



University of Kentucky
UKnowledge

University of Kentucky Doctoral Dissertations

Graduate School

2008

EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS

Sungbeom Lee

University of Kentucky, slee5@uky.edu

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Lee, Sungbeom, "EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS" (2008). *University of Kentucky Doctoral Dissertations*. 587.
https://uknowledge.uky.edu/gradschool_diss/587

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

ABSTRACT OF DISSERTATION

Sungbeom Lee

The Graduate School
University of Kentucky

2008

EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY
OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
Department of Plant and Soil Sciences at the University of Kentucky

By
Sungbeom Lee

Lexington, Kentucky

Director: Dr. Joseph Chappell, Professor

Lexington, Kentucky

2008

Copyright © Sungbeom Lee 2008

ABSTRACT OF DISSERTATION

EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS

Southern Magnolia (*Magnolia grandiflora*) is a primitive tree species that has attracted attention because of its horticultural distinctiveness, the wealth of natural products associated with it, and its evolutionary position as a basal angiosperm. Terpenoid constituents were determined from Magnolia leaves and flowers. Magnolia leaves constitutively produced two major terpenoids, β -cubebene and germacrene A. However, upon wounding Magnolia leaves biosynthesized a significant array of mono- and sesquiterpenoids, including β -pinene, *trans*- β -ocimene, α -gurjunene, β -caryophyllene and β -cubebene, along with fatty acid derivatives such as *cis*-jasnone, for up to 19 hours after treatment. Flowers were also examined for their emission of terpene volatiles prior to and after opening, and also in response to challenge by Japanese beetles. Opened and un-opened flowers constitutively emitted a blend of monoterpenes dominated by β -pinene and *cis*- β -ocimene. However, the emission levels of monoterpenes such as verbenone, geraniol, and citral, and sesquiterpenes such as β -cubebene, α -farnesene, and β -caryophyllene were significantly elevated in the emissions of the beetle-challenged flowers.

Three cDNAs corresponding to terpene synthase (TPS) genes expressed in young Magnolia leaves were isolated and the corresponding enzymes were functionally characterized *in vitro*. Recombinant Mg25 converted FPP (C_{15}) predominantly to β -cubebene, while Mg17 converted GPP (C_5) to α -terpineol. Efforts to functionally characterize Mg11 were unsuccessful. Transcript levels for all 3 genes were prominent in young leaf tissue and significantly elevated for *Mg25* and *Mg11* mRNAs in stamens. A putative N-terminal signal peptide of Mg17 targeted the reporter GFP protein to both chloroplasts and mitochondria when transiently expressed in epidermal cells of *Nicotiana tabacum* leaves. Phylogenetic analyses indicated that Mg25 and Mg11 belonged to the angiosperm sesquiterpene synthase subclass TPS-a, while Mg17 aligned more closely to the angiosperm monoterpene synthase subclass TPS-b. Unexpectedly, intron/exon organizations for the three Magnolia *TPS* genes were different from one another and from other well characterized terpene synthase gene sets. The *Mg17* gene consists of 6 introns

arranged in a manner similar to many other angiosperm sesquiterpene synthases, but *Mg11* contains only 4 introns, and *Mg25* has only a single intron near the 5 terminus of the gene. Our results suggest that much of the structural diversity observed in the *Magnolia TPS* genes may have occurred by means other than intron-loss from a common ancestor *TPS* gene.

Costunolide is a sesquiterpene lactone widely recognized for its diverse biological activities, including its bitter taste in lettuces, and as a precursor to the more potent pharmacological agent parthenolide. A lettuce EST database was screened for cytochrome P450 genes that might be associated with sesquiterpene hydroxylation. Five ESTs were selected based on sequence similarity to known sesquiterpene hydroxylases and three of them (*Ls7108*, *Ls3597* and *Ls2101*) were successfully amplified as full-length cDNAs. To functionally characterize these cDNAs, they were co-expressed along with a germacrene A synthase and a cytochrome P450 reductase in yeast. Based on product profile comparisons between the three different lines to the control line, only the *Ls7108*-harboring line produced unique compounds. Neither of the other lines showed a new product peak. The more abundant, polar product generated by the *Ls7108*-containing line was purified and identified as a 12-acetoxy-germacrene by NMR analysis. *In vitro* studies using *Ls7108* microsomal proteins did not yield the 12-acetoxy-germacrene A, but the putative germacra-1(10),4,11(13)-trien-12-ol intermediate. Catalytic activity of the *Ls7108* microsomal enzyme was NADPH, pH and time dependent. Our results demonstrate that *Ls7108* is a lettuce cytochrome P450 which catalyzes the hydroxylation of a methyl group of the isopropenyl substituent of germacrene A, generating germacra-1(10),4,11(13)-trien-12-ol, and that when this mono-hydroxylated sesquiterpene is synthesized in yeast, an endogenous yeast enzyme further modifies the germacrenol compound by acetylation of the alcohol group at the C-12 position.

KEYWORD: *Magnolia grandiflora* terpene synthases, Lettuce cytochrome P450s, Exon/intron organizations, Terpene synthase gene evolution

Sungbeom Lee

February, 2008

EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY
OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS

By

Sungbeom Lee

Joseph Chappell
Director of Dissertation

Arthur Hunt
Director of Graduate Studies

February, 2008

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertation submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name

Date

EXPLORING THE BIOCHEMICAL AND EVOLUTIONARY DIVERSITY
OF TERPENE BIOSYNTHETIC ENZYMES IN PLANTS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture at the University of Kentucky

By
Sungbeom Lee

Lexington, Kentucky

Director: Dr. Joseph Chappell, Professor

Lexington, Kentucky

Copyright © Sungbeom Lee 2008

DISSERTATION

Sungbeom Lee

The Graduate School
University of Kentucky

2008

ACKNOWLEDGEMENTS

I want to express my gratitude to Dr. Joseph Chappell for the scientific training and the research opportunities provided in his laboratory. He always encouraged and supported me to complete my Ph.D. degree. I also thank my committee members Drs. Art Hunt, Robert Houtz, and Sharyn Perry for their critical comments and encouragement, and especially my outside examiner Carol Baskin. My achievements would not have been possible without the help and encouragement of all the Chappell laboratory members, particularly Shuiqin Wu and Shunji Takahashi, and our neighboring lab members, especially, Carol Von Lanken, Randy Dinkins and Nihar R Nayak. I also thank our collaborators Drs. Robert Coates and Juan Faraldos for their chemical expertise and insights. All of these people assisted me in making this thesis a reality. I also thank my family for their continual support of my educational dreams. Last but not least, I express my greatest gratitude to my wife, Jeongwha Myoung. Her unwavering support and patience have served to inspire me to complete my thesis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF FILES	viii
CHAPTER I BACKGROUND.....	1
CHAPTER II INNATE AND INDUCED CHEMICAL CONSTITUENTS IN MAGNOLIA LEAVES AND FLOWERS	27
INTRODUCTION.....	27
MATERIALS AND METHODS.....	28
RESULTS AND DISCUSSION.....	30
CHAPTER III CLONING AND FUNCTIONAL CHARACTERIZATION OF NOVEL TERPENE SYNTHASE GENES FROM MAGNOLIA GRANDIFLORA	41
INTRODUCTION.....	41
MATERIALS AND METHODS.....	44
RESULTS.....	52
DISCUSSION.....	58
CHAPTER IV FUNCTIONAL IDENTIFICATION OF A C-12 HYDROXYLASE FROM LETTUCE, THE FIRST COMMITTED MONOOXYGENASE IN COSTUNOLIDE BIOSYNTHESIS	80
INTRODUCTION.....	80
MATERIALS AND METHODS.....	81
RESULTS.....	88
DISCUSSION.....	93
CHAPTER V SUMMARY AND PERSPECTIVES.....	104
REFERENCES	107
VITA.....	121

LIST OF TABLES

Table 2.1. Terpenoid profiles from <i>M. grandiflora</i> leaves.	38
Table 2.2. Terpenoid profiles from <i>M. grandiflora</i> flowers	39
Table 2.3. A summary of the volatile compounds emitted from <i>M. grandiflora</i> leaves and flowers as reported by Azuma et al. (1997).	40
Table 3.1. A summary of terpene synthase genes used to infer structural and functional features of the <i>M. grandiflora</i> terpene synthase genes.	74
Table 3.2. Phases of introns within Magnolia terpene synthase genes examined, and their comparison to those of known angiosperms and gymnosperms TPS genes.	75

LIST OF FIGURES

Figure 1.1. Examples of isoprenoids found in bacteria, fungi, insects, animals and plants that served primary and secondary physiological roles in these organisms	15
Figure 1.2. Isoprenoids are synthesized from isoprene, a repeating five carbon building block.....	16
Figure 1.3. The IPP biosynthetic pathways of a typical plant cell localized to the cytosol (A, the acetate/mevalonate (MEV) pathway) and the plastid (B, the methyl erythritol phosphate (MEP) pathway) compartments.....	17
Figure 1.4. Terpene biosynthesis occurs via independent pathways localized to different intracellular compartments.....	18
Figure 1.5. Biosynthesis of representative monoterpenes from geranyl diphosphate (Bohlmann, 1998). LPP represents linalyl diphosphate.....	19
Figure 1.6. Biosynthesis of representative sesquiterpenes from farnesyl diphosphate. FPP and NPP represent farnesyl diphosphate and nerolidyl diphosphate, respectively.....	20
Figure 1.7. Floral structures of a primitive flower from <i>M. grandiflora</i> (A) and a representative recent flower in angiosperm (B).....	21
Figure 1.8. The annual variation of sesquiterpene lactones in leaves of Southern <i>Magnolia virginiana</i> (Song et al., 1998).....	22
Figure 1.9. A classification of terpene synthases, TPS, based on the organization of exons/introns within their genes as proposed by Trapp and Croteau (2001).....	23
Figure 1.10. The biosynthetic pathway for allene oxide (Schuler, 1996).....	24
Figure 1.11. Examples of well known hydroxylation patterns of monoterpene (A) and sesquiterpene (B) products.....	25
Figure 1.12. The biosynthesis of the plant growth regulatory gibberellins, like GA12, appear to require multiple hydroxylation/oxidation type reactions, possibly as many as 6 independent reactions (Schuler 1996)	26
Figure 2.1. Interrelationships of the orders and some families in plant kingdom supported by jackknife or bootstrap frequencies above 50% in large-scale analyses of angiosperms	34
Figure 2.2. Comparison of chemical profiles in untreated- and wound-treated leaves	35
Figure 2.3. Comparison of chemical profiles in control and beetle-feeding flowers	36
Figure 2.4. A simplified cartoon depicts that volatiles were collected by head-space trapping	37
Figure 3.1. An illustration of the sesquiterpenes comprising the eremophilene class of sesquiterpenes	63
Figure 3.2. A comparison of the deduced amino acid sequences for terpene synthases cloned from <i>M. grandiflora</i>	64
Figure 3.3. Total ion chromatograms of the sesquiterpene products generated by Mg25	65
Figure 3.4. A proposed reaction mechanism to explain how Mg25 might catalyze the formation of several different sesquiterpene reaction products from the substrate farnesyl diphosphate	66
Figure 3.5. Total ion chromatograms of the monoterpene (A) and sesquiterpene (B) products generated by Mg17.....	67

Figure 3.6. Proposed reaction mechanisms to explain how Mg17 might catalyze the formation of different monoterpene (A) and sesquiterpene (B) reaction products from the substrates geranyl diphosphate (GPP) and farnesyl diphosphate (FPP)	68
Figure 3.7. A measurement of the steady-state mRNA levels for the <i>M. grandiflora</i> terpene synthase genes <i>Mg25</i> , <i>Mg17</i> and <i>Mg11</i> in different tissues.....	69
Figure 3.8. Confocal images of tobacco leaf epidermal cells showing the localization of reporter proteins to the chloroplasts and mitochondria when fused with the N-terminal signal peptide (43 aa) from the <i>M. grandiflora</i> monoterpene synthase Mg17. 70	
Figure 3.9. Secondary structure and helical wheel predictions of the Mg17 transit peptide..	71
Figure 3.10. An un-rooted phylogenetic tree of terpene synthases analyzed by the neighbor-joining method and depicting the estimation of pair-wise distances at the amino acid level	72
Figure 3.11. Comparison of the genomic organization of Magnolia terpene synthase genes to well known and characterized terpene synthase genes.....	73
Supplementary Figure 3.1. Enzymatic properties of Mg25	76
Supplementary Figure 3.2. Additional MS spectra from GC peaks shown in Figs. 3.3 and 3.5 in comparison to NIST standards.....	77
Supplementary Figure 3.2 (continued). Additional MS spectra from GC peaks shown in Figs. 3.3 and 3.5 in comparison to NIST standards	78
Supplementary Figure 3.3. Nucleotide sequence alignments of <i>Mg25</i> , <i>Mg11</i> , and <i>Mg17</i> in order to design gene-specific PCR primers	79
Figure 4.1. An illustration of the sesquiterpenes accumulating constitutively in the latex of lettuce and chicory (<i>Cichorium intybus</i>)	95
Figure 4.2. A proposed biosynthetic pathways for the sesquiterpene costunolide based on the work of de Kraker et al. (2001).....	96
Figure 4.3. A deduced amino acid sequence alignment of 4 putative P450 genes isolated from <i>Lactuca sativa</i> in comparison to 2 well characterized sesquiterpene hydroxylases, 5-epi-aristolochene dihydroxylase (CYP71D20) and amorphadiene oxidase (CYP71AV1)	97
Figure 4.4. Functional identification of Ls7018 by its co-expression along with germacrene A synthase in yeast cells.....	98
Figure 4.5. Mass spectral patterns for the compounds generated in yeast as a consequence of co-expressing Ls7018 along with germacrene A synthase.....	99
Figure 4.6. <i>In vitro</i> characterization of the cytochrome P450 encoded by Ls7018	100
Figure 4.7. The pH-dependence of the enzyme activity encoded by the Ls7108 cDNA	101
Figure 4.8. The time-dependence of the reaction catalyzed by the enzyme encoded by the Ls7108 cDNA	102
Figure 4.9. Possible conformation of 12-acetoxy-germacrene A	103

LIST OF FILES

LeePhDthesis.pdf.....3.78 MB

CHAPTER I BACKGROUND

1. Isoprenoids are important compounds to all living organisms, especially plants

All forms of life synthesize isoprenoid natural products (Fig. 1.1) (Klayman, 1985; Cane, 1999; Ridley, 1999). This includes all prokaryotes and eukaryotes from bacteria, fungi, plants and animals. For example, archaeobacteria accumulate hopenes, sterol-like compounds that are common in higher eukaryotes like man. Streptomycetes and fungi produce a range of extremely toxic 15 carbon isoprenoids known as mycotoxins, which are highly regulated as contaminants of our food stocks by federal regulatory agencies. Plants produce a similar array of 15 carbon isoprenoids, yet some of these are considered vital sources for new therapeutic treatments, like artemisinin for malaria (Klayman, 1985; Ridley, 1999). While the practical utility of isoprenoids may be well appreciated, the physiological importance for many of these compounds to the organism synthesizing them remains unclear (Harrewijn et al., 2001).

Isoprenoids are generally classified as being associated with primary or secondary metabolism. Primary metabolites are those essential for survival, and in plants isoprenoids serving in these roles include plant growth regulators (gibberellins, brassinosteroids, cytokinin and abscisic acid) (Hoffmann-Benning and Kende, 1992; Binns, 1994; Clouse, 1998), photosynthetic pigments (carotenoids) (Cunningham and Gantt, 1998), structural constituents of membranes (phytosterols) (Clouse, 2000), molecular tethers for glycosylation metabolism of proteins and other small molecules (dolichols) (Carroll et al., 1992), and side chains for electron transport carriers (ubiquinone and plastoquinone) (Disch et al., 1998). The isoprenoids associated with secondary metabolism do not participate directly in plant growth and development, but instead appear to play roles in defense mechanisms and in providing ecological adaptations for plants to interface with their environment, like attracting insect pollinators and predatory insects to feeding herbivores, and plant-plant interactions like those between normal and parasitic plants (Harborne and Tomas-Barberan, 1991; Runyon et al., 2006).

2. The isoprenoid biosynthetic pathways

Intense competition in the late 1800s by chemists to define the structure of isoprenoids, particularly monoterpenes, led the German chemist Otto Wallach to define the “isoprene rule” (Wallach, 1914). His idea or unifying principle was based on the rationale that a vast array of terpenoids could be derived by uniting isoprene units in a repetitive head-to-tail pattern, for which he received the 1914 Nobel Prize in Chemistry (Fig. 1.2). Several decades later, Ruzicka received the 1939 Nobel Prize in Chemistry for his revised concept of the isoprene rule known as the “biogenetic isoprene rule”. Ruzicka’s proposal included a much greater emphasis on mechanistic considerations rather than structural ones, and was especially exciting for suggesting a role of carbocations in the biosynthetic mechanisms (Ruzicka, 1953) . The biogenetic isoprene rule helped resolved a classification conundrum. A very large number of compounds known at the time had many features suggesting that they were in fact terpenes (i.e. consisted of 10 or 15 or 20 carbons), but all the carbon atoms within the structures were not arranged in a simple isoprene repeating pattern. Ruzicka reasoned the compounds had undergone a re-arrangement and his biogenic rule and the idea of carbocations allowed for such re-arrangements.

Isopentenyl diphosphate (IPP), the universal precursor of terpenoids, and dimethylallyl diphosphate (DMAPP), its closely related isomeric form (Fig. 1.3), can be derived from two distinct pathways in plants, the cytosolic acetate/mevalonate pathway (Chappell, 1995; Cane, 1999) and/or the plastidic 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway (Lichtenthaler, 1999; Rohmer, 1999) (Fig. 1.3). In the cytosolic pathway, the two-step condensation of three molecules of acetyl-CoA is required to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) catalyzed by thiolase and HMG-CoA synthase. HMG-CoA is then reduced by HMG-CoA reductase, which is the rate-limiting step of this pathway operating in man (Goldstein and Brown, 1990; Chappell et al., 1995), to form mevalonic acid. This resulting product is then modified by two sequential ATP-dependent phosphorylations and a decarboxylation, generating the basic five-carbon unit, IPP. In the cytoplasm, IPP is readily isomerized to DMAPP by an enzyme favoring IPP formation. The other pathway for IPP and DMAPP biosynthesis occurs in the photosynthetic compartment, the chloroplast or plastidic compartment (Fig. 1.3B). In this pathway, the synthesis initiates from pyruvate and glyceraldehyde 3-phosphate (GAP),

proceeds through the five-carbon intermediates, 1-deoxy-D-xylulose 5-phosphate and 2-C-methyl-D-erythritol-4-phosphate (MEP) and ends with the biosynthesis of both IPP and DMAPP in stoichiometric amounts. The necessity for an IPP/DMAPP isomerase in the MEP pathway has yet to be resolved.

Terpenoids are biosynthesized by the condensation of various numbers of IPP (C_5) units and one molecule of the allylic ester dimethylallyl diphosphate (DMAPP, C_5) in typical head (diphosphate moiety) to tail (dimethyl substituents) fashion, although other orientations are possible (i.e. head to head). The head-to-tail joining of the one molecule of IPP and DMAPP generates geranyl diphosphate (GPP, C_{10}); farnesyl diphosphate (FPP, C_{15}) is derived from 2 IPP units joined with 1 DMAPP; and geranylgeranyl diphosphate (GGPP, C_{20}) is from 3 IPPs and 1 DMAPP (Fig. 1.4). Monoterpene synthases then utilize GPP as a substrate to synthesize linear and cyclized monoterpenes; sesquiterpene synthases use FPP for sesquiterpene biosynthesis; and diterpene synthases generate diterpenes from GGPP. Notably, sesquiterpenes are more structurally diverse than monoterpenes because of greater cyclization and rearrangement reactions possible with the larger precursor, and likewise diterpenes represent an even greater structural diversity than either monoterpene and sesquiterpenes (Cane, 1990; Chappell, 1995; Davis and Croteau, 2000; Christianson, 2006).

3. Uniqueness of the cyclization reactions of terpene biosynthesis

GPP, the natural substrate for monoterpene synthases, can be used directly for the biosynthesis of linear monoterpenes, but must be isomerized around the C2-C3 bond to allow for cyclization (Fig. 1.5) (Croteau, 1987). Hence, GPP ionizes by virtue of a divalent metal ion (Mn^{2+} or Mg^{2+}) producing the allylic cation-diphosphate anion pair. This allows for rearrangement of the C2-C3 double bond to the C1-C2 position and formation of the tertiary allylic isomer 3*R*- or 3*S*-linalyl diphosphate (LPP). After rotation about the C2-C3 single bond to the cisoid form, a second round of ionization and carbocation formation can lead to cyclization formation of either 4*R*- or 4*S*- α -terpinyl cation, a universal cyclic intermediate in monoterpene metabolism. The regio- and stereo-specific monoterpene products, such as limonene (Rajaonarivony et al., 1992), α - or β -pinene (Gambliel and Croteau, 1984), phellandrenes (Lafever and Croteau, 1993),

thujanes (Croteau, 1987), α -terpineol (Martin and Bohlmann, 2004) and as reported later in this study, 1,8-cineol (Croteau et al., 1994), are generated from this intermediate by one of several routes (i.e. deprotonation, hydride shift, water capture and heterocyclization).

FPP, like GPP, is the precursor to linear and cyclic terpenes. However, unlike GPP, the ionized form of *trans*,*-trans* FPP can undergo direct cyclization to generate a series of so-called all-*trans* sesquiterpenes, or undergo an isomerization of the terminal isoprene unit to overcome a stereochemical barrier for the formation of another entire class of sesquiterpenes, the so-called cisoid sesquiterpenes. Fig. 1.6 illustrates how *trans*,*-trans*-farnesyl pyrophosphate can be isomerized via ionization to the sesquiterpene analog, nerolidyl pyrophosphate (NPP), which has additional conformational flexibility (Fig. 1.6). Both the cisoid and transoid conformers can then undergo a large number of transformations including cyclizations, hydride shifts, and methyl migrations (Wagner-Meerwein rearrangements) to yield such compounds as α -bisabolene (Bohlmann et al., 1998b), germacrene A (Bertea et al., 2006), germacrene D (Picaud et al., 2006), germacrene C (Colby et al., 1998), valencene (Lucker et al., 2004), γ -humulene (Steele et al., 1998), δ -cadinene (Chen et al., 1995), δ -selinene (Steele et al., 1998), 5-*epi*-aristolochene (Facchini and Chappell, 1992), premnaspirodiene (Back and Chappell, 1995), and β -cubebene (described in the present study).

4. The structural diversity of terpenes is enhanced by secondary modifications

While the terpene synthases provide for hydrocarbon skeletons and in some cases hydroxylated forms arising from water quenching of the catalytic cascades, additional layers of structural diversity arise from secondary modification to these hydrocarbon scaffolds. These include modifications such as hydroxylation, oxidation, reduction, methylation, acetylation, aroyl substitution and peroxide ring formation (Schuler and Werck-Reichhart, 2003). These secondary, tertiary and further modifications often serve to impart important physical attributes onto the terpene molecules. For example, the all hydrocarbon skeletons are typically quite volatile and tend to be lost to the gas phase. Hydroxylation can reduce the volatility of these compounds dramatically. Also, because the secondary modifications occur with regio- and stereo-specific orientations, these too

impart unique structural and biological features. For instance, (+)-carvone has the odor of caraway and (-)-carvone smells of spearmint.

5. Regulation of terpene metabolism is complex

The biosynthesis and accumulation of terpenoids in plants are not only regulated temporally, but also spatially. In gymnosperms such as conifers, fir, and pine, large amount of monoterpenes and diterpenes are produced and sequestered in the highly specialized structures known as resin ducts, blisters, and scattered resin cells (Mcgarvey and Croteau, 1995). In angiosperms, secreting glandular trichomes and laticifers are utilized as the site of accumulation and storage of phytotoxic hydrophobic terpenoids, such that they serve as first defenders at the surface of plants (Wagner, 1991; Hulskamp and Schnittger, 1998; Simmons and Gurr, 2005). Site-specific accumulation of terpenes seems to correspond to a localization of the biosynthetic machinery as well. Carvone (C₁₀) biosynthesis in spearmint (*Mentha spicata*), for example, is restricted in the glandular trichomes, where all of the cyclase and hydroxylase activities involved in this biosynthesis have been measured (Gershenzon et al., 1989). Aside from trichomes, other types of compartments have been reported. The ultrastructure of secretory oil cells, into which terpenoids preferentially accumulate during development, have been described for leaves of *Magnolia grandiflora* (Postek and Tucker, 1983), and ontogeny of the articulated or non-articulated laticifers, a type of elongated secretory cell, has also been described in leaves and vascular structures of the latex-producing plant, *Lactuca sativa* (Olson et al., 1969).

