHER PUBLICATIONS


GenBank Accession No. A25676.

GenBank Accession No. AAC3505.

GenBank Accession No. AAD44150.

GenBank Accession No. AAD44151.

GenBank Accession No. AB015762.

GenBank Accession No. A007575.

GenBank Accession No. CAC24711.


Porter et al., “Strategies to enhance the coexpression of cytochrome P450 2E1 and reductase in bacteria,” Drug Metab. Rev. 31(1):159-74 (Feb. 1999).


Takahashi et al., “Functional Characterization of Premnaasiophiociode Oxyxagenase, a Cytochrome P450 Catalyzing Regio- and Stereo-
(56) References Cited

OTHER PUBLICATIONS


Molot et al., “Relations between capsidiol concentration, speed of fungal invasion and level of induced resistance in cultivars of pepper (Capsicum annuum) susceptible or resistant to Phytophthora capsici,” Physiol. Plant Pathol. 18 (3):379-89 (May 1981).


* cited by examiner
FIG. 1

1-deoxycapsidiol (3-hydroxy-5-epi-aristolochene)
3-deoxycapsidiol (1-hydroxy-5-epi-aristolochene)
capsidiol

5-epi-aristolochene

farnesyl diphosphate

EAS

+ NADPH

+ O₂
FIG. 2

Enzyme activity (% of maximum) vs. Time after elicitation (h)

- Graph showing enzyme activity over time.
FIG. 4A

- N-term
- C-term

FIG. 4B

<table>
<thead>
<tr>
<th>Primer/Sequence</th>
<th>Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>KETLRLH-for</td>
<td>5'-AARGARACIYTIMGTYTICA-3'</td>
</tr>
<tr>
<td>KETLRLY-for</td>
<td>5'-AARGARACIYTIMGTYTITA-3'</td>
</tr>
<tr>
<td>KETLRLR-for</td>
<td>5'-AARGARACIYTIMGTYTIMG-3'</td>
</tr>
<tr>
<td>FXPERF-for</td>
<td>5'-TTYIICCGARMGITTY-3'</td>
</tr>
<tr>
<td>FXPERF-rev</td>
<td>5'-RAAICKYTCIGIIIRAA-3'</td>
</tr>
<tr>
<td>GRRXCP(A/G)-for</td>
<td>5'-GGIMGIMGIIITGYCCIGS-3'</td>
</tr>
<tr>
<td>PFGXGRR-rev</td>
<td>5'-Crickicciicccraaig-3'</td>
</tr>
<tr>
<td>T7</td>
<td>5'-GTAATACGACTCACTATAGGG-3'</td>
</tr>
<tr>
<td>T3</td>
<td>5'-CTTAACCCTACTAAGGG-3'</td>
</tr>
</tbody>
</table>

FIG. 4C

[Image of gel electrophoresis with bands at 600, 500, 400, and 300 base pairs]
FIG. 5

<table>
<thead>
<tr>
<th>Time (hours)</th>
<th>Control</th>
<th>Elicitor-treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- CYP71D
- CYP73A
- CYP82E
- CYP92A
- EAS

Loading control
Figure 8B
Figure 8C
Figure 8D
### CYTOCHROME P450S AND USES THEREOF

**RELATED APPLICATIONS**

This application is a continuation of co-pending U.S. patent application Ser. No. 14/243,778, filed Apr. 2, 2014, which is a continuation of U.S. patent application Ser. No. 13/986,446, filed May 3, 2013 now issued U.S. Pat. No. 8,722,363, which is a continuation of U.S. patent application Ser. No. 13/199,349, filed Aug. 26, 2011 (now issued U.S. Pat. No. 8,445,231), which is a continuation of U.S. patent application Ser. No. 12/182,000, filed Jul. 29, 2008 (now issued U.S. Pat. No. 8,263,362), which is a continuation of U.S. patent application Ser. No. 10/097,559, filed Mar. 8, 2002 (now issued U.S. Pat. No. 7,405,057), which claims the benefit of U.S. Provisional Application Nos. 60/274,421 and 60/275,597, filed on Mar. 9, 2001 and Mar. 13, 2001, respectively, all of which are hereby incorporated by reference.

**INCORPORATION BY REFERENCE OF SEQUENCE LISTING PROVIDED ELECTRONICALLY**

An electronic version of the Sequence Listing is filed herewith, the contents of which are incorporated by reference in their entirety. The electronic file was created on Apr. 1, 2014, is 65 kilobytes in size, and titled 2017SEQ001.txt.

**FIELD OF THE INVENTION**

This invention relates to cytochrome P450s and uses thereof.

**BACKGROUND OF THE INVENTION**

Cytochrome P450s encompass a superfamily of oxidases responsible for the oxidation of numerous endobiotics and thousands of xenobiotics. In addition, in plants, cytochrome P450s play important roles in wound healing, pest resistance, signaling, and anti-microbial and anti-fungal activity.

Capsidol is a bicyclic, dihydroxylated sesquiterpene produced by many *Solanaceous* species in response to a variety of environmental stimuli, including exposure to UV (Bock et al., Plant Cell Physiol. 389:899-904, 1998) and infection by microorganisms (Molot et al., Physiol. Plant Pathol. 379:389, 1981; Stolle et al., Phytopathology 78:1193-1197, 1988; Keller et al., Planta. 205:467-476, 1998). It is the primary antibiotic or phytoalexin produced in tobacco in response to fungal elicitation, and it is derived from the isoprenoid pathway via its hydrocarbon precursor, 5-epi-aristolochene (FIG. 1). Several of the biosynthetic enzymes leading up to 5-epi-aristolochene formation have been studied (Chappell, Annu. Rev. Plant Physiol. Plant Mol. Biol. 46:521-547, 1995), especially 5-epi-aristolochene synthase (BAS) (Vogeli and Chappell, Plant Physiol. 98:1291-1296, 1998; Bock and Chappell, Proc. Natl. Acad. Sci. U.S.A. 93:6841-6845, 1996; Mathis et al., Biochemistry 36:8340-8348, 1997; Starks et al., Science 277:1815-1820, 1997). BAS commits carbon to sesquiterpene metabolism by catalyzing the cyclization of farnesyl diphasphate (FPP) to 5-epi-aristolochene. However, until the present invention, the enzyme(s) responsible for the conversion of 5-epi-aristolochene to capsidol has yet to be fully identified and characterized.

Biochemical evidence from previous studies in tobacco (Whitehead et al., Phytochemistry 28:775-779, 1989) and green pepper (Hoshino et al., Phytochemistry 38:609-613, 1995) have suggested that the oxidation of 5-epi-aristolochene to capsidol occurs in a two step process with one of the hydroxylation steps being constitutive and the other being mediated by an elicitor-inducible cytochrome P450 (FIG. 1). Because 1-deoxycapsidol had been isolated from natural sources (Watson et al., Biochem. Soc. Trans. 11:589, 1983), Whitehead et al. (Physiochemistry 28:775-779, 1989), surmised that perhaps the biosynthesis of this intermediate was due to pathogen induction of a corresponding hydroxylase. They therefore prepared synthetic 1-deoxycapsidol and reported a modest conversion of this compound to capsidol when fed to control or unelicited tobacco cell cultures. This was further supported by their observation that radiolabeled 5-epi-aristolochene was only converted to capsidol when fed to elicitor-induced cell cultures but not control cultures. Whitehead et al. (Physiochemistry 28:775-779, 1989) therefore concluded that the 3-hydroxylase, responsible for hydroxylation of 5-epi-aristolochene at C3 to generate 1-deoxycapsidol, was pathogen/elicitor inducible, while the 1-hydroxylase, responsible for hydroxylation of 1-deoxycapsidol at the C1 to generate capsidol, was constitutive. Hoshino et al. (Physiochemistry 38:609-613, 1995) added to the observations of Whitehead et al. (Physiochemistry 28:775-779, 1989) by directly measuring 3-hydroxylase activity in microsomal preparations of archidonic acid-elicited *Capsicum annum* fruits and seedlings. These assays consisted of incubating 5-epi-aristolochene with microsome preparations and subsequently determining the amount of 1-deoxycapsidol generated by a combination of thin-layer chromatography (TLC) separations and gas chromatography (GC). Their evidence demonstrated that the conversion of 5-epi-aristolochene to 1-deoxycapsidol was dependent on both NADPH and O2, and that 1-deoxycapsidol accumulation in vitro was arrested by the P450 antagonists carbon monoxide (Omura and Sato, J. Biol. Chem. 239:2370-2378, 1964), avenocim (Coolbaugh et al., Plant Physiol. 62:571-576, 1978), and ketoconazole (Rademacher, Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:501-531, 2000).

Recent results suggest that the hydroxylation of 5-epi-aristolochene is an important regulated step in capsidol biosynthesis. In studies to evaluate the effectiveness of methyl-jasmonate as an inducer of capsidol biosynthesis in tobacco cell cultures, Mandujano-Chavez et al. (Arch. Biochem. Biophys. 381:285-294, 2000), reported that the modest accumulation of this phytoalexin was accompanied by a strong induction of EAS. This result implied that steps before or after the sesquiterpene cyclase reaction were limiting. Using an in vivo assay measuring the conversion rate of radiolabeled 5-epi-aristolochene to capsidol, a very limited induction of the hydroxylase activity was observed in cells treated with methyl jasmonate relative to that in fungal elicitor-treated cells. This result pointed to the hydroxylase reactions as a potentially limiting step in capsidol biosynthesis.

**SUMMARY OF THE INVENTION**

In one aspect, the invention features several isolated cytochrome P450 polypeptides (such as CYP71D20, CYP71D21, CYP73A27, CYP73A28, and CYP92A5, and P450s having substantial identity to these polypeptides), as well as isolated nucleic acid molecules that encode these P450s.

In related aspects, the invention features a vector (such as an expression vector) including an isolated nucleic acid
molecule of the invention, wherein the nucleic acid molecule is expressed in the transgenic plant or the transgenic plant component. Preferably, the transgenic plant or transgenic plant component is an angiosperm (for example, a monocot or dicot). In preferred embodiments, the transgenic plant or transgenic plant component is a *solanaceous* plant, such as *tomato*, *potato*, or *maize*. The invention further includes a seed produced by the transgenic plant or transgenic plant component, or progeny thereof.

In another aspect, the invention features a method of providing an increased level of resistance against a disease caused by a plant pathogen in a transgenic plant. The method involves: (a) producing a transgenic plant cell including the nucleic acid molecule of the invention integrated into the genome of the transgenic plant cell and positioned for expression in the plant cell; and (b) growing a transgenic plant from the plant cell wherein the nucleic acid molecule is expressed in the transgenic plant and the transgenic plant is thereby provided with an increased level of resistance against a disease caused by a plant pathogen.

In another aspect, the invention features a method for producing an altered compound, the method including the steps of contacting the compound with one or more of the isolated polypeptides disclosed herein under conditions allowing for the hydroxylation, oxidation, demethylation, or methylation of the compound and recovering the altered compound.

In still another aspect, the invention features a hydroxylating agent including any of the isolated polypeptides disclosed herein.

In yet another embodiment, the invention features an isolated nucleic acid molecule that specifically hybridizes under highly stringent conditions to the complement of any one of the sequences described in SEQ ID NO:2 (CYP71D20), SEQ ID NO:4 (CYP71D21), SEQ ID NO:6 (CYP73A27), SEQ ID NO:8 (CYP73A28), or SEQ ID NO:12 (CYP92A5), wherein such a nucleic acid molecule encodes a cytochrome P450 505 polypeptide.

In another aspect, the invention features a host cell expressing a recombinant isoprenoid synthase and a recombinant cytochrome P450. In preferred embodiments, the host cell further expresses, independently or in combination, a recombinant acetyltransferase, a recombinant methyltransferase, or a recombinant fatty acyltransferase. In other preferred embodiments, the host cell expresses an endogenous or recombinant cytochrome reductase. Preferably, the host cell is a yeast cell, a bacterial cell, an insect cell, or a plant cell.

In a related aspect, the invention features a method for producing an isoprenoid compound, the method including the steps of: (a) culturing a cell that expresses a recombinant isoprenoid synthase and a recombinant cytochrome P450 under conditions wherein the isoprenoid synthase and the cytochrome P450 are expressed and catalyze the formation of an isoprenoid compound not normally produced by the cell; and (b) recovering the isoprenoid compound. In preferred embodiments, the host cell further expresses a recombinant acetyltransferase, a recombinant methyltransferase, or a recombinant fatty acyltransferase. In other preferred embodiments, the host cell expresses an endogenous or recombinant cytochrome reductase. Preferably, the host cell is a yeast cell, a bacterial cell, an insect cell, or a plant cell.

In yet another aspect, the invention features a recombinant nucleic acid compound produced according to the above-mentioned methods.

By “P450 polypeptide,” “cytochrome P450,” or “P450” is meant a polypeptide that contains a heme-binding domain and shows a CO absorption spectra peak at 450 nm according to standard methods, for example, those described herein. Such P450s may also include, without limitation, hydroxylase activity, dual hydroxylase activity, demethylase activity, or oxidase activity. Such enzymatic activities are determined using methods well known in the art.

By “polypeptide” is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By “substantially identical” is meant a polypeptide or nucleic acid exhibiting at least 80 or 85%, preferably 90%, more preferably 95%, and most preferably 97%, or even 98% identity to a reference amino acid sequence (for example, the amino acid sequence shown in SEQ ID NO: 1, 3, 5, 7 and 11) or nucleic acid sequence (for example, the nucleic acid sequences shown in SEQ ID NO:S:2, 4, 6, 8 and 12, respectively). For polypeptides, the length of comparison sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids. For nucleic acids, the length of comparison sequences will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 110 nucleotides.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine, valine, isoleucine, leucine, aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

By an “isolated polypeptide” is meant a P450 polypeptide (for example, a CYP71D20 (SEQ ID NO:1), CYP71D21 (SEQ ID NO:3), CYP73A27 (SEQ ID NO:5), or CYP92A5 (SEQ ID NO:11) polypeptide) that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a P450 polypeptide. An isolated P450 polypeptide may be obtained, for example, by extraction from a natural source (for example, a plant cell); by expression of a recombinant nucleic acid encoding a P450 polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By “derived from” or “obtained from” is meant isolated from or having the sequence of a naturally-occurring sequence (e.g., cDNA, genomic DNA, synthetic, or combination thereof).

By “isolated nucleic acid molecule” is meant a nucleic acid molecule, e.g., a DNA molecule, that is free of the nucleic acid sequence(s) which, in the naturally-occurring
genome of the organism from which the nucleic acid molecule of the invention is derived, flank the nucleic acid molecule. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. The term “isolated nucleic acid molecule” also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequences.

By “specifically hybridizes” is meant that a nucleic acid sequence is capable of hybridizing to a DNA sequence at least under low stringency conditions, and preferably under high stringency conditions. For example, high stringency conditions may include hybridization at approximately 42° C. in about 50% formamide, 0.1 mg/mL sheared salmon sperm DNA, 1% SDS, 2xSSC, 10% Dextran sulfate, a first wash at approximately 65° C. in about 2xSSC, 1% SDS, followed by a second wash at approximately 65° C. in about 0.1xSSC. Alternatively high stringency conditions may include hybridization at approximately 42° C. in about 50% formamide, 0.1 mg/mL sheared salmon sperm DNA, 0.5% SDS, 5xSSPE, 1xDenhardt’s, followed by two washes at room temperature in 2xSSC, 0.1% SDS, and two washes at between 55-60° C. in 0.2xSSC, 0.1% SDS. Reducing the stringency of the hybridization conditions may involve lowering the wash temperature and/or washing at a higher concentration of salt. For example, low stringency conditions may include washing in 2xSSC, 0.1% SDS at 40° C.

By “transformed cell” is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) a P450 polypeptide.

By “positioned for expression” is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (i.e., facilitates the production of, for example, a P450 polypeptide, a recombinant protein, or an RNA molecule).

By “reporter gene” is meant a gene whose expression may be assayed; such genes include, without limitation, beta-glucuronidase (GUS), luciferase, chloramphenicol transacetylase (CAT), green fluorescent protein (GFP), beta-galactosidase, herbicide resistant genes, and antibiotic resistance genes.

By “expression control region” is meant any minimal sequence sufficient to direct transcription. Included in the invention are promoter elements that are sufficient to render promoter-dependent gene expression controllable for cell-, tissue-, or organ-specific gene expression, or elements that are inducible by external signals or agents (for example, light-, pathogen-, wound-, stress-, or hormone-inducible elements or chemical inducers such as salicylic acid (SA) or 2,2-dichloroisonicotinic acid (DIA)); such elements may be located in the 5’ or 3’ regions of the native gene or engineered into a transgene construct.

By “operably linked” is meant that a gene and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (for example, transcriptional activator proteins) are bound to the regulatory sequence(s).

By “plant cell” is meant any self-propagating cell bounded by a semi-permeable membrane and typically is one containing a plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

By “plant component” is meant a part, segment, or organ obtained from an intact plant or plant cell. Exemplary plant components include, without limitation, somatic embryos, leaves, stems, roots, flowers, tendrils, fruits, scions, and rootstocks.