Temporal regulation of terpene biosynthesis has been well documented in plants as well. In snapdragon flowers, regulated emission of floral scent, consisting of the monoterpenes myrcene and (*E*)-beta-ocimene, was reported to follow a diurnal pattern with the highest emission rate between 11 AM and 6 PM, and entrained by a circadian clock (Dudareva et al., 2003). Emission of the acyclic monoterpene linalool has also been coordinated with the timing of flower opening in *Clarkia breweri* (Pichersky et al., 1994).

Terpene biosynthesis can be induced by various treatments and signal molecules. Monoterpene and diterpene biosynthesis is induced in gymnosperms by wounding and herbivore feeding (Arimura et al., 2004b). Perhaps more dramatic, sesquiterpene

biosynthesis is rapidly induced in Solanaceous plants in response to fungal elicitors and various signal molecules like methyl jasmonates (Choi et al., 1992; Zook et al., 1992; Singh et al., 1998). Induction of sesquiterpene metabolism in tobacco has been shown to result from induced expression of the corresponding genes, for instance the induction of a sesquiterpene synthase and a sesquiterpene hydroxylase (Mandujano-Chavez et al., 2000; Ralston et al., 2001; Devarenne et al., 2002; Facchini and Chappell, 1992; Vogeli and Chappell, 1988; Chappell and Nable, 1987).

6. Value of an extant primitive angiosperm, *Magnolia grandiflora*

Fossils have been useful, but limited in tracing the evolutionary history and dating the origin of taxonomic groups. The difficulty is derived from the frequency of parallel evolution, which leads to the formation of very similar organs and structures in groups that are otherwise not very similar or not closely related. Fossils of plants are usually found in the form of dispersed pollen, leaves, fruits, seeds, and occasionally an intact flower, however often not in a connection to another. Without the connected forms of fossils, it is hard to determine if a postulated evolutionary pathway or reconstruction is correct (Cronquist, 1988). With the availability of advanced molecular technologies, the age of species within the genus Magnoliales, extant and those in the fossil records, based on the sequences for the *matK* and *rbcL* loci plus and two intergenic regions consisting of *psbA-trnH* and *atpB-rbcL*, have been dated to 20-25 million years ago (Golenberg et al., 1990; Azuma et al., 2001; Berry, 2003). A recent update in the angiosperm molecular phylogenetic analysis reconstructed by the sequences of *rbcL*, rDNA and *atpB* gene predicted that the archaic Magnoliales is located at the root of angiosperm lineages allied with Laurales and Piperales (Bremer et al., 2003), placing the origins of Magnoliales between 20 to 125 million years ago (Sun et al., 2002).

Aside from the fossil records and the molecular phylogeny, the method of pollination in angiosperms has also been correlated with their rate of evolution. The primitive *Magnolia* flower for instance has numerous spirally arranged, rather large and firm tepals, numerous spirally arranged laminar stamens and unsealed carpels (Fig. 1.7). All of these characteristics can be found individually among modern members of the Magnoliales, suggesting that species like *Magnolia grandiflora* are like windows to the

past and may offer a better appreciation for how genes might have looked 10s of millions of years ago (Cronquist, 1988).

Corollary to the presumed primitive nature of genes within *M. grandiflora* is its pollination by beetles. Ancestral angiosperms have been pollinated by beetles, while flowers of modern angiosperms are most often pollinated by wind, water, insects, and even flying animals such as birds (Richards, 1986; Westerkamp, 1990; Cunningham, 1991). The most basal order of Magnoliales is mainly pollinated by beetles (Coleoptera). Having evolved before bees appeared, the tough carpels of Magnolia flowers are thought to have arisen in order to protect the flower from the damage by beetle pollinators (Cronquist, 1988). Another primitive aspect of the Magnolia flower is the lack of distinct sepals or petals. The term tepal was proposed in reference to the intermediate nature of the Magnolia floral parts that resemble, in part, petals yet have features similar to sepals (Jones and Luchsinger, 1979).

Southern Magnolia (*Magnolia grandiflora*) is widely used as a major landscape tree species, but is probably less appreciated for its uses to the perfumery and aromatic therapy commodity's market (Sparrow and Hanly, 2002). *M. grandiflora* has attracted special attention from phytochemists and pharmaceutical researchers because of its many biologically active phytochemicals, including sesquiterpene lactones. Parthenolide and costunolide are seasonally regulated. (Fig. 1.8) (Song et al., 1998), along with biphenolics (honokiol and magnolol) (Elferaly and Li, 1978) and alkaloids (magnofluorine, magnocurarine, and salicifoline) (Tomita and Nakano, 1952) in the leaves and barks of this tree. Given the regulation of lactone contents, seasonal variations of lactones were examined in another Magnolia species, *Magnolia virginiana*. With the leaves from a single *M. virginiana*, planted on 1025 Highland Park Drive, East Baton Rouge Parish, Louisiana, three different sesquiterpene lactones were isolated at the times indicated throughout the year (Fig. 1.8). The air-dried leaves (940 g) of *M. virginiana* were ground and extracted with dichloromethane (CH_2Cl_2) for 24 hrs and chromatography of crude extracts on silica gel isolated costunolide, parthenolide, and costunolact-12-ol. The levels of the lactones in these extracts were quantified by HPLC and identified by NMR comparison with known standards. Parthenolide had previously been isolated from *M. grandiflora* as well (Castaneda-Acosta et al., 1993). All of these compounds have been

proposed for pharmaceutical and clinical usages in animals (Wiedhopf et al., 1973; Elferaly and Li, 1978; Mellado et al., 1980; Yang et al., 1994; Castanedaacosta et al., 1995; Koo et al., 2001; Feltenstein et al., 2004). For example, parthenolide has a tumor inhibitory effect (Wiedhopf et al., 1973), and shows anti-inflammatory and anti-hyperanalgesic effects in the rat carrageenan inflammation model (Feltenstein et al., 2004). Costunolide, a possible biosynthetic precursor to parthenolide, also can suppress tumor growth, possibly by inhibiting NF-kappa β activation and DNA-binding activity of NF-kappa β (Koo et al., 2001). Other biologically active sesquiterpene lactones found in *M. grandiflora* include peroxy-derivatives like peroxycostunolide (Verlotrin) and peroxyparthenolide (Elferaly et al., 1979b), and magnolialide (Elferaly et al., 1979a).

7. Molecular evolution of terpene synthase genes in plants

Since the discovery of introns in the late 1970s (Williamson, 1977), they have been suspected of playing some role(s) in gene evolution (Blake, 1978; Gilbert, 1978; Cavalier-smith, 1991; Palmer and Logsdon, 1991). The genes from eukaryotic organisms are discontinuous, showing exon/intron structure, in comparison to the continuous gene structures in prokaryotes (Lewin, 2000). In general, exon sequences and structures are highly conserved in homologous genes from different organisms, while the introns vary in sequence and in length, but seem to maintain their position within genes (de Souza et al., 1996; Betts et al., 2001; Trapp and Croteau, 2001).

The exon/intron structures of eukaryotic genes have been suggested to represent important clues in two putative pathways of gene evolution (Naora and Deacon, 1982; Frugoli et al., 1998; Trapp and Croteau, 2001), the intron-late theory (Palmer and Logsdon, 1991) and the intron-early (exon shuffling) theory (Long et al., 1995; de Souza et al., 1996). The introns-early hypothesis suggests that exon shuffling between homologous and orthologous genes was important for gene and protein evolution. If this were so, then one should expect the so-called phasing of introns between exons to be non-randomized. In contrast, the introns-late hypothesis suggests random insertion of introns within genes generating exons to be a random event and resulting in a random distribution of intron phasing across a gene or locus. Long et al. (1995) examined the relationship of intron phasing by analyzing the number of introns, their positions, phasing

and the overall sizes of the corresponding exons in 9,276 genes/loci in plants and animals. The occurrence rates of phase 0, 1, and 2 introns for the internal exons of plant genes were 56%, 23%, and 21% out of a total 2790 introns, respectively. Animals also showed similar rates of introns phase bias, phase 0 (46%), phase 1 (31%), and phase 2 (23%). Long et al. (1995) concluded that introns were distributed in a non-random pattern in eukaryotic genes, rather than randomly, and suggested that these supported the notion of exon shuffling as a means for the dispersal of genetic mutants amongst genes and providing another mechanism for molecular evolution.

The first suggestions for introns/exon organization being important for terpene synthase (TPS) genes in plants were presented by Mau and West (Mau and West, 1994) and Back and Chappell (Back and Chappell, 1996). Both these groups observed an unusually high conservation in the number and position of introns within sesquiterpene and diterpene synthase genes. Back and Chappell went on to demonstrate that the reciprocal exchange of exons between two sesquiterpene synthase genes yielded catalytically competent enzymes which took on the catalytic specificity of the synthase contributing the last 3 exons. These observations have been extended by many other investigators performing so-called domain swapping experiments (Cseke et al., 1998; Back et al., 2000; Davis and Croteau, 2000; Katoh et al., 2004), but the greatest success of this approach has been achieved by delimiting the swapping domains to exonic units (Katoh et al., 2004).

Several authors have extended the notion of intron/exon organization to the evolution of terpene synthase genes by detailed comparisons from large gene and genome sequence databases (Bohlmann et al., 1998a; Trapp and Croteau, 2001; Aubourg et al., 2002). Given a wide range of over 20 *TPS* genes collected from angiosperms and gymnosperms, Trapp and Croteau (2001) were able to divide these into three classes (class I, II, and III), simply based upon the number of exons and introns (Fig. 1.9). The class I *TPS* genes contained 12-14 introns and consisted mainly of diterpene synthase genes from both gymnosperms and angiosperms. The class II *TPS* genes contained 9 introns and were comprised of monoterpene and sesquiterpene synthase genes found in gymnosperms, but not in angiosperms. Trapp and Croteau (2001) also noted that the majority of mono-, sesqui-, and di-terpene synthase genes from angiosperms contained 6

introns, all conserved with regards to their position within the genes, and referred to these genes as class III. The *TPS* genes in the class II and the class III characteristically lack introns I and II found in the class I *TPS* genes. All the plant *TPS* genes falling into class I, II or III, nonetheless, showed conservation of intron position, phasing and the ultimate exon size (Fig. 1.9B).

Interestingly, a unique N-terminal domain (about 210 amino acids), known as the conifer diterpene internal sequence (CDIS) domain, was found in all diterpene synthases (Bohlmann et al., 1998a) except casbene synthase (Mau and West, 1994), as well as in a sesquiterpene synthase (grand fir (*E*)- α -bisabolene synthase) (Bohlmann et al., 1998b), and a monoterpene synthase (*Clarkia* linalool synthase). The observation of this domain in *TPS* genes other than diterpene synthases is perhaps a signature sequence remaining in a member of an earlier evolutionary event that ultimately gave rise to a new *TPS* gene family (Dudareva et al., 1996; Bohlmann et al., 1998a).

This classification system was extended to an analysis of all *TPS* genes in the *Arabidopsis thaliana* (At) genome. Most of the *AtTPS* genes readily fall into class I (3 genes) or class III (29 genes), with 8 pseudogenes remaining as non-classifiable (Aubourg et al., 2002). While the number and relative position of the introns are well conserved in the *AtTPS* genes, equally impressive is that the phasing of introns between paralogues are also highly conserved in this organism, as observed in other plant species (Trapp and Croteau, 2001; Aubourg et al., 2002).

8. Cytochrome P450s (CYPs) and sesquiterpene hydroxylases in plants

More than 300 distinct sesquiterpene skeletal-types have been identified to date, and over 1,000 sesquiterpene derivatives have been isolated from plants, insects, microbial sources (Devon and Scott, 1972; Croteau, 1987; Cane, 1990; Davis and Croteau, 2000; Phillips et al., 2003). This diversity of sesquiterpenes arises in large part by the decoration of the hydrocarbon scaffolds with hydroxyl substituents and their further oxidation to aldehydes, ketones and acid functionalities. All these chemical modifications can arise from the action of cytochrome P450s. Cytochrome P450s carry out a variety of modifications to small molecules, including hydroxylations, demethylations, epoxidations, decarboxylations, and deaminations with the participation of NADPH-

dependent P450 reductase which shuttle electrons from NADPH into the catalytic site of the CYPs (Bak et al., 1997; Lupien et al., 1999; Ralston et al., 2001; Morant et al., 2003; Schuler and Werck-Reichhart, 2003). Interestingly, cytochrome P450s appear highly conserved in their secondary and tertiary structures as defined by crystal structures of bacterial P450s (Park et al., 1997; Podust et al., 2001) and mammal P450s (Williams et al., 2000b, a), even though their primary sequences are relatively divergent from one another (<20% identity) (Schuler and Werck-Reichhart, 2003). P450 monooxygenases, for example, range in molecular mass from 45 to 62 kDa, and all possess a ferriprotoporphyrin IX prosthetic group covalently linked to a cysteine in the highly conserved FxxGxRxCxG motif located near C-terminus.

There are two general classification groups of P450s, one known as the “classical” CYPs and the other as the “nonclassical P450s” (Schuler, 1996). The P450s in the former class require flavoproteins to transfer electrons and incorporate molecular oxygen into their final products. The majority of these P450s exist as membrane-anchored proteins in the endoplasmic reticulum (ER). A small number of mitochondrial, classical P450s have been observed in mammals (Pechurskaya et al., 2007), but their existence in plants is unknown. Unlike the soluble bacterial P450s (Taylor et al., 1999; Munro et al., 2002), microsomal and mitochondrial P450s are synthesized on cytosolic ribosomes and transported into the ER or posttranslationally into the mitochondria. However, the sizes of mature products in both organelles are different because a transit peptide of the mitochondrial directed P450s is removed by proteolytic cleavage during insertion into the inner mitochondrial membrane. Microsomal P450s retain their signal sequences even after insertion into the ER membranes (Schuler, 1996).

The “nonclassical P450s” are different from the “classical P450s” in terms of flavoprotein utilization. They do not utilize flavoproteins for reductive activation of dioxygen. Hence, neither NADPH-dependent P450 reductase nor cytochrome b₅/NADH-dependent cytochrome b₅ reductase are involved in electron transfer. Allene oxide synthases, for example, convert 13-hydroperoxide to allene oxide in the absence of an NADPH-dependent P450 reductase or oxygen (Fig. 1.10) (Song and Brash, 1991). To date, no non-classical CYPs have been identified that use terpenes as substrates.

Many terpenes take on biological activities as a consequence of stereo- and regio-specific hydroxylation by P450s. For example, limonene is hydroxylated at the C3, C6 and C7 positions in a species-specific manner (Fig. 1.11) (Karp et al., 1990; Schuler, 1996; Lupien et al., 1999). Several terpene hydroxylases have been described that catalyze successive hydroxylation/oxidation reactions. For example, 5-*epi*-aristolochene dihydroxylase hydroxylates 5-*epi*-aristolochene at C1, followed by hydroxylation at C3 (Takahashi et al., 2005) (Fig. 1.11B). In contrast, a CYP recently isolated from *Hyoscyamus muticus* hydroxylates premnaspirodiene at C2, followed by a second hydroxylation at the same position which spontaneously oxidizes to a ketone. The biosynthesis of gibberellic acid (GA), a diterpene plant growth regulator, occurs by a series of successive oxidations at methyl substituents to yield carboxylic acids (Fig. 1.12) (Schuler, 1996; Hedden and Kamiya, 1997). Very surprisingly, the successive oxidations at C19 and C7 appear to be catalyzed by single, multi-function cytochrome P450s.

9. Engineering terpene production platforms in yeast

With the availability of genetic information and advanced molecular biology techniques, intensive and extensive efforts have been made by various research groups to enhance the production of valuable plant natural products *in vivo* (Wenzel and Muller, 2005; Liu et al., 2006). These attempts have been directed not only to the genetic engineering of plant tissue for natural product production (Abdin et al., 2003), including the improvement of extraction and cultivating methods of the host (Kumar et al., 2004; Lapkin et al., 2006), but also to the development of heterologous expression host like bacteria or novel plant platforms (DeJong et al., 2006; Wu et al., 2006).

As a practical matter, obtaining sufficient quantities of terpenes and hydroxylated terpene products from plants for chemical identification and biological testing is often a very serious limitation. Artemisinin is currently obtained from *Artemisia annua* (Kumar et al., 2004; Lapkin et al., 2006), a slow growing plant that does not provide a consistent yield of artemisinin because of genetic and environmental effects (Charles et al., 1990). Production of these compounds *in planta* is generally low (0.01 % – 0.8 % of DW), which is exacerbated by losses during complex purification schemes (Abdin et al., 2003). The desire for more robust production means for terpenes is exemplified by the

increasing desire for the anti-malaria drug artemisinin (Liu et al., 2006; Ro et al., 2006) and the chemo-therapeutic drug taxol (DeJong et al., 2006).

Recently, bacterial and yeast production platforms for terpenes have been advanced. The bacterial system was first exemplified by introducing both an amorpha-4,11-diene synthase and a gene cassette of the mevalonate pathway to supply sufficient common precursors of the isoprenoid pathway in *E. coli* (Martin et al., 2003). This system could produce amorphadiene up to 24 μg equivalent of caryophyllene per milliliter. Using the similar strategy and system, lycopene, a most potent carotenoid antioxidant, also was over-produced up to 102 mg/L (Yoon et al., 2006). The next step for the natural product engineering would be the secondary modification of over-produced hydrocarbons. However, to express eukaryotic cytochrome P450s, a most potent biocatalyst for stereo- and regio-specific modifications, in bacterial systems is difficult because of the absence of membrane structures necessary for tethering the N-terminus of the P450 enzyme moieties. Nonetheless, a recent report provided an exciting means for the successful expression of plant P450s in an *E. coli* system for the production of functionalized terpenoids (Chang et al., 2007)

To achieve coupled modification of hydrocarbon *in vivo*, yeast cells have been considered. In an earlier attempt, DeJong et al. introduced eight genes out of ten of the taxol biosynthetic pathway including a taxadiene synthase, three regiospecific hydroxylases, and three acyl/aroyl-CoA dependent transferases into yeast. After 65 hrs incubation, taxadiene was readily measurable (0.7 mg/L) in the engineered cells, however, only a very little amount of first hydroxylated intermediate taxadien-5 α -ol ($\leq 25 \mu\text{g/L}$), was produced (DeJong et al., 2006).

Successful examples in natural product engineering in yeast was recently provided by several research groups (Ro et al., 2006; Takahashi et al., 2007b). Engineered yeast cells over-producing artemisinic acid, which can readily be converted to artemisinin at relatively low cost, have been achieved (Ro et al., 2006). Artemisinic acid, the precursor to the anti-malarial sesquiterpene lactone artemisinin, was successfully produced in large scale (100 mg/L culture) by introducing a multifunctional P450 (CYP71AV1) into the artemisinin biosynthetic pathway (Ro et al., 2006; Teoh et al., 2006). More recently, the metabolic engineering of phytoalexin capsidiol was employed

in yeast cells (Takahashi et al., 2007b). A coordinated expression of both the sesquiterpene synthase and the sesquiterpene hydroxylase elevated the production of capsidiol significantly (> 50 mg/L culture), suggesting that large-scale production has been broadly extended to other valuable natural products. Moreover, these platforms will be valuable *in vivo* measures in identifying novel hydrocarbon and hydroxylated natural products.

Origin	Isoprenoids	Function	Reference
<u>bacteria</u>	hopenes	lipid constituent (cyanobacteria)	Boudou et al., 1986
	geosmin	Earthy odorant (<i>Streptomyces coelicolor</i>)	Jiang et al., 2006
<u>fungi</u>	lycopene	carotenoid singlet oxygen quencher (<i>Blakeslea trispora</i>)	Di mascio et al., 1989 Estrella et al., 2004
	Aflatoxins	mycotoxin (<i>Aspergillus flavus</i>)	Davis et al., 1966
	Trichothecenes	mycotoxin (<i>Fusarium graminearum</i>)	Miller et al., 1991
	Fumonisin	mycotoxin (<i>Fusarium verticillioides</i>)	Nelson et al., 1991
<u>plant</u>	Limonene	insect repellent	Hollingsworth, 2005
	β -farnesene	attract natural enemies	Schnee, 2006
	Capsidol	phytoalexin	Bailey et al., 1975
	Gibberellins	growth hormone	Hoffmann-Benning and Kende, 1992
	Brassinosteroids	vascular differentiation, cell division and expansion, and pollen tube elongation	Clouse, 1998
	Cytokinin	promoting cell division and differentiation	Binns, 1994
	abscisic acid	growth hormone	Hoffmann-Benning and Kende, 1992
	carotenoids	photosynthetic pigments	Cunningham and Gantt, 1998
	phytosterols	structural constituents of membranes	Clouse, 2000
	dolichols	tethers for glycosylation metabolism of proteins and other small molecules	Carroll et al., 1992
	ubiquinone and plastoquinone	side chains for electron transport carriers	Disch et al., 1998
<u>animal</u>	coenzymeQ	cofactor in electron transport chain, antioxidant	Turunen et al., 2004
	retinoids	vision and bone growth	Blomhoff, 2006
	cholesterol	composition of cell membranes, the synthesis of steroid hormones	Brown and Goldstein, 1997

Figure 1.1. Examples of isoprenoids found in bacteria, fungi, insects, animals and plants that serve primary and secondary physiological roles in these organisms.

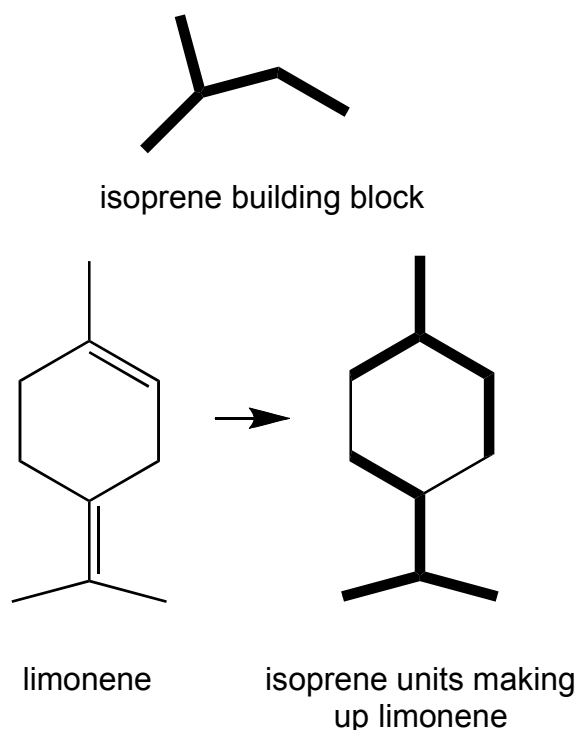


Figure 1.2. Isoprenoids are synthesized from isoprene, a repeating five carbon building block.

Wallach received the Nobel prize in 1914 for recognizing the repeating nature of isoprene in many terpenoid compounds, while Ruzicka was honored with the Nobel prize in 1939 for suggesting a biosynthetic or rational mechanism for how the isoprene units could condense together and subsequent rearrange to generate the immense structural diversity observed in terpenes. The essence of Ruzicka's proposal was the suggestion of carbocations, molecular species of carbon having only three electrons to fill their sp^3 orbitals rather than four.

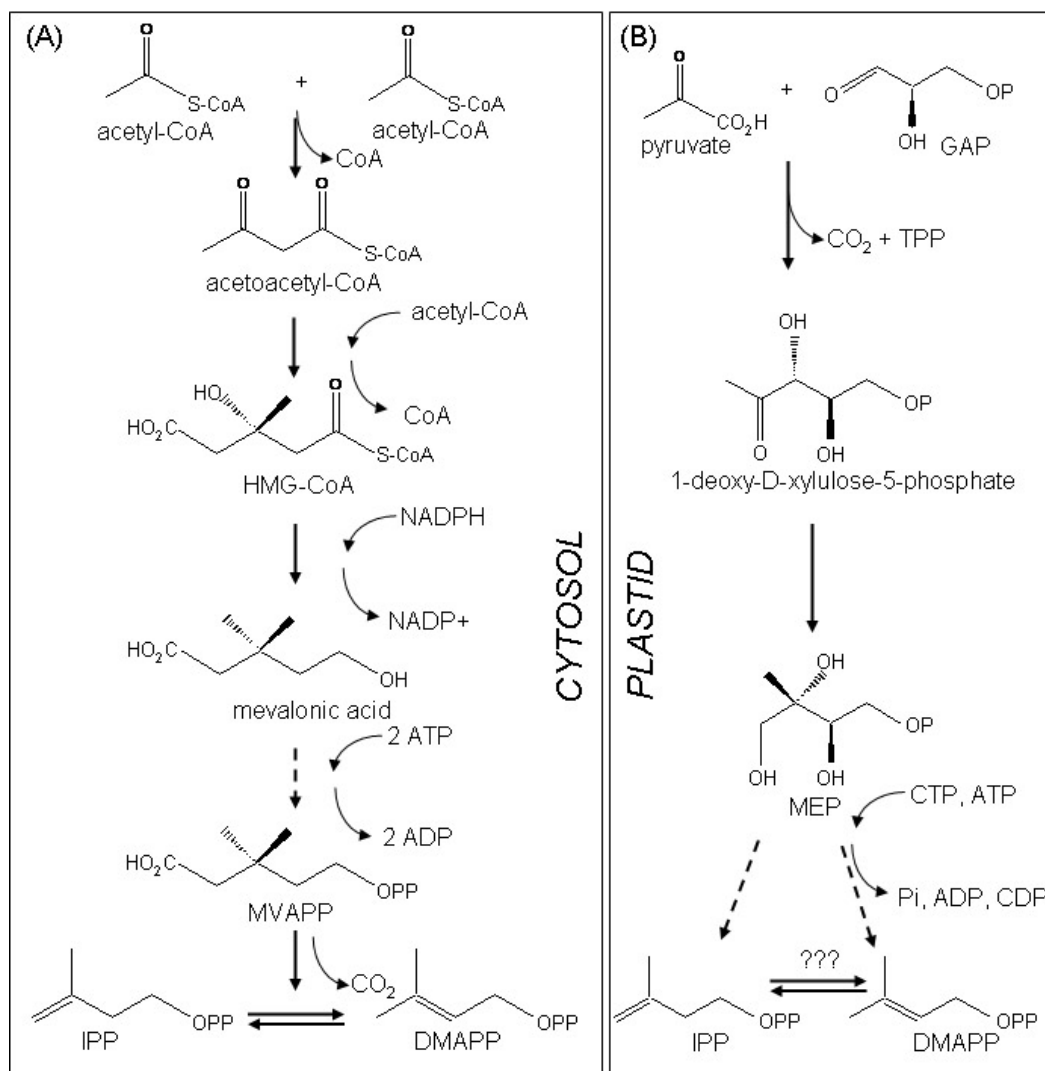


Figure 1.3. The IPP biosynthetic pathways of a typical plant cell localized to the cytosol (A, the acetate/mevalonate (MEV) pathway) and the plastid (B, the methyl erythritol phosphate (MEP) pathway) compartments. HMG-CoA, β -hydroxy- β -methylglutaryl-CoA; MVAPP, mevalonic acid diphosphate; GAP, D-glyceraldehyde-3-phosphate; TPP, thiamine diphosphate; MEP, 2-C-methylerythritol-4-phosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate.

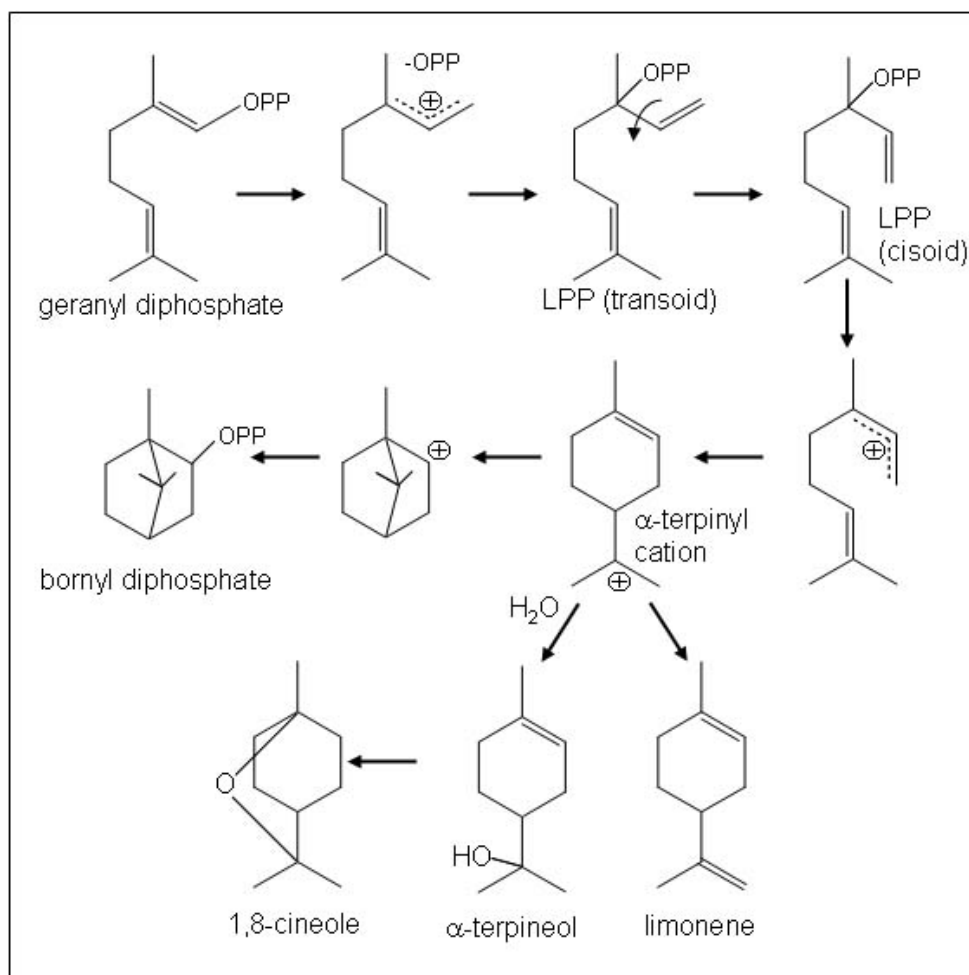


Figure 1.5. Biosynthesis of representative monoterpenes from geranyl diphosphate (Bohlmann, 1998). LPP represents linalyl diphosphate.

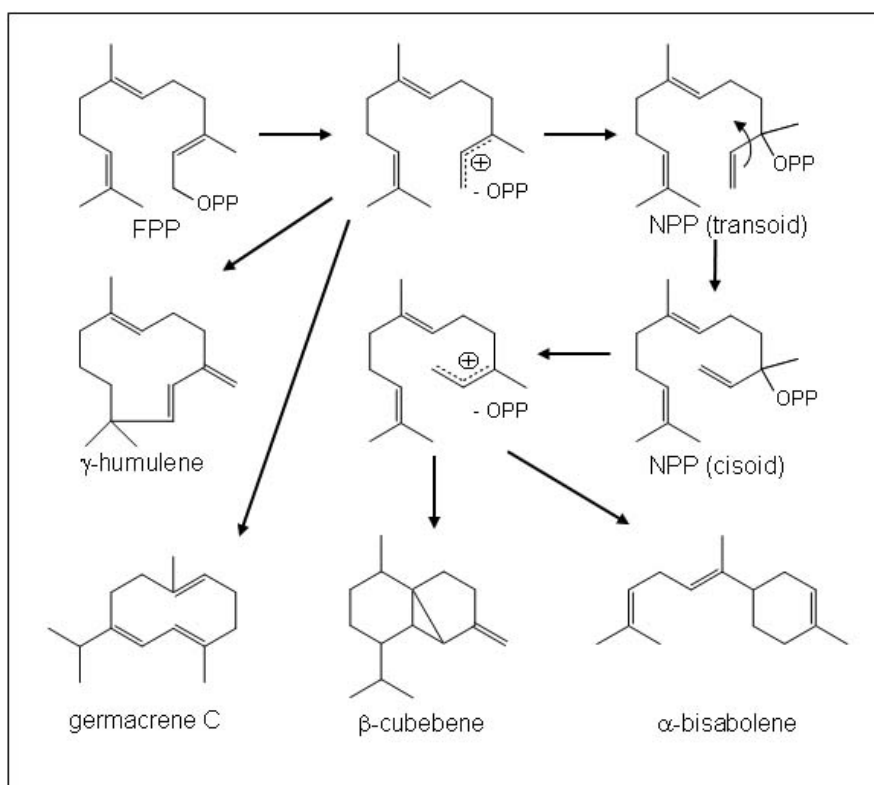


Figure 1.6. Biosynthesis of representative sesquiterpenes from farnesyl diphosphate. FPP and NPP represent farnesyl diphosphate and nerolidyl diphosphate, respectively.

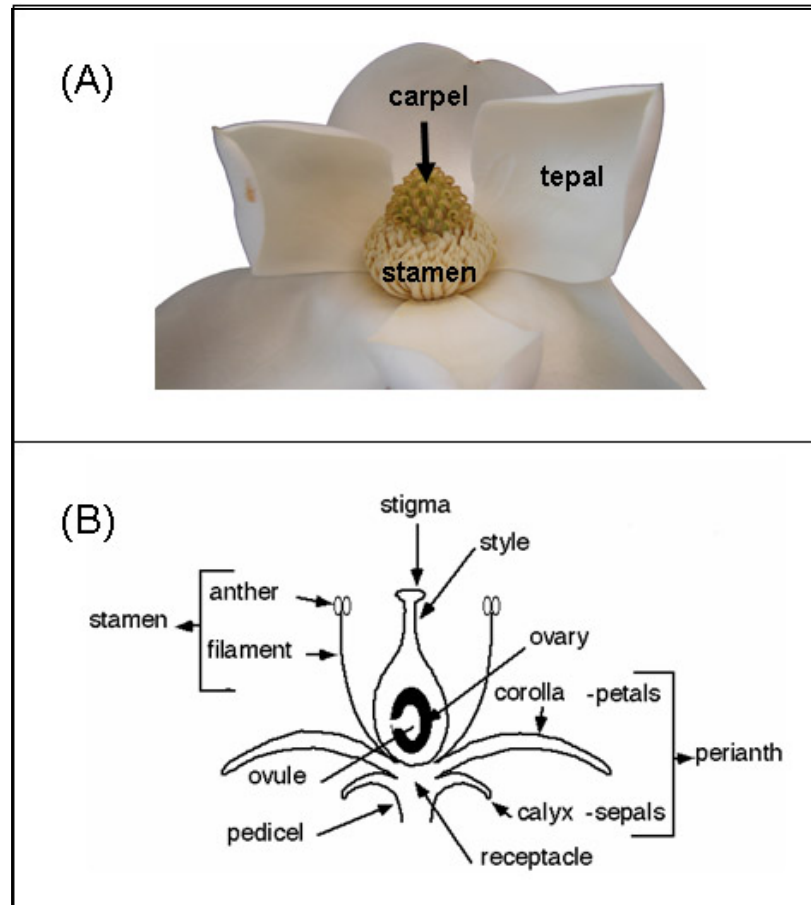


Figure 1.7. Floral structures of a primitive flower from *M. grandiflora* (A) and a representative recent flower in angiosperm (B). The primitive *Magnolia* flower has numerous spirally arranged, rather large and firm tepals, numerous spirally arranged laminar stamens and unsealed carpels. There is no fusion of parts and little differentiation of parts.

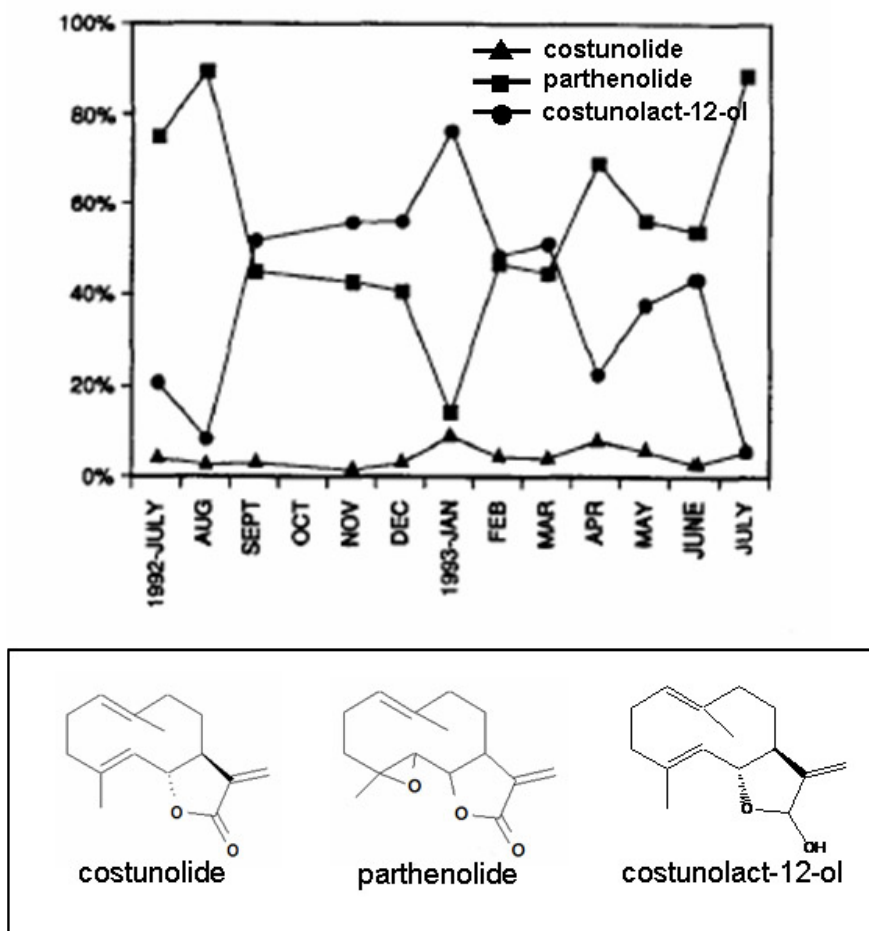


Figure 1.8. The annual variation of sesquiterpene lactones in leaves of Southern *Magnolia virginiana* (Song et al., 1998). Leaves from a single *M. virginiana*, planted on 1025 Highland Park Drive, East Baton Rouge Parish, Louisiana, were collected at the times indicated throughout the year. The air-dried leaves (940 g) of *M. virginiana* were ground and extracted with dichloromethane (CH_2Cl_2) for 24 hrs and the chromatography of crude extracts on silica gel isolated costunolide, parthenolide, and costunolact-12-ol. The levels of the lactones in these extracts were quantified by HPLC and identified by NMR comparison with known standards. Parthenolide had previously been isolated from *M. grandiflora* as well (Castaneda-Acosta et al., 1993).

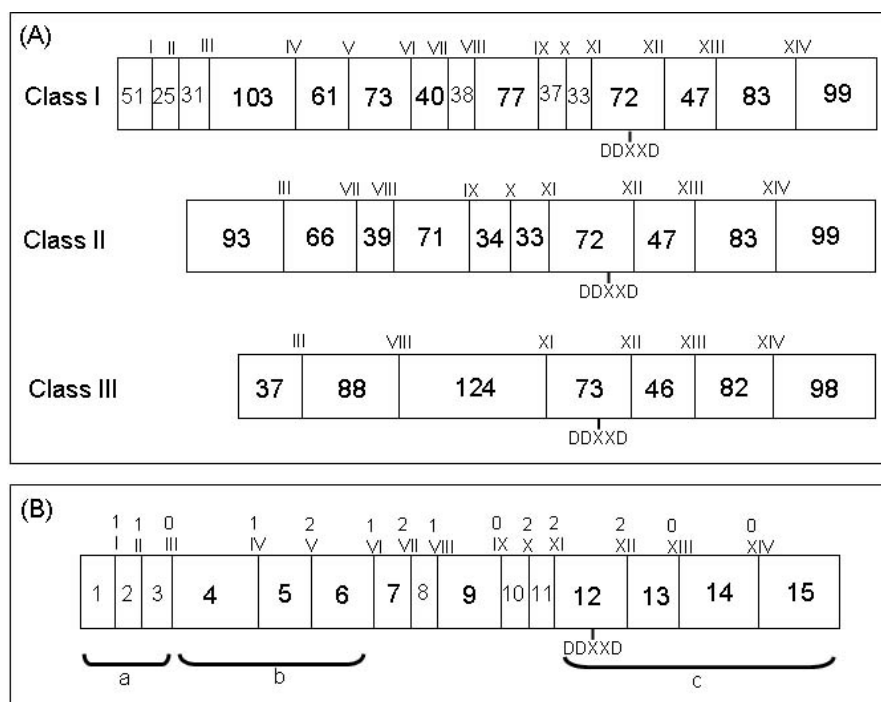


Figure 1.9. A classification of terpene synthases, TPS, based on the organization of exons/introns within their genes as proposed by Trapp and Croteau (2001). (panel A) All the TPS genes examined to date can be assigned to one of these three classes of gene architectures depending on the number of introns and their relative insertion sites within the genes (roman numbers). The size of exons is represented by the number of amino acids encoded by each as noted within each exon. The TPS used to illustrate Class I TPS is abietadiene synthase, a diterpene synthase from grand fir (Trapp and Croteau, 2001); (-)-limonene synthase, a monoterpene synthase, from grand fir (Trapp and Croteau, 2001) for Class II; and 5-*epi*-aristolochene synthase, a sesquiterpene synthase from tobacco (Facchini and Chappell, 1992) for Class III. The Class I-III genes are further annotated in panel B with the regards to the type of intron insertion, either without disrupting the reading frame (0) or within a codon after the first (1) or second (2) base pair. The gene region denoted as “a” generally encodes for an intracellular targeting signal sequences, region “b” represents an internal sequence of unknown function associated predominantly with diterpene synthases, and region c denotes the active region of the protein, including a highly conserved aspartate rich-domain that coordinates metal and substrate binding (Bohlmann et al., 1998).

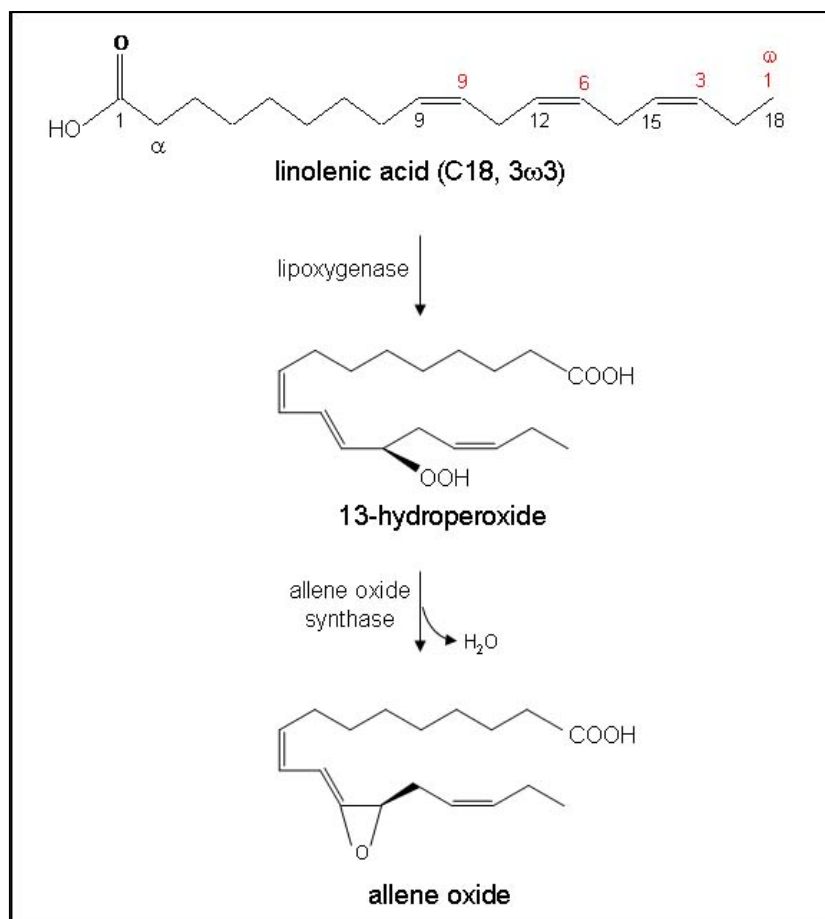


Figure 1.10. The biosynthetic pathway for allene oxide (Schuler, 1996).

The non-classic P450, allene oxide synthase, converts 13-hydroperoxide to allene oxide in the absence of an NADPH-dependent P450 reductase or molecular oxygen.

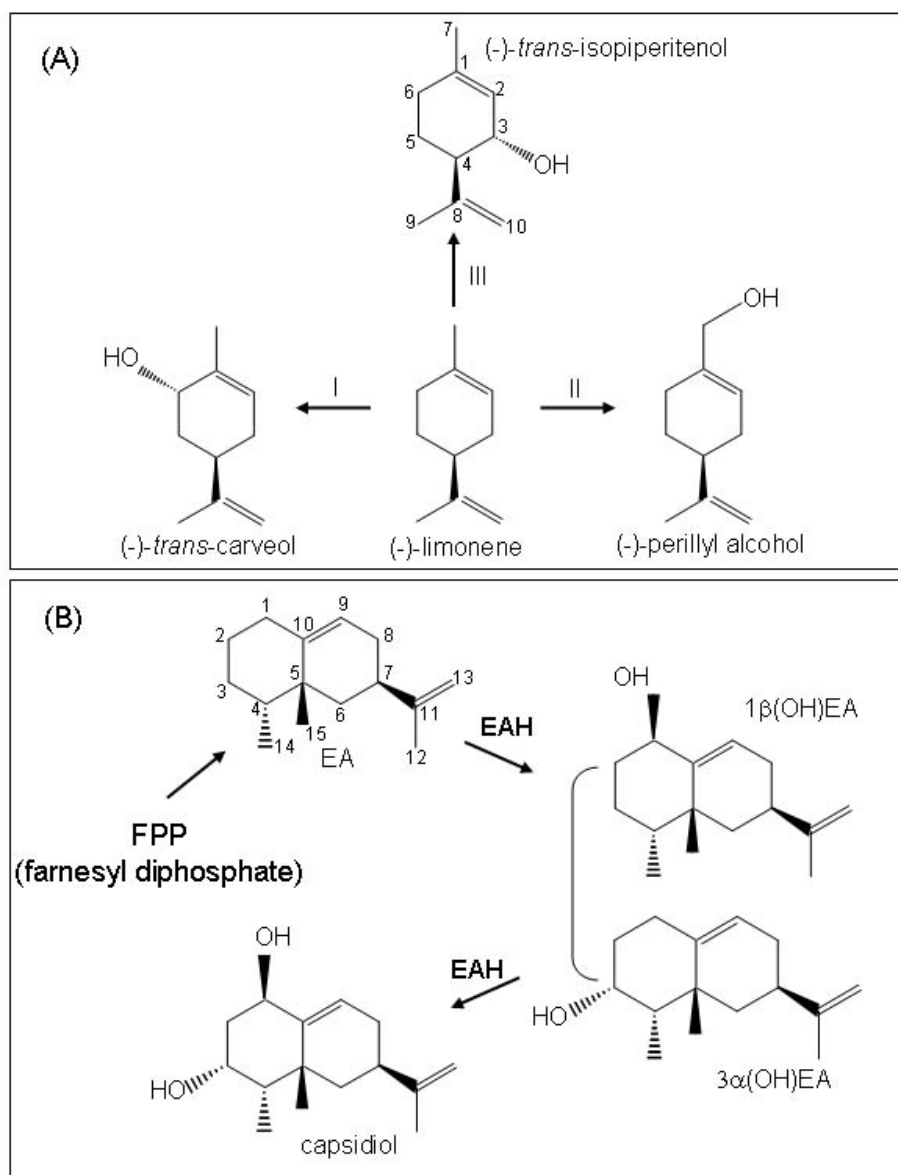


Figure 1.11. Examples of well known hydroxylation patterns of monoterpene (A) and sesquiterpene (B) products. (A) Species-specific hydroxylation of (-)-limonene (C_{10}) by regiospecific hydroxylases (Schuler, 1996): reaction I represents C6 hydroxylation in *Mentha spicata* (Bouwmeester et al., 1999); reaction II, C7 hydroxylation in *Perilla frutescens*; and reaction III, C3 hydroxylation in *Mentha piperita*. (B) Regio- and stereo-specific hydroxylations of 5-*epi*-aristolochene (EA) catalyzed by EA hydroxylase, CYP71D20 (Takahashi, 2005).