By “transgene” is meant any piece of DNA which is inserted by artifice into a cell and typically becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By “transgenic” is meant any cell which includes a DNA sequence which is inserted by artifice into a cell and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic plants and the DNA (transgene) is inserted by artifice into the nuclear or plastidic genome. A transgenic plant according to the invention may contain one or more engineered traits.

By “pathogen” is meant an organism whose infection of viable plant tissue elicits a disease response in the plant tissue. Such pathogens include, without limitation, bacteria, mycoplasmas, fungi, insects, nematodes, viruses, and viroids. Plant diseases caused by these pathogens are described in Chapters 11-16 of Agrios, Plant Pathology, 3rd ed., Academic Press, Inc., New York, 1988.

By “increased level of resistance” is meant a greater level of resistance to a disease-causing pathogen in a transgenic plant (or cell or seed thereof) of the invention than the level of resistance relative to a control plant (for example, a non-transgenic plant). In preferred embodiments, the level of resistance in a transgenic plant of the invention is at least 20% (and preferably 30% or 40%) greater than the resistance of a control plant. In other preferred embodiments, the level of resistance to a disease-causing pathogen is greater, 60% greater, and more preferably even 75% or 90% greater than a control plant; with up to 100% above the level of resistance as compared to a control plant being most preferred. The level of resistance is measured using conventional methods. For example, the level of resistance to a pathogen may be determined by comparing physical features and characteristics (for example, plant height and weight, or by comparing disease symptoms, for example, delayed lesion development, reduced lesion size, leaf wilting and curling, water-soaked spots, and discoloration of cells) of transgenic plants.

By “purified antibody” is meant antibody which is at least 60%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 95%, by weight, antibody, for example, an acquired resistance polypeptide-specific antibody. A purified P450 antibody may be obtained, for example, by affinity chromatography using a recombinantly-produced P450 polypeptide and standard techniques.

By “specifically binds” is meant an antibody which recognizes and binds a P450 protein but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample, which naturally includes a P450 protein such as CYP71D20, CYP71D21, CYP73A27, CYP73A28, or CYP92A5.
Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic diagram of a proposed alternative pathway for the biosynthesis of capsidiol in elicitor-treated *Nicotiana tabacum* cells. 5-epi-aristolochene is synthesized from FPP by the action of a sesquiterpene cyclase, 5-epi-aristolochene synthase (EAS), and is subsequently hydroxylated at C1 and C3 to form capsidiol.

FIG. 2 is a graph showing an induction time course for sesquiterpene cyclase enzyme activity and sesquiterpene hydroxylase activity in cell-culture-treated cell cultures. Sesquiterpene cyclase (5-epi-aristolochene synthase, EAS) enzyme activity was determined in extracts prepared from control (open squares) and elicitor-treated (closed squares) cells collected at the indicated time points. Sesquiterpene hydroxylase activity was determined using an indirect assay for control (open circles) and elicitor-treated (closed circles) cells. Cell cultures were incubated with [3H]-5-epi-aristolochene for 3 hours ending at the indicated time points before quantifying the incorporation of radioactivity into extracellular capsidiol, a dihydroxylated form of aristolochene (Mandujano-Chávez et al., Arch. Biochem. Biophys. 381:285-294, 2000).

FIGS. 3A-3B are a series of graphs showing the dose dependent inhibition of 5-epi-aristolochene hydroxylase activity by ancamidol and ketoconazole. Cell cultures were incubated in the presence of cellulase (0.5 µg/mL) plus the indicated concentrations of ancamidol (A) or ketoconazole (B) for 12 hours prior to measuring the in vivo 5-epi-aristolochene hydroxylase activity in the cell suspension cultures (squares), or the EAS enzyme activity in extracts prepared from the collected cells (triangles). The in vitro activity of a purified EAS preparation (Bock and Chappell, J. Biol. Chem. 270:7375-7381, 1995) was also measured at the indicated inhibitor concentrations as an additional test for non-specific effects of these inhibitors (circles).

FIG. 4A is a schematic diagram of the primary structure of a generalized cytochrome P450 with conserved domains used for the design of PCR primers highlighted (SEQ ID NOS:26-29).

FIG. 4B is a list of the degenerate P450-specific primers (SEQ ID NOS:30-36) that were used in various combinations with vector specific primers in the amplification of cytochrome P450 cDNA fragments.

FIG. 4C is a scanned image of an ethidium bromide-stained agarose gel showing the PCR products amplified from a directional cDNA library prepared with mRNA isolated from elicitor-treated cells using the degenerate primer GRRXCP(A/G)—for (SEQ ID NO:35) and the T7 vector-specific primer (SEQ ID NO:37). The T3 vector-specific primer is also shown (SEQ ID NO:38).

FIG. 5 is a series of Northern blots showing the induction time course for CYP71D1, CYP73A1, CYP82E1, CYP82A1, and EAS transcript accumulation in elicitor-treated cells. Total RNA was extracted from tobacco suspension cells incubated with the cellulase elicitor for the indicated durations, size fractionated by agarose gel electrophoresis under denaturing conditions, and transferred to a nylon membrane before probing with the respective full-length cDNAs. The uniformity of sample loading was verified by ethidium bromide staining of ribosomal RNA (Loading control).

FIGS. 6A-6B are a series of graphs showing carbon monoxide (CO) difference spectra of the microsomal fraction isolated from yeast expressing the CYP92A5 (A) and CYP71D20 (B) cDNAs. Expression of the respective plasmid constructs engineered into the yeast (WAT11) cells was induced by a galactose treatment, followed by isolation of microsomal preparations. The difference adsorption spectra of microsomes incubated in the presence (solid lines) and absence (broken lines) of carbon monoxide was determined.

FIGS. 7A-7D are a series of gas chromatograms of the reaction products formed upon incubation of microsomes isolated from WAT11 yeast cells containing the CYP71D20 expression construct (A and C) or vector control DNA (B and D) with sesquiterpene substrates. Microsomes isolated from the indicated yeast lines were incubated with 5-epi-aristolochene (A and B) or 1-deoxycapsidiol (C and D) in the presence (solid lines) or absence (dashed lines) of NADPH. The identities of 5-epi-aristolochene, 1-deoxycapsidiol, and capsidiol were verified by mass spectrometry.

FIGS. 8A-8D provide a sequence comparison of the amino acid sequence of *Aicottiana tabacum* 5-epi-aristolochene (sesquiterpene) hydroxylase NtCYP71D20 (SEQ ID NO:1) with other plant terpene hydroxylases (SEQ ID NOS:39-43). NtCYP71A5v1 (GenBank accession number AAA70575) catalyzes the mono-hydroxylation of neral and geraniol, linear monoterpenes, while NtCYP71A1 (AAS5867) catalyzes the epoxidation of these substrates (Hallal et al., Biochem. Biophys. Acta. 1201:94-100, 1994). MsCYP71D18 (AAD44150) and MpcYP71D13 (AAD44151) catalyze the mono-hydroxylation at C6 and C3 of limonene, a cyclic monoterpen, respectively (Lupien et al., Arch. Biochem. Biophys. 368:181-192, 1999). AtCYP701A3 (AAC39505) encodes for kaurene oxidase, which catalyzes a 3-step reaction including a hydroxylation followed by oxidation of a diterpene (Helliwell et al., Plant Physiol. 119:507-510, 1999). Shown are sequences from *Mentha piperita* (MpcYP71D13; SEQ ID NO:39), *Mentha spicata* (MsCYP71D18; SEQ ID NO:40), *Nepea racemosa* (NtCYP71A5v1; SEQ ID NO:41), *Nicotiana tabacum* (NtCYP71D20; SEQ ID NO:42), *Peroea americana* (PaCYP71A1; SEQ ID NO:43), and *Arabidopsis thaliana* (CYP701A3; SEQ ID NO:44). Conserved residues are shaded.

DETAILED DESCRIPTION

Capsidiol is a bicyclic, dihydroxylated sesquiterpene produced by several *Solanaceous* species in response to a variety of environmental stimuli. It is the primary antimicrobial compound produced by *Nicotiana tabacum* in response to fungal elicitation, and it is formed via the isoprenoid pathway from 5-epi-aristolochene. Much of the biosynthetic pathway for the formation of this compound has been elucidated, except for the enzyme(s) responsible for the conversion of the allylic sesquiterpene 5-epi-aristolochene to its dihydroxylated form, capsidiol.

Accordingly, an in vivo assay for 5-epi-aristolochene hydroxylase activity was developed and used to demonstrate a dose dependent inhibition of activity by ancamidol and ketoconazole, two well-characterized inhibitors of cytochrome P450 enzymes. Using degenerate oligonucleotide primers designed to the well-conserved domains found within most P450 enzymes, including the heme binding domain, cDNA fragments representing four distinct P450 families (CYP71, CYP73, CYP82, and CYP92) were amplified from a cDNA library prepared against mRNA from elicitor-treated cells using PCR. The PCR fragments were subsequently used to isolate full-length cDNAs (CYP71D20 (SEQ ID NO:2) and D21 (SEQ ID NO:4), CYP73A27 (SEQ ID NO:5) from elicitor-treated cells. Further sequence analysis and expression studies will be needed to determine the complete role of these enzymes in the biosynthesis of capsidiol.
About half of the sequenced DNAs contained signature kingdoms were used to identify conserved regions to which
50 were rapidly and transiently induced with slightly different
tions. 50 were rapidly and transiently induced with slightly different
the formation of 5-epi-aristolochene from FPP (FIG. 5).
In addition, Figs. 8A-8D shows an amino acid alignment of
several terpene cytochrome P450s. Alignments were performed
using the algorithm of the MACVECTOR software suite.

Inhibition of the 5-epi-aristolochene to Capsidiol Conversion

US 10,184,119 B2

9
ID NO:6) and A28 (SEQ ID NO:8), CYP82E1 (SEQ ID
NO:10), and CYP92A5 (SEQ ID NO:12)), and these in turn
were used to demonstrate that the corresponding mRNAs
were all induced in elicitor-treated cells, albeit with different
induction patterns.

EXPERIMENTAL

There now follows a description of the cloning of several
P450s from Nicotiana tabacum. These examples are provided
for the purpose of illustrating the invention, and are not to be
considered as limiting.

Inhibition of the 5-epi-aristolochene to Capsidiol Conversion

Using an indirect assay, a detailed induction time course
of 5EAH activity in elicitor-induced cell cultures was
determined relative to that of EAS activity (FIG. 2), the well-
characterized sesquiterpene cyclase activity that catalyzes
the formation of 5-epi-aristolochene from FPP (FIG. 1).
Using assays for EAS and 5EAH, EAS activity is not
detectable in control cell cultures, but is induced signifi-
cantly within 3 hours and reaches its maximal level within
15 to 18 hours of elicitor-treatment. Similar to the EAS
enzyme activity, 5EAH activity was negligible in control
cell cultures. Nonetheless, after an apparent lag phase of 8
hours, a rapid induction of hydroxylase activity was
observed 10 to 15 hours post elicitor addition to the cell
cultures, reaching a maximum by 18 hours followed by a
rather gradual decline of 10 to 20% over the next 8 hours.

Tobacco cell suspension cultures treated with cellulase
plus varying concentrations of ancyamol or ketoconazole
were pre-incubated for 12 hours before measuring the cells'
ability to convert exogenous supplied [1H] labeled 5-epi- 
aristolochene to radioisotopically labeled 5-epi-aristolochene
during a subsequent 3 hour incubation period (FIGS. 3A-3B).
Apparent activity of 5EAH was inhibited in a dose-dependent manner with
approximately 50% inhibition by either 25 µM ancyamol or
ketoconazole, and more than 80% by 75 µM ancyamol and
95% by 100 µM ketoconazole (FIGS. 3A and 3B). Import-
antly, neither the in vitro activity of recombinant EAS nor
the induction of EAS in the elicitor-treated cell cultures was
significantly affected by ancyamol at concentrations as high
as 100 µM (FIG. 3A). Ketoconazole also does not appear to
affect the in vitro activity of EAS. However, the inducibility
of cyclase activity in elicitor-treated cell extracts was in-
bhibited by ketoconazole at concentrations above 50 µM (FIG.
3B). Therefore, the specificity of ketoconazole as an inhibi-
tor of P450 type reactions should be assessed at or below
a concentration of 50 µM under these experimental condi-
tions.

Isolation of Elicitor-Inducible Cytochrome P450 cDNAs

A two-step approach for the isolation of candidate P450
cDNAs was followed. A PCR strategy was first employed
using a directionally cDNA library prepared against mRNA
isolated from elicitor-induced cells as the template and
degenerate PCR primers (FIGS. 4A-4C). Sequence align-
ments of cytochrome P450s from many families across
kingdoms were used to identify conserved regions to which
a series of degenerate primers were prepared (FIGS. 4A
and 4B). In cloning experiments, 450 to 550 bp products
were expected from reactions utilizing the primer prepared to the
home-binding domain (GRXXCPAVG) (SEQ ID NOS:27
and 28) and the T7 vector primer (FIG. 4C). The mixtures
of reaction products were shotgun cloned, and approxi-
mately 100 of the cloned PCR fragments were sequenced.
About half of the sequenced DNAs contained signature
sequences typical of P450 enzymes as revealed by BlastX
database searches, and these corresponded to typical plant
P450 family members of the CYP71, CYP73, CYP92 and
CYP82 classes. Each of these PCR fragments was isolated
multiple times in separate experiments. In addition, we
isolated full-length cDNAs for these P450 family members.
Table 1 compares the similarity and identity of the full-
length cDNAs of P450 family members with those of their
nearest family member in the GenBank database. In addi-
tion, Figs. 8A-8D shows an amino acid alignment of
several terpene cytochrome P450s. Alignments were performed
using the algorithm of the MACVECTOR software suite.

<table>
<thead>
<tr>
<th>Cytochrome P450</th>
<th>Nearest relative/ accession number</th>
<th>% Identity</th>
<th>% Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP71D20</td>
<td>CYP71D7 (S. chacoense) Gen EMBL U48435</td>
<td>76.5</td>
<td>88.8</td>
</tr>
<tr>
<td>CYP71D21</td>
<td>CYP71D7 (S. chacoense) Gen EMBL U48435</td>
<td>76.3</td>
<td>88.8</td>
</tr>
<tr>
<td>CYP73A27</td>
<td>CYP73A15 (P. vulgaris) Gen EMBL Y09447</td>
<td>79.4</td>
<td>92.6</td>
</tr>
<tr>
<td>CYP73A28</td>
<td>CYP73A15 (P. vulgaris) Gen EMBL Y09447</td>
<td>79.2</td>
<td>92.4</td>
</tr>
<tr>
<td>CYP82E1</td>
<td>CYP82E1 (N. tabacum) Gen EMBL AB015762</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>CYP92A5</td>
<td>CYP92A3 (N. tabacum) Gen EMBL X69764</td>
<td>95.5</td>
<td>98.6</td>
</tr>
</tbody>
</table>

The cloned fragments were used in a second step to isolate
full-length clones from the cDNA library. Screening the
cDNA library by hybridization with the CYP71 and CYP73
gene fragments yielded four full-length cDNAs, two
CYP71Ds and two CYP73As. The former clones were
designated CYP71D20 and CYP71D21, and the latter were
designated CYP73A27 and CYP73A28. The other two
cDNA fragments corresponded to tobacco cDNAs already
found in the GenBank database, CYP82E1 and CYP92A3.
These two cDNAs were cloned using specific primers
designed with the help of the available sequence informa-
tion to amplify the full-length cDNA.

Induction of Cytochrome P450 mRNAs in Elicitor-Treated Cells

To correlate a biochemical role for P450s in sesquiterpene
metabolism, RNA blot analyses were used to determine the
steady-state levels of the mRNAs coding for all four of the
cytochrome P450 clones and EAS in control and elicitor-
treated cells (FIG. 5). The mRNAs for all four of the P450s
were rapidly and transiently induced with slightly different
time courses relative to one another and to the EAS mRNA.
CYP73A27 mRNA, for instance, displayed an induction
pattern similar to that of EAS with the maximum mRNA
level occurring 9 to 12 hours after elicitation. While the EAS
mRNA remained high throughout the duration of the experi-
ment, the CYP73A27 mRNA was negligible in cells 24
hours after elicitor-treatment. In contrast, the CYP71D
mRNA was more rapidly induced than the EAS mRNA,
reached its maximum 6 to 9 hours after elicitation, and was
decaying by 12 hours when the EAS mRNA level was still
very high.

Functional Identification of CYP71D20 as 5-epi-aristo-
lochene Hydroxylase

To ascribe functional identity to the various P450 transcripts,
full-length cDNAs for CYP71D20, CYP82E1 and CYP92A5 were inserted into the yeast expression vector
pCDP600 (Urban et al., Biochimie 72:463-472, 1990, Pom-
then the other, in either order. Moreover, such hydroxylases
the following materials and methods.