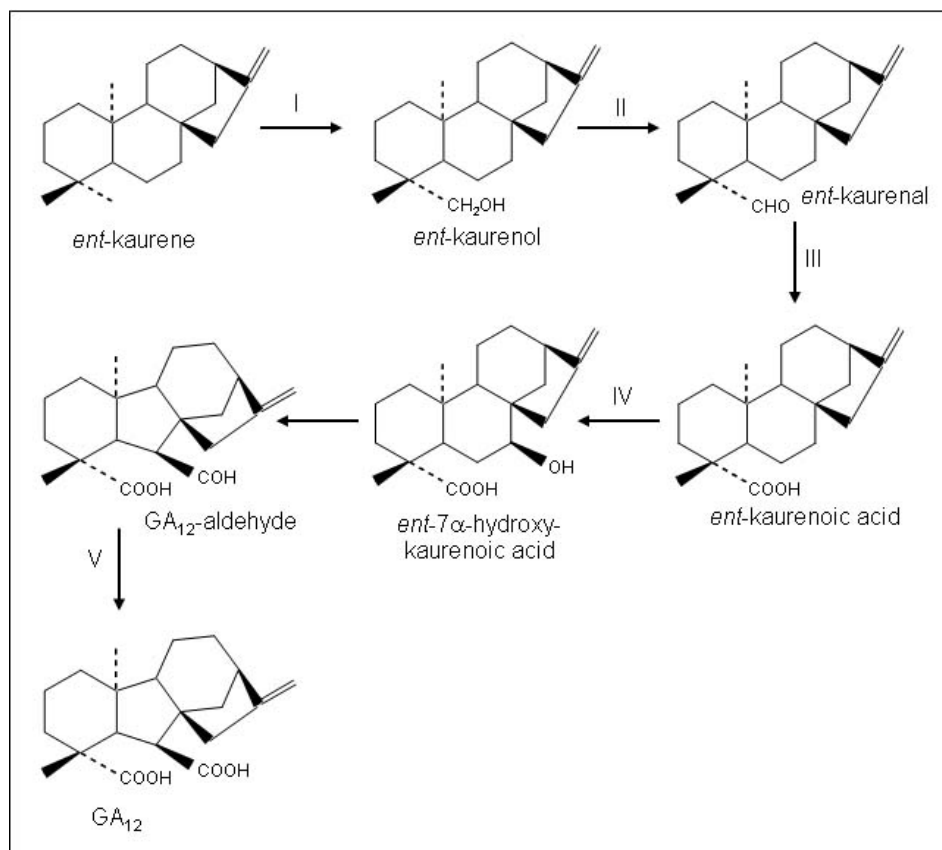


Figure 1.12. The biosynthesis of the plant growth regulatory gibberellins, like GA₁₂, appear to require multiple hydroxylation/oxidation type reactions, possibly as many as 6 independent reactions (Schuler 1996). In fact, in higher plants the introduction of these carboxyl substituents appears to be mediated by two P450s acting to introduce the hydroxyl groups, followed by successive oxidations to yield the carboxylated functionality.

CHAPTER II INNATE AND INDUCED CHEMICAL CONSTITUENTS IN MAGNOLIA LEAVES AND FLOWERS

INTRODUCTION

All plants produce terpenes, sometime in very large amounts as a single chemical species and other times as very diverse suites of compounds. Some of the terpene derivatives like carotenoids, sterols and growth hormones are essential for general plant growth and development (Giuliano et al., 1993; Li et al., 1996; Clouse, 2000; McCourt et al., 2005), and hence are classified as primary metabolites. Other terpene constituents are non-essential in the sense that genetic lesions abolishing their biosynthesis do not prevent overall growth of the plant. These compounds are nonetheless important for many ecological interactions between plants and their environment. For instance, a broad range of terpenes and terpene derivatives play roles as phytoalexins, insect attractants, repellents, pheromone-like compounds, and as attractants of predators for feeding herbivores (Elferaly and Chan, 1978; Lambers and Schepers, 1978; Facchini and Chappell, 1992; Beale et al., 2006; Schnee et al., 2006). A second layer of complexity concerning these secondary terpenes relates to their modes of regulation. Some of these terpenes are produced constitutively, while others are produced only in response to external stimuli. For example, when forest tent caterpillars feed on the poplar leaves, a mixture of volatile monoterpenes (terpenes containing 10 carbons, β -ocimene and linalool) and sesquiterpenes (terpenes containing 15 carbons terpenes including (-)-germacrene D and (*E,E*)- α -farnesene) are produced and their emission is regulated diurnally (Arimura et al., 2004a).

The terpene constituents in leaves and flowers of *M. grandiflora* have been examined previously (Elferaly and Chan, 1978; Halim et al., 1984; Yang et al., 1994; Azuma et al., 1997b). Azuma et al. (1997b) reported that the floral scent of *M. grandiflora* consisted largely of monoterpenes dominated by geraniol followed by lesser amounts of (*E*)- β -ocimene and β -myrcene, along with a few of sesquiterpenes including a 4,8-dimethyl-1,3(*E*),7-nonatriene. Interestingly, the levels of specific components and the overall terpene profile differed between attached and detached flowers. Overall, less volatiles were released from the intact flowers than those detached. Azuma et al. (1997a)

also described the inducible release of terpenes from physically damaged, but attached leaves of the *M. grandiflora*. 4,8-dimethyl-1,3(*E*),7-nonatriene was immediately emitted from the leaves after damage treatment along with β -elemene, caryophyllene, and α -bisabolene. In contrast, undamaged leaves did not appear to emit these terpene compounds in appreciable levels.

Because of our interest in understanding the regulation of terpene biosynthesis, the current experiments were designed to determine if novel or the absolute amounts of terpenes emitted from detached leaves of *M. grandiflora* changed in response to a wounding-treatment. These experiments were also coupled with an assessment of the terpenes found in leaf extracts. If volatile emissions arise directly from terpenes stored in the leaf cells, then one might expect cell extracts from control and wounded leaves to reflect this profile and this profile to remain constant. If the induced emissions come about via an induction in the biosynthesis of particular terpenes, then the terpene profile from cell extracts may not reflect the emission profile. Lastly, we have also investigated the inducibility of terpene emission from flowers in response to beetle-feeding as well as the developmental dependence of terpene emissions from detached flowers.

MATERIALS AND METHODS

Reagents and Standards

All solvents were from Fisher Scientific Inc. (Pittsburgh, PA) unless stated otherwise. Authentic terpene standards were from Sigma Chemical Company (St. Louis, MO, for α -terpineol), Bedoukian Research, Inc. (Danbury, CT, for bisabolenes), and Firmenich, Ltd. (Geneva, Switzerland, for β -cubebene).

Plant Materials and Treatments

Southern Magnolia (*Magnolia grandiflora*) leaves and flowers were collected from June through August, 2005-2007, from trees grown on the University of Kentucky (Lexington, KY) campus at variable times throughout the day. Light-green leaves with white pubescence on the abaxial side were classified as young leaf tissue. Leaves with dark-green color were considered as the more mature leaves, which usually had a brown pubescence on the abaxial side. Flowers were classified as immature (unopened petals

prior to blooming) or mature (petals fully opened and stamens leaning or fallen away from the stigma).

To mechanically wound leaves, each leaf detached was gently clamped with standard pliers for approximately one second at about 20 independent sites per leaf, then placed on a moistened paper towel in a suitable plastic dish. To examine the inducible emissions from flowers, immature flowers were collected directly from trees and the cut pedicle placed in a small beaker of water. To simulate beetle attack, 4 to 5 Japanese beetles (*Popillia japonica*) were placed onto the detached flower. The Japanese beetles were collected directly from Magnolia tree flowers. The collected leaves and flowers were incubated in open bell jars prior to collecting volatile emissions for the indicated time periods.

Terpene Analyses

To extract terpene compounds, two grams of plant material were pulverized in liquid nitrogen with a mortar and pestle. Terpenes were then extracted sequentially with 2x-5 mL of hexane:ethyl acetate mixture, the extracts combined and further purified by 2 rounds of silica column chromatography. The extract was applied to a silica column (146 mm x 7 mm, length x diameter) followed by washing with 6 ml of fresh hexane:ethyl acetate mixture. The collected eluent was then applied to a second silica column (146 mm x 7 mm), followed by 3 ml of the hexane:ethyl acetate mixture to elute hydrocarbons and oxygenated compounds. The final eluant (~20 mL) was concentrated to 2 mL under nitrogen gas flow before analysis by GC-MS. For the quantification of compounds, α -cedrene was used as an external standard.

Headspace gas analyses were performed similar to that described by Wu et al. (2006). Leaves or flowers were placed into a closed bell jar system at specified times after detaching them from the trees and volatiles were trapped onto 150 mg of Tenax TA Mesh 60/80 resin (Supelco, St. Louis, MO) for the indicated time intervals (5 to 24 hrs) with a gas flow of 500mL/min through the chamber (Fig. 2.4). Compounds trapped onto the Tenax adsorbent were eluted with 3 mL of 100% ethyl acetate and concentrated to 100 μ L.

One microliter aliquots of the various samples were analyzed for terpene constituents with a Thermo Finnigan DSQ GC/MS system (Thermo Fisher Scientific, Waltham, MA) equipped with a Restec Rtx-5 capillary column (30 m x 0.32 mm, 0.25- μ m phase thickness). For all analyses, the injector port was maintained at 220°C in the splitless mode and the initial oven temperature of 40°C (0.5 min) was increased a 4°C/min gradient to 200°C followed by a 20°C/min gradient to 300°C. Mass spectra (MS) were recorded at 70 eV, scanning from 35 to 270 atomic mass units. All compounds detected from extracts or volatile emissions were confirmed by comparing retention time and mass spectra with authentic standards, or by comparison to mass spectra reported in the NIST MS library version 2.0 (National Institute of Standards and Technology, Gaithersburg, MD) and MassFinder 2.3 software (edited by Dr. D.H. Hochmuth). For the quantification of compounds, α -cedrene was used as an external standard.

Experiments were performed twice with identical results, and representative GC traces are shown.

RESULTS AND DISCUSSION

Terpenes Associated with *M. grandiflora* Leaves

Because our aim was to determine how the terpenes accumulating within leaf and floral tissues of *M. grandiflora* correlated the emission profiles from these same tissues, the initial analysis examined the terpene profile for an extract prepared from a control tissue compared to the volatile constituents released from this same tissue type into headspace gases. For the headspace gas analysis, mature and immature leaves were detached from a locally grown tree by mid-morning, and the cut petiole packed with moistened cotton and the cotton-covered petiole wrapped with saran-wrap to prevent evaporation. Three leaves were then placed in bell jar plumbed with a purified air flow of approximately 500 mL/min and the vented gas routed through a Tenax trap to collect released hydrocarbon volatiles. Interestingly, no volatile hydrocarbons were detected from control leaves regardless of their maturity (Fig. 2.2C). In contrast, mono- and sesquiterpene (hydrocarbons and few oxygenated species) were readily detectable in extracts prepared from the control leaf tissues (Fig. 2.2A). As noted by Azuma et al (1997a), sesquiterpenes dominated the extract profiles with β -elemene and β -cubebene

making up approximately 50% of the total. Important to note, the β -elemene reported here likely reflects the endogenous level of germacrene A, not β -elemene. Under the GC temperature conditions used here, germacrene A undergoes a thermal conversion to β -elemene via a Cope re-arrangement mechanism (Takeda, 1974; de Kraker et al., 1998). Moreover, there is no biochemical precedence for β -elemene biosynthesis, yet many reports of germacrene A synthesis in plants (Rising et al., 2000; de Kraker et al., 2001; Bennett et al., 2002a; de Kraker et al., 2003). Monoterpenes were also found in the leaf extracts but at much lower amounts, like β -pinene at approximately 2.9% of the total.

In contrast to the absence of emissions from control leaves, wounded leaves emitted a very significant array of mono- and sesquiterpenes in addition to compounds like fatty acid derivatives (Fig. 2.2D and Table 2.1). As reported first by Azuma et al. (Azuma et al., 1997a), volatile emissions from wounded tissues of *M. grandiflora* leaves commences almost immediately upon wounding. Within the first 5 hours of wounding, the monoterpenes β -pinene, cineole and (*E*)- β -ocimene comprise upwards of 20% of the volatile emissions, with the sesquiterpenes β -elemene, α -gurjunene, β -caryophyllene and β -cubebene making up greater than 60% of the total volatiles (Table 2.1). The overall profile of volatiles does not change substantially over the next 19 h except for an increase in (*E*)- β -ocimene from 3.5 to greater than 20% of the total hydrocarbon emissions. The emission level of *cis*-jasmonone, a fatty acid derivative, also increased ~4-fold during this time period from 3.5 to greater than 15% of the total released compounds.

Extracts from 24 h wounded leaves generally reflected the volatile emission profiles, but not completely (Table 2.1). For instance, the monoterpene β -pinene observed in the emissions of wounded leaves was also dramatically elevated in the extracts prepared from the wounded leaves. Alpha-gurjunene, a sesquiterpene, exhibited a similar trend – present in control extracts, dramatically increased upon wounding and becoming a dominate volatile in the emission profile. Interestingly, the relative amount of compounds like β -elemene and β -cubebene, evident at relatively constant amounts in extracts from control and wounded leaves, tended to decline in the volatile phase over the time course of these experiments, perhaps compensated by the increased levels of α -gurjunene. Equally important to note, although we observed 13 to 20% of the volatile

terpenes as β -cubebene, Azuma et al. (Peigler, 1988; Azuma et al., 1997a; Dieringer et al., 1999) failed to report any such compound. Conversely, they reported on the volatile release of the irregular sesquiterpene dimethyl nonatriene from leaves, which we were not able to confirm (Table 2.3). These differences may be due to differences in the growth of *M. grandiflora* at the geographic distinct locales, or perhaps a true genetic difference between the various *M. grandiflora* lines.

Volatiles Associated with *M. grandiflora* Flowers

Magnolia flowers develop over the course of several weeks from their bud stage to the completely open stage. For our purposes, we chose to compare the volatile emissions from fully opened flowers to those just prior to opening, a process requiring about 4 to 6 days for the petals to uncurl from surrounding the reproductive cone. These 2 stages were also chosen in part because at no time in our monitoring of the campus trees were Japanese beetles (*Popillia japonica*) associated with the fully open flowers. However, beetles did colonize onto the earlier, un-opened flowers. No control over natural pollination or discrimination between pollinated or non-pollinated flowers was attempted. Somewhat surprising however, volatile emissions over a 24 h period from fully opened and closed control (no beetles) flowers differed only by the levels of minor terpenes making up a few percent of the total emissions (Fig 2.3A-B and Table 2.2). While both types of flowers emitted very substantial amounts of β -pinene and *cis*- β -ocimene, the un-opened flowers emitted a small amount of a novel suite of mono- and sesquiterpenes (Fig. 2.3B and Table 2.2). These included α -phellandrene, pinene oxide, terpineol and geraniol. Un-opened flowers colonized by beetles emitted many of the same terpenes, but also emitted a distinct profile of terpenes as well (Fig. 2.3C). *Cis*- β -ocimene emission from the infested flowers was approximately half that of control flowers. And while many of the minor terpenes found in unopened, beetle-free flowers were lost, the levels of citral, β -elemene, verbenone, caryophyllene, cubebene and farnesene were very significantly elevated in the emissions of the beetle infested flowers (Fig. 2.3C and Table 2.2).

The overall picture emerging from this analysis of *M. grandiflora* flowers is that the immature flowers emit minor amounts of novel terpenes like terpineol and thujenone

as well as some found in multiple stages of flower development like verbenone. These compounds might in fact act to serve in attracting beetles like *Popillia japonica*, which are considered important possible pollinators for this species of flowering trees (Peigler, 1988; Dieringer et al., 1999). Japanese beetles have previously been shown to respond to volatile cues (Ladd, 1980; Loughrin et al., 1996; Kappers et al., 2005; Schnee et al., 2006), including terpenes, and it would be interesting to determine if they were able to discriminate between the terpene profiles emitted from the immature and fully mature *M. grandiflora* flowers. The appearance of novel, volatile terpenes, plus the elevation of others, in the emissions of beetle-infested flowers are reminiscent of other herbivore-induced volatile emissions. However, in many of these previous reports, the feeding induced volatiles serve to attract predator wasps and hence, the volatile emissions are considered plant-mediated, defense mechanism. In the current instance, the minor terpene emissions by the immature flowers may be important for attracting pollinating insect species like the Japanese beetles. The role of the beetle-induced terpenes is not obvious, but could serve to attract predatory species (Kappers et al., 2005; Schnee et al., 2006), or may play some role as a feeding deterrent.

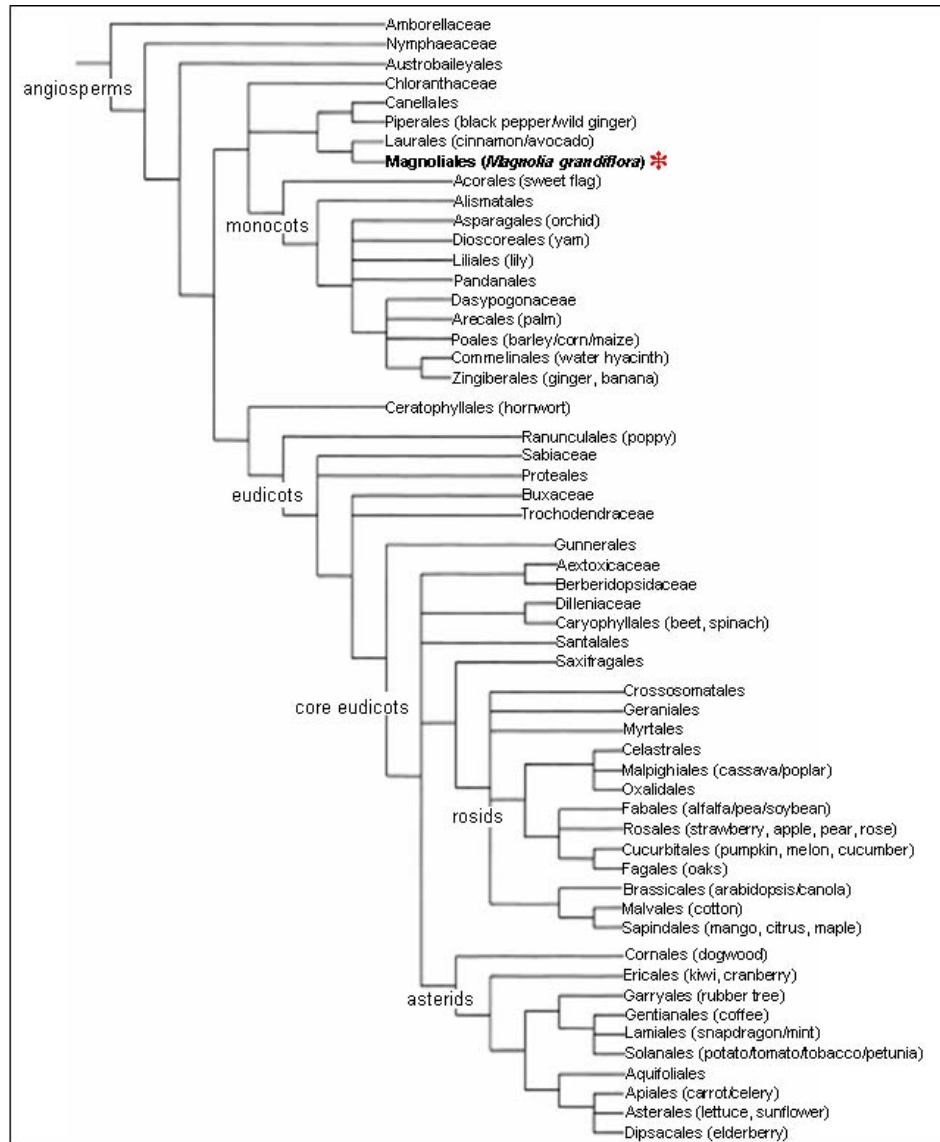


Figure 2.1. Interrelationships of the orders and some families in plant kingdom supported by jackknife or bootstrap frequencies above 50% in large-scale analyses of angiosperms. All except five of the clades are supported by the Soltis et al. (2000) analysis of 18S rDNA, *rbcL*, and *atpB* sequences from wide sample of angiosperms. *Magnolia grandiflora* signified with an asterisk positioned at the basal clade with the order of Laurales. (adopted and re-drawn from APG II, 2003)

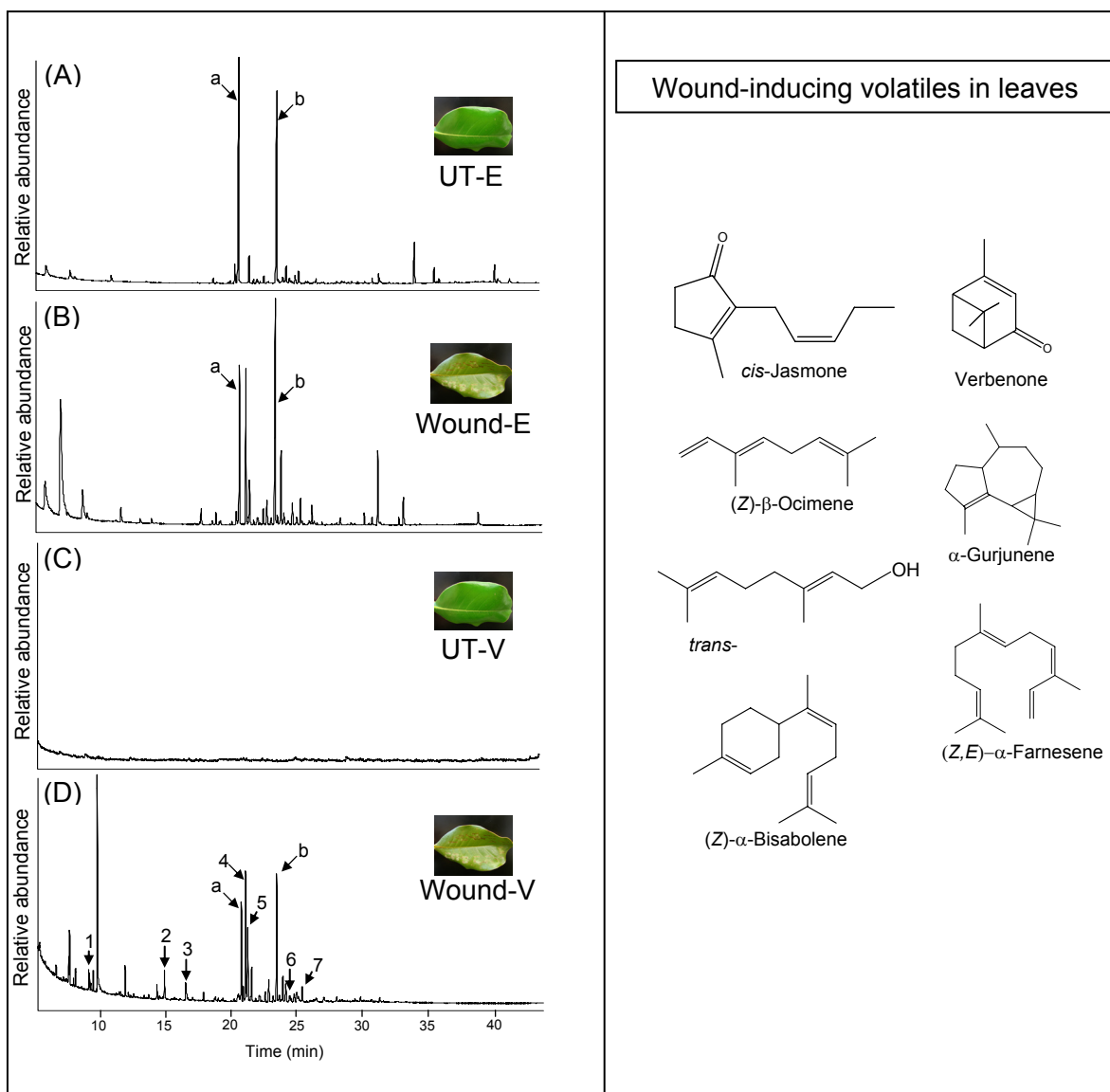


Figure 2.2. Comparison of chemical profiles in untreated- and wound-treated leaves. Chemicals either increased over 2% in relative amounts or *de novo* accumulated were drawn in the right column. UT-E (A) indicates a chemical profile of the extract from untreated leaves; Wound-E (B), a chemical profile of the leaf extract at 24 hrs after wound-treatment; UT-V (C), a chemical profile of the volatiles collected from untreated leaves; Wound-V (D), a chemical profile of the volatiles collected at 24 hrs after wound-treatment. 1, (*Z*)- β -ocimene; 2, verbenone; 3, *trans*-geraniol; 4, *cis*-jasmone; 5, α -gurjunene; 6, (*Z,E*)- α -farnesene; 7, (*Z*)- α -bisabolene. Peaks a (β -elemene) and b (β -cubebene) were indicated to synchronize retention times in the figures. Entire lists of components and their quantities were presented in Table 2.1.