Using the extinction coefficient of 91
these cDNAs was under the control of the glucose-repres­
metabolized to only one product with the same retention
was detectable in the 5-epi-aristolochene incubations (FIGS.
identical to that for the capsidiol standard (EIMS m/z 236,
was determined that approximately 107 pmol
5-epi-aristolochene hydroxy lase-activity was measured as
in the yeast cells per milligram of total yeast protein.
method of YeDP60 vector (control) alone.
After induction with galactose for approximately 16
controls and cells containing the various P450
5-epi-aristolochene and 1-deoxycapsidiol were
both showed characteristic CO difference
it was determined that approximately 100,000 dpm at 2.5 nM) for 3 hour periods at various
both were inducible with the membrane anchoring sequence of
three hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
addition of these hydroxyl groups may occur, for example,
All experiments were replicated in several independent trials. While the absolute values presented may have varied between experiments by as much as 50%, the trends and time courses were consistent throughout.

Construction of an Elicitor-Induced cDNA Library

Cell cultures were incubated with fungal elicitor (0.5 µg cellulase/mL) for 6 hours before collecting the cells by filtration. The cells were kept frozen at 80°C until total RNA was extracted from them using Trizol (Life Technologies, Rockville, Md.) according to the manufacturer’s instructions. Poly (A)* RNA was purified by two rounds of oligo (dT) cellulose column chromatography (Life Technologies, Rockville, Md.). cDNA synthesis and library construction were subsequently carried out using the UNI-ZAP XR library kit (Stratagene, La Jolla, Calif.), according to manufacturer’s instructions.

PCR Cloning Strategy

Cytochrome P450 cDNA fragments were amplified from the elicitor-induced cDNA library using various combinations of degenerate forward and reverse primers with the vector-specific T3 and T7 primers. The template DNA was prepared from a 500 µL aliquot of the elicitor-induced cDNA library (3x10⁷ pfu/µL) by heat denaturation at 70°C for 10 minutes, followed by phenol/chloroform extraction, ethanol precipitation and re-suspension in 500 µL of sterile, deionized water. Amplification reactions were performed in 50 µL volumes containing 50 mM KCl; 10 mM Tris-HCl, pH 8.8; 1.5 mM MgCl₂; 200 µM of each dNTP; 2 µL template DNA; 20 pmol each of forward and reverse primer; and 1 unit Taq Polymerase (Life Technologies, Rockville, Md.). Reactions were preheated at 94°C for 2 minutes, followed by thirty-five cycles of denaturing at 94°C for 1 minute, annealing at 50°C for 1 minute 30 seconds, and polymerization at 72°C for 2 minutes. The reactions were completed by a 10-minute extension at 72°C. Aliquots of the reaction products were examined for DNA products by agarose gel fractionation, and ligated directly into the pGEM-T Easy vector (Promega, Madison, Wis.). Resulting recombinant plasmids containing insert DNAs within the expected size range were sequenced using T7 and Sp6 primers.

DNA Sequencing

All the DNA sequencing reactions were performed using the BIGDYE™ Terminator Cycle sequencing kit (Perkin-Elmer, Wellesley, Mass.) with the sequences being read on an automated ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, Calif.). Computer assessment of the DNA sequencing reactions were performed using the MACVECTOR (Oxford Molecular, Madison, Wis.) software package.

cDNA Library Screening

The cDNA library was screened with digoxigenin labeled probes. A 258 bp DNA fragment amplified from the pGEM-deg6.4 clone using gene-specific forward (5'-GGCGGGA-GAATTTGTGCTGAATGTATGGTTTAG-3' (SEQ ID NO:13)) and reverse (5'-GTACAAATGTGAGGTGATCTGACAATG-3' (SEQ ID NO:14)) primers; and a 374 bp DNA fragment amplified from the pBKs-CYPB3.843 clone with specific forward (5'-GGTGTTGTAATTGATG-3' (SEQ ID NO:15)) and reverse (5'-TTATATCGCAATAG-GGCTTAGAAACA-3' (SEQ ID NO:16)) primers, were used to screen for CYP71Ds. The probes were labeled with digoxigenin-11-dUTP using the PCR DIG Labeling Mix (Roche Molecular Biochemicals, Indianapolis, Ind.), hybridized to plaque lifts of the cDNA library plated at approximately 10,000 PFUs per 150 mm plate, and was hybridization detected with the DIG detection system according to the manufacturer’s instructions (Roche Molecular Biochemicals, Indianapolis, Ind.). Plaques exhibiting strong hybridization were plaque purified, auto-subcloned to their plasmid forms according to the manufacturer’s recommendations (Stratagene, La Jolla, Calif.), and then subjected to DNA sequencing as described above.

RNA Analysis

RNA gel blot analysis was carried out using 10 µg aliquots of total RNA. RNA samples were heat-denatured at 70°C for 15 minutes in sample buffer (1x MOPS, 50% formamide, 16% formaldehyde, 30% glycerol, and 3% ethidium bromide), and size fractionated on a 1.2% agarose gel containing 1xMOPS and 18.1% formaldehyde. Uniformity of sample loading was determined by visual inspection of the gel for RNA bands. The RNAs were then transferred to a Zeta Probe nylon membrane (Bio-Rad Laboratories, Hercules, Calif.) and hybridized according to the manufacturer’s recommendations. Full-length cDNA probes were labeled with [32P]-dCTP (PRIME-IT Kit, Stratagene, La Jolla, Calif.) prior to hybridization. After hybridization, the membranes were washed in 2xSSC/0.1% SDS once at room temperature followed by sequential washes in 0.2xSSC/0.1% SDS at 42°C and 65°C. Hybridization was detected with a Phosphoimager (Molecular Dynamics, model 445 SI).

Construction of Yeast Expression Vectors

The coding regions of the P450 cDNAs were cloned into the pYcDP60 expression vector (Urban et al., J. Biol. Chem. 272:19176-19186, 1990; Pompon et al., Methods Enzymol. 272:51-64, 1996). Appropriate BamHI, EcoRI, and SstI restriction sites (underlined) were introduced via PCR primers containing these sequences either upstream of the translation start site (ATG) or downstream of the stop codon (TAA or TGA). The primers used to amplify the CYP71D20 cDNA were 5'-GGG GATATCCATGCGAACATCTTCAGTGTGTTTCGTT-3' (SEQ ID NO:17) and 5'-GGG GAATTCTAATCTCTCGAAGAAGATGTTGAGAAGG-3' (SEQ ID NO:18); for the CYPB2E1 cDNA 5'-GGC GGATATCCATGACCTCCTCCTCATAGAAAAACCCT-3' (SEQ ID NO:19) and 5'-GGG GAATTCTAATCTCTCGAAGAAGATGTTGAGAAGG-3' (SEQ ID NO:20); and for the CYP92A3 cDNA 5'-GGC GGATATCCATGACCTCCTCCTCATAGAAAAACCCT-3' (SEQ ID NO:21) and 5'-GGG GAATTCTAATCTCTCGAAGAAGATGTTGAGAAGG-3' (SEQ ID NO:22). Two long, overlapping (italicized) primers 5'-GCCATATGCGGCAATACTACTCCAATCTCCGCGCTCTTAAATTCAGC-3' (SEQ ID NO:23) and 5'-GGG GGATATCCATGACCTCCTCCTCATAGAAAAACCCT CGTGGCCTATATCCGGCCATTATCGCGGCAATCAGTATGGTTTAG-3' (SEQ ID NO:24) coding for the N-terminal sequence of CYP73A27 (GenEMBL Zl 7369) up to the hinge region were used for the modification of the membrane anchoring segment of CYP73A27 to avoid possible problems with intracellular targeting due to the unusual N-terminus (Nedelkina et al., 1999); the reverse primer used for both amplifications was 5'-GGG GAATTCTAATCTCTCGAAGAAGATGTTGAGAAGG-3'. The primers were used to amplify the P450 cDNAs directly from the cDNA library template. Amplifications were performed in 50 µL reactions containing 1x Pfx amplification buffer; 1 mM MgSO₄; 300 µM of each dNTP; 10 ng template DNA; 20 pmol each of forward and reverse primer; and 1.25 units PLATINUM®
Pfx Polymerase (Life Technologies, Rockville, Md.). Reactions were preheated at 94°C for 2 minutes, followed by thirty-five cycles of denaturing at 94°C for 15 seconds, annealing at 55°C for 30 seconds, and elongating at 68°C for 1.5 minutes. PCR products were ligated into the pGEM-T EASY vector (Promega, Madison, Wis.) and subcloned into the pYeDP60 vector. The resulting constructs were validated by a combination of PCR and DNA sequencing.

Yeast Expression Studies

Verified pYeDP60-P450 cDNA constructs were introduced into the yeast WAT11 line, a derivative of the W303-1B strain (MATa; ade-2-1; his 3-11; leu 2-3, -112; ura 3-1; can4; cyr), provided by Dr. P. Urban (Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette, France). The endogenous NADPH-cytochrome P450 reductase (CPR1) locus has been replaced with ATR1, a NADPH-cytochrome P450 reductase from Arabidopsis thaliana (Pompon et al., Methods Enzymol. 272:51-64, 1996; Urban et al., J. Biol. Chem. 272:19176-19186, 1997), in the WAT11 line. Yeast was grown overnight in a 30°C shaker in YPAD (1 g/1 yeast extract; 1 g/L peptone; 20 g/L glucose; 200 mg/L adenine) media. Cultures were harvested at an A600 between 0.5 and 1.5. Cells were collected by centrifugation at 2,500 x g for 5 minutes at 4°C, and resuspended in ice-cold, sterile dH2O. Cells were pelleted again as above and resuspended in 1M sorbitol. Forty µL of yeast suspension was mixed with 0.5 to 1 µg plasmid DNA (in <5 µL dH2O) in a pre-chilled 0.5 mL tube, and transferred to a chilled cuvette with a 0.2 cm electrode gap. One pulse at 1.5

5-epi-aristolochene-1,3-hydroxylase Assays

5-epi-aristolochene-1,3-hydroxylase assays were performed in 0.5 mL polyethylene tubes in 100 µL volumes. 5-epi-aristolochene or 1-deoxycapsidiol dissolved in hexane was added to the tube, and the organic solvent was removed by incubation of the open tube at 30°C. 5-epi-aristolochene and 1-deoxycapsidiol were resuspended in 2 µL dimethyl sulfoxide before adding the reaction mixture. Reactions were carried out in 100 mM Tris-HCl, pH 7.5, to which microsomal protein was added to a final concentration of 1 mg/mL. Reactions were initiated by the addition of 2 mM NADPH. The final concentration of 5-epi-aristolochene and 1-deoxycapsidiol in these assays varied from 20 to 50 µM. After incubations for variable lengths of time at 30°C, the reactions were extracted with two volumes of ethyl acetate. The organic extracts were concentrated and evaluated by GC-MS along with standards of 5-epi-aristolochene (Whitehead et al., Phytochemistry 28:775-779, 1989; Rising et al., J. Am. Chem. Soc. 122:1861-1866, 2000), 1-deoxycapsidiol (Whitehead et al., Phytochemistry 29:19176-19186, 1997), in the WAT11 line. Yeast was grown overnight in a 30°C shaker in YPAD (1 g/1 yeast extract; 1 g/L peptone; 20 g/L glucose; 200 mg/L adenine) media. Cultures were harvested at an A600 between 0.5 and 1.5. Cells were collected by centrifugation at 2,500 x g for 5 minutes at 4°C, and resuspended in ice-cold, sterile dH2O. Cells were pelleted again as above and resuspended in 1M sorbitol. Forty µL of yeast suspension was mixed with 0.5 to 1 µg plasmid DNA (in <5 µL dH2O) in a pre-chilled 0.5 mL tube, and transferred to a chilled cuvette with a 0.2 cm electrode gap. One pulse at 1.5

5-epi-aristolochene-1,3-hydroxylase Assays

5-epi-aristolochene-1,3-hydroxylase assays were performed in 0.5 mL polyethylene tubes in 100 µL volumes. 5-epi-aristolochene or 1-deoxycapsidiol dissolved in hexane was added to the tube, and the organic solvent was removed by incubation of the open tube at 30°C. 5-epi-aristolochene and 1-deoxycapsidiol were resuspended in 2 µL dimethyl sulfoxide before adding the reaction mixture. Reactions were carried out in 100 mM Tris-HCl, pH 7.5, to which microsomal protein was added to a final concentration of 1 mg/mL. Reactions were initiated by the addition of 2 mM NADPH. The final concentration of 5-epi-aristolochene and 1-deoxycapsidiol in these assays varied from 20 to 50 µM. After incubations for variable lengths of time at 30°C, the reactions were extracted with two volumes of ethyl acetate. The organic extracts were concentrated and evaluated by GC-MS along with standards of 5-epi-aristolochene (Whitehead et al., Phytochemistry 28:775-779, 1989; Rising et al., J. Am. Chem. Soc. 122:1861-1866, 2000), 1-deoxycapsidiol (Whitehead et al., Phytochemistry 29:479-182, 1990), and capsidiol (Whitehead et al., Phytochemistry 30:2171-2173, 1991). GC analysis was routinely performed with an HP5890 GC equipped with a Hewlett-Packard HP-5 capillary column (30 mm x 0.25 mm, 0.25 µm phase thickness) and FID as described previously (Rising et al., J. Am. Chem. Soc. 122:1861-1866, 2000). GC-MS analysis was performed at the University of Kentucky Mass Spectrometry Facility using a Varian 3400 gas chromatograph and a Finnigan INCOS 50 quadrupole mass selective detector. The GC was equipped with a J&W DB-5 ms capillary column (15 mm x 0.25 mm, 0.25 µm phase thickness) and run with He as the carrier gas (10 psi). Splitless injections were done at an injection port temperature of 280°C. The column temperature was maintained at 40°C for 1 minute and then increased to 280°C at 10°C per minute. Following separation by the GC column, samples were introduced directly into the electron impact ionization source. Mass spectra were acquired at 70 eV, scanning from 40-440 Da in 1 second.
Production of Cytochrome P450s

Using the standard molecular techniques described herein, the isolation of additional cytochrome P450 coding sequences is readily accomplished. For example, using all or a portion of the amino acid sequence of any of the disclosed P450s, one may readily design P450-specific oligonucleotide probes, including P450 degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either DNA strand and any appropriate portion of the P450 nucleotide sequence. General methods for designing and preparing such probes are provided, for example, in Ausubel et al., 2000, Current Protocols in Molecular Biology, Wiley Interscience, New York, and Berger and Kimmel, Guide to Molecular Cloning Techniques, 1987, Academic Press, New York. These oligonucleotides are useful for P450 gene isolation, either through their use as probes capable of hybridizing to a P450 complementary sequence, or as primers for various amplification techniques, for example, polymerase chain reaction (PCR) cloning strategies.

Hybridization techniques and screening procedures are well known to those skilled in the art and are described, for example, in Ausubel et al. (supra); Berger and Kimmel (supra); Chen et al. (Arch. Biochem. Biophys. 324:255, 1995); and Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York). If desired, a combination of different oligonucleotide probes may be used for the screening of a recombinant DNA library. The oligonucleotides may be detectably-labeled using methods known in the art and used to probe filter replicas from a recombinant DNA library. Recombinant DNA libraries are prepared according to methods well known in the art, for example, as described in Ausubel et al. (supra), or they may be obtained from commercial sources.

As discussed above, P450 oligonucleotides may also be used as primers in a polymerase chain reaction (PCR) amplification cloning strategy. PCR methods are well known in the art and are described, for example, in PCR Technology, Erlich, ed., Stockton Press, London, 1989; PCR Protocols: A Guide to Methods and Applications, Innis et al., eds., Academic Press, Inc., New York, 1990; and Ausubel et al. (supra). Primers are optionally designed to allow cloning of the amplified product into a suitable vector, for example, by including appropriate restriction sites at the 5' and 3' ends of the amplified fragment (as described herein). If desired, a P450 gene may be isolated using the PCR “RACE” technique, or Rapid Amplification of cDNA Ends (see, e.g., Innis et al. (supra)). By this method, oligonucleotide primers based on a P450 sequence are oriented in the 3' and 5' directions and are used to generate overlapping PCR fragments. These overlapping 5'- and 3'-end RACE products are combined to produce an intact full-length cDNA. This method is described in Innis et al. (supra) and Frohman et al. (Proc. Natl. Acad. Sci. U.S.A. 85:8998, 1988).

Additional methods for identifying sequences encoding P450s are provided in Maughan et al. (Arch. Biochem. Biophys. 341:104-111, 1997) and Clark et al. (Plant Mol. Biol. 33:875-885, 1997).

Useful P450 sequences may be isolated from any appropriate organism. Confirmation of a sequence’s relatedness to a P450 polypeptide disclosed herein may be accomplished by a variety of conventional methods, for example, by comparing the sequence with a known P450 sequence found in a database. In addition, the activity of any P450 may be evaluated according to any of the techniques described herein.

P450 Polypeptide Expression

P450 polypeptides may be produced by transformation of a suitable host cell with all or part of a P450 DNA (for example, anyone of the P450 cDNAs described herein) in a suitable expression vehicle or with a plasmid construct engineered for increasing the expression of a P450 polypeptide in vivo.

Those skilled in the field of molecular biology will appreciate that any of a wide variety of expression systems may be used to provide the recombinant protein. The precise host cell used is not critical to the invention. The P450 protein may be produced in a prokaryotic host, for example, E. coli TB 1, or in an eukaryotic host, for example, Saccharomyces cerevisiae, insect cells, mammalian cells (for example, COS 1 or NIH 3T3 cells), or any of a number of plant cells including, without limitation, algae, tree species, ornamental species, temperate fruit species, tropical fruit species, vegetable species, legume species, monocots, dicots, or in any plant of commercial or agricultural significance. Particular examples of suitable plant hosts include, but are not limited to, Conifers, Potania, Tomato, Potato, Tobacco, Rape, Arabidopsis, Lettuce, Sunflower, Oilseed rape, Flax, Cotton, Sugarbeet, Celery, Soybean, Alfalfa, Medicago, Lotus, Vigna, Cucumber, Carrot, Eggplant, Cauliflower, Horseradish, Morning Glory, Poplar, Walnut, Apple, Asparagus, Grape, Rice, Maize, Millet, Onion, Barley, Orchard grass, Oat, Rye, Tobacco and Wheat.

Such cells are available from a wide range of sources including: the American Type Culture Collection (Rockland, Md.); or from any of a number of seed companies, for example, W. Atlee Burpee Seed Co. (Warminster, Pa.), Park Seed Co. (Greenwood, S.C.), Johnny Seed Co. (Albion, Me.), or Northrup King Seeds (Harstville, S.C.). Descriptions and sources of useful host cells are also found in Pouwels et al. (supra) or Ausubel et al. (supra). Additional methods for identifying sequences encoding P450s, one may readily design P450-specific oligonucleotide primers based on any of the disclosed P450 cDNAs described herein) in a suitable expression vehicle or with a plasmid construct engineered for increasing the expression of a P450 polypeptide in vivo.

For prokaryotic expression, DNA encoding a P450 polypeptide is carried on a vector operably linked to control signals capable of affecting expression in the prokaryotic host. If desired, the coding sequence may contain, at its 5' or 3' end, a sequence encoding any of the known signal sequences capable of affecting secretion of the expressed protein into the periplasmic space of the host cell, thereby facilitating recovery of the protein and subsequent purification. Prokaryotes most frequently used are various strains of E. coli; however, other microbial strains may also be used. Plasmid vectors are used which contain replication origins, selectable markers, and control sequences derived from a species compatible with the microbial host. Examples of such vectors are found in Pouwels et al. (supra) or Ausubel et al. (supra). Commonly used prokaryotic control sequences (also referred to as “regulatory elements”) are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences. Promoters commonly used to direct protein expression include the beta-lactamase (penicillinase), the lactose (lac), the tryptophan (Trp) (Goedde et al., Nucl. Acids Res. 8:4057, 1980), and the tac promoter systems, as well as the lambda-derived P.sub.L promoter and N-gene ribosome binding site (Simatake et al., Nature 292:128, 1981).
One particular bacterial expression system for P450 production is the E. coli pET expression system (Novagen). According to this expression system, DNA encoding a P450 is inserted into a pET vector in an orientation designed to allow expression. Since the P450 gene is under the control of the T7 regulatory signals, P450 expression is dependent on inducing the expression of T7 RNA polymerase in the host cell. This is typically achieved using host strains which express T7 RNA polymerase in response to IPTG induction. Once produced, recombinant P450 is then isolated according to standard methods known in the art, for example, those described herein.

Another bacterial expression system for P450 production is the pGEX expression system (Pharmacia). This system employs a GST gene fusion system that is designed for high-level expression of a gene or fragment as a fusion protein with rapid purification on glutathione-agarose beads. Glutathione S-transferase domain from the fusion protein is readily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Fusion proteins can be recovered under mild conditions by elution with glutathione. Cleavage of the glutathione S-transferase domain from the fusion protein is facilitated by the presence of recognition sites for site-specific proteases upstream of this domain. For example, proteins expressed in pGEX-2T plasmids may be cleaved with thrombin; those expressed in pGEX-3X may be cleaved with factor Xa.

Other prokaryotic systems useful for expressing eukaryotic P450s are described by Cooper (Mutat. Res. 454:45-52, 2000) and Dong et al. (Arch. Biochem. Biophys. 327:254-259, 1996). In addition, strategies for enhancing the prokaryotic expression of a cytochrome P450 in combination with cytochrome reductase are described in Porter et al. (Drug. Metab. Rev. 31:159-174, 1999).


One preferred eukaryotic expression system is the mouse 3T3 fibroblast host cell transfected with a pMAMneo expression vector (Clontech). pMAMneo provides: an RSV-LTR enhancer linked to a dexamethasone-inducible MMTV-LTR promoter, an SV40 origin of replication which allows replication in mammalian systems, a selectable neomycin gene, and SV40 splicing and polyadenylation sites. DNA encoding a P450 is inserted into the pMAMneo vector in an orientation designed to allow expression. The recombinant P450 is then isolated as described below. Other preferable host cells which may be used in conjunction with the pMAMneo expression system include COS cells and CHO cells (ATCC Accession Nos. CRL 1650 and CCL 61, respectively).

Alternatively, if desired, a P450 is produced by a stably-transfected mammalian cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public, e.g., see Pouwels et al. (supra); methods for constructing such cell lines are also publicly available, e.g., in Ausubel et al. (supra). In one example, cDNA encoding the P450 is cloned into an expression vector which includes the dihydrofolate reductase (DHFR) gene. Integration of the plasmid and, therefore, the P450-encoding gene into the host cell chromosome is selected for by inclusion of 0.01-300 μM methotrexate in the cell culture medium (as described in Ausubel et al., supra). This dominant selection can be accomplished in most cell types. Recombinant protein expression can be increased by DHFR-mediated amplification of the transfected gene. Methods for selecting cell lines bearing gene amplifications are described in Ausubel et al. (supra); such methods generally involve extended culture in medium containing gradually increasing levels of methotrexate. DHFR-containing expression vectors commonly used for this purpose include pCVSEII-DHrF and pAd26SV(A) (described in Ausubel et al., supra). Any of the host cells described above or, preferably, a DHFR-deficient CHO cell line (for example, CHO DHFR cells, ATCC Accession Number CRL 9096) are among the host cells preferred for DHFR selection of a stably-transfected cell line or DHFR-mediated gene amplification.

A cytochrome P450 may also be produced in insect cells, such cells include, without limitation, Spodoptera frugiperda (SI)-9, SI-21, or Drosophila melanogaster Schneider (SI-2) cells. For P450 production, insect cells are typically infected with a baculovirus, for example, Autographa californica Multiple Nuclear Polyhedrosis Virus (AcMNPV) containing an expression cassette for such a protein, e.g., cytochrome P450, at a multiplicity of infection of 1 to 10. The infected cells are generally cultured in a standard insect cell culture medium for 24 to 48 hours prior to recovering the protein using standard molecular biology techniques. If desired, a P450 polypeptide may also be produced in insect cells directly transfected with a DNA construct containing an expression cassette encoding the P450.

Furthermore, any of the cytochrome P450s described herein may be produced in yeast, for example, Pichia pastoris. In order to produce the P450, yeast cells are transformed with an expression cassette containing, for example, a promoter such as the AOX1 or phosphoglycerate kinase gene promoter, the P450 gene to be expressed, and a terminator. Such an expression cassette may contain an origin of replication or may be integrated into the yeast genomic DNA. The expression cassette is generally introduced by lithium acetate transformation or by the use of spheroplasts. In order to select for successfully transformed cells, the yeast are plated, for example, on minimal media which only allows yeast carrying the introduced expression cassette to grow.

In addition, expression of recombinant proteins in yeast using a Hansenula polymorpha expression system is described in U.S. Pat. Nos. 5,741,674 and 5,672,487.

A P450 may also be produced by a stably-transfected plant cell line or by a transgenic plant. Such genetically-engineered plants are useful for a variety of industrial and agricultural applications as discussed below. Importantly, this invention is applicable to gymnosperms and angiosperms.
sporns, and will be readily applicable to any new or improved transformation or regeneration method.

A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are described in Poulwels et al. (supra), Weissbach and Weissbach (supra), and Gelvin et al. (supra). Methods for constructing such cell lines are described in, e.g., Weissbach and Weissbach (supra), and Gelvin et al. (supra). Typically, plant expression vectors include (1) a cloned P450 gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive expression, or environmentally- or developmentally-regulated, or pathogen- or wound-inducible, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

The P450 DNA sequence of the invention may, if desired, be combined with other DNA sequences in a variety of ways. The P450 DNA sequence of the invention may be employed with all or part of the gene sequences normally associated with a P450. In its component parts, a DNA sequence encoding a P450 is combined in a DNA construct having a transcription initiation control region capable of promoting transcription and translation in a host cell.

In general, the constructs will involve regulatory regions functional in plants which provide for production of a P450 as discussed herein. The open reading frame coding for the P450, or a functional fragment thereof, will be joined at its 5' end to a transcription initiation regulatory region such as the sequence naturally found in the 5' upstream region of a P450 structural gene, for example, a CYP71D20 (SEQ ID NO:2) or CYP71D21 (SEQ ID NO:4) gene. Numerous other transcription initiation regions are available which provide for constitutive or inducible regulation.

For applications when developmental, cell, tissue, hormonal, environmental, or pathogen-inducible expression are desired, appropriate 5' upstream non-coding regions are obtained from other genes; for example, from genes regulated during seed development, embryo development, leaf development, or in response to a pathogen.

Regulatory transcript termination regions may also be provided in DNA constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding a P450 or any convenient transcription termination region derived from a different gene source. The transcript termination region will contain preferably at least 1-3 kb of sequence 3' to the structural gene from which the termination region is derived.

An example of a useful plant promoter according to the invention is a caulimovirus promoter, such as, a cauliflower mosaic virus (CaMV) promoter. These promoters confer high levels of expression in most plant tissues, and the activity of these promoters is not dependent on virally encoded proteins. CaMV is a source for both the 35S and 19S promoters. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odell et al., Nature 313:810, 1985). The CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., Plant Cell 2:591, 1990; Terada and Shimamoto, Mol. Genet. 220:389, 1990). Moreover, activity of this promoter can be further increased (i.e., between 2-10 fold) by duplication of the CaMV 35S promoter (see, e.g., Kay et al., Science 236:1299, 1987; Ow et al., Proc. Nutl. Acad. Sci. U.S.A. 84:4870, 1987; and Fang et al., Plant Cell 1:141, 1989).

Other useful plant promoters include, without limitation, the nopalin synthase promoter (An et al., Plant Physiol. 88:547, 1988) and the octopine synthase promoter (Fromm et al., Plant Cell 1:977, 1989).

For certain applications, it may be desirable to produce the P450 gene product in an appropriate tissue, at an appropriate level, or at an appropriate developmental time. For this purpose, there is an assortment of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, which have been shown to be regulated in response to the environment, hormones, and/or developmental cues. These include gene promoters that are responsible for heat-regulated gene expression (see, e.g., Callis et al., Plant Physiol. 88:965, 1988; Takahashi and Komedak, Mol. Genet. 219:365, 1989; and Takahashi et al., Plant J. 2:751, 1992); light-regulated gene expression (e.g., the pea rbcS-3A described by Kuhlmeier et al. (Plant Cell 1:471, 1989); the maize rbcS promoter described by Schaffner and Sheen (Plant Cell 3:997, 1991); or the chlophyll a/b-binding protein gene found in pea described by Simpson et al. (EMBO J. 4:2723, 1985}); hormone-regulated gene expression (for example, the abscisic acid (ABA) responsive sequences from the Em gene of wheat described by Marcotte et al. (Plant Cell 1:969, 1989); the ABA-inducible HVA1 and HVA22, and the rd29A promoters described for barley and Arabidopsis by Straub et al. (Plant Cell 6:617, 1994), Shen et al. (Plant Cell 7:295, 1994)); and wound-induced gene expression (for example, of wunl described by Siebertz et al. (Plant Cell 1:961, 1989); or organ-specific expression (for example, of the tuber-specific storage protein gene described by Roshal et al. (EMBO J. 6:1155, 1987); the 23-kDa zein gene from maize described by Schierthianer et al. (EMBO J. 7:1249, 1988); or the French bean beta-phaseolin gene described by Bustos et al. (Plant Cell 1:839, 1989); and pathogen-inducible gene expression described by Chappell et al. in U.S. Ser. Nos. 08/471,983; 08/443,639; and 08/577,483; hereby incorporated by reference.

Plant expression vectors may also optionally include RNA processing signals, for example, introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., Genes and Dev. 1:1183, 1987). The location of the RNA splice sequences can dramatically influence the level of transgene expression in plants. In view of this fact, an intron may be positioned upstream or downstream of a P450-encoding sequence in the transgene to modulate levels of gene expression.

In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally present in the 3' regions of plant genes (Thomburg et al., Proc. Nutl. Acad. Sci. U.S.A. 84:744, 1987; An et al., Plant Cell 1:115, 1989). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the Pl-11 terminator region of potato. In addition, other commonly used terminators are derived from the octopine or nopalin synthase signals.

The plant expression vector also typically contains a dominant selectable marker gene used to identify those cells that have become transformed. Useful selectable genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin, or spectinomycin. Genes required for photosynthesis may also be used as selectable markers in photosynthetic-deficient strains. Alternatively, the green-fluorescent protein from the...
jellyfish *Aequorea victoria* may be used as a selectable marker (Sheen et al., Plant J. 8:777, 1995; Chiu et al., Current Biology 6:325, 1996). Finally, genes encoding herbicide resistance may be used as selectable markers; useful herbicide resistance genes include the bar gene encoding the enzyme phosphinothricin acetyltransferase and conferring resistance to the broad-spectrum herbicide BASTA (Hoechst AG, Frankfurt, Germany).

Efficient use of selectable markers is facilitated by a determination of the susceptibility of a plant cell to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100 µg/mL (kanamycin), 20-50 µg/mL (hygromycin), or 5-10 µg/mL (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., by Vasil et al., supra.

It should be readily apparent to one skilled in the art of molecular biology, especially in the field of plant molecular biology, that the level of gene expression is dependent, not only on the combination of promoters, RNA processing signals, and terminator elements, but also on how these elements are used to increase the levels of selectable marker gene expression.

**Plant Transformation**

Upon construction of the plant expression vector, several standard methods are available for introduction of the vector into a plant host, thereby generating a transgenic plant. These methods include (1) *Agrobacterium*-mediated transformation (*A. tumefaciens* or *A. rhizogenes*) (see, e.g., Lichtenstein and Fuller, In: Genetic Engineering, vol. 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C. P., and Draper, J., In: DNA Cloning, Vol II, D. M. Glover, ed, Oxford, IRI Press, 1985); (2) the particle delivery system (see, e.g., Gordon-Kamm et al., Plant Cell 2:603, 1990; or BioRad Technical Bulletin 1687, supra); (3) microinjection protocols (see, e.g., Green et al., supra); (4) polyethylene glycol (PEG) procedures (see, e.g., Drummer et al., Plant Cell Physiol. 23:451, 1982; or e.g., Zhang and Wu, Theor. Appl. Genet. 76:835, 1988); (5) liposome-mediated DNA uptake (see, e.g., Freeman et al., Plant Cell Physiol. 25:1353, 1984); (6) electroporation protocols (see, e.g., Gelvin et al., supra; Dekeyser et al., supra; Fromm et al., Nature 319:791, 1986; Sheen, Plant Cell 2:1027, 1990; or Jang and Sheen, Plant Cell 6:1665, 1994); and (7) the vortexing method (see, e.g., Kindle, supra). The method of transformation is not critical to the present invention. Any method which provides for efficient transformation may be employed. As newer methods are available to transform crops or other host cells, they may be directly applied.

The following is an example outlining one particular technique, an *Agrobacterium*-mediated plant transformation. By this technique, the general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, cloning and DNA modification steps are carried out in *E. coli*, and the plasmid containing the gene construct of interest is transferred by conjugation or electroporation into *Agrobacterium*. Second, the resulting *Agrobacterium* strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains an origin of replication that allows it to replicate in *Agrobacterium* and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to *Agrobacterium* for subsequent introduction into plants. Resistance genes can be carried on the vector, one for selection in bacteria, for example, streptomycin, and another that will function in plants, for example, a gene encoding kanamycin resistance or herbicide resistance. Also present on the vector are restriction endonuclease sites for the addition of one or more transgenes and directional T-DNA border sequences which, when recognized by the transfer functions of *Agrobacterium*, delimit the DNA region that will be transferred to the plant.

In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad) used for the shooting, a gunpowder charge (22 caliber Power Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The latter is fired at an acrylic stopping plate that has a hole through it that is too small for the macroprojectile to pass through. As a result, the plastic macroprojectile smashes against the stopping plate, and the tungsten microprojectiles continue toward their target through the hole in the plate. For the present invention, the target can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast.

In general, transfer and expression of transgenes in plant cells are now routine practices to those skilled in the art, and have become major tools to carry out gene expression studies in plants and to produce improved plant varieties of agricultural or commercial interest.