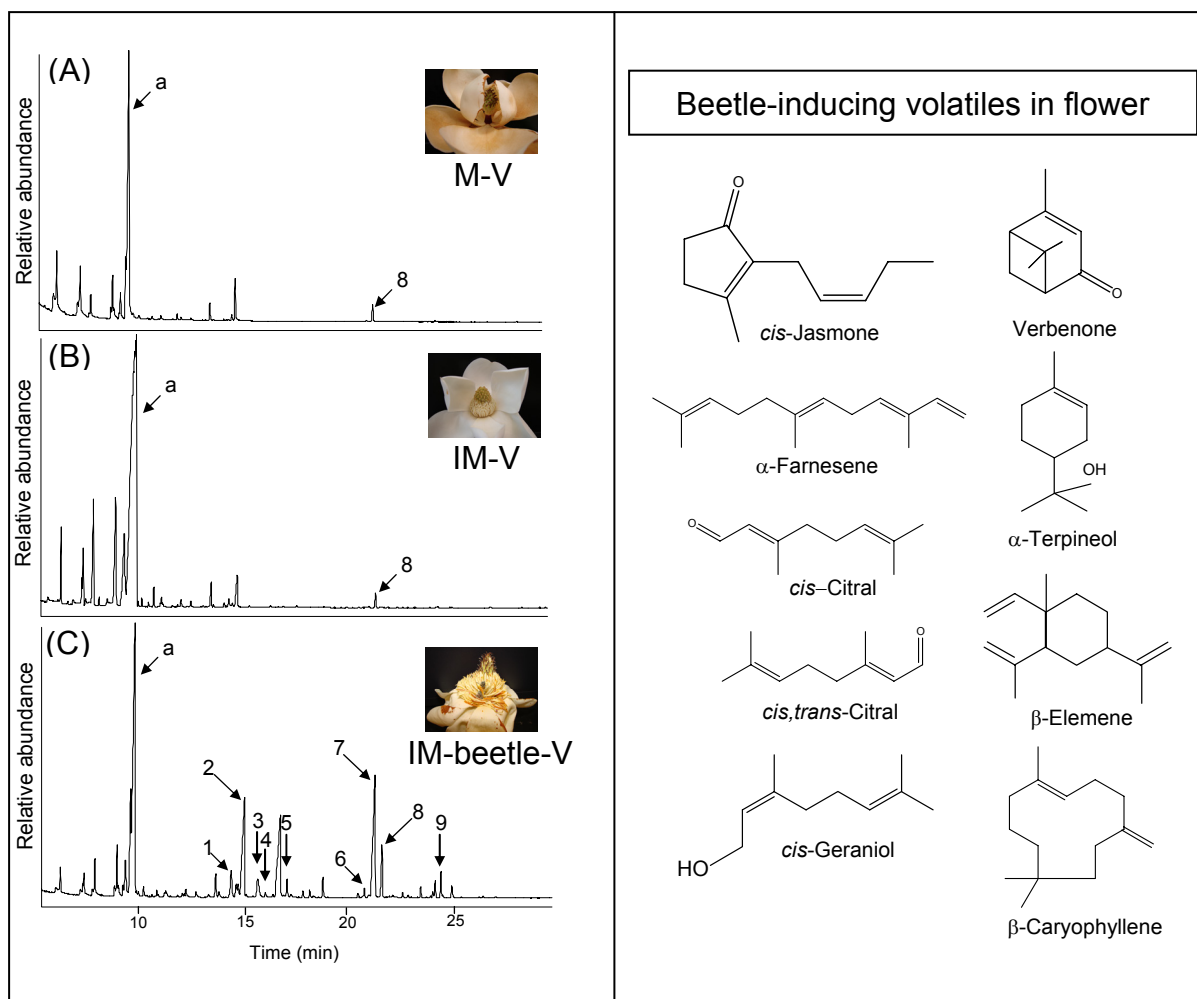


Figure 2.3. Comparison of chemical profiles in control and beetle-feeding flowers.

Chemicals either increased over 2% in relative amounts or *de novo* accumulated were drawn in the right column. For the M-V (A), volatiles were collected by headspace trapping from mature flower, which means all petals were fully opened and stamens were fallen down; for the IM-V (B), volatiles were trapped by following same procedures used for the M-V from immature flower, which has all stamens and closed petals; and for the IM-beetle-V (C), volatiles were trapped from immature flower challenged with 4 to 5 beetles. The image for the IM-V was taken after opening the petals to show stamens attached. 1, α -terpineol; 2, verbenone; 3, *cis*-geraniol; 4, *cis*-citral; 5, *cis,trans*-citral; 6, β -elemene; 7, *cis*-jasmone; 8, β -caryophyllene; 9, α -farnesene. Peaks a ((*E*)- β -ocimene) and 8 (β -caryophyllene) were represented to synchronize retention times in the figures. Entire lists of components and quantities were presented in Table 2.2.

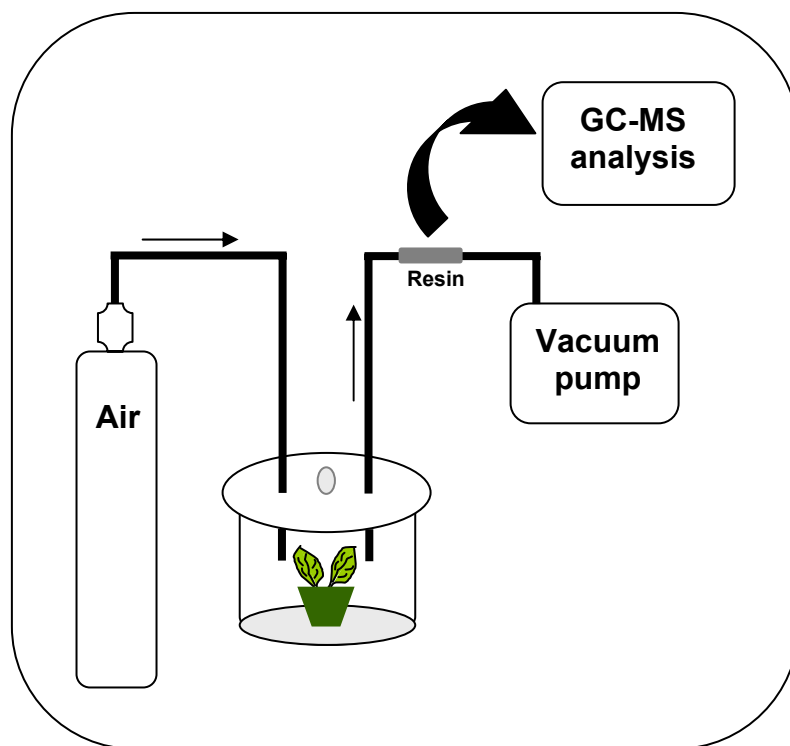


Figure 2.4. A simplified cartoon depicts how volatiles were collected by head-space trapping. Purified air was forced into a bell jar where plant materials were placed and the plant volatiles trapped onto a Tenax resin inserted into vacuum line. The volatiles trapped onto the Tenax resin for specific lengths of time were release by solvent washes and examined by GC-MS.

Table 2.1. Terpenoid profiles from *M. grandiflora* leaves.

Extracts were prepared from control, untreated (UT) leaves or those incubated for 24 hrs after wounding (W-24 hrs), and the terpene constituents identified and quantified by GC-MS. The volatiles emitted from leaves were collected immediately after wounding for 5 hrs (W-5 hrs), then changing out the trap and continuing to collect volatiles for an additional 19 hr. The traps were eluted with solvent and the terpene profiles determined by GC-MS. All compounds were identified by AMDIS using the NIST MS library version 2.0 and quantified relative to external and internal standards. Experiments were performed twice with identical results, and representative results are shown.

COMPOUND	Leaf			
	Extract ($\mu\text{g/g}$ fresh wt)		Volatile ($\mu\text{g/g}$ fresh wt)	
	UT	W-24 hrs	W-5 hrs	W-24 hrs
1R- α -Pinene	122	728	-	-
β -Pinene	35.4	261.1	6.6	1.7
Cineole	24.2	53.8	4.6	1.5
(Z)- β -Ocimene	6.3	-	2.1	6.7
Mytenol	-	5.3	-	-
α -Terpineol	3.0	-	1.0	0.5
Verbenone	-	-	0.6	1.1
trans-Geraniol	-	-	-	0.9
τ -Elemene	8.8	12.5	0.5	-
Aciphyllene	2.1	7.6	0.5	-
α -Cubebene	3.2	0.6	0.3	0.1
β -Bourbonene	-	-	0.3	0.2
Camphene	-	-	1.0	-
β -Elemene	379.7	146.7	7.5	3.2
cis-Jasmone	-	-	2.2	5.0
α -Gurjunene	38.2	168.9	12.9	3.3
β -Caryophyllene	5.8	44.5	3.5	1.2
β -Bergamotene	10.0	-	0.6	-
α -Guaiene	10.4	13.3	0.6	0.4
α -Caryophyllene	6.3	14.5	0.9	0.4
β -Cubebene	306.2	217.9	12.2	4.2
Longifolene	-	7.8	-	0.2
Aromadendrene	-	16.0	-	-
(Z,E)- α -Farnesene	-	-	-	0.5
δ -Cadinene	14.8	22.1	0.5	0.3
α -Patchoulene	19.8	-	-	-
cis- α -Bisabolene	-	30.1	0.6	0.5
Palustrol	-	2.2	-	-
Nerolidol	4.1	6.4	-	-
Germaene D-4-ol	6.5	17.8	-	-
Dehydrosaussurea lacton	67.7	32.4	-	-
Costunolide	33.0	17.0	-	-
Tomentosin	214.0	-	-	-
TOTAL	1211.5	1171.3	59.2	32.23

Table 2.2. Terpenoid profiles from *M. grandiflora* flowers.

Flower volatiles were prepared from immature (IM-V) and mature (M-V) for 24 hrs or incubated for 24 hrs after feeding of beetles into immature flower (IM-beetle-V) and the terpene constituents were identified and quantified by GC-MS. The traps capturing volatiles were eluted with solvent and the terpene profiles determined by GC-MS. All compounds were identified by AMDIS using the NIST MS library version 2.0 and quantified relative to external and internal standards. Experiments were performed twice with identical results, and representative results are shown.

COMPOUND	Flower		
	IM-V	IM-beetle-V	M-V
	µg/flower/ 24hr	µg/flower/ 24hr	µg/flower / 24hr
1R- α -Pinene	35.1	48.7	99.7
β -Pinene	127.5	147.5	189.2
α -Phellandrene	3.5	-	-
(+)-Limonene	101.8	186.9	144.1
(Z)- β -Ocimene	35.1	107.1	39.3
(E)- β -Ocimene	773.4	1836.3	603.3
(Z)- β -Terpineol	3.5	21.6	4.7
α -Terpinene	-	25.9	10.0
α -Pinene oxide	2.3	-	-
Terpinolene	9.4	-	6.8
α -Pinene epoxide	8.2	-	0.6
Chrysanthone	1.2	-	5.3
3-Thujen-2-one	1.2	-	1.8
Pinocamphone	18.7	53.5	13.2
(-)-4-Terpineol	1.2	15.6	-
α -Terpineol	2.3	65.5	-
Myrtenol	5.9	49.1	1.5
Verbenone	29.3	334.1	5.1
cis-Geraniol	1.2	63.2	-
cis-Citral	-	6.8	-
cis,trans-Citral	-	20.9	-
β -Elemene	-	17.2	-
cis-Jasmone	-	229.6	-
α -Gurjunene	-	18.6	-
β -Caryophyllene	9.4	92.5	1.6
β -Cubebene	-	21.7	-
α -Farnesene	-	31.3	-
TOTAL	1170.6	3393.7	1126.2

Table 2.3. A summary of the volatile compounds emitted from *M. grandiflora* leaves and flowers as reported by Azuma et al. (1997).

No.	Leaf	
	Terpenes	Amount ($\mu\text{g}/12\text{hr}/100\text{cm}^2$ leaf area)
1	β -Myrcene	0.24
2	Limonene	0.09
3	4,8-Dimethyl-1,3(<i>E</i>),7-nonatriene	2.48
4	Bicyclo germacrene	4.18
5	β -Elemene	7.65
6	Caryophyllene	2.73
7	α -Humulene	0.15
8	γ -Cadinene	2.5
9	α -Bisabolene	6.23

No.	Flower	
	Terpenes	Relative amount
1	β -Myrcene	12.7
2	(<i>Z</i>)- β -Ocimene	0.7
3	(<i>E</i>)- β -Ocimene	14.4
4	Perillene	1.5
5	Linalool	0.6
6	Nerol	0.7
7	Geraniol	19.7
8	Neral	0.6
9	Limonene	9.9
10	Sabinene+ α -Pinene	3.3
11	Terpinolene	2.1
12	β -Pinene	1.5
13	α -Terpineol	1.0
14	Isopinocampone	1.0
15	Verbenone	5.1
16	4,8-Dimethyl-1,3(<i>E</i>),7-nonatriene	0.5

CHAPTER III CLONING AND FUNCTIONAL CHARACTERIZATION OF NOVEL TERPENE SYNTHASE GENES FROM MAGNOLIA GRANDIFLORA

INTRODUCTION

Almost any report on the identification of a new terpene compound or the characterization of a terpene biosynthetic enzyme begins with a comment about the structural complexity and diversity of terpene metabolism in general. But rarely do such comments sufficiently capture what is meant by the term complexity, nor how this complexity arises. The bulk of this arises from the terpene synthases that initiate the biosynthesis of different classes of terpenes (i.e. mono-, sesqui- and diterpenes), and more importantly, how the terpene synthases impart exquisite regio- and stereo-specificity into the biosynthesis of specific members within a terpene family. Valencene synthase and 5-*epi*-aristolochene synthase, for example, are related to one another based on the biosynthesis of sesquiterpenes within the eremophilene class of compounds. Sesquiterpenes in the eremophilane class are fused (bicyclic), decalin ring structures which harbor methyl and isopropenyl substituents in the same regio positions, but with alternative stereo configurations. Valencene also contains a double bond in the A-ring, which is mirrored in the B-ring of aristolochene. One can imagine other possible ring structures (class designations) and alternative arrangements of the various substituents and double bonds, hence creating a wealth of chemical complexity just within this single class of sesquiterpenes. Equally intriguing to consider are the prospects that the synthase enzymes responsible for the biosynthesis of these similar yet distinct sesquiterpenes are evolutionarily related. As a first approximation, one might assume that these two enzymes arose from a common ancestral gene, which underwent a duplication event followed by divergent evolution to give rise to the unique biochemical specificities. If this were so, and we could delineate an evolutionary pathway for how these enzyme specificities arose, then we should, in principle, be able to use this same pathway/strategy to evolve alternative and new terpene synthase specificities.

One approach to study the evolution of terpene synthases has been to define the structural features of these enzymes that impart catalytic specificities to the respective enzymes. If regions or domains controlling catalysis can be identified and these in fact

serve as “hotspots” for the evolution of catalytic specificities, then these regions of the genes might show different rates of mutation over evolutionary time relative to other regions of the protein not directly related to catalysis. First, such regions or domains must be identified. Two recent studies illustrate efforts in this regard. Yoshikuni et al. used a random approach to mutate all 19 amino acids within the putative active site of a multifunctional terpene synthase and observed that select mutants in fact reduced the product diversity of these enzymes, but did not alter or uncover new product specificities (Yoshikuni et al., 2006). In contrast, Greenhagen et al. evaluated the role of second-tier residues on terpene synthase specificities (Greenhagen et al., 2006). Second-tier residues were defined as those residues outside the active site, yet oriented in such a manner to influence active site residues, the contour of the active site pocket, and the catalytic cascade yielding reaction product specificities. These investigators were able to demonstrate that reciprocal mutations of 8-9 residues were sufficient to interconvert the catalytic specific of 5-*epi*-aristolochene synthase to that of premnaspirodiene synthase and vice versa. Together, these two studies would suggest that evolution of reaction specificity within the terpene synthase gene families could arise as a consequence of mutations within and outside the active site. Moreover, the residues outside the active site that influence catalysis were clearly defined as falling within specific regions and not the consequence of a random distribution throughout the entire protein.

The notion of divergent evolution rather than convergent evolution for the terpene synthases is supported by the conservation of sequence similarity and gene structure amongst the diverse terpene synthases isolated from gymnosperms and angiosperms to date. Bohlmann et al. (1998) initially categorized terpene synthases into six subfamilies, *TPS*-a through *TPS*-f, based on amino acid sequence similarities and assuming all terpene synthases originated from a common origin. More recently, Trapp and Croteau (2001) extended this comparison by examining the intron-exon structures for 37 terpene synthases representing 18 *TPS* genes in gymnosperm species and 19 genes in angiosperm species. Their analyses included documenting the overall number and position of introns, as well as intron phasing. Class I *TPS* genes were described as containing 12-14 introns and consisting mainly of diterpene synthase genes from gymnosperms and angiosperms. Class II *TPS* genes contain 9 introns inserted in a conserved pattern and are comprised

predominately of monoterpene and sesquiterpene synthase genes found in gymnosperms. The Class III *TPS* genes harbor 6 introns and include mono-, sesqui-, and diterpene synthase genes all from angiosperm species (Trapp and Croteau, 2001). These authors also suggested that the first progenitor *TPS* genes contained 14 introns and a CDIS domain, a specific DNA sequence found internally in diterpene synthase genes of unknown function. Furthermore, they proposed that all *TPS* genes of extant angiosperms evolved from such an ancient gymnosperm gene through an 'intron loss' mechanism. A more recent classification of 40 *TPS* genes annotated in the *Arabidopsis thaliana* genome, categorizing 29 *TPS* genes into Class III, 3 genes into Class I, and 8 as pseudogenes, has served to extend this model of *TPS* gene evolution (Aubourg et al., 2002).

To date, many of terpene genes used in any of the phylogenetic analyses have been isolated from gymnosperms and several relatively advanced families of angiosperms. Moreover, the angiosperm families represent a very limited taxonomic range without much consideration for basal angiosperms (Cronquist, 1988). A more complete phylogeny for the *TPS* gene families would thus likely benefit from inclusion of *TPS* genes from putative ancient and diverse lineages. Basal angiosperms include plants within the Magnoliaceae, which are considered ancient because of the primitive organization of their floral parts (Crane et al., 1995), because of the molecular phylogeny of their *rbcL*, *atpB*, and 18S rDNA loci (Bremer et al., 2003), and because of their representation within fossil records (Golenberg et al., 1990). For instance, the fossil records suggest that the genus Magnoliaceae existed 20-25 million years ago, yet the present day *rbcL* gene found in *Magnolia grandiflora* exhibits only 22 bp differences relative to the likely progenitor gene found in the fossil species *Magnolia latahensis*. Such information thus provides an evolutionary context for comparisons of extant *TPS* genes in *Magnolia* species to the *TPS* genes found in other angiosperms and gymnosperms (Golenberg et al., 1990).

The aim of the work described here was hence to isolate and functionally characterize terpene synthase genes from *M. grandiflora*, and then to use this information to assess prevailing models for the evolution of *TPS* gene structure and function according to the previous studies (Bohlmann et al., 1998a; Trapp and Croteau, 2001;

Greenhagen et al., 2006). Given these models, we anticipated that the *TPS* genes of *M. grandiflora* would more closely resemble the Class I genes of gymnosperms in possessing many introns, up to 14, and to encode for terpene synthases catalyzing multiple reaction product profiles. What we found is that the three *M. grandiflora* *TPS* genes characterized here do not fit with these prevailing models in terms of overall intron-exon organization, and biochemically these Magnolia enzymes catalyze reactions as sophisticated as any other *TPS* reported to date. Our data, hence, argue for alternative origins and pathways in the evolution of these highly conserved genes and biosynthetic pathways.

MATERIALS AND METHODS

Plant Material, Reagents, and Standards

Plant leaves and flowers were collected from June through August, 2005 to 2007 from Southern Magnolia trees (*Magnolia grandiflora*) grown on the University of Kentucky campus at variable times throughout the day. Light-green leaves with white pubescence on the abaxial side were classified as developing leaf tissue. Larger leaves with dark-green color were considered as the more mature leaves, which had a brown pubescence on the abaxial side. All solvents were from Fisher Scientific Inc. (Pittsburgh, PA) unless stated otherwise. Authentic terpene standards were from Sigma Chemical Company (St. Louis, MO, for α -terpineol), Bedoukian Research, Inc. (Danbury, CT, for bisabolenes), and Firmenich, Ltd. (Geneva, Switzerland, for β -cubebene).

RNA Isolation

Leaves were collected, frozen immediately in liquid nitrogen, and then pulverized to make a fine powder using a chilled mortar and pestle. Total RNA was isolated by using a SIA (sodium isoascorbate acid:guanidium-phenol-chloroform) method with minor modifications (Suzuki et al., 2003) and 200 to 250 μ g of total RNA was usually obtained from 1 g of leaf tissues. One gram of a leaf was pulverized in chilled mortar and pestle. Five volumes of extraction buffer (500 mM sodium isoascorbate, 100 mM Tris-HCl, pH 8.0, 10 mM EDTA, 5% (v/v) β -mercaptoethanol, and 2% (w/v) SDS) were added directly into the powder making a buffer-powder slush. About 500 μ L of the slush

was transferred to a fresh 1.5 mL eppendorf tube and equal volume of chloroform:isoamylalcohol (24:1, v/v) added. After phase separation by centrifugation for 20 min at 4°C, the supernatant was extracted a second time using an equal volume of chloroform:isoamylalcohol (24:1, v/v). The supernatant was then mixed with 0.5 volume of 66.7% (w/v) guanidine isothiocyanate, equal volume of water-saturated phenol, and 0.15 volume of 3 M sodium acetate, pH 5.2. After 3 min incubation at room temperature, 0.2 volume (of the added phenol) of chloroform:isoamylalcohol (24:1, v/v) was added and shaken vigorously by hand. After incubation on ice for 15 min, the mixture was centrifuged for 20 min at 4°C. To remove polysaccharides from the supernatant, one-third volume of 1.2 M NaCl-0.8 M sodium citrate was added, incubated shortly, then centrifuged for 10 min at full speed. Total RNA was precipitated from the resulting supernatant with two-thirds volume of isopropanol. After 10 min incubation, total RNA was pelleted by centrifugation for 20 min at 4°C. The pellet was washed with 75% ethanol twice and dissolved in RNase-free water. Total RNA was quantified by spectrophotometry. When necessary, mRNA was purified from 100 µg of total RNA isolated using Oligotex mRNA mini kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations.

Isolating Initial *TPS* cDNA Fragments

First-strand cDNAs were synthesized from Poly A⁺ RNA using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA) and Oligo-(dT) primer (Invitrogen, Carlsbad, CA) in a total volume of 20 µL at 42°C for 50 min according to the manufacturer's recommendations. Polymerase chain reactions were performed in a total volume of 50 µL reaction containing 2 µL of first-strand cDNA, 0.25 mM primers, 0.2 mM dNTPs, 1x *Pfu* PCR buffer, and 2.5 units *Pfu* Ultra High-fidelity DNA Polymerase (Stratagene, La Jolla, CA). Several combinations of degenerate oligonucleotides were used to amplify terpene synthase cDNA fragments. Three forward oligonucleotides were designed: PrimerA-F, 5'-GAG CTT AGC (C/G)A(A/T) (C/G)TT TCA AAG TGG TGG-3'; PrimerB-F, 5'-TAG C(C/G)A (A/T)(C/G)T TTC AAA GTG GTG GAA-3'; PrimerC-F, 5'-TGG ATA TTA GGA GT(C/G) TAC TT(C/T) GAG CC-3' and three reverse primers: PrimerA-R, 5'-GTA GCA TA(A/G) GCA TCA TA(A/C/T) GTG TCG TC-3';

PrimerB-R, 5'-ATC (A/C)AT (A/G)G CA(C/G) TAA TTT CCC ACC T-3'; PrimerC-R, 5'-GCA ATC TTC ATA TAC TC(A/C/T) TCA ATG TTG G-3'. Amplicons were (dA)-tailed and ligated in pGEM[®]-T Easy vector (Promega, Madison, WI), transformed into DH5- α competent cells. Cloned inserts were sequenced in both directions using the BigDye Terminator Cycle sequencing Kit (Perkin-Elmer, Waltham, MA) with an automated sequencing analyzer (ABI PRISM[®] 310 Genetic Analyzer, Applied Biosystems, Foster city, CA). DNA sequence information was analyzed by BLASTX searches to the NCBI database (<http://www.ncbi.nlm.nih.gov/BLAST>). Fragments exhibiting significant matches (E values $\leq e^{-4}$) to terpene synthases in Genbank database were considered significant. All custom oligonucleotides were synthesized by Integrated DNA technologies (Integrated DNA Technologies, Coralville, IA).

5'- and 3'-RACE for *Mg25*, *Mg17* and *Mg11* Genes

A BD SMART[™] RACE cDNA Amplification Kit (Clontech, Palo Alto, CA) was used to obtain the sequence information necessary to obtain full-length cDNA clones for the prioritized magnolia terpene synthase cDNA fragments. Adaptor-ligated double-strand cDNAs was generated and adaptor-specific element provided by the kit was used for amplifications combined with gene specific oligonucleotides. In some cases, a gene specific first-strand cDNA (GSD) was synthesized to increase the corresponding cDNA pools using gene-specific oligonucleotides instead of 5'- or 3'-CDS. The 3'- and 5'-ends were amplified with Advantage[®] 2 PCR Kit (Clontech, Palo Alto, CA). The reaction mixtures in 50 μ L included 5 μ L of 10X PCR buffer, 200 nM dNTP mix, 1 μ L of Advantage[®] 2 Polymerase Mix, 200 μ M of adaptor-specific primer, 200 μ M of gene-specific primer, and 2.5 μ L of 100 to 300-fold diluted RACE-ready cDNA. The first amplification was completed in a condition as follows: 95°C for 3 min; 30 cycles of 30 sec at 95°C, 40 sec at 60 to 65°C, and 2 to 3 min at 72°C. Sequentially then, nested PCR was performed with 5 μ L of 200-fold diluted cDNA from the first amplification combined with nested oligonucleotides sets at 2°C to 3°C lower annealing temperature. The amplified products were cloned and sequenced as described above.