**Transgenic Plant Regeneration**

Plants cells transformed with plant expression vectors can be regenerated, for example, from single cells, callus tissue, or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil (supra), Green et al. (supra), Weissbach and Weissbach (supra) and Gelvin et al. (supra). In one particular example, a cloned P450, under the control of the EAS4 promoter and the nopaline synthase terminator and carrying a selectable marker (for example, kanamycin resistance), is transformed into *Agrobacterium*. Transformation of leaf discs (for example, of tobacco leaf discs), with vector-containing *Agrobacterium* is carried out as described by Horsch et al. (Science 227:1229, 1985). Putative transformants are selected after a few weeks (for example, 3 to 5 weeks) on plant tissue culture media containing kanamycin (e.g., 100 µg/mL). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin-resistant plants are then selected for greenhouse growth. If desired, seeds from self-fertilized transgenic plants can then be sowed in soil-less medium and grown in a greenhouse. Kanamycin-resistant progeny are selected by sowing surface sterilized seeds on hormone-free kanamycin-containing media. Analysis for the integration of the transgene is accomplished by standard techniques (see, for example, Ausbel et al. (supra); Gelvin et al. (supra)).

Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA by standard immunoblot and DNA detection techniques. Each positive transgenic plant and its transgenic progeny is unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic DNA is in most cases random, and the site of integration can profoundly affect the levels and the
The invention also includes engineering host cells to include novel isoprenoid metabolic pathways useful in the production of new isoprenoid compounds. By introducing genes encoding an isoprenoid synthase (as disclosed in U.S. Pat. No. 5,824,774 and WO 00/17327) and a cytochrome P450, an acetyltransferase, a methyl transferase, a fatty and an acyltransferase, or a combination thereof, various isoprenoid reaction products may be modified, controlled, or manipulated, resulting in enhancement of production of numerous isoprenoid reaction products, for example, the production of novel monoterpenes, diterpenes, and sesquiterpenes. Such compounds are useful as phytalexins, insecticides, perfumes, and pharmaceuticals such as anti-bacterial and fungal agents.

In one working example, an isoprenoid synthase or a chimeric isoprenoid synthase (as disclosed in U.S. Pat. No. 5,824,774 and WO 00/17327) and a P450 gene are introduced into yeast, for example, using any of the procedures described herein. If desired, such cells may also express, either independently or in combination, an acetyltransferase (see, for example, Walker et al., Proc. Natl. Acad. Sci. U.S.A. 18:583-587, 2000), a methylase transferase gene (see, for example, Diener et al., Plant Cell 12:853-870, 2000), or a fatty acyltransferase gene, as well as a cytochrome reductase. Cells are then cultured under standard conditions and the production of isoprenoid compounds is assayed according to methods known in the art. Isoprenoid compounds are further purified according to methods well known in the art. Cells expressing novel isoprenoid compounds are taken as useful in the invention.

Such methods provide a unique approach for producing novel isoprenoid starting materials and end products. Either prokaryotic or eukaryotic cell transformed with any of the aforementioned enzymes (or combinations thereof) may be used. Moreover, isoprenoid compounds may be produced in any number of ways known in the art including an in vitro combination of purified enzymes with an appropriate substrate or direct fermentation using a host cell which expresses any combination of the aforementioned enzymes and the appropriate substrates sufficient to drive production of isoprenoid compounds.

The invention is also useful for the production of insect attractants and deterrents, which may either deter insect pests or attract insect predators. In addition, the invention is also useful for generating novel flavorings and perfumes.

Other Embodiments

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

All publications and patents mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.
<160> NUMBER OF SEQ ID NOS: 43
<210> SEQ ID NO 1
<211> LENGTH: 473
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 1
Met Gln Phe Phe Ser Leu Val Ser Ile Phe Leu Phe Leu Ala Phe Leu
1   5   10  15  
Phe Leu Leu Arg Lys Trp Lys Asn Ser Asn Ser Glu Ser Lys Lys Leu
20  25  30  
Pro Pro Gly Pro Trp Lys Ile Pro Ile Leu Gly Ser Met Leu His Met
35  40  45  
Ile Gly Gly Glu Pro His His Val Leu Arg Asp Leu Ala Lys Lys Tyr
50  55  60  
Gly Pro Leu Met His Leu Gin Leu Gly Glu Ile Ser Ala Val Val Val
65  70  75  80  
Thr Ser Arg Asp Met Ala Lys Glu Val Leu Lys Thr His Asp Val Val
85  90  95  
Phe Ala Ser Arg Pro Lys Ile Val Ala Met Asp Ile Ile Cys Tyr Asn
100 105 110  
Gln Ser Asp Ile Ala Phe Ser Pro Tyr Gly Asp His Trp Arg Gin Met
115 120 125  
Arg Lys Ile Cys Val Met Gin Leu Leu Asn Ala Lys Asn Val Arg Ser
130 135 140  
Phe Ser Ser Ile Arg Arg Gin Val Gin Gin Leu Arg Leu Ile Gin Gin Ile
145 150 155 160  
Gly Leu Val Asp Ser Gin Met Gin Gin Gin Gin Gin Gin Gin Gin Gin
165 170 175  
Ile Trp Phe Ala Ser Ser Met Thr Cys Arg Ser Ala Phe Gly Gin Val
180 185 190  
Leu Lys Gly Gin Asp Ile Phe Ala Lys Lys Gin Gin Met Gin Gin Gin
195 200 205  
Leu Ala Glu Gly Phe Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
210 215 220  
Leu Gin Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
225 230 235 240  
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
245 250 255  
Leu Ala Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
260 265 270  
Gly Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
275 280 285  
Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
290 295 300  
Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
305 310 315 320  
Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
325 330 335  
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
340 345 350  
Gly Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
355 360 365  

US 10,184,119 B2
Leu Gly Arg Asp Pro Lys Tyr Trp Asp Asp Ala Glu Ser Phe Lys Pro
270
275
280
Glu Arg Phe Glu Gln Cys Ser Val Asp Phe Phe Gly Asn Asn Phe Glu
385
390
395
400
Phe Leu Pro Phe Gly Gly Arg Arg Ile Cys Pro Gly Met Ser Phe
405
410
415
Gly Leu Ala Asn Leu Tyr Leu Pro Leu Ala Gin Leu Leu Tyr His Phe
420
425
430
Asp Trp Lys Leu Pro Thr Gly Ile Met Pro Arg Asp Leu Asp Leu Thr
445
450
455
460
Glu Leu Ser Gly Ile Thr Ile Ala Arg Lys Gly Asp Leu Tyr Leu Ann
465
470
Ala Thr Pro Tyr Gin Pro Ser Arg Glu
485
490
<210> SEQ ID NO 2
<211> LENGTH: 1660
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 2

```
ggatgtccta ataatctccc attatatcc aaaaatgcaa ttcttcagct tggtttccat
tttctcttct ctagctcc tatttttttt gaggaaatgg aagaactcca atagccaaag
caaaaattttg ccacccaggtct cagggaaaat accaatactaa ggaagtatgc ttcatatgat
ttggtggaag ccgcaccatg tccttagaga tttagccaaa aaatatggac cacttatgca
ccttcagtta ggtgaaatttt ctgcagttgt ggtacttctc caaggtgatgg caaaagaagt
cgttataac cagtcgagca ttgctcttag ccctcgtcgc gcaccatgga gacaaatgcg
taxattttt ggctcagagcc aaccagcttc aaccagcttc gctcgcagtg 480
tggatgagc gttggtttgct tctattgact ctctctctct tctggtcactttag 540
agttaaattt gccgcagagtct cttctgctttg tggcacttcc aacccaggtatgcat 600
tggggcaagtt ctaaagggg ccaccaggtatgcat 720
tggggtggct gattgctgatgct gatctgcttctagg 840
ttttattgtttt gtttatttggct gcacgtcgct gatctgcttctagg 960
ttttatttttt gattgcctgga atgtgctgta gatctgcttctagg 1080
ttttatttttt tattggctgc attggttcgct gatctgcttctagg 1200
ttttatttttt gattgcctgga attggttcgct gatctgcttctagg 1320
ttttatttttt tattggctgc attggttcgct gatctgcttctagg 1440
ttttatttttt tattggctgc attggttcgct gatctgcttctagg 1560
```
atattgaca ataatcattc ttctggtgtt atagcttta tcgatttcca atacatgtat
atattttag ggttaactgt tccatcttta aaaaaaaaaa

<210> SEQ ID NO 3
<211> LENGTH: 474
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 3

Met Glu Phe Phe Ser Leu Val Ser Ile Phe Leu Phe Leu Ser Phe Leu
1  5  10  15
Phe Leu Leu Arg Lys Cys Lys Asn Ser Asn Ser Glu Thr Lys Glu Leu
20 25 30
Pro Pro Gly Pro Trp Lys Ile Pro Ile Leu Gly Ser Met Leu His Met
35 40 45
Leu Gly Gly Glu Pro His His Ile Leu Arg Asp Leu Ala Lys Lys Tyr
50 55 60
Gly Pro Ile Met His Leu Glu Phe Gly Ile Ser Ala Val Val
65 70 75 80
Thr Ser Arg Glu Met Ala Lys Glu Val Leu Lys Thr His Asp Val Val
85 90 95
Phe Ala Ser Arg Pro Lys Ile Val Ala Met Asp Ile Ile Cys Tyr Asn
100 105 110
Gln Ser Asp Ile Ala Phe Ser Pro Tyr Gly Asp His Trp Arg Glu Met
115 120 125
Arg Lys Ile Cys Val Met Glu Leu Leu Asn Ala Lys Asn Val Arg Ser
130 135 140
Phe Ser Ser Ile Arg Arg Glu Val Arg Leu Ile Asp Ser Ile
145 150 155 160
Arg Ser Asp Ser Ser Ser Gly Leu Val Asn Phe Thr Glu Arg Ile
165 170 175
Ile Trp Phe Ala Ser Ser Ser Met Thr Cys Arg Ser Ala Phe Gly Glu Val
180 185 190
Leu Lys Gly Gln Asp Val Phe Ala Lys Lys Ile Arg Glu Val Ile Gly
195 200 205
Leu Ala Glu Gly Phe Asp Val Ala Asp Ile Phe Pro Ser Tyr Lys Phe
210 215 220
Leu His Val Leu Ser Gly Met Lys Arg Lys Leu Leu Asn Ala His Leu
225 230 235 240
Lys Val Asp Ala Ile Val Glu Asp Val Ile Asn Glu His Lys Asn
245 250 255
Leu Ala Thr Gly Lys Thr Asn Gly Ala Leu Gly Glu Asp Met Phe Ala
260 265 270
Ala Gly Thr Glu Thr Ser Ser Thr Thr Thr Val Try Ala Met Ala Glu
275 280 285
Met Met Lys Asn Pro Asn Val Phe Asn Lys Ala Glu Ala Glu Val Arg
290 295 300
Glu Thr Phe Lys Asp Lys Val Thr Phe Asp Glu Ile Asp Ala Glu Glu
305 310 315 320
Leu Glu Tyr Leu Leu Leu Val Ile Lys Glu Thr Thr Arg Leu His Pro
325 330 335
Pro Ser Pro Leu Leu Val Pro Arg Glu Cys Arg Glu Asp Thr Asp Ile
Asn Gly Tyr Thr Ile Pro Ala Lys Thr Lys Val Met Val Val Val Trp
355 360 365
Ala Leu Gly Arg Asp Pro Lys Tyr Trp Asp Asp Ala Glu Ser Phe Lys
370 375 380
Pro Glu Arg Phe Glu Gln Cys Ser Val Asp Phe Phe Gly Asn Asn Phe
385 390 395 400
Glu Phe Leu Pro Phe Gly Gly Gly Arg Arg Ile Cys Pro Gly Met Ser
405 410 415
Phe Gly Leu Ala Asn Leu Tyr Leu Pro Leu Ala Glu Leu Leu Leu Tyr His
420 425 430
Thr Glu Leu Ala Gly Ile Ile Ala Arg Lys Gly Asp Tyr Leu Tyr Leu
450 455 460
Met Ala Thr Pro Tyr Glu Pro Ser Arg Glu
465 470

<210> SEQ ID NO 4
<211> LENGTH: 1614
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 4

```
ggatggtcta ataatcctcc atttatctcc caaaatggaa ttcttcagct tggtttccat
     60
attctatcct atatcctcc tctttttgta aagaaatgt aagaaatcct atagccttaa
     120
caaacaagct cctctacggtc cagtctttgc acaatctctc cgaagatctgc tctcatatgct
     180
tggtggagaa ccacaccata tccttaggga tttagccaaa aaatatggac caattatgca
     240
cctcagcttt ggtgaatatt tctgaacttt cttctattc atgttgacct aagaggtctgc caaagaagct
     300
gctaaaaact cagcagctag ttttattctc tggctctaac attgggctcc cgggcatcctt
     360
cggattataac cagcagctag tgcctttagg gttgcttttt ctttcatagct gcgactttct gccttatctg
     420
ctacattatc cagcagctag tgcctttagt gttgcttttt ctttcatagct gcgactttct gccttatctg
     480
agctctcctg cgcctcctcc ctcctctctt ctcctctctt ctcctctctt ctcctctctt
     540
ctacattatc cagcagctag tgcctttagt gttgcttttt ctttcatagct gcgactttct gccttatctg
     600
ctacattatc cagcagctag tgcctttagt gttgcttttt ctttcatagct gcgactttct gccttatctg
     660
agctctcctg cgcctcctcc ctcctctctt ctcctctctt ctcctctctt ctcctctctt
     720
tggaatgaag cgtgtagcct gttggcctag gctgtagcct gttggcctag gctgtagcct
     780
tggaatgaag cgtgtagcct gttggcctag gctgtagcct gttggcctag gctgtagcct
     840
tggaatgaag cgtgtagcct gttggcctag gctgtagcct gttggcctag gctgtagcct
     900
tggaatgaag cgtgtagcct gttggcctag gctgtagcct gttggcctag gctgtagcct
     960
tggaatgaag cgtgtagcct gttggcctag gctgtagcct gttggcctag gctgtagcct
     1020
```

This sequence continues from the previous page.
attagctgga atacacattg ctgaaaaggg tgacctttac ttaatggtca ctccttatca 1440
acacctcgcga atgcaatcag gtttttaaat tccatgtgca acctcactat 1500
tgacaagct tctgatgtt tcaggttgg caatatgtag ttttataat 1560
atgtatcata ccagttcaga agagggacgata taatagtt gtaaaaaaaa aaaa 1614

<210> SEQ ID NO 5
<211> LENGTH: 537
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 5

Met Lys Asn Met Ala Lys Leu Leu Asn Lys Thr Ile Phe Cys Ile Leu
1  5 10 15
Phe Thr Ile Ala Phe Leu Ser Phe Ala Lys Leu Leu Ser Ser Tyr Leu
20 25 30
Ser Met Pro Phe Pro Leu Lys Tyr Met Ser Leu Ile Val Pro Leu Leu
35 40 45
Pro Leu Ile Asn Phe Leu Tyr Val Pro Gln Asn Asn Leu Pro
50 55 60
Pro Gly Pro Thr Ala Val Pro Ile Phe Gly Asn Trp Leu Gln Val Gly
65 70 75 80
Asn Asp Leu Asn His Gln Leu Ala Thr Met Ser Gln Thr Tyr Gly
85 90 95
Pro Ile Phe Leu Leu Lys Leu Gly Ser Lys Asn Leu Ala Val Val Ser
100 105 110
Asn Pro Glu Leu Ala Asp Gln Val Leu His Thr Gln Gly Val Glu Phe
115 120 125
Gly Ser Arg Pro Arg Asn Val Val Phe Asp Ile Phe Thr Gly Asn Gly
130 135 140
Gln Asp Met Val Phe Ile Tyr Gly Asp His Trp Arg Lys Met Arg
145 150 155 160
Arg Ile Met Thr Leu Pro Phe Phe Thr Asp Lys Val Val His Gln Tyr
165 170 175
Ser Asp Met Trp Glu Asn Asn Met Asp Leu Val Val Asp Leu Lys
180 185 190
Lys Asn Glu Lys Val Lys Tyr Gln Gly Ile Val Ile Arg Lys Arg Leu
195 200 205
Gln Leu Met Leu Tyr Asn Ile Met Tyr Arg Met Met Phe Asp Ala Lys
210 215 220
Phe Glu Ser Gln Asn Asp Pro Leu Phe Ile Glu Ala Thr Lys Phe Asn
225 230 235 240
Ser Glu Arg Ser Arg Leu Ala Glu Ser Phe Asp Tyr Asn Tyr Gln Asp
245 250 255
Phe Ile Pro Leu Leu Arg Phe Leu Arg Gly Tyr Leu Asn Lys Cys
260 265 270
Lys Asp Leu Gln Thr Arg Arg Leu Ala Phe Phe Asn Asp Tyr Phe Val
275 280 285
Glu Lys Arg Gly Ile Met Asp Glu Asn Gly Lys His Lys Ile
290 295 300
Ser Cys Ala Ile Asp His Ile Ile Asp Ala Glu Met Lys Gly Glu Ile
305 310 315
Asn Glu Gln Asn Val Leu Tyr Ile Val Gln Asn Ile Asn Val Ala Ala
325 330 335
<210> SEQ ID NO 6
<211> LENGTH: 1745
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<table>
<thead>
<tr>
<th>Ile Glu Thr Thr Leu Trp Ser Met Glu Trp Ala Ile Ala Glu Leu Val</th>
<th>340</th>
<th>345</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asn His Pro Ile Val Gln Gln Lys Ile Arg Asp Ile Ser Thr Val</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Leu Lys Gly Arg Ser Val Thr Glu Ser Asn Leu His Glu Leu Pro Tyr</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Leu Leu Ala Thr Val Asn Glu Thr Arg Leu His Thr Pro Ile Pro</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Leu Leu Val Pro His Met Asn Leu Glu Ala Lys Leu Gly Gly Tyr</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Thr Ile Pro Lys Glu Thr Lys Val Val Val Asn Ala Trp Trp Leu Ala</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Asn Asn Pro Ala Trp Trp Lys Asn Pro Glu Phe Arg Pro Glu Arg</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Phe Leu Glu Glu Asp Ser Ser Thr Glu Ala Ala Val Ala Gly Gly Lys</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Val Phe Arg Tyr Leu Pro Phe Gly Met Gly Arg Arg Ser Cys Pro</td>
<td>465</td>
<td>470</td>
<td>475</td>
</tr>
<tr>
<td>Gly Ile Ile Leu Ala Leu Pro Ile Leu Gly Leu Val Ile Ala Lys Leu</td>
<td>485</td>
<td>490</td>
<td>495</td>
</tr>
<tr>
<td>Val Ser Asn Phe Glu Met Gin Gly Pro Pro Gin Val Gly Lys Val Asp</td>
<td>500</td>
<td>505</td>
<td>510</td>
</tr>
<tr>
<td>Thr Ser Glu Arg Gly Gin Phe Ser Leu His Ile Ala Lys His Ser</td>
<td>515</td>
<td>520</td>
<td>525</td>
</tr>
<tr>
<td>Thr Val Val Phe Lys Pro Ile Ala Ala</td>
<td>530</td>
<td>535</td>
<td></td>
</tr>
</tbody>
</table>

<400> SEQUENCE:

```
cctctagcta atgaaaaaca tggccaaact tctcaacaag accatctttt gcattctctt
tacaattgca tttctttcat ttgccaagtt actgtcctcc tacctatcta tgcctttccc
tcttaagtac atgtcactta ttgtcccttt acttcccctt ataatcaact tcctctatgt
taagccccaa aacaacctcc cacctggtcc aacagcagtc ccaatatttg gtaattggct
tcaagttggc aatgacttga accatcaact ccttgccacc atgtcacaaa cctacggtcc
tatattttta cttaaacttg gttcaaaaaa cctagctgtg gtatatcaacc cagagcagc
tgaccaagtt ctacacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
cgacatatt actgttaatg gacaaagcat ggtgtccacc attatagttg accattggcg
taacaaagtt ctcacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgaccaagtt ctacacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgcagatagc tttgacatag ccaaaagca cctagttcag ttcagatgct ctttgacatag
taacaaagtt ctcacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgaccaagtt ctacacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgcagatagc tttgacatag ccaaaagca cctagttcag ttcagatgct ctttgacatag
taacaaagtt ctcacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgaccaagtt ctacacacac aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgcagatagc tttgacatag ccaaaagca cctagttcag ttcagatgct ctttgacatag
```
aggagaata astgagcgaas atgtactctta tatgtgtggag aatatacaatgt tgcagcaaat
1020
tgaaacact ctattgtcga tgcagatgggc ctaagctgaa cttgtaaatct atccccatgt
1080
tcaacagag ataggaggtg aatctctcaac agtcgtctcaa ggcagatcag tcaacagact
1140
aaaccttcat gactgtcctg acctcctacg aacagtaaat gcacatcataa gactcacaac
1200
accatatact tcactttgac cccatataag cttgaagag aagagagttg atggtagacac
1260
tattctaaaa gcacatcag tgggttggtgaa gtggtagtcg tgcagtcaaca aacccgtggt
1320
gtggaaaaac ccgaatgaat ggttcgctctttg gaggagagata gcacacacat
1380
agacagtgtgt gtcgggagca agagtagattt caggtactta ccattcgcgtt tggggagcg
1440
gagcgccccc gcacatcaca tcgcacgcgg aatcttcgggg cttgctcag gaaactctgt
1500
gtcaacctttt gaaagccagg gtcggcagag cttgtaaaga agtgataacag atggaaaggg
1560
aggccagttt aogtcgcaac ttgaaaaac ttccagcggt gttttcaagc atactgtgct
1620
atataatat tcctgagcta aatctatattt aadtatattt gttttcaagac aagagcaaac
1680
tactagttta ctogaaaaaatttctaatg ataactacgttttgttttaaaa aaaaaaa
1740
aaaa
1745

<210> SEQ ID NO 7
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 7
Met Ala Lys Leu Leu Asn Asn Thr Ile Phe Cys Ile Leu Phe Ser Ser Ile
1 5 10 15
Val Phe Leu Ser Phe Ala Lys Leu Ser Ser Tyr Leu Ser Ile Pro
20 25 30
Phe Pro Leu Glu Tyr Ile Ser Leu Ile Val Leu Leu Pro Leu Ile
35 40 45
Ile Asn Phe Leu Cys Val Lys Pro Gin Asn Asn Leu Pro Pro Gly Pro
50 55 60
Thr Ala Val Pro Ile Phe Gly Asn Trp Leu Gin Val Gly Asn Asp Leu
65 70 75 80
Asn His Gin Leu Leu Ala Thr Met Ser Gin Thr Tyr Gly Pro Ile Phe
85 90 95
Leu Leu Lys Leu Gly Ser Lys Asn Leu Ala Val Ser Asn Pro Glu
100 105 110
Leu Ala Asn Gin Val Leu His Thr Gin Gly Val Glu Phe Gly Ser Arg
115 120 125
Pro Arg Asn Val Val Phe Asp Ile Phe Thr Gly Gin Asn Gin Asp Met
130 135 140
Val Phe Thr Ile Tyr Gly Asp His Trp Arg Lys Met Arg Arg Ile Met
145 150 155 160
Thr Leu Pro Phe Phe Thr Asn Lys Val Val His Gin Tyr Ser Asp Met
165 170 175
Trp Glu Asn Gin Met Asp Leu Val Val Asp Asp Lys Asn Gin Leu
180 185 190
Lys Val Lys Tyr Asp Gly Ile Val Ile Arg Lys Arg Leu Gin Leu Met
195 200 205
Pro Leu Tyr Asn Ile Met Tyr Arg Met Met Phe Asp Ala Lys Phe Glu Ser
210 215 220
Gln Asp Asp Pro Leu Phe Ile Glu Ala Thr Lys Phe Asn Ser Glu Arg
Ser Arg Leu Ala Gln Ser Phe Asp Tyr Asn Tyr Gly Asp Phe Ile Pro
Leu Leu Arg Pro Phe Leu Lys Gly Tyr Leu Asn Lys Cys Lys Asp Leu
Gln Thr Arg Arg Leu Ala Phe Phe Asn Tyr Phe Val Gly Lys Arg
Arg Lys Ile Met Gly Glu Asn Gly Glu Lys His Lys Ile Cys Cys Ala
Ile Asp His Ile Ile Asp Ala Glu Met Lys Gly Glu Ile Ser Glu Gln
Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Ala Ala Ile Glu Thr
Thr Leu Trp Ser Met Glu Trp Ala Ile Ala Glu Leu Val Asn His Pro
Ile Val Gln Gln Lys Ile Arg Asp Glu Ile Ser Thr Val Leu Lys Gly
Lys Ser Val Lys Glu Ser Asn Leu His Glu Leu Pro Tyr Leu Leu Ala
Thr Val Asn Glu Thr Leu Arg Leu His Thr Pro Ile Pro Leu Leu Val
Pro His Met Asn Leu Glu Ala Lys Leu Gly Gly Tyr Thr Ile Pro
Lys Glu Thr Lys Val Val Asn Ala Trp Trp Leu Ala Asn Asn Pro
Ala Trp Trp Lys Asn Gln Asn Glu Phe Arg Pro Glu Arg Phe Leu Glu
Glu Asp Ser Ser Thr Glu Ala Val Ala Gly Lys Val Asp Phe
Arg Tyr Leu Pro Phe Gly Met Gly Arg Arg Ser Cys Pro Gly Ile Ile
Leu Ala Leu Pro Ile Leu Gly Leu Val Ile Ala Lys Leu Val Ser Asn
Phe Glu Met Gln Ala Pro Pro Gly Val Gly Lys Val Asp Thr Ser Glu
Lys Gly Gly Lys Phe Ser Leu His Ile Ala His Ser Thr Val Val
Phe Lys Pro Ile Ala Ala

<210> SEQ ID NO 8
<211> LENGTH: 1693
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 8
cctctagct aattaaaaaa tacgcaasct ttctcaascac acacatcttt ttgatccttt 60
ttcasattgta ttcttcttcg tggcaasatt actatactcc tacctctota taccttctcc 120
tctttgatc atttcaactc taatcctttt acttccccctta atataacact ttctctgtcttg 180
taagccccaa aacaacctcc cacctggtcc tcaagttggc aatgacttga 240
actatcctcc tacctctcta acttccccta ataatcaact 300
aacagcagtc ccaaatttttg 360
tctcaacaac accatctttt 420
tactatcctcc tacctctcta 480
gtaattggct cctatggtcc 540
tctgagctagc gcattctcttt 600
tacctttcccc tcctctgtgtg 660
gtaattggct cctatggtcc 720
tctgagctagc gcattctcttt 780
tacctttcccc tcctctgtgtg 840
gtaattggct cctatggtcc 900
tctgagctagc gcattctcttt 960
tacctttcccc tcctctgtgtg 1020
gtaattggct cctatggtcc 1080
tctgagctagc gcattctcttt 1140
tacctttcccc tcctctgtgtg 1200
gtaattggct cctatggtcc 1260
tctgagctagc gcattctcttt 1320
tacctttcccc tcctctgtgtg 1380
gtaattggct cctatggtcc 1440
tctgagctagc gcattctcttt 1500
tacctttcccc tcctctgtgtg 1560
gtaattggct cctatggtcc 1620
tctgagctagc gcattctcttt 1680

<table>
<thead>
<tr>
<th>Met</th>
<th>Tyr</th>
<th>His</th>
<th>Leu</th>
<th>Leu</th>
<th>Ser</th>
<th>Pro</th>
<th>Ile</th>
<th>Glu</th>
<th>Ala</th>
<th>Ile</th>
<th>Val</th>
<th>Gly</th>
<th>Leu</th>
<th>Val</th>
<th>Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
<td>Leu</td>
<td>Tyr</td>
<td>Leu</td>
<td>Trp</td>
<td>Thr</td>
<td>Lys</td>
<td>Lys</td>
<td>Gin</td>
<td>Ser</td>
<td>Lys</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Aan</td>
<td>Pro</td>
<td>Leu</td>
<td>Pro</td>
<td>Pro</td>
<td>Lys</td>
<td>Ile</td>
<td>Pro</td>
<td>Gly</td>
<td>Gly</td>
<td>Tsp</td>
<td>Pro</td>
<td>Val</td>
<td>Ile</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Leu</td>
<td>Phe</td>
<td>Tyr</td>
<td>Phe</td>
<td>Aan</td>
<td>Aan</td>
<td>Aan</td>
<td>Gly</td>
<td>Asp</td>
<td>Asp</td>
<td>Asp</td>
<td>Arg</td>
<td>Arg</td>
<td>His</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gin</td>
<td>Lys</td>
<td>Leu</td>
<td>Gly</td>
<td>Asp</td>
<td>Leu</td>
<td>Ala</td>
<td>Asp</td>
<td>Asp</td>
<td>Lys</td>
<td>Tyr</td>
<td>Gly</td>
<td>Pro</td>
<td>Val</td>
<td>Phe</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Leu</td>
<td>Gly</td>
<td>Phe</td>
<td>Arg</td>
<td>Arg</td>
<td>Phe</td>
<td>Leu</td>
<td>Ala</td>
<td>Val</td>
<td>Ser</td>
<td>Tyr</td>
<td>Glu</td>
<td>Ala</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>95</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Glu</td>
<td>Cys</td>
<td>Phe</td>
<td>Ser</td>
<td>Thr</td>
<td>Aen</td>
<td>Aen</td>
<td>Ile</td>
<td>His</td>
<td>Phe</td>
<td>Aan</td>
<td>Asp</td>
<td>Arg</td>
<td>Pro</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Tyr</td>
<td>Gly</td>
<td>Glu</td>
<td>Tyr</td>
<td>Leu</td>
<td>Cys</td>
<td>Tyr</td>
<td>Aen</td>
<td>Aen</td>
<td>Ala</td>
<td>Met</td>
<td>Leu</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Tyr</td>
<td>Gly</td>
<td>Pro</td>
<td>Tyr</td>
<td>Trp</td>
<td>Lys</td>
<td>Lys</td>
<td>Aen</td>
<td>Arg</td>
<td>Lys</td>
<td>Leu</td>
<td>Val</td>
<td>Aen</td>
<td>Gin</td>
</tr>
</tbody>
</table>

SEQ ID NO 9
LENGTH: 519
TYPE: PRT
ORGANISM: Nicotiana tabacum
SEQUENCE: 9

```
taaccagtt ctacacaocg aaggggtcga gtttgggtcc cgtccacgta acgttgtctt
tgatatatt actgtaatgc gagaagaatc cttcactaat tttcactaac aaagtggtgc
aattatatgt gagaatgaga tggacctagt tgttgatgac ttgaagaaga ataaggtata
aatatgatgg gttttgtgc caattctatg ctagatattg gaaattgacg gggggggtac
ctttgagcgt tggcagcgtt ctactgttatt tctcactaaattttttttttttttaaa
```
<table>
<thead>
<tr>
<th>Glu</th>
<th>Leu</th>
<th>Leu</th>
<th>Ser</th>
<th>Val</th>
<th>Ser</th>
<th>Arg</th>
<th>Glu</th>
<th>Lys</th>
<th>Phe</th>
<th>Lys</th>
<th>His</th>
<th>Val</th>
<th>Arg</th>
<th>Phe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Continued...**

**SEQ ID NO 10**

**LENGTH:** 1578

**TYPE:** DNA

**ORGANISM:** *Nicotiana tabacum*
<400> SEQUENCE: 10

atgtatcatc tctttttccc catagaagcc attgtgaggc tttgtaacct ttgcatctca
ctttaacctgc tattgaaaaa aaaaatctaa aacatctgcc tccaaaaatc
ccaggtggtg gcaggcaagt ttggtgaggc ttttatttca acaatggaaga atggatgatc
ttcgccctgg cggtggtttg cttggcggtg agtagttatg aagctatgaa agaatgcttc
tataacatct cctgtttgcc tcggcgccaa atttgcgcct acttgaaaaa aaatggaaga
tggtcggctc tccgctggtc acctccatttgc agagctggag aatgcccgct taccctcttg
tataataatt ttacaagtct taaagagctt tcttgaaat gtagttgtgg
aagacgatgat tttgcaagac tttccagaag ttttccagaag gttgctgtgt
aatgtgatgtgc aatggtttct cttgccgctt cttggcggtg agtagttatg aagctatgaa agaatgcttc
tataacatct cctgtttgcc tcggcgccaa atttgcgcct acttgaaaaa aaatggaaga
tggtcggctc tccgctggtc acctccatttgc agagctggag aatgcccgct taccctcttg
<210> SEQ ID NO 11
<211> LENGTH: 509
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 11