In order to obtain 3'-end amplicons, gene-specific forward primers were designed: for *Mg25*, 5'-ACC TAT GAC GTA TAC GGC ACA TTG GAG G-3' and 5'-

GAA ATA TAT ACT ACT GCT ATT GAG AGA TGG G-3'; for *Mg17*, 5'-CTT GCA TTG GAG GGT ACT AAG ATT AGA GGC-3'; and 5'-CAA TTT CAG TGC AAT ATT AAG CAG CGA CAC ACA CG-3' for *Mg11*. To obtain specific fragments, nested gene-specific primers were considered: 5'-GGG CTT CTG AAT GCA GTC GAT CAA ATC-3' for *Mg25*; for *Mg17*, 5'-AAC ATC GTT CAA AAT GTA TAC CAG GGA CAA GTC AG-3'; and for *Mg11*, 5'-TCT TTC ACC ACA GTG GCA CCT CAA ATC TGC-3'. Based on the sequence information from 3'-end amplicons, three gene-specific reverse oligonucleotides were designed for 5'-RACE: 5'-CAT CGT CCT TCC CAT TGA TGT GGT CAT CAT AC-3' for *Mg25*, 5'-GTA GCC TTT GGC CAA ATT CGT AAT AAC TTT CTT TAT GTA GGG-3' for *Mg17*, and 5'-GTC AAA ACC ATC ATC ATT CCC ATT GAT GTG G-3' for *Mg11*. For nested gene-specific primers, 5'-TAG GCT ACA CCA AGG CGT TGA ATT GC-3' for *Mg25*, 5'-GTA TTC GGT AGG ATT TTC CTT CCC TGC-3' for *Mg17*, and 5'-CGC ATA AAG GAC AGG GTG AAG AAC TGA AAG G-3' for *Mg11*. Methionine codons served to locate the most upstream start site, and the various ORFs were assessed according to those putative start sites. The first stop codon in any ORF was regarded as the translation stop codon.

Isolating Full-Length *Mg25*, *Mg17* and *Mg11* cDNAs

Mg25 full-length cDNA was amplified in the reaction containing *Pfu* Ultra High-fidelity DNA Polymerase with forward (5'-CACC ATG GAT AGT CCC ACT ACT CAA AGG CCA AAC ATG GAG-3') and reverse primers (5'-TTA AAG TGG AAT AGG GTC CAC AAA CAG TGA TG-3') using 3'-RACE Ready cDNA as a template. PCR condition was as follows: 3 min at 95°C, 30 cycles of 30 sec at 95°C, 30 sec at 52°C, 100 sec at 72°C, and final extension for 5 min at 72°C. The amplified cDNA (1.6 kb) was ligated into pET100/D-TOPO vector (Invitrogen, Carlsbad, CA) and transformed into DH5- α competent cells for plasmid amplification. Insert DNAs were sequenced in both directions using automated DNA sequencing. For *Mg17* harboring a putative signal sequence, two different cDNA fragments were amplified and cloned. Full-length *Mg17* (1.8 kb) was amplified with forward primer 5'-ATG GCA CTT AAG CTC CTC TTC CAA TGC-3' and reverse primer 5'-TTA GCT CTC CAT TAA TTG AAT AGG TTC C-3'; for signal sequence truncated cDNA (1.6 kb), forward primer 5'-ATG CGA CGC

TCG GCT AAT TAC C-3' was used with the same reverse primer used for full length amplification. The signal sequence domain was hence removed and a methionine was added to the N-terminus of the truncated ORF fragment creating an artificial translation start site. Both cDNAs were ligated into pET100/D-TOPO or pET43(a) for functional characterization. For *MgII* (1.6 kb), forward primer 5'-ATG GAT AGT CCT GCT ACT CAA AGG-3' and reverse primer 5'-CTA AAG TGG AAT AGG GTC CAC-3' were used to amplify full-length cDNA, and site-directed mutagenesis was performed using the QuikChange methodology (Stratagene, La Jolla, CA) to modify an internal stop codon near the N-terminus to tryptophan (W), a conserved amino acid at this position amongst terpene synthases. The mutagenized cDNA was cloned and expressed in three different expression vectors including pET100/D-TOPO, pET28(a) and pET43(a). Sequence alignments were performed with ClustalW (<http://www.ch.embnet.org/software/CustalW.html>) and visualized with GENEDOC (<http://www.psc.edu/biomed/genedoc>).

Heterologous Expression of TPS cDNAs in *E. coli* and Enzyme Assays

Full-length cDNAs cloned into various expression vectors were transformed into BL21 StarTM(DE3) (Invitrogen, Carlsbad, CA). Cell cultures initiated from single transformation colonies were grown to an OD₆₀₀ equal to 0.6 at 37°C before addition of 0.4 mM IPTG for pET43(a), or 1 mM IPTG for pET100/D-TOPO and pET28(a), then incubated overnight at 22°C. The cells were separated from the media by centrifugation and the pellets stored at -80°C, if not used immediately. The pellets were resuspended and sonicated six times for 10 sec in lysis buffer (80 mM potassium phosphate, pH 7.0; 10 mM meta bisulfate; 10 mM sodium ascorbic acid; 1 mM PMSF) and the clarified lysate was collected by centrifugation at 12,000g for 15 min. The lysate was then loaded onto a Ni²⁺ affinity resin (Novagen, Madison, WI) to purify His-tagged recombinant protein. Purified proteins were dialyzed against a buffer containing 20 mM HEPES, pH 7.5, 5 mM MgCl₂, 1 mM DTT. Protein concentrations were estimated by the Bradford assay (Bio-Rad) using BSA as the standard and enzyme assays performed, or otherwise stored as a 50% glycerol stock at -80°C. Bacterial lysates and purified enzyme samples were also examined on SDS-polyacrylamide gels (10%) stained with Coomassie Blue.

Enzymatic properties evaluated included incubation time, temperature, pH, and cofactor requirement in an assay using radio-labeled substrates. Typically, cell lysates or purified enzymes were incubated with labeled substrates, 0.5 μCi of 1- ^3H -GPP (20 Ci mmol^{-1} , American Radiolabeled Chemicals, St. Louis, MO), 1- ^3H -FPP (26 Ci mmol^{-1} , Perkin-Elmer, Waltham, MA), or 1- ^3H -GGPP (23 Ci mmol^{-1} , Perkin-Elmer, Waltham, MA) with reaction conditions specific for monoterpene (25 mM HEPES, pH 7.2, 100 mM KCl, 10 mM MnCl_2), sesquiterpene (25 mM Tris-HCl, pH 7.5, 20 mM MgCl_2), and diterpene (25 mM HEPES, pH 7.2, 100 mM KCl, 10 mM MgCl_2 , 10 μM MnCl_2) (Rising et al., 2000; Martin et al., 2004) biosynthesis. For kinetic analyses, purified enzyme was incubated with unlabeled FPP (2 mg mL^{-1} , Sigma Chemical, St. Louis, MO) ranging from 0.115 to 23 μM in 1 mL reaction volumes. Reactions were incubated for 10 min at 25°C before adding 100 ng of α -cedrene as an internal standard, and then extracting twice with 2 mL of pentane. Extracts were combined, concentrated to 50 μL under nitrogen, and then analyzed by GC-MS as previously described (Takahashi et al., 2005). For Mg17 enzyme assays, both FPP (2 mg/mL, Sigma Chemical, St. Louis, MO) and GPP (1 mg/mL, Sigma Chemical, St. Louis, MO) were evaluated.

Steady-State mRNA Detection by RT-PCR

Steady-state transcript levels for the three genes were examined in various tissues; young and mature leaves, and specified floral organs such as tepals, stamens, and carpels. Total RNA was isolated from each tissue using the same method described above and first strand cDNAs synthesized using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA) and Oligo-(dT) primer (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. Oligonucleotides for PCR were designed to unique, gene-specific regions for each of the three genes. Gene-specific primers were verified by showing that primers designed for one *Magnolia* gene could not be used to amplify DNA fragments for either of the other 2 cDNA clones. The amount of 18S ribosomal RNA was used as the quality and quantitative control.

Phylogenetic Analysis

To examine the relatedness of genes within phylogenetic trees, the deduced amino acid sequences of full-length Magnolia *TPS* genes were aligned with those of known TPS genes (Trapp and Croteau, 2001; Martin et al., 2004) in ClustalW (<http://www.ch.embnet.org/software/ClustalW.html>), and the phylogenetic tree re-drawn by the neighbor-joining method in TREEVIEW program (Page, 1996). For monoterpene synthases, signal sequences were removed prior to analysis because these sequences show low similarities amongst all the genes.

Reporter Protein Localization Studies

The amino-terminal sequence of Mg17 was analyzed with computer-based prediction servers such as ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>) (Emanuelsson et al., 1999), PSORT (Old version, <http://psort.nibb.ac.jp>) (Nakai and Horton, 1999), and PREDOTAR (<http://www.inra.fr/predotar/>) (Small et al., 2004). The secondary structure of the Mg17 transit peptide was predicted with the PredictProtein program (Rost et al., 2004), and the helical wheel structure drawn with PEPWHEEL (<http://bioweb.pasteur.fr/seqanal/interfaces/pepwheel.html>).

The putative signal sequence was amplified with *Pfu* Ultra High-fidelity DNA Polymerase with 10 pmol of forward primer (*Hind*III) 5'-ATG GCA CTT AAG CTC CTC TTC-3' and reverse primer (*Sst*I) 5-AGT ACT CGC ACA GCA CCG-3'. The amplicon was digested with *Hind*III/*Sst*I restriction enzymes. The fragments were gel purified and cloned into pGEM[®]-T Easy vector, followed by DNA sequencing analysis. All inserts were then digested and integrated in pKYLX80-GFP, a pKYLX71 derivative. This vector was generated from pBluescript vector armed with a 35S² promoter, a multiple cloning site, a GFP reporter gene, and an *rbcS* 3'-terminator. In order to examine localization sites, constructs of Mg17 and targeting-control constructs were used alone or premixed together and co-bombarded onto abaxial side of tobacco leaves (*Nicotiana tabacum* L. cv KY160) according to Dinkins et al. (2003). For a chloroplast- and a mitochondria-targeting control, the Arabidopsis signal sequences of *rbcS* and *CoxIV* were fused with green fluorescent protein (GFP) and *Discosoma* red fluorescent protein (DsRFP), respectively (Dinkins et al., 2003).

The fluorescence of transiently-expressed reporter proteins was monitored within 24 to 48 hrs after bombardment with an Olympus FV1000 laser-scanning confocal microscope (Olympus America, Center Valley, PA). An inter-vein section (0.5 cm x 1.5 cm) of leaf tissue was placed on glass slides in water and covered with a glass cover-slip and 6 different cells in regions of interest were examined. Initial focus was performed with an Olympus water immersion PLAP40X WLSM-NA1.0, and 488 nm (GFP) or 543 nm (RFP) laser lines from a multi-line argon laser set at 10% or 22% respectively. Images were acquired with a resolution of 512 x 512 pixels and a scan rate of 4 μ s pixel⁻¹. Acquired files were exported as TIFF files from Olympus FluoView software version 1.5 and final images rendered with Adobe Photoshop 7.0.

DNA Extraction and Structural Gene Analysis

Prior to extraction, the pubescence material of young leaf tissue was removed by scrubbing because it was inhibitory to subsequent DNA manipulations. Genomic DNA was extracted using a method adapted from Li et al. (Li et al., 2002). Briefly, 100 mg (1 volume) of frozen leaf was pulverized in a pre-chilled mortar to fine powder and immediately 5 volumes of extraction buffer (25 mM Tris-HCl, pH 8.0, 50 mM EDTA, 0.5% (w/v) SDS, 2% (v/v) β -mercaptoethanol, and insoluble PVPP) were added to the powder and mixed. After 15 min incubation at 56°C with gentle mixing by inversion, 5 volumes of phenol:chloroform (1:1, v/v) were added and mixed gently again. Debris was removed by centrifugation and the aqueous phase was extracted a second time. An equal volume of chloroform:isoamylalcohol (24:1, v/v) was mixed with the aqueous phase and incubated on ice for 15 min. The aqueous phase was then separated by centrifugation; mixed with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ice-cold isopropanol; incubated at -20°C for 15 min; and the precipitated nucleic acids collected by centrifugation for 15 min. The precipitated nucleic acids were washed with 75% ethanol twice and dissolved in dH₂O or TE buffer followed by a RNase A treatment. The quality and yield of genomic DNA were estimated by spectrophotometry.

Initial polymerase chain reactions were performed using primers that annealed to the start and stop sites of the respective cDNAs. If these amplification reactions failed, the ORFs (open reading frame) of the respective cDNAs were divided into 2 to 3

segments (>500 bp) which overlapped each other minimally 100 nucleotides. The PCR reactions were performed with 50 to 500 ng of genomic DNA, 10 pmol of primers, and *Pfu* Ultra High-fidelity DNA polymerase in a total volume of 50 μ L with typical reaction condition of 94°C for 3 min, 7 cycles of 94°C for 30 sec, 58 to 61°C for 30 sec, 72°C for 1 to 2 min, and 30 cycles of 94°C for 30 sec, 56 to 59°C for 30 sec, and 72°C for 2 to 3 min, with a final extension for 5 min at 72°C. The amplicons were purified with QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and cloned into pGEM[®]-T Easy vector (Promega, Madison, WI) for DNA sequencing. DNA sequences were assembled and aligned with corresponding cDNA sequences using Vector NTI 10.3.0 (Invitrogen, Carlsbad, CA). The 3 *Magnolia* genes were classified by following the classification method of Trapp and Croteau (2001).

RESULTS

Isolation of Terpene Synthase cDNAs from a Basal Angiosperm *Magnolia grandiflora*

The chemical profile of terpenes found in *Magnolia* leaves was first assessed to facilitate the design of PCR primers for the isolation of terpene synthases catalyzing the biosynthesis of particular classes of terpenes. Interestingly, 2 terpenes dominated the profile with β -cubebene accounting for 31.3% of the total and β -elemene for 25.3% (Fig. 3.3A). β -Elemene is not the actual terpene accumulating in the *Magnolia* leaves. Instead, germacrene A extracted from the leaf tissue undergoes a rapid Cope re-arrangement to β -elemene upon injection into the heated sample port of the GC (data not shown) (Takeda, 1974; de Kraker et al., 1998). Hence, the dominant terpenes accumulating in *Magnolia* leaves are β -cubebene and germacrene A, which are biosynthetically related by a predicted germacrenyl cation intermediate (Fig. 3.4)

The initial PCR primers were designed to conserved regions near the carboxy-terminal aspartate-rich motif (DDXXD) found in association with germacrene A-D synthases functionally characterized from plants in the Solanaceae and Asteraceae (Colby et al., 1998; van der Hoeven et al., 2000; Bennett et al., 2002b; Bouwmeester et al., 2002; Prosser et al., 2002; Arimura et al., 2004a; Iijima et al., 2004; Prosser et al., 2004). Seven fragments between 100 and 250 bp in length were amplified using these degenerate

primers in RT-PCR assays with mRNA extracted from young leaf tissue. All 7 fragments exhibited sequence similarities to terpene synthases (E values $\leq e^{-19}$) and were considered further. 5'- and 3'-RACE approaches were then used to isolate three full-length cDNAs and the other four cDNAs were subsequently disregarded because they possessed several internal stop codons, were unusually short length (1 to 1.2 kb), or were missing obvious 5'-start sequences.

Mg25 (551 residues, 1,653 bp) and *Mg11* (555 residues, 1,779 bp) showed considerable sequence similarities to one another with 91.7% identity at the nucleotide level, and 91.5% and 87.2% similarity and identity, respectively, at the predicted amino acid level (Fig. 3.2). No additional evidence for N-terminal sequences upstream of the R(R)X₈W motif was obtained for either of these cDNAs even after multiple 5'-RACE attempts. Highly conserved regions amongst sesquiterpene synthases were also found in these putative sesquiterpene synthases from a basal angiosperm *Magnolia*. For example, the aspartate-rich substrate binding motif known as DDXXD (Fig. 3.2) and the R(R)X₈W motif (Fig. 3.2) were found in spatially conserved positions. *Mg25* and *Mg11* were most similar to the germacrene D synthase gene from grape vine (*Vitis vinifera*) with amino acid identity in excess of 48% and similarity greater than 66% (Lucker et al., 2004). *Mg11* contained an internal stop codon corresponding to residue 25, which was mutagenized to a tryptophan codon, a conserved amino acid at this position amongst sesquiterpene synthase, prior to functional analysis. In contrast, *Mg17* (592 residues, 1,779 bp) harbored not only the DDXXD motif and the RRX₈W (residues 54-64), but an additional N-terminus sequences (43 aa) characteristic of known monoterpene synthases (Bohlmann et al., 1998a; Williams et al., 1998; Lucker et al., 2002). Moreover, the *Mg17* N-terminus sequence upstream of the RRX₈W motif was predicted as a putative transit peptide for chloroplast targeting by several computational algorithms (ChloroP, PREDOTAR and PSORT). *Mg17* most closely resembled the grape vine (*Vitis vinifera*) α -terpineol synthase with an amino acid sequence identity of 60% and similarity score of 75% (Martin and Bohlmann, 2004).

Functional Characterization of *Mg25* as a β -Cubebene Synthase

To determine the functional activities encoded by each of the putative terpene synthase cDNAs, each was expressed in bacteria and bacterial lysates were tested for their ability to convert GPP, FPP and GGPP into terpene reaction products. Hydrocarbon and oxygenated reaction products generated by incubations with the recombinant enzymes were analyzed in GC-MS and identified by a combination of retention time and mass spectra in comparison to authentic standards and NIST library matches (Rising et al., 2000; Martin et al., 2004). Mg25 was capable of utilizing FPP as a substrate, but not GPP (negligible activity) or GGPP (no activity). The dominant reaction products generated by Mg25 were β -cubebene (24.5%), α -muurolene (19.3%), δ -cadinol (18.6%), δ -elemene (16.0%), τ -muurolene (10.8%), β -elemene (10.8%) (Fig. 3.3B). The multi-functional nature of Mg25 is not unique (Steele et al., 1998), and the multiple reaction products can readily be rationalized as arising from a common reaction intermediate, nerolidyl diphosphate, through a series of hydride shifts, ring closures, deprotonations, and a hydroxide capture (Fig. 3.4). Not surprisingly, the mixture of reaction products was pH dependent (Supplementary Fig. 3.1A). Maximal activity was observed at pH 7.5 and dominated by four products; β -cubebene, α -muurolene, δ -cadinol, and τ -muurolene. However, at pH 5.5, no β -cubebene was evident and only detected from pH 6.5 to pH 8.5. In comparison, δ -cadinol was readily detected from pH 5.5 to pH 7.5, but not at pHs above this range. Maximal activity was observed at 24°C to 26°C with an apparent K_m value for FPP of $1.07 \pm 0.22 \mu\text{M}$ and a K_{cat} value of $0.35 \times 10^{-3} \pm 0.02 \text{ S}^{-1}$ (Supplementary Fig. 3.1B). Mg25 also exhibited a divalent cation requirement for activity, with maximal activities observed at Mg^{2+} concentrations of 1 to 10 mM and Mn^{2+} concentrations of 1 mM or less (Supplementary Fig. 3.1C).

Functional Characterization of Mg11 Failed

Although full-length cDNAs for Mg11 were isolated repeatedly in separate experimental attempts, a stop codon was invariably observed at the position corresponding to codon 25. In an attempt to express and functionally characterize Mg11, the stop codon was substituted with that for tryptophan (W), which is a highly conserved residue at this position amongst terpene synthases. Unfortunately, no enzyme activity was ever associated with over-expression of the modified Mg11 cDNA in bacteria, despite

attempts with a variety of expression vectors (pET100/D-TOPO, pET28, pET32 and pET42) and expression conditions, including the induction treatment, the induction time period and the temperature at which the bacteria were grown. In all cases, the Mg11 protein was detected in the insoluble fraction (inclusion bodies) and even attempts to solubilize the Mg11 protein using chaotropic reagents like urea were unsuccessful in uncovering any terpene synthase activity with GPP, FPP or GGPP as substrates.

Functional Characterization of Mg17 as an α -Terpineol Synthase

When over-expressed in *E. coli*, amino-terminal truncated Mg17 encoded for a bifunctional terpene synthase that readily utilized GPP and FPP as substrates, producing α -terpineol and (*E*)- α -bisabolene, respectively, as the major reaction products (Fig. 3.5). Deletion of the DNA sequence coding for the putative signal peptide (43 amino acids) at the N-terminus was necessary in order to obtain soluble enzyme activity, similar to what has been observed for many monoterpene synthases (Bohlmann et al., 1998a), otherwise an insoluble form of the Mg17 protein was observed. With FPP as the substrate, bisabolenoid sesquiterpenes were produced by the Mg17 enzyme. These included (*E*)- α -bisabolene (33.1%) as the principal product, α -bisabolene (18.7%), β -sesquiphellandrene (15.8%), β -bergamotene (12.6%), (*Z*)- α -farnesene (10.4%), and (*Z*)- α -bisabolene (9.4%) (Fig. 3.5). With GPP as the substrate, only a single monocyclic alcohol, α -terpineol, was biosynthesized. This contrasts with the results obtained with α -terpineol synthases isolated from grape vines (Martin and Bohlmann, 2004) and pine (Phillips et al., 2003). *Vitis vinifera* α -terpineol synthase catalyzes the conversion of GPP into 14 monoterpene products, including both enantiomers of α -terpineol together comprising greater than 50% of the total reaction products (Martin and Bohlmann, 2004). Phillips et al. likewise demonstrated that the loblolly pine monoterpene synthase *Pt10* converted GPP into at least 6 reaction products with multiple enantiomeric forms of α -terpineol accounting for 57% of the total reaction products (Phillips et al., 2003).

Because of Mg17's higher sequence similarity to monoterpene synthases than sesquiterpene synthases and its putative amino-terminal chloroplast targeting sequence, Mg17 was classified as a monoterpene synthase. This was in spite of reports that

monoterpene synthases typically have a more restricted substrate specificity (Steele et al., 1998). Nonetheless, plausible mechanisms for how Mg17 might utilize either GPP or FPP as substrates can be reconciled with chemical rationalization for partial reactions including initial ionization followed by isomerization in common with both substrates (Fig. 3.6).

Steady-State mRNA Accumulation in Various Tissues

A quantitative RT-PCR method was employed to measure the steady-state transcript levels for the 3 terpene synthase genes in leaves at different developmental stages and in the flower organs including tepals, stamens and carpels. PCR primer pairs were designed to selectively amplify each gene (Supplementary Fig. 3.3) and normalized relative to an amplicon for 18S rRNA in each sample (Fig. 3.7). Transcripts for all 3 genes, *Mg25*, *Mg17* and *Mg11*, were much more evident in young developing leaf tissues than in older more mature leaves (Fig. 3.7). And while the level of transcripts for all 3 terpene synthases were low or below detection limits in tepal and carpel tissues, the levels of *Mg25* and *Mg11* mRNA were greatly elevated in stamens.

The N-terminal Sequence of Mg17 Functions as a Dual Targeting Signal Sequence

Several signal sequence analysis programs, including PREDOTAR, PSORT and ChloroP, predicted that the amino-terminal extension of Mg17 could be targeting this enzyme to both the chloroplast and the mitochondrial compartments. The Mg17 transit peptide (Mg17tp) sequence also satisfied the criteria that were derived from the previously characterized transit peptides (Roise et al., 1988; Akashi et al., 1998). Mg17tp lacks acidic residues (0%), but has a significant proportion of basic (13.3%), hydrophilic (31.1%) and hydrophobic (37.7%) residues making this peptide sequence potentially amphiphilic. The amphiphilic nature of signal sequences was noted earlier by Cunillera et al. (1997) to give targeting sequences sidedness when projected in helical formats. Secondary structure predictions for the 43 residues making up the putative signal sequence of Mg17 likewise suggested two possible amphiphilic structures, from residues 1 to 9 and residues 24 to 34 with each displaying hydrophobic residues on one-face of the helix and polar/basic residues on the other (Fig. 3.9).