Met Glu Gly Thr Asn Leu Thr Thr Tyr Ala Ala Val Phe Leu Gly Thr
1 5 10 15
Leu Phe Leu Leu Leu Leu Ser Lys Phe Leu Arg Gln Arg Lys Leu Arg
20 25 30
Leu Pro Pro Gly Pro Lys Pro Trp Pro Ile Ile Gly Asn Leu Asn Leu
35 40 45
Ile Gly Asn Leu Pro His Arg Ser Ile His Glu Leu Ser Leu Lys Tyr
50 55 60
Gly Pro Ile Met Gln Leu Gln Phe Gly Thr Phe Pro Val Val Val Gly
65 70 75 80
Ser Ser Val Glu Met Ala Lys Val Phe Leu Lys Ser Met Asp Ile Asn
  85  90  95
Phe Val Gly Arg Pro Lys Thr Ala Ala Gly Lys Tyr Thr Thr Tyr Asn
  100  105  110
Tyr Ser Asp Ile Thr Trp Ser Pro Tyr Gly Pro Tyr Trp Arg Glu Ala
  115  120  125
Arg Arg Met Cys Leu Met Glu Leu Phe Ser Thr Lys Arg Leu Asp Ser
  130  135  140
Tyr Glu Tyr Ile Arg Ala Glu Leu His Ser Leu Leu His Asn Leu
  145  150  155  160
Asn Lys Ile Ser Gly Lys Pro Ile Val Leu Lys Asp Tyr Leu Thr Thr
  165  170  175
Leu Ser Leu Asn Val Ile Ser Arg Met Val Leu Gly Lys Arg Tyr Leu
  180  185  190
Asp Glu Ser Glu Asn Ser Ile Val Thr Pro Glu Glu Phe Lys Lys Met
  195  200  205
Leu Asp Glu Leu Phe Leu Leu Asn Gly Val Leu Asn Ile Gly Asp Ser
  210  215  220
Ile Pro Trp Ile Asp Phe Met Asp Leu Gln Gly Tyr Val Lys Arg Met
  225  230  235  240
Lys Phe Val Ser Lys Phe Phe Lys Phe Leu Glu His Val Ile Asp
  245  250  255
Glu His Asn Val Arg Arg Asn Gly Val Glu Asn Tyr Ile Ala Lys Asp
  260  265  270
Met Val Asp Val Leu Leu Gln Leu Ala Asp Asp Pro Thr Leu Glu Val
  275  280  285
Lys Leu Glu Arg His Gly Val Lys Ala Phe Thr Glu Asp Met Leu Ala
  290  295  300
Gly Gly Thr Glu Ser Ser Ala Val Thr Val Glu Trp Ala Ile Ser Glu
  305  310  315  320
Leu Leu Lys Lys Pro Glu Ile Phe Lys Lys Ala Thr Glu Leu Asp
  325  330  335
Arg Val Ile Gly Glu Asn Arg Thr Val Glu Glu Lys Asp Ile Pro Asn
  340  345  350
Leu Pro Tyr Ile Glu Ala Ile Val Lys Glu Thr Met Arg Leu His Pro
  355  360  365
Val Ala Pro Met Leu Val Val Pro Arg Glu Cys Arg Glu Asp Cys Lys Val
  370  375  380
 Ala Gly Tyr Asp Val Lys Gly Thr Arg Val Leu Val Ser Val Trp
  385  390  395  400
Thr Ile Gly Arg Asp Pro Thr Leu Trp Asp Glu Pro Glu Ala Phe Lys
  405  410  415
Pro Glu Arg Phe His Glu Tyr Ser Ile Asp Val Lys Gly His Asp Phe
  420  425  430
Glu Leu Leu Pro Phe Gly Ala Gly Arg Arg Met Cys Pro Gly Tyr Asn
  435  440  445
Leu Gly Leu Lys Val Ile Gin Ala Ser Leu Ala Asn Leu Ile His Gly
  450  455  460
Phe Asn Trp Ser Leu Pro Asp Asn Met Thr Pro Glu Asp Leu Asp Met
  465  470  475  480
Asp Glu Ile Phe Gly Leu Ser Thr Pro Lys Lys Phe Pro Leu Ala Thr
  485  490  495
Val Ile Glu Pro Arg Leu Ser Pro Lyn Leu Tyr Ser Val
500 505
<210> SEQ ID NO 12
<211> LENGTH: 1530
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 12

atggaaggta ccaaacttgac tacatatgca gcagtttctc ttggtactct ttcctttcag
50 100
ctctttccca aatttttctg ccataaaggg aatacttttta aattttttctt gcggggggtgc
100 150
ccttcatcgg gaactctaa cccattcggc atcactttct aatgctcaaat ccaacagtct
150 200
ttattctttcctt ggtttcttttgc aagaagggggg ttgaagagtttt cagaaaggggt
200 250
tgcagcagcgc taatgcttttgc ccatttttttt tttggggggttc cactattttgc
250 300
tttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt
300 350
ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt
350 400
<210> SEQ ID NO 13
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence

<400> SEQUENCE: 13

ggcggagaat ttgtcctgga atgtcatttg gtttag
10 15
<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA

<400> SEQUENCE: 14

ggcggagaat ttgtcctgga atgtcatttg gtttag
gtcaaatagt ggggtgaca atg

seq id no 15
length: 18
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 15

number of characters: 18
number of bases: 18

ttatgcagca ataggcttga agaca

seq id no 16
length: 25
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 16

number of characters: 25
number of bases: 25

ggggatatca tgggtgtgtg aag

seq id no 17
length: 33
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 17

number of characters: 33
number of bases: 33

cccggatcca tgtatcatct tctttctccc

seq id no 18
length: 33
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 18

number of characters: 33
number of bases: 33

ggggatatct tactctcgag aaggttgata agg

seq id no 19
length: 30
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 19

number of characters: 30
number of bases: 30

cccggatcca tgtatcatct tctttctccc

seq id no 20
length: 33
type: dna
organism: artificial sequence
feature:
other information: derived from nicotiana tabacum p450 gene

sequence: 20

number of characters: 33
number of bases: 33
<210> SEQ ID NO 21
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Derived from Nicotiana tabacum p450 gene
<400> SEQUENCE: 21
cccgatccattgcaatcctt cagcttggtt tee
33

<210> SEQ ID NO 22
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Derived from Nicotiana tabacum p450 gene
<400> SEQUENCE: 22
ggggagctct cactcgcaag aagattgata agg
33

<210> SEQ ID NO 23
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Derived from Nicotiana tabacum p450 gene
<400> SEQUENCE: 23
gcattatcg gogcaacta aatctccaa aatgoggtta aaaaatctca aatocccacttc 60
gtccgaaccag cgtc 75

<210> SEQ ID NO 24
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Derived from Nicotiana tabacum p450 gene
<400> SEQUENCE: 24
ggggagctct tatgcagcaa taggcttgaa gac
33

<210> SEQ ID NO 25
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Derived from Nicotiana tabacum p450 gene
<400> SEQUENCE: 25
gggagctcttatgcagcaataggcttggaagac
33

<210> SEQ ID NO 26
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<223> OTHER INFORMATION: Nicotiana tabacum p450 protein
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (7) .. (7)
<223> OTHER INFORMATION: Xaa = any amino acid
<400> SEQUENCE: 26
any amino acid
Lys Glu Thr Leu Arg Leu Xaa
1 5

<210> SEQ ID NO 27
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<222> OTHER INFORMATION: Nicotiana tabacum p450 protein
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (4) .. (4)
<223> OTHER INFORMATION: Xaa = any amino acid
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (8) .. (8)
<223> OTHER INFORMATION: Xaa = any amino acid

<400> SEQUENCE: 27

Pro Phe Gly Xaa Gly Arg Arg Xaa Cys Pro Ala
1 5 10

<210> SEQ ID NO 28
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<222> OTHER INFORMATION: Nicotiana tabacum p450 protein
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (4) .. (4)
<223> OTHER INFORMATION: Xaa = any amino acid
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (8) .. (8)
<223> OTHER INFORMATION: Xaa = any amino acid

<400> SEQUENCE: 28

Pro Phe Gly Xaa Gly Arg Arg Xaa Cys Pro Gly
1 5 10

<210> SEQ ID NO 29
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum
<220> FEATURE:
<222> OTHER INFORMATION: Nicotiana tabacum p450 protein
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: (2) .. (2)
<223> OTHER INFORMATION: Xaa = any amino acid

<400> SEQUENCE: 29

Phe Xaa Pro Glu Arg Phe
1 5

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<222> OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 30

Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
1 5 10 15

Thr Ile Cys Ala
20
SEQ ID NO 31
LENGTH: 20
TYPE: PRT
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein
SEQUENCE:
Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
1  5  10  15
Thr Ile Thr Ala
20

SEQ ID NO 32
LENGTH: 20
TYPE: PRT
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein
SEQUENCE:
Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
1  5  10  15
Thr Ile Met Gly
20

SEQ ID NO 33
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein
SEQUENCE:
Thr Thr Tyr Ile Ile Ile Cys Cys Ile Gly Ala Arg Met Gly Ile Thr
1  5  10  15
Thr Tyr

SEQ ID NO 34
LENGTH: 18
TYPE: PRT
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein
SEQUENCE:
Arg Ala Ala Ile Cys Lys Tyr Thr Cys Ile Gly Gly Ile Ile Ile Arg
1  5  10  15
Ala Ala

SEQ ID NO 35
LENGTH: 20
TYPE: PRT
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from Nicotiana tabacum p450 protein
SEQUENCE:
Gly Gly Ile Met Gly Ile Met Gly Ile Ile Ile Thr Gly Tyr Cys
1  5  10  15
Cys Ile Gly Ser
20

SEQ ID NO 36
LENGTH, 21
TYPE: DNA
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from T7 bacteriophage promoter
SEQUENCE: 37

```
gtaatacgac tcactatagg g
```

LENGTH, 21
TYPE: DNA
ORGANISM: Artificial sequence
OTHER INFORMATION: Derived from T3 bacteriophage promoter
SEQUENCE: 38

```
caattaaccc tcactaaagg g
```

LENGTH, 500
TYPE: PRT
ORGANISM: Mentha piperita
SEQUENCE: 39

```
Met Glu Leu Gln Ile Ser Ser Ala Ile Ile Ile Leu Val Thr Tyr
      1 5 10 15
Thr Ile Ser Leu Leu Ile Lys Gln Trp Arg Lys Pro Lys Pro Gln
    20 25 30
Glu Asn Leu Pro Pro Gly Pro Pro Lys Leu Pro Leu Ile Gly His Leu
    35 40 45
His Leu Leu Trp Gly Lys Leu Pro Gln His Ala Leu Ala Ser Val Ala
   50 55 60
Lys Gln Tyr Gly Pro Val Ala His Val Gln Leu Gly Glu Val Phe Ser
   65 70 75 80
Val Val Leu Ser Ser Arg Glu Ala Thr Lys Glu Ala Met Lys Leu Val
   85 90 95
Asp Pro Ala Cys Ala Asp Arg Phe Glu Ser Ile Gly Thr Lys Ile Met
  100 105 110
Trp Tyr Asp Asn Asp Ile Ile Phe Ser Pro Tyr Ser Val His Trp
  115 120 125
Arg Gln Met Arg Lys Ile Cys Val Ser Glu Leu Ser Ala Arg Asn
  130 135 140
Val Arg Ser Phe Gly Phe Ile Arg Gin Asp Glu Val Ser Arg Leu Leu
  145 150 155 160
Gly His Leu Arg Ser Ser Ala Ala Ala Gly Glu Ala Val Asp Leu Thr
  165 170 175
Glu Arg Ile Ala Thr Leu Thr Cys Ser Ile Ile Cys Arg Ala Ala Phe 180 185 190
Gly Ser Val Ile Arg Asp His Glu Leu Val Glu Leu Val Lys Asp 195 200 205
Ala Leu Ser Met Ala Ser Gly Phe Glu Leu Ala Asp Met Phe Pro Ser 210 215 220
Ser Lys Leu Leu Asn Leu Leu Cys Trp Ann Lys Ser Lys Leu Trp Arg 225 230 235 240
Met Arg Arg Arg Val Asp Ala Ile Leu Glu Ala Ile Val Glu Glu His 245 250 255
Lys Leu Lys Lys Ser Gly Glu Phe Gly Gly Glu Asp Ile Ile Asp Val 260 265 270
Leu Phe Arg Met Gln Lys Asp Ser Gln Ile Lys Val Pro Ile Thr Thr 275 280 285
Ann Ala Ile Lys Ala Phe Ile Phe Asp Thr Phe Ser Ala Gly Thr Glu 290 295 300
Thr Ser Ser Thr Thr Thr Thr Leu Trp Val Met Ala Glu Leu Met Arg Ann 305 310 315 320
Pro Glu Val Met Ala Lys Ala Gln Ala Glu Val Arg Ala Ala Leu Lys 325 330 335
Gly Lys Thr Asp Trp Asp Asp Asp Val Gin Glu Leu Lys Tyr Met 340 345 350
Lys Ser Val Val Lys Glu Thr Met Arg Met His Pro Pro Ile Pro Leu 355 360 365
Ile Pro Arg Ser Cys Arg Glu Cys Glu Val Ann Gly Tyr Thr Ile 370 375
Pro Ann Lys Ala Arg Ile Met Ile Ann Val Trp Ser Met Gly Arg Ann 385 390 395 400
Pro Leu Tyr Trp Glu Lys Pro Glu Thr Phe Trp Pro Glu Arg Phe Asp 405 410 415
Gln Val Ser Arg Asp Phe Met Gly Ann Asp Phe Glu Phe Ile Pro Phe 420 425 430
Val Glu Val Pro Leu Ala Gln Leu Leu Tyr His Phe Asp Trp Lys Leu 450 455 460
Ala Glu Gly Met Ann Pro Ser Arg Met Asp Met Ser Glu Ala Glu Gly 465 470 475 480
Leu Thr Gly Ile Arg Lys Ann Ann Leu Leu Val Val Thr Pro Tyr 485 490 495
Asp Pro Ser Ser 500

<210> SEQ ID NO 40
<211> LENGTH: 496
<212> TYPE: PRT
<213> ORGANISM: Mentha spicata

<400> QUENTENCE: 40
Met Glu Leu Asp Leu Leu Ser Ala Ile Ile Leu Val Ala Thr Tyr 1 5 10 15
Ile Val Ser Leu Ile Ile Gln Trp Arg Lys Ser Lys Ser Gln Gln 20 25 30
Ann Leu Pro Pro Ser Pro Pro Lys Leu Pro Val Ile Gly His Leu His 35 40 45
Phe Leu Trp Gly Gly Leu Pro Gln His Val Phe Arg Ser Ile Ala Gln
50 55 60
Lys Tyr Gly Pro Val Ala His Val Gln Leu Gly Gln Tyr Ser Val
60 70 75 80
Val Leu Ser Ser Ala Glu Ala Ala Lys Gln Ala Gln Met Lys Val Leu Asp
85 90 95 100
Pro Asn Phe Ala Asp Arg Phe Asp Gly Ile Gly Ser Arg Thr Met Trp
100 105 110
Tyr Asp Lys Asp Ile Ile Phe Ser Pro Tyr Asn Asp His Trp Arg
110 115 120 125
Gln Met Arg Arg Ile Cys Val Thr Leu Leu Ser Ser Pro Tyr Asn Val
125 130 135 140
Arg Ser Phe Gly Tyr Ile Arg Glu Glu Ile Glu Arg Leu Ile Arg
140 145 150 155 160
Leu Leu Gly Ser Ser Gly Gly Ala Pro Val Asp Val Thr Glu Glu Val
160 165 170 175
Ser Lys Met Ser Cys Val Val Cys Arg Ala Ala Phe Gly Ser Val
175 180 185 190
Leu Lys Asp Gln Gly Ser Leu Ala Glu Leu Val Lys Glu Ser Leu Ala
190 195 200 205
Leu Ala Ser Gly Phe Glu Leu Ala Asp Leu Tyr Pro Ser Ser Trp Leu
205 210 215 220
Leu Asn Leu Leu Ser Leu Asn Tyr Arg Leu Gln Arg Met Arg Arg
220 225 230 235 240
Arg Leu Asp His Ile Leu Asp Gly Phe Leu Glu Glu His Arg Glu Lys
240 245 250 255
Lys Ser Gly Glu Phe Gly Glu Asp Ile Val Asp Val Leu Phe Arg
255 260 265 270
Met Gln Lys Gly Ser Asp Ile Lys Ile Pro Ile Thr Ser Asn Cys Ile
270 275 280 285
Lys Gly Phe Ile Phe Asp Thr Phe Ser Ala Gly Ala Glu Thr Ser Ser
285 290 295 300
Thr Thr Ile Ser Trp Ala Leu Ser Glu Leu Met Arg Asn Pro Ala Lys
300 305 310 315 320
Met Ala Lys Val Gln Ala Glu Val Arg Glu Ala Leu Lys Gln Lys Thr
320 325 330 335
Val Val Asp Leu Ser Glu Val Gin Glu Leu Lys Tyr Leu Arg Ser Val
335 340 345 350
Leu Lys Glu Thr Leu Arg Leu His Pro Pro Phe Pro Leu Ile Pro Arg
350 355 360 365
Gln Ser Arg Glu Cys Val Gin Asp Gly Thr Ile Pro Ala Lys
365 370 375 380
Thr Arg Ile Phe Ile Asn Val Trp Ala Ile Gly Arg Asp Pro Gin Tyr
380 385 390 395 400
Trp Glu Asp Pro Asp Thr Phe Arg Pro Glu Arg Phe Asp Glu Val Ser
400 405 410 415
Arg Asp Phe Met Gln Asn Asp Phe Glu Phe Ile Pro Phe Gly Ala Gly
415 420 425 430
Arg Arg Ile Cys Pro Gly Leu His Phe Gly Leu Ala Asn Val Glu Ile
430 435 440 445
Pro Leu Ala Glu Leu Leu Tyr His Phe Asp Trp Lys Leu Leu Pro Gin Gly
445 450 455 460
<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Thr Asp Ala Asp Leu Asp Met Thr Glu Thr Pro Gly Leu Ser Gly</td>
</tr>
<tr>
<td>465</td>
</tr>
<tr>
<td>Pro Lys Lys Lys Asn Val Cys Leu Val Pro Thr Leu Tyr Lys Ser Pro</td>
</tr>
<tr>
<td>485</td>
</tr>
</tbody>
</table>

**<210> SEQ ID NO 41**

**<211> LENGTH, 509**

**<212> TYPE: PRT**

**<213> ORGANISM: Nepeta racemosa**

**<400> SEQUENCE: 41**

<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Val Ser Leu Ser Tyr Phe Leu Ile Ala Leu Leu Cys Thr Leu Pro</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Phe Leu Leu Phe Leu Asn Lys Trp Arg Arg Ser Tyr Ser Gly Lys Thr</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>Pro Pro Pro Ser Pro Pro Leu Pro Val Ile Gly Asn Leu His Gln</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>Leu Gly Leu Tyr Pro His Arg Tyr Leu Gin Ser Leu Ser Arg Arg Tyr</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>Gly Pro Leu Met Gin Leu His Phe Gly Ser Val Pro Val Leu Val Ala</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>Ser Ser Pro Glu Ala Ala Arg Glu Ile Met Lys Asn Gln Asp Ile Val</td>
</tr>
<tr>
<td>85</td>
</tr>
<tr>
<td>Phe Ser Asn Arg Pro Lys Met Ser Ile Ala Asn Arg Leu Phe Phe Asn</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>Asn Arg Asp Val Ala Phe Thr Gin Tyr Gly Tyr Tryp Arg Gin Ile</td>
</tr>
<tr>
<td>115</td>
</tr>
<tr>
<td>Arg Ser Ile Cys Val Leu Gin Leu Leu Ser Asn Lys Leu Arg Val Gin Ser</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>Phe Arg Arg Val Arg Glu Glu Glu Thr Ser Ile Met Val Glu Lys Ile</td>
</tr>
<tr>
<td>145</td>
</tr>
<tr>
<td>Met Gin Leu Gly Ser Ser Ser Thr Pro Val Asn Leu Ser Glu Leu</td>
</tr>
<tr>
<td>165</td>
</tr>
<tr>
<td>Leu Leu Ser Leu Thr Asn Asp Val Val Cys Arg Val Thr Leu Gly Lys</td>
</tr>
<tr>
<td>180</td>
</tr>
<tr>
<td>Lys Tyr Gly Gly Asn Gly Ser Glu Glu Val Asp Lys Leu Lys Glu</td>
</tr>
<tr>
<td>195</td>
</tr>
<tr>
<td>Met Leu Thr Glu Ile Gin Asn Leu Met Gly Ile Ser Pro Val Trp Glu</td>
</tr>
<tr>
<td>210</td>
</tr>
<tr>
<td>Phe Ile Pro Trp Leu Asn Trp Thr Arg Arg Phe Asp Gly Val Asp Gin</td>
</tr>
<tr>
<td>225</td>
</tr>
<tr>
<td>Arg Val Asp Arg Ile Val Lys Ala Phe Asp Gly Phe Leu Gin Ser Val</td>
</tr>
<tr>
<td>245</td>
</tr>
<tr>
<td>Ile Gin Glu His Lys Gin Arg Asp Gly Gin Lys Arg Gin Asp Gin Gin</td>
</tr>
<tr>
<td>260</td>
</tr>
<tr>
<td>Gly Ala Leu Asp Phe Val Asp Ile Leu Leu Gin Phe Gin Arg Gin Glu</td>
</tr>
<tr>
<td>275</td>
</tr>
<tr>
<td>Lys Gin Arg Ser Pro Val Gin Gin Arg Thr Val Gin Gin Arg Gin Gin</td>
</tr>
<tr>
<td>290</td>
</tr>
<tr>
<td>Asp Met Phe Val Ala Gly Thr Asp Thr Thr Thr Thr Ala Leu Gin Trp</td>
</tr>
<tr>
<td>305</td>
</tr>
<tr>
<td>Ala Val Ala Gin Leu Ile Lys Gin Pro Gin Gin Met Gin Gin Gin Gin</td>
</tr>
<tr>
<td>325</td>
</tr>
<tr>
<td>Asn Gin Val Arg Gin Val Val Gin Gin Gin Gin Gin Gin Gin Gin</td>
</tr>
<tr>
<td>345</td>
</tr>
</tbody>
</table>
Asp Leu Glu Lys Met Pro Tyr Leu Lys Ala Ser Ile Lys Gly Ser Leu
355 360 365
Arg Leu His Val Pro Val Val Leu Val Pro Arg Glu Ser Thr Arg
370 375 380
Asp Thr Asn Val Leu Gly Tyr Asp Ile Ala Ser Gly Thr Arg Val Leu
385 390 395 400
Ile Asn Ala Trp Ala Ile Ala Arg Pro Ser Val Trp Gly Asn Pro
405 410 415
Glu Glu Phe Leu Pro Glu Arg Phe Leu Asp Ser Ser Ile Asp Tyr Lys
420 425 430
Gly Leu His Phe Glu Leu Leu Ser Phe Gly Ala Gly Arg Arg Gly Cys
435 440 445
Pro Gly Ala Thr Phe Ala Val Ala Ile Asp Glu Leu Ala Leu Lys
450 455 460
Leu Val His Lys Phe Asp Phe Gly Leu Pro Asn Gly Ala Arg Met Glu
465 470 475 480
Glu Leu Asp Met Ser Glu Thr Ser Gly Met Thr Val His Lys Lys Ser
485 490 495
Pro Leu Leu Leu Leu Pro Ile Pro His His Ala Ala Pro
500 505

<210> SEQ ID NO 42
<211> LENGTH: 471
<212> TYPE: PRT
<213> ORGANISM: Persea americana

<400> SEQUENCE: 42
Met Ala Ile Leu Val Ser Leu Leu Phe Leu Ala Ile Ala Leu Thr Phe
1 5 10 15
Phe Leu Leu Lys Leu Asn Glu Lys Arg Glu Lys Lys Pro Asn Leu Pro
20 25 30
Pro Ser Pro Pro Asn Leu Pro Ile Ile Gly Asn Leu His Gln Leu Gly
35 40 45
Asn Leu Pro His Arg Ser Leu Arg Ser Leu Ala Asn Glu Leu Gly Pro
50 55 60
Leu Ile Leu Leu His Leu Gly His Ile Pro Thr Leu Ile Val Ser Thr
65 70 75 80
Ala Glu Ile Ala Glu Ile Leu Lys Thr His Asp Leu Ile Phe Ala
85 90 95
Ser Arg Pro Ser Thr Ala Ala Arg Arg Ile Phe Tyr Asp Cys Thr
100 105 110
Asp Val Ala Phe Ser Pro Tyr Gly Glu Tyr Trp Arg Glu Val Arg Lys
115 120 125
Ile Cys Val Leu Glu Leu Ser Ile Lys Arg Val Asn Ser Tyr Arg
130 135 140
Ser Ile Arg Glu Glu Val Gly Leu Met Met Glu Arg Ile Ser Gin
145 150 155 160
Ser Cys Ser Thr Gly Glu Ala Val Asn Leu Ser Glu Leu Leu Leu
165 170 175
Leu Ser Ser Gly Thr Ile Thr Arg Val Ala Phe Gly Lys Tyr Glu
180 185 190
Gly Glu Glu Arg Lys Asn Lys Phe Ala Asp Leu Ala Thr Glu Leu
195 200 205
Thr Thr Leu Met Gly Ala Phe Phe Val Gly Asp Tyr Phe Pro Ser Phe
Ala Trp Val Asp Val Leu Thr Gly Met Asp Ala Arg Leu Lys Arg Asn
His Gly Leu Asp Ala Phe Val Asp His Val Ile Asp Asp His Leu
Leu Ser Arg Lys Ala Asn Gly Ser Asp Gly Val Glu Gln Lys Asp Leu
Val Asp Val Leu Leu His Leu Gln Asp Ser Ser Leu Gly Val His
Leu Asn Arg Asn Asn Leu Lys Ala Val Ile Leu Asp Met Phe Ser Gly
Gly Thr Asp Thr Thr Ala Val Thr Leu Glu Trp Ala Met Ala Glu Leu
Ile Lys His Pro Asp Val Met Glu Ala Gin Gin Glu Val Arg Arg
Val Val Gly Lys Lys Ala Val Glu Glu Glu Asp Leu His Gln Leu
His Tyr Leu Lys Leu Ile Lys Glu Thr Leu Arg Leu His Pro Val
Ala Pro Leu Val Pro Arg Glu Ser Thr Arg Asp Val Val Ile Arg
Gly Tyr His Ile Pro Ala Lys Thr Arg Val Phe Ile Asn Ala Trp Ala
Ile Gly Arg Asp Pro Lys Ser Trp Glu Asn Ala Glu Phe Leu Pro
Glu Arg Phe Val Asn Asn Ser Val Asp Phe Lys Gin Asp Phe Gin
Leu Ile Pro Phe Gly Ala Gin Arg Arg Cys Pro Gly Ile Ala Phe
Gly Ile Ser Ser Val Glu Ile Ser Leu Ala Asn Leu Val Tyr Trp Phe
Asn Trp Glu Leu Pro Gly Ile

<210> SEQ ID NO 43
<211> LENGTH: 509
<212> TYPE: PRT
<213> ORGANISM: Arabidopsis thaliana
<400> SEQUENCE: 43
Met Ala Phe Phe Ser Met Ile Ser Ile Leu Leu Gly Phe Val Ile Ser
1       5       10      15
Ser Phe Ile Phe Ile Phe Phe Phe Lys Lys Leu Ser Phe Ser Arg
20      25      30
Lys Asn Met Ser Glu Val Ser Thr Leu Pro Ser Val Pro Val Val Pro
35      40      45
Gly Phe Pro Val Ile Gly Asn Leu Leu Gin Leu Lys Glu Lys Pro
50      55      60
His Lys Thr Phe Thr Arg Trp Ser Glu Ile Tyr Gly Pro Ile Tyr Ser
65      70      75      80
Ile Lys Met Gly Ser Ser Ser Leu Ile Val Leu Asn Ser Thr Glu Thr
95      90      95
Ala Lys Glu Ala Met Val Thr Arg Phe Ser Ser Ile Ser Thr Arg Lys
100     105     110
| 115 | Leu Ser Asn Ala Leu Thr Val Leu Thr Cys Asp Lys Ser Met Val Ala |
| 120 | 125 |
| 130 | Thr Ser Asp Tyr Asp Asp Phe His Lys Leu Val Lys Arg Cys Leu Leu |
| 135 | 140 |
| 145 | Asn Gly Leu Leu Gly Ala Asn Ala Glu Lys Arg Lys Arg His Tyr Arg |
| 150 | 155 | 160 |
| 165 | Asp Ala Leu Ile Glu Asn Val Ser Ser Lys Leu His Ala His Ala Arg |
| 170 | 175 |
| 180 | Asp His Pro Gln Glu Pro Val Asn Phe Arg Ala Ile Phe Glu His Glu |
| 185 | 190 |
| 190 | Leu Phe Gly Val Ala Leu Lys Gln Ala Phe Gly Lys Asp Val Glu Ser |
| 195 | 200 | 205 |
| 210 | Ile Tyr Val Lys Glu Leu Gly Val Thr Leu Ser Lys Arg Glu Ile Phe |
| 215 | 220 |
| 225 | Lys Val Leu Val His Asp Met Met Glu Gly Ala Ile Asp Val Asp Trp |
| 230 | 235 | 240 |
| 245 | Arg Asp Phe Phe Pro Tyr Leu Lys Trp Ile Pro Asn Lys Ser Phe Glu |
| 250 | 255 |
| 260 | Ala Arg Ile Gln Gln Lys His Arg Arg Leu Ala Val Met Asn Ala |
| 265 | 270 |
| 275 | Leu Ile Gln Asp Arg Leu Lys Gln Asp Ser Glu Ser Asp Asp Asp |
| 280 | 285 |
| 290 | Cys Tyr Leu Asn Phe Leu Met Ser Gly Ala Lys Thr Leu Thr Lys Glu |
| 295 | 300 |
| 305 | Gln Ile Ala Ile Leu Val Trp Glu Thr Ile Ile Glu Thr Ala Asp Thr |
| 310 | 315 | 320 |
| 325 | Thr Leu Val Thr Thr Glu Trp Ala Ile Tyr Glu Leu Ala Lys His Pro |
| 330 | 335 |
| 340 | Ser Val Gln Asp Arg Leu Cys Lys Glu Ile Gln Asn Val Cys Gly Gly |
| 345 | 350 |
| 355 | Glu Lys Phe Lys Glu Glu Leu Ser Glu Val Pro Tyr Leu Asn Gly |
| 360 | 365 |
| 370 | Val Phe His Lys Thr Leu Arg Lys Tyr Ser Pro Ala Pro Leu Val Pro |
| 375 | 380 |
| 385 | Ile Arg Tyr Ala His Glu Asp Thr Gln Ile Gly Gly Tyr His Val Pro |
| 390 | 395 | 400 |
| 405 | Ala Gly Ser Glu Ile Ala Ile Tyr Gly Cys Asn Met Asp Lys |
| 410 | 415 |
| 420 | Lys Arg Trp Glu Arg Pro Glu Asp Trp Trp Pro Glu Arg Phe Leu Asp |
| 425 | 430 |
| 435 | Asp Gly Lys Tyr Glu Thr Ser Asp Leu His Lys Thr Met Ala Phe Gly |
| 440 | 445 |
| 450 | Ala Gly Lys Arg Val Cys Ala Gly Ala Leu Glu Ala Ser Leu Met Ala |
| 455 | 460 |
| 465 | Gly Ile Ala Ile Gly Arg Val Leu Val Glu Phe Glu Trp Lys Leu Arg |
| 470 | 475 | 480 |
| 485 | Asp Gly Glu Glu Glu Asn Val Asp Thr Tyr Gly Leu Thr Ser Glu Lys |
| 490 | 495 |
| 500 | Leu Tyr Pro Leu Met Ala Ile Ile Asn Pro Arg Arg Ser |
The invention claimed is:

1. A method for producing in a recombinant host cell an isoprenoid compound not endogenously produced by a non-recombinant host cell, comprising:
   a) providing a recombinant host cell that comprises heterologous nucleic acid encoding an isoprenoid synthase, and a heterologous nucleic acid encoding one or more protein(s) comprising a CYP71, CYP73, CYP82 or a CYP92 family cytochrome P450 polypeptide, wherein the nucleic acid encoding a CYP71, CYP73, CYP82 or a CYP92 family cytochrome P450 polypeptide can be amplified with degenerate primers based on any one of SEQ ID NOs: 26-29; wherein the isoprenoid synthase catalyzes production of an isoprenoid compound; wherein the isoprenoid synthase is a diterpene synthase; and wherein the cytochrome P450 polypeptide(s) catalyzes dual hydroxylation, oxidation, demethylation or methylation of the isoprenoid compound; and
   b) culturing the recombinant host cell under conditions suitable for expressing the isoprenoid synthase and the cytochrome P450 polypeptide(s) under conditions for producing the isoprenoid compound; wherein the isoprenoid synthase is a diterpene synthase; and wherein the cytochrome P450 polypeptide(s) catalyzes dual hydroxylation, oxidation, demethylation or methylation of the isoprenoid compound; and

2. The method of claim 1, wherein at least one cytochrome P450 polypeptide has oxidase activity.

3. The method of claim 1, wherein at least one cytochrome P450 polypeptide has dual hydroxylase activity.

4. The method of claim 1, wherein the host cell is a yeast cell, a bacterial cell, an insect cell or a plant cell.

5. The method of claim 1, wherein the host cell is a yeast cell.

6. The method of claim 5, wherein the yeast is Saccharomyces cerevisiae.

7. The method of claim 1, wherein at least one of the cytochrome P450 polypeptides is 5-epi-aristolochene hydroxylase or kaurene oxidase.

8. The method of claim 1, wherein the isoprenoid compound is a diterpene.

9. The method of claim 1, further comprising c) isolating the isoprenoid compound.

10. A host cell, comprising nucleic acid encoding an isoprenoid synthase and nucleic acid encoding one or more protein(s) comprising a CYP71, CYP73, CYP82 or a CYP92 family cytochrome P450 polypeptide encoded by nucleic acid that can be amplified with degenerate primers based on any one of SEQ ID NOs: 26-29, wherein:
    the nucleic acid encoding the synthase and the nucleic acid encoding the P450 polypeptides are heterologous to the host cell;
    the synthase catalyzes production of an isoprenoid compound;
    the isoprenoid synthase is a diterpene synthase;
    the cytochrome P450 polypeptides catalyze hydroxylation, oxidation, demethylation or methylation of the isoprenoid compound;
    at least one of the P450 polypeptide(s) catalyzes the dual hydroxylation, oxidation, demethylation or methylation of the isoprenoid compound whose production is catalyzed by the synthase; and
    the isoprenoid compound produced by the host cell is not endogenously produced by a non-recombinant host cell.

11. The host cell of claim 10 that is a yeast cell, a bacterial cell, an insect cell or a plant cell.

12. The host cell of claim 10 that is a yeast cell.

13. The host cell of claim 10, wherein at least one of the cytochrome P450 polypeptide(s) is 5-epi-aristolochene hydroxylase or kaurene oxidase.

14. The method of claim 1, wherein at least one of the cytochrome P450 polypeptide(s) is selected from among polypeptides comprising at least 80% identity to an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and SEQ ID NO: 11.

15. The host cell of claim 10, wherein at least one of the cytochrome P450 polypeptide(s) is selected from among polypeptides comprising at least 80% identity to an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and SEQ ID NO: 11.