To functionally test the targeting nature of the Mg17 leader sequence, the coding sequence for the 43 amino-terminal residues were fused to the GFP (Green Fluorescent Protein) reporter gene, and transient expression of the construct was monitored after particle bombardment into epidermal cells of tobacco leaves by laser confocal microscopy (Fig. 3.8). Expression controls included the Arabidopsis *rbcS* transit peptide fused with GFP as a plastidic marker (Fig. 3.8A) and the Arabidopsis *CoxIV* signal sequence fused with DsRFP (*Discosoma* Red Fluorescent Protein) as a mitochondrial signal sequence control (Fig. 3.8B) (Dinkins et al., 2003). When the recombinant Mg17tp:GFP and the *CoxIV*:DsRFP were co-expressed, Mg17tp-GFP appeared localized to two populations of organelles resembling mitochondria and chloroplasts (Fig. 3.8D, G and J). Evidence for mitochondrial targeting was provided by co-localization of *CoxIV*-RFP to the smaller organelles (Fig. 3.8E, H and K), which over-lapped with the Mg17tp-GFP images (Fig. 3.8F, I and L). Evidence that the fluorescence localized to the larger organelles in cells expressing Mg17tp-GFP was associated with chloroplasts was evident from the auto-fluorescence (red organelles in panels F, I and L in Fig. 3.8) coincident with the GFP fluorescence and independent localization of *rbcS*-GFP and Mg17tp-GFP to these same organelles in separate determinations (data not shown).

Genomic Organization of Terpene Synthase Genes in Magnolia

To examine the intron-exon organization of the Magnolia terpene synthase genes, over-lapping amplicons for genomic clones were assembled and aligned with each of the corresponding cDNAs. The genomic clones were then oriented relative to the highly conserved DDXXD motif to illustrate relative length and position of exons and introns, and arrayed along with archetypical representatives of Class I to III terpene synthase genes defined according to Bohlmann et al. (Bohlmann et al., 1998a) and Trapp and Croteau (2001) (Fig. 3.11 and Table 3.1). Interestingly, all 3 Magnolia genes exhibited different intron-exon organizations. *Mg17* resembled a typical Class III terpene synthase gene with 7 exons and 6 introns, much like 5-epi-aristolochene synthase from tobacco (Facchini and Chappell, 1992) or (+)- δ -cadinene synthase from cotton (Chen et al., 1995). Also important to note, the exon and intron sizes are relatively well conserved between these Class III genes and *Mg17*, except for exon 1 of *Mg17*, which contains an

amino-terminal signal sequence of 129 extra bp. In contrast, *Mg11* consists of 5 exons and 4 introns with the first 4 exons organized similar to monoterpene synthase genes found within the Class III, but differing by having a single carboxy-terminal exon encompassing the equivalent of the last 3 exons of the Class III genes. Because the *Mg11* gene structure may represent a variant of the Class III organization it was designated as a Class III-a member. The *Mg25* genomic gene represents a more unique exon/intron organization with only two exons interrupted by a single large intron. While the size of the 5' exon does resemble those for other sesquiterpene synthase genes in the Class III family, no similar single-intron containing terpene synthase gene from higher plants has yet to be reported. A fungal terpene synthase gene harboring a single introns has, however, been observed (Trapp et al., 1998).

All three *Magnolia* genes contained intron-exon junction sites that followed standard GT/AG boundary rules (Brown and Simpson, 1998). Moreover, the phasing of the introns insertion sites appeared equally well conserved. Intron phasing is analogous to open-reading frames, refers to how codons are split across introns boundaries, and defined as phase 0, 1 and 2 depending on the number of nucleotides split between two exons (Long et al., 1995). Based on spatial positioning and phasing of each intron, the introns in all 3 *Magnolia* genes appear very well conserved relative to one another, as well as with all other angiosperm and gymnosperm terpene genes examined (Table 3.2).

DISCUSSION

Functional Characterization of *Magnolia* Terpene Synthases

In this study, three new terpene synthase genes were isolated from *M. grandiflora*, an ancestral or basal angiosperm, and further characterized biochemically and genetically. All three *magnolia* genes were able to form clades with known terpene synthases, consistent with other studies for this class of genes (Bohlmann et al., 1998a). In an un-rooted phylogenetic tree of terpene synthases assembled by the nearest neighbor-joining method, *Mg25*, a β -cubebene synthase, and *Mg11*, a highly expressed, likely pseudogene, were most closely related to a group of well-characterized angiosperm sesquiterpene synthases comprising the TPS-a clade (Fig. 3.10) (Bohlmann et al., 1998a). In contrast, the *Magnolia* α -terpineol synthase (*Mg17*) fell into the TPS-b clade of

synthases, a clade dominated by monoterpene synthases from angiosperms. This is also consistent with the designation of Mg17 as a monoterpene synthase, rather than a sesquiterpene synthase, even though it could utilize both monoterpene (GPP) and sesquiterpene (FPP) substrates. As noted above, designating Mg17 as a monoterpene synthase is also more consistent with the signal sequence it harbors at its amino terminus. No such signal sequences have been reported for any sesquiterpene synthases, and there are no reports of FPP biosynthesis occurring naturally in chloroplasts. Hence, Mg17 targeted to the chloroplast would not have access to FPP, and likely be associated only with monoterpene biosynthesis. However, Mg17 targeted to the mitochondria could be associated with sesquiterpene biosynthesis there because FPP synthesis is known to occur in this organelle to some extent (Cunillera et al., 1997; Martin et al., 2007).

Intriguing observations of dual targeting of a single gene product to both mitochondria and chloroplasts have been reported previously, and two mechanisms have been proposed to account for this (Karniely and Pines, 2005). In one case, alternative translation start sites can provide different targeting signal sequences, yielding different peptides that are targeted to different intracellular locales or organelles. The Arabidopsis TH11 gene associated with thiamine biosynthesis utilizes such an alternative targeting signal sequences system to achieve differential targeting (Chabregas et al., 2001). The other case is for a single signal sequence to serve as a signal for two different targets, a so-called ‘ambiguous targeting signal’ that receptors of two organelles can recognize. Both aminoacyl-tRNA synthetase and methioninyl-tRNA synthetase from Arabidopsis, which play a critical role in protein translation, rely on an ‘ambiguous targeting signal’ to direct these proteins to the mitochondria and chloroplasts (Akashi et al., 1998; Menand et al., 1998). With the available sequence information, Mg17 does not appear to contain alternative translation start sites, so if this enzyme does co-localize to the mitochondria and chloroplast, then it might occur via the action of an ambiguous signal sequence mechanism.

The catalytic abilities of Mg17 to generate both the monoterpene and the sesquiterpenes may not be related to any physiological targeting of the enzyme to multiple intracellular compartments, but may simply reflect the evolutionary recruitment of a sesquiterpene biosynthetic gene to monoterpene metabolism. This could have entailed a

molecular exchange of the original amino terminal exon of Mg17 with that for another monoterpene synthase, creating a “chimeric” gene – exon 1 from a monoterpene synthase and exons 2 to 7 from the original Mg17 sesquiterpene synthase. The ability of sesquiterpene synthases to utilize GPP as a substrate has been documented and makes sense from the perspective that GPP simply represents a smaller substrate than the normal FPP substrate. The converse does not appear to have precedence. Monoterpene synthases have not been reported to utilize FPP as a substrate, possibly because the active site cannot accommodate the larger substrate. If the suggestion of exon exchange is correct in accounting for the origin of Mg17, then one might suspect additional examples of exon 1 swapping to be found in large phylogenetic types of analyses where putative “progenitor” terpene synthase genes found in ancient, basal plant species are compared to “similar” genes found in evolutionarily advanced genomes.

Evolutionary Inferences from the Magnolia Terpene Synthase Genes

The trichodiene synthase gene found in the fungus *Myrothecium roridum* contains a single intron (Genbank accession no. AF009416) and, until now, no counterpart TPS gene in the plant kingdom has been described (Trapp et al., 1998). Trapp and Croteau (2001) used this type of information, the intron/exon organization, to suggest that plant and fungal terpene synthase genes might have had different evolutionary origins (Trapp and Croteau, 2001). The β -cubebene synthase gene, *Mg25*, described here from *Magnolia* possesses a single intron positioned near the 5'-region of the gene similar to the first intron in all other Class III terpene synthases from plants. The intron found in the fungal trichodiene synthase gene is inserted into the middle of the trichodiene synthase gene and not spatially oriented similar to the insertion site of the first intron in any of the plant genes, including *Mg25*. Hence, our data is consistent with those of Trapp and Croteau (2001) in arguing for separate evolutionary pathways for the fungal and plant terpene synthase gene families.

Bohlmann et al. (1998) and Trapp and Croteau (2001) also hypothesized earlier that angiosperm terpene synthases might have evolved from an ancestral terpene synthase gene in common with gymnosperms and containing 14 introns. This is an appealing hypothesis because it is easier to envision the loss of introns from a common ancestral

gene giving rise to the apparently conserved intron-exon organization and intron phasing observed in the extant terpene synthase genes of angiosperms and gymnosperms, rather than the conservation of a mechanism able to insert introns into spatially conserved positions over evolutionary time and between plant families. However, the unique exon-intron organization of Mg11 and Mg25 observed here would suggest that such intron loss proceeded relatively more rapidly in *Magnolia* than other branches of the angiosperms, which is not wholly consistent with the relatively modest changes observed at other genetic loci within the *Magnolia* species (Goldenberg et al., 1990).

The findings reported here suggest that alternative mechanisms for terpene synthase gene evolution might also need to be considered. For example, if the angiosperm *TPS* genes evolved from gene in common with gymnosperms, sequence similarities and overall conservation of gene structure might be anticipated, especially with genes from a basal angiosperm like *Magnolia* in comparison to those in gymnosperms. None of the *Magnolia TPS* genes described here fall into associations with the Class I or II *TPS* genes, classes where all the gymnosperm terpene synthase genes fall. Instead, the *Magnolia* genes map to Class III, which includes most of angiosperm terpene synthase genes, or appear to represent distinct or unique classes.

One final point supporting alternative origin(s) for the gymnosperm and angiosperm *TPS* genes is the presence of the unique CDIS motif in all the gymnosperm genes, but not in any of the *Magnolia* genes. If the *TPS* genes in *Magnolia* arose from a progenitor gene in common with gymnosperms, then the absence of the CDIS motif in these genes must mean this sequence was deleted or lost prior to when the Magnoliaceae, and by inference all angiosperms, diverged from the gymnosperms. And if this were the case, then one would not expect to observe the CDIS motif in any angiosperm *TPS* genes. In fact, this motif has been reported in a monoterpene synthase gene, a linalool synthase, found in the angiosperm *Clarkia* (Dudareva et al., 1996).

The greatest obstacle in elucidating the evolution pathway of present day genes is that what we can document today is only the final format to have arisen by the gain or loss of introns. Moreover, with the limited datasets that we now possess, one can easily envision angiosperm *TPS* genes evolving independently from gymnosperm genes, and by an intron gain rather than loss mechanism as suggested by Trapp and Croteau (2001).

Perhaps a progenote *TPS* gene had a single long intron (600-900 bp) in a specific position, for example the position occupied by present day intron 3. Through the course of evolution, this intron might have undergone modifications allowing it to be partitioned into small pieces and inserted into the neighbor *TPS* genes, creating multiple introns. The basic 6 intron organization of *TPS* genes observed today in angiosperms thus formed could then have multiplied and diverged by gene duplication and exon shuffling to create diverse *TPS* functionalities. This alternative hypothesis to “the introns loss” model of Trapp and Croteau (2001) might be referred to as the ‘intron partitioning and distribution theory’. Without a doubt, such a discussion based on so few genes from so few taxa makes any such suggestion highly speculative. Hence, to further sort through these possibilities, *TPS* genes from other extant basal angiosperms (Laurales and Piperales, for examples) and intermediate taxa (between Magnolia and advanced species in the molecular phylogeny tree) must be examined, as well as even earlier land plants, such as mosses, liverworts and hornworts.

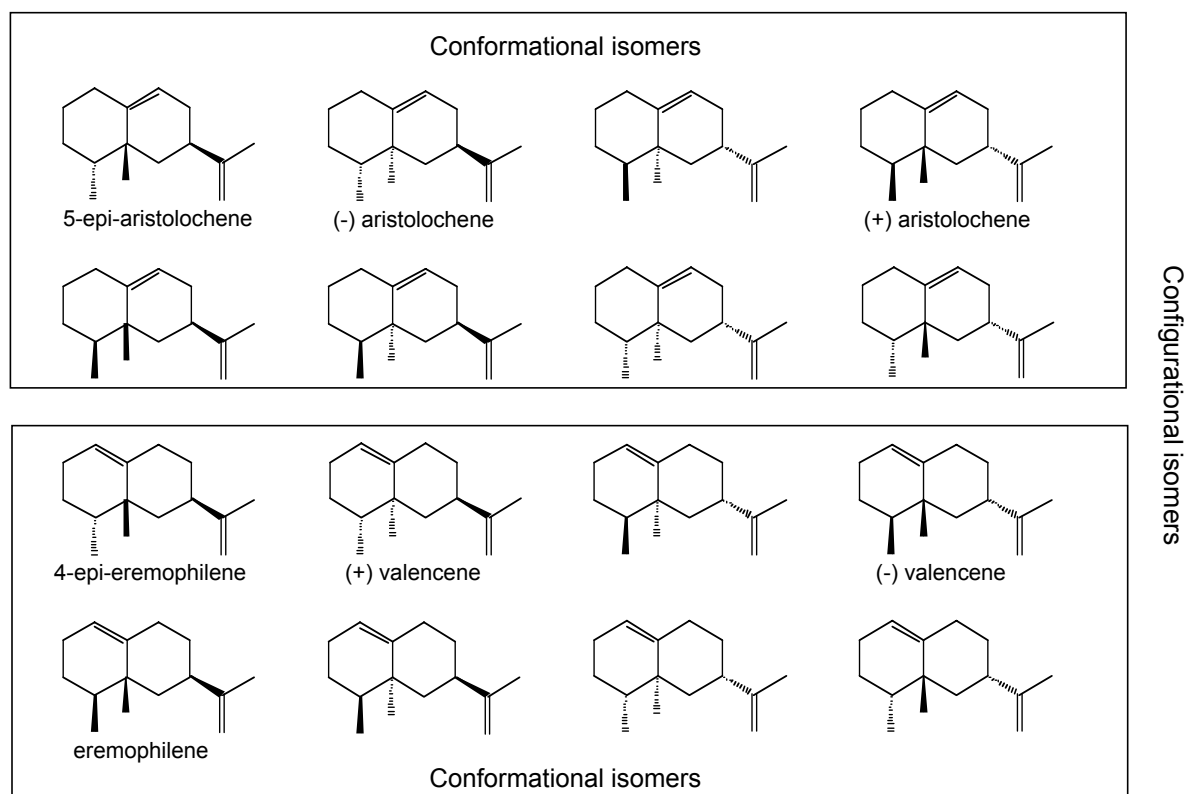


Figure 3.1. An illustration of the sesquiterpenes comprising the eremophilene class of sesquiterpenes. One of the major challenges in terpene metabolism is to understand how these specific chemical species arise biochemically, then to use this information to infer how unique and distinct enzymes for different eremophilene species may have arisen over evolutionary time.

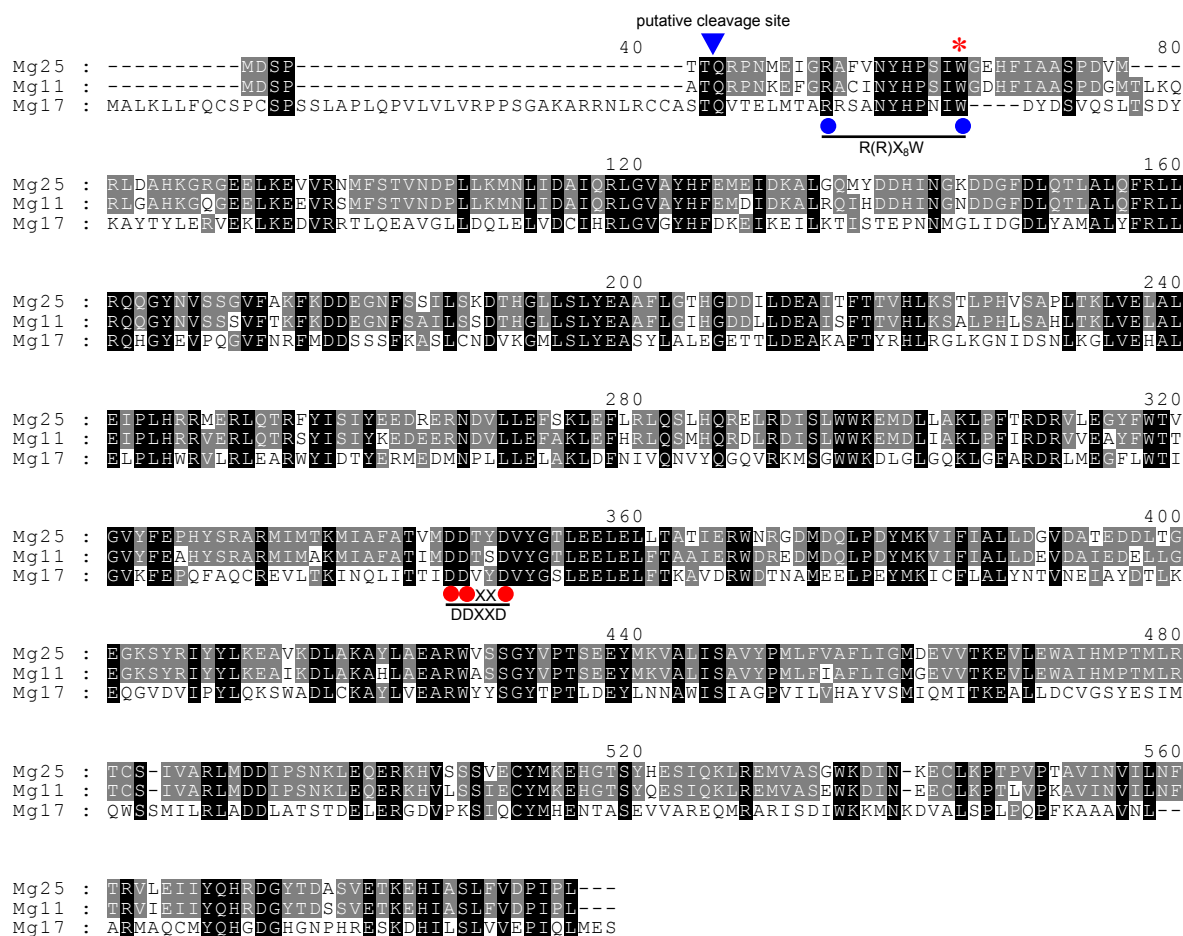


Figure 3.2. A comparison of the deduced amino acid sequences for terpene synthases cloned from *M. grandiflora*. All three genes have a conserved aspartate-rich domain, DDXXD, which coordinates substrate binding via the formation of divalent cation salt bridges (Starks et al. 1997). The monoterpene synthase, Mg17, contains a R(R)X₈W motif which is a characteristic domain for all monoterpene synthase, as well as a putative amino terminal signal sequence (approximately 43 aa at the N-terminal region) which putatively targets the mature protein to the plastid organelle. The asterisk denotes the location of an internal stop codon in Mg11 that was removed by site-directed mutagenesis, creating a tryptophan (W) residue instead. The putative cleavage site was predicted by the ChloroP server (<http://www.cbs.dtu.dk/services/ChloroP/>), the alignment performed by ClustalW (<http://www.ch.embnet.org/software/ClustalW.html>), and the alignment visualized with GENEDOC (<http://www.psc.edu/biomed/genedoc>).

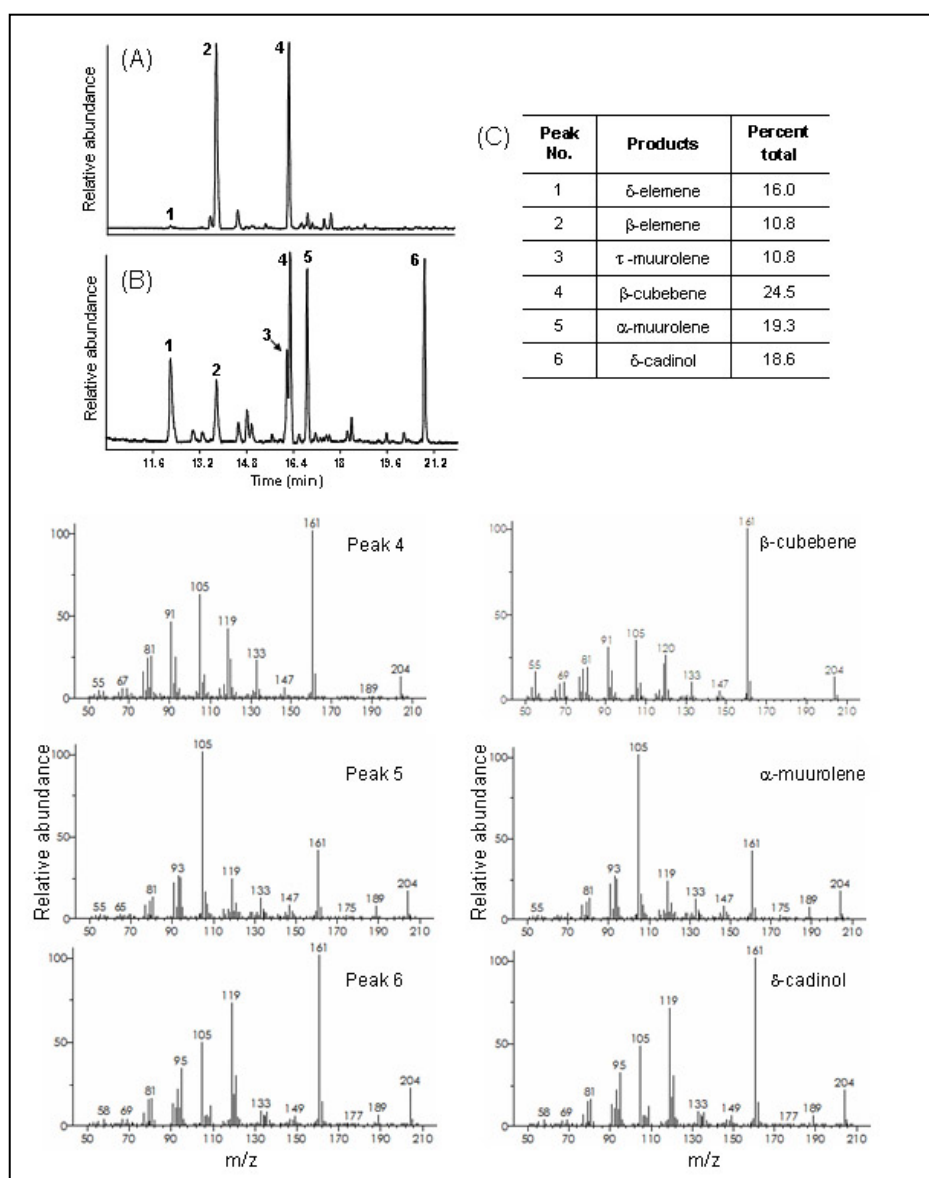


Figure 3.3. Total ion chromatograms of the sesquiterpene products generated by Mg25. The *Mg25* cDNA was expressed in bacteria and bacterial lysates used as the source of enzyme for *in vitro* assays, which consisted of incubations with FPP and profiling the reaction products by GC-MS. (A) GC trace of terpenes extracted from *M. grandiflora* leaves, (B) GC trace of *in vitro* reaction products generated by Mg25 enzyme, and (C) a table of relative amounts generated by the Mg25 enzyme, taken from the chromatography in (B). The peaks labeled in GC traces (A) and (B) were identified by comparison of their MS spectra to those in the NIST MS library and authentic standards, as illustrated in the lower panels for β -cubebene, α -muurolene, and δ -cadinol, and in the Supplementary Fig. 3.2 for δ -elemene, β -elemene, and τ -muurolene.

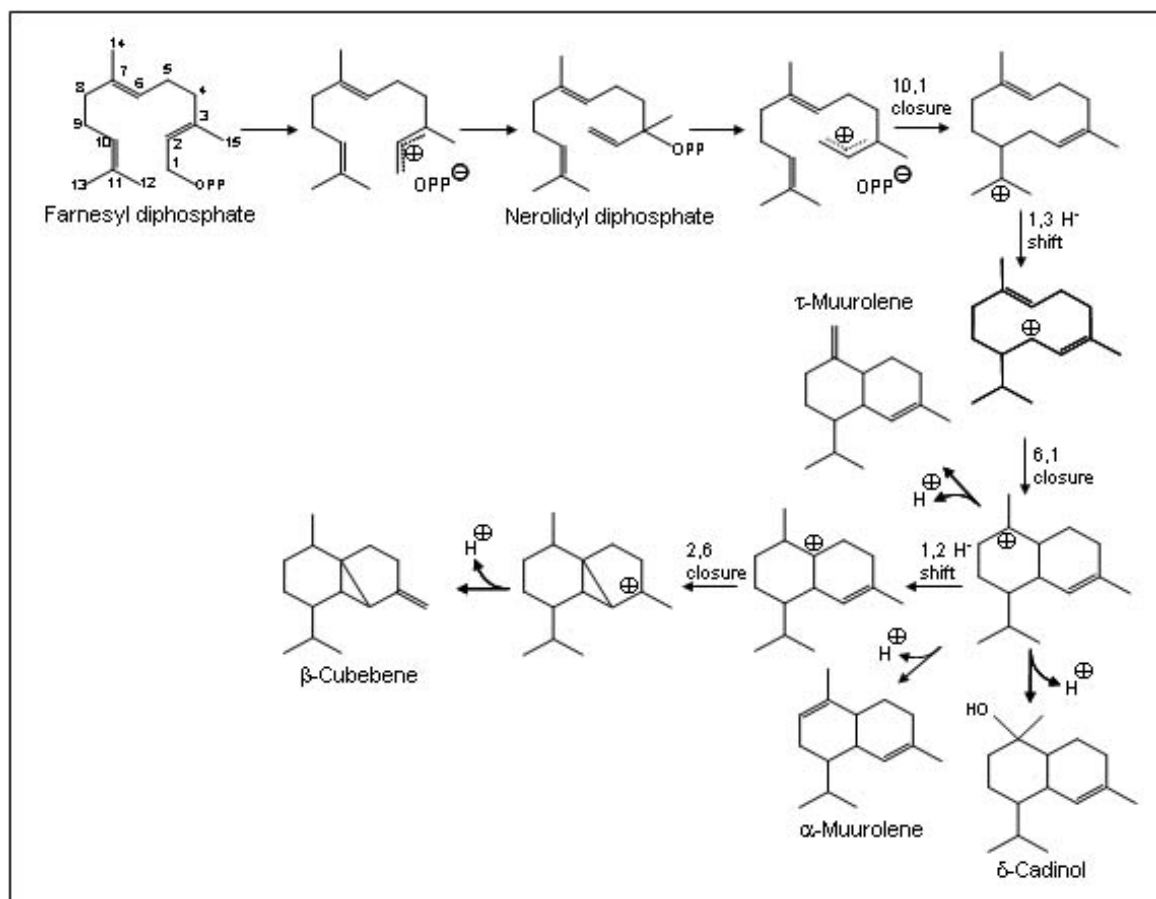


Figure 3.4. A proposed reaction mechanism to explain how Mg25 might catalyze the formation of several different sesquiterpene reaction products from the substrate farnesyl diphosphate.

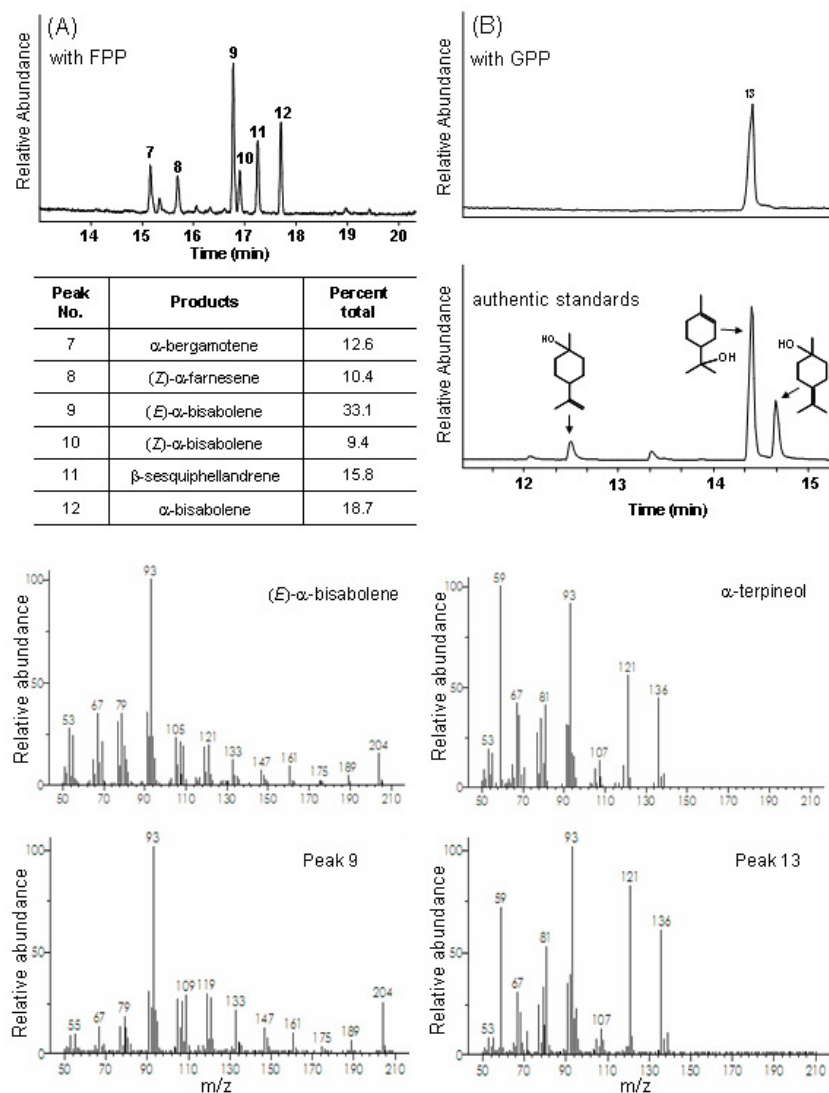


Figure 3.5. Total ion chromatograms of the monoterpene (A) and sesquiterpene (B) products generated by Mg17. An amino-terminal signal sequence corresponding to 53 amino acids were deleted from the *Mg17* cDNA before expressing this gene in bacteria. Bacterial lysates were subsequently used as the source of enzyme for *in vitro* assays, which consisted of incubations with GPP (A) or FPP (B), and profiling the reaction products by GC-MS. Labeled peaks were identified by MS comparisons with either authentic standards or matched to the MS of compounds reported in the NIST library: 7, α -bergamotene; 8, (Z)- α -farnesene; 9, (E)- α -bisabolene; 10, (Z)- α -bisabolene; 11, β -sesquiphellandrene; 12, α -bisabolene; 13, α -terpineol. MS spectra for the dominant products formed with FPP or GPP are shown in the lower panel, while the MS for the other peaks are shown in the Supplementary Fig. 3.2.

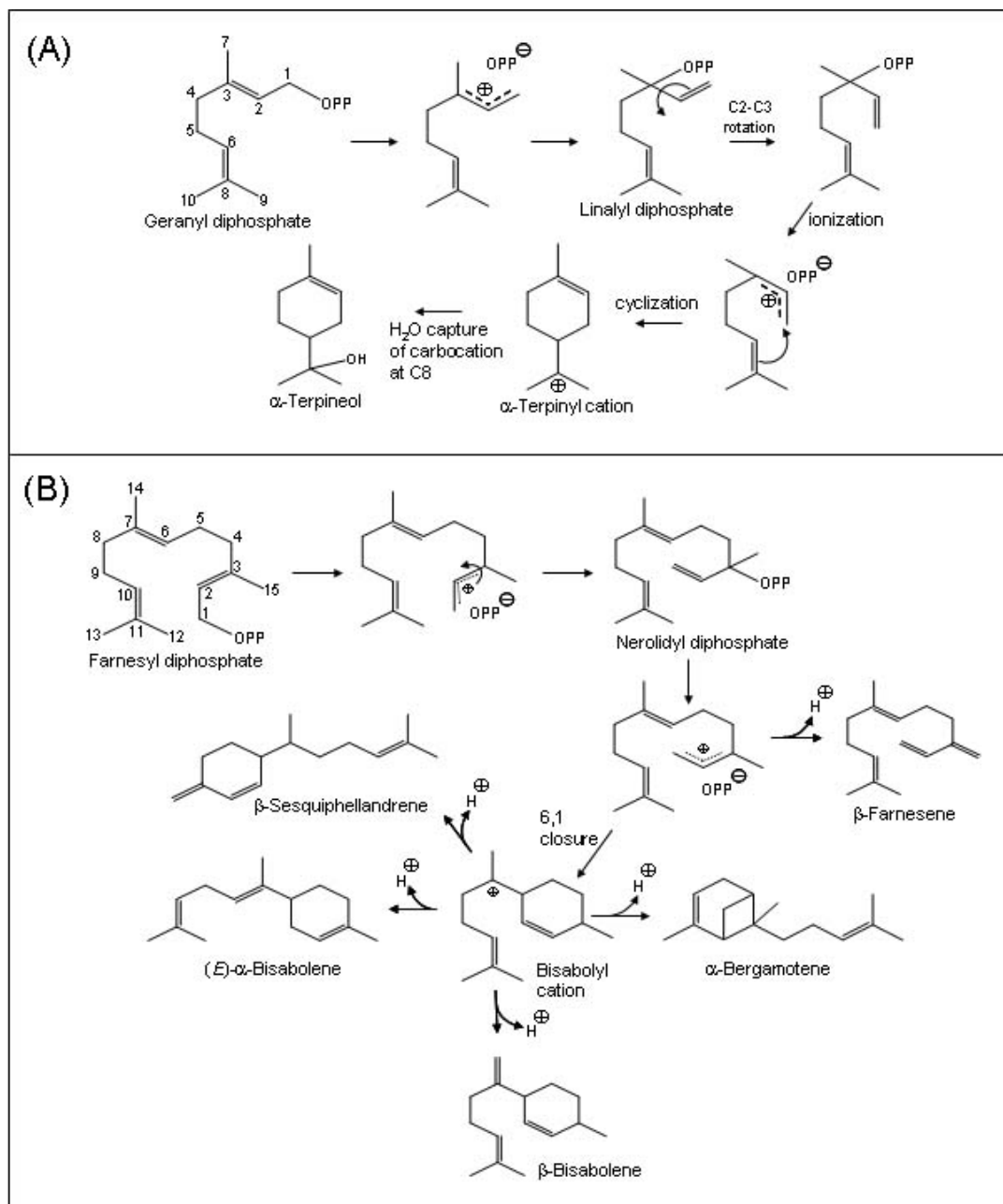


Figure 3.6. Proposed reaction mechanisms to explain how Mg17 might catalyze the formation of different monoterpene (A) and sesquiterpene (B) reaction products from the substrates geranyl diphosphate (GPP) and farnesyl diphosphate (FPP).

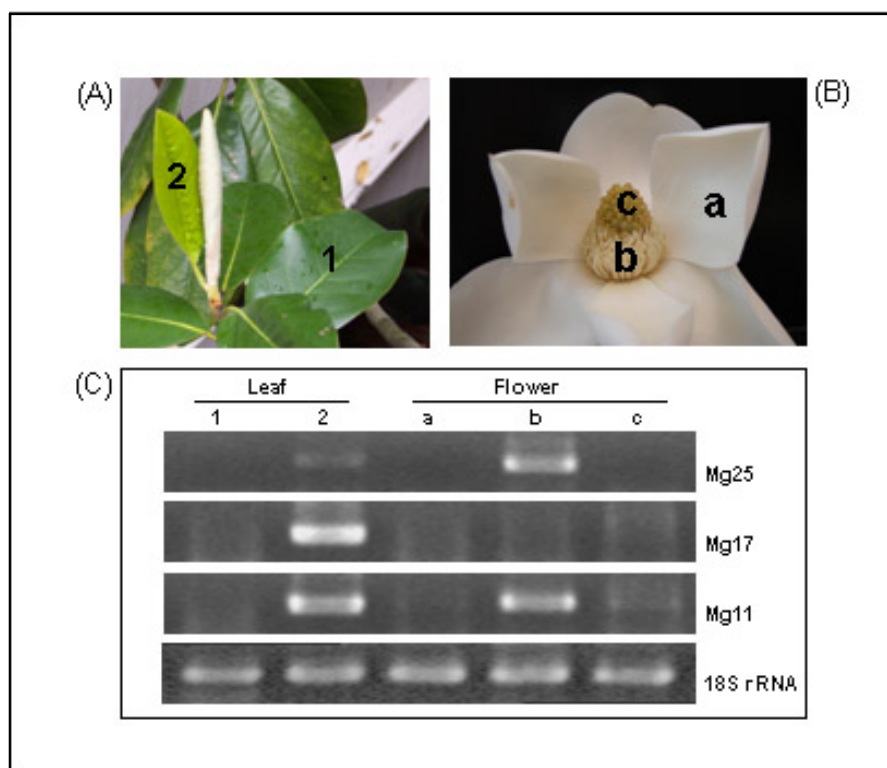


Figure 3.7. A measurement of the steady-state mRNA levels for the *M. grandiflora* terpene synthase genes *Mg25*, *Mg17* and *Mg11* in different tissues. Total RNAs were isolated from the indicated tissues and used as the template for gene-specific RT-PCR assays. Gene-specific primers were designed and evaluated by sequencing of the respective amplicons (data not shown). Amplification of an 18S ribosomal RNA fragment was used as the PCR control. Panel (A) shows the different developmental stages of leaves; 1, mature and 2, developing, while panel (B) shows the primitive floral organs of *M. grandiflora*; a, tepals (sepal+petal); b, stamens; and c, carpels.

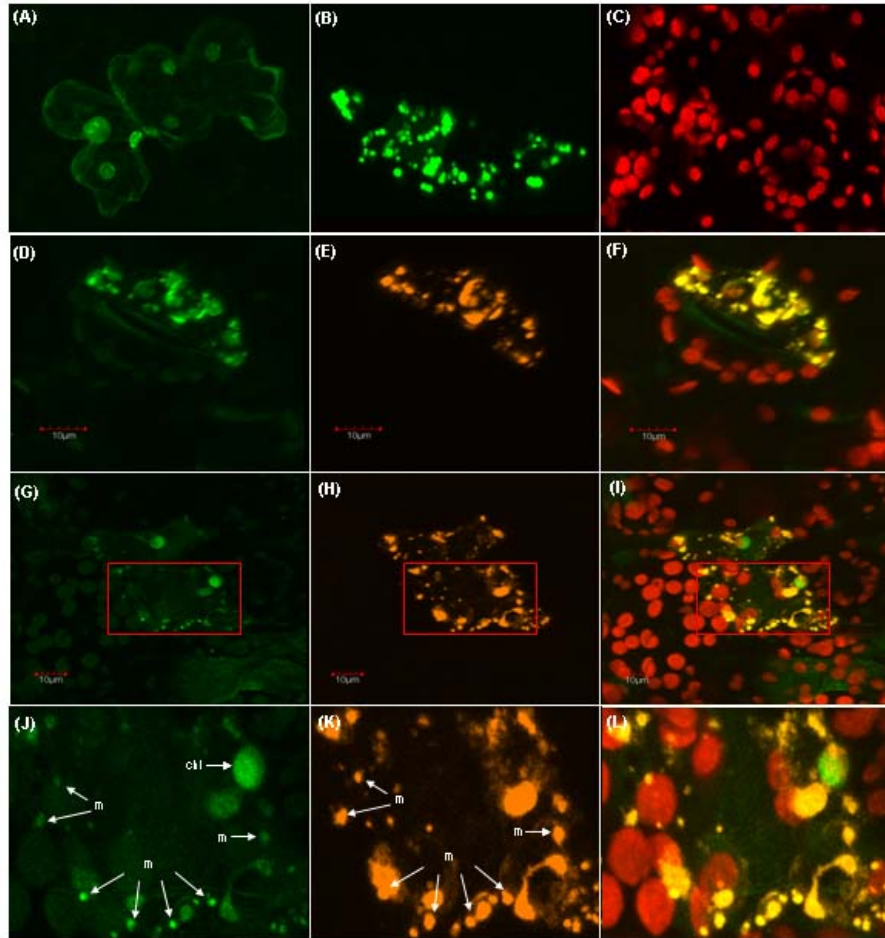


Figure 3.8. Confocal images of tobacco leaf epidermal cells showing the localization of reporter proteins to the chloroplasts and mitochondria when fused with the N-terminal signal peptide (43 aa) from the *M. grandiflora* monoterpene synthase Mg17. At least eight different cells were examined for each construct, but only one representative image for targeting marker and two independent images for Mg17tp are presented here. A, a chloroplast targeting marker transit peptide from *A. thaliana rbcS* fused with GFP; B, a mitochondrial targeting marker transit peptide from *A. thaliana CoxIV* fused with GFP; C, a background auto-fluorescence resulting from untransformed chloroplasts; D and G, a Mg17 transit peptide fused with GFP; E and H, a mitochondrial targeting marker fused with DsRFP; F, I and L, overlaid images (D+E, G+H and J+K respectively); J-L, magnified images of corresponding zones indicated in G-I. Arrows labeled 'chl' and 'm' point to chloroplasts and mitochondria, respectively. Boxes on G through I were positioned on the identical area of each image in G-I.

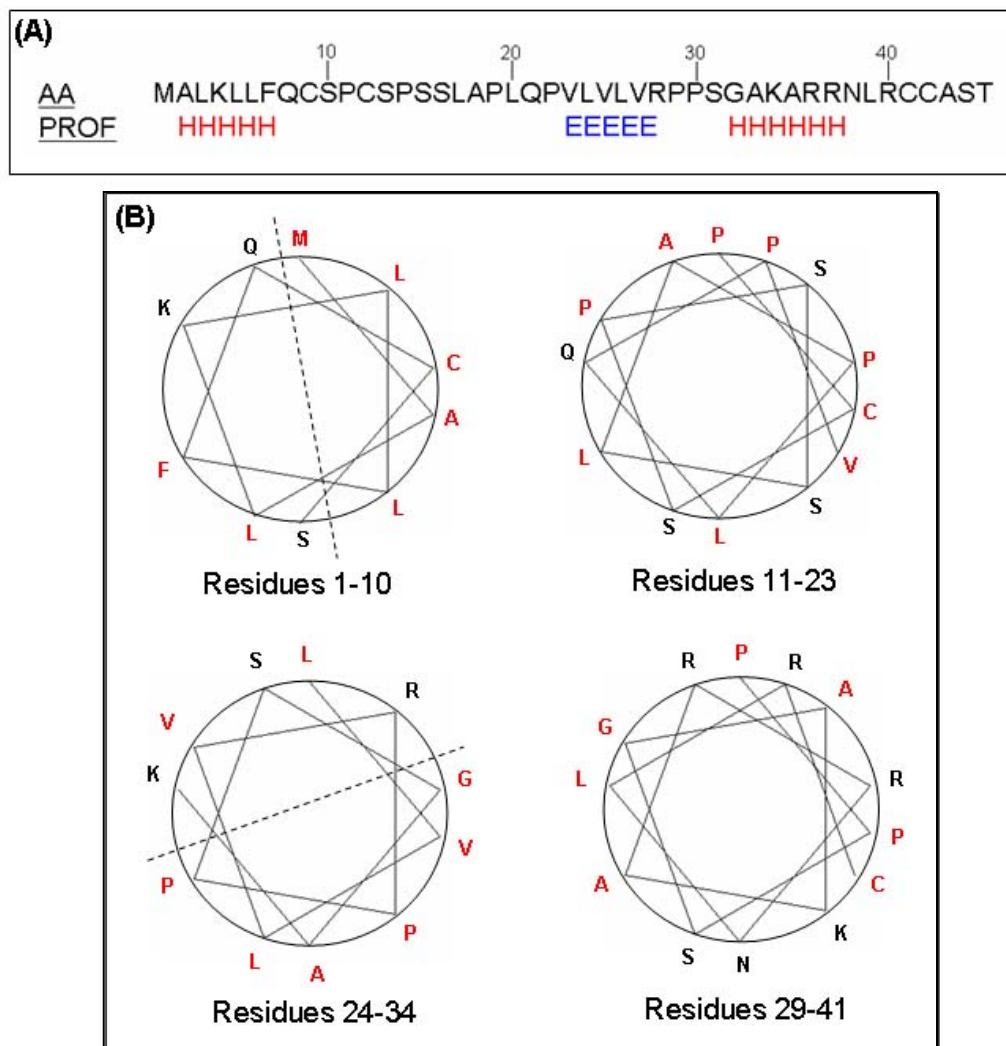


Figure 3.9. Secondary structure and helical wheel predictions of the Mg17 transit peptide. A: Secondary structures of the Mg17 transit peptide were predicted by PredictProtein (Rost et al., 2004) and are noted below the amino acid sequence. H, helical structure; E, sheet; and blank, loop structure. B: Helical wheel prediction of Mg17 signal peptide drawn according to PEPWHEEL (<http://bioweb.pasteur.fr/seqanal/interfaces/pepwheel.html>). Letters in red color in helical wheel predictions represent hydrophobic residues.

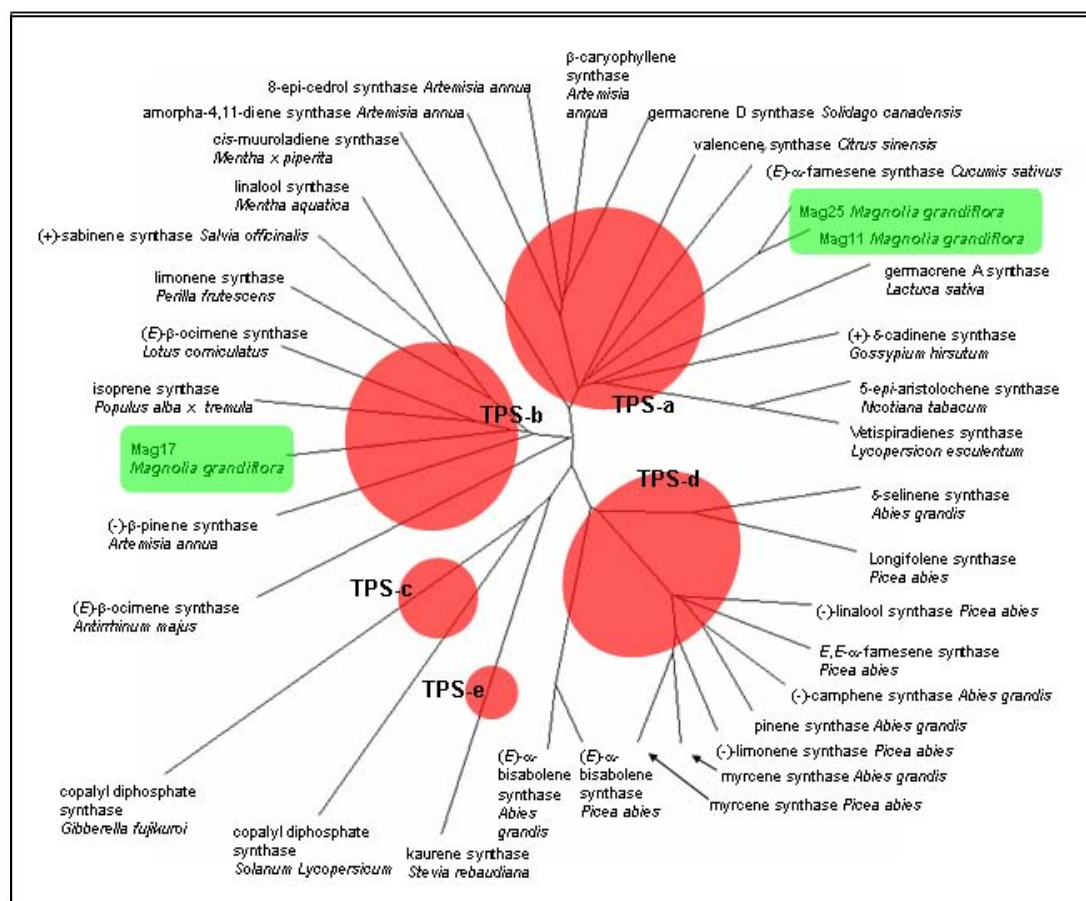


Figure 3.10. An un-rooted phylogenetic tree of terpene synthases analyzed by the neighbor-joining method and depicting the estimation of pair-wise distances at the amino acid level.

The red circles represent groups, families or clades of related genes. The *M. grandiflora* genes characterized in this study are highlighted by green rectangular boxes. For the monoterpenes, putative signal sequences were deleted for the analysis. Classification of TPS a-e (terpene synthases) followed the previous analyses of Bohlmann *et al.*, 1998a. The analysis was generated with ClustalW (<http://www.ch.embnet.org/software/ClustalW.html>) and the tree was drawn with the TREEVIEW program (Page, 1996). Aa, *Artemisia annua*; Ag, *Abies grandis*; Am, *Antirrhinum majus*; Cs, *Cucumis sativus*; Gf, *Gibberella fujikuroi*; Gh, *Gossypium hirsutum*; Lc, *Lotus corniculatus*; Le, *Lycopersicon esculentum*; Ls, *Lactuca sativa*; Ma, *Mentha aquatica*; Nt, *Nicotiana tabacum*; Pa, *Picea abies*; Pf, *Perilla frutescens*; Sc, *Solidago canadensis*; Sl, *Solanum lycopersicum*; So, *Salvia officinalis*; Sr, *Stevia rebaudiana*.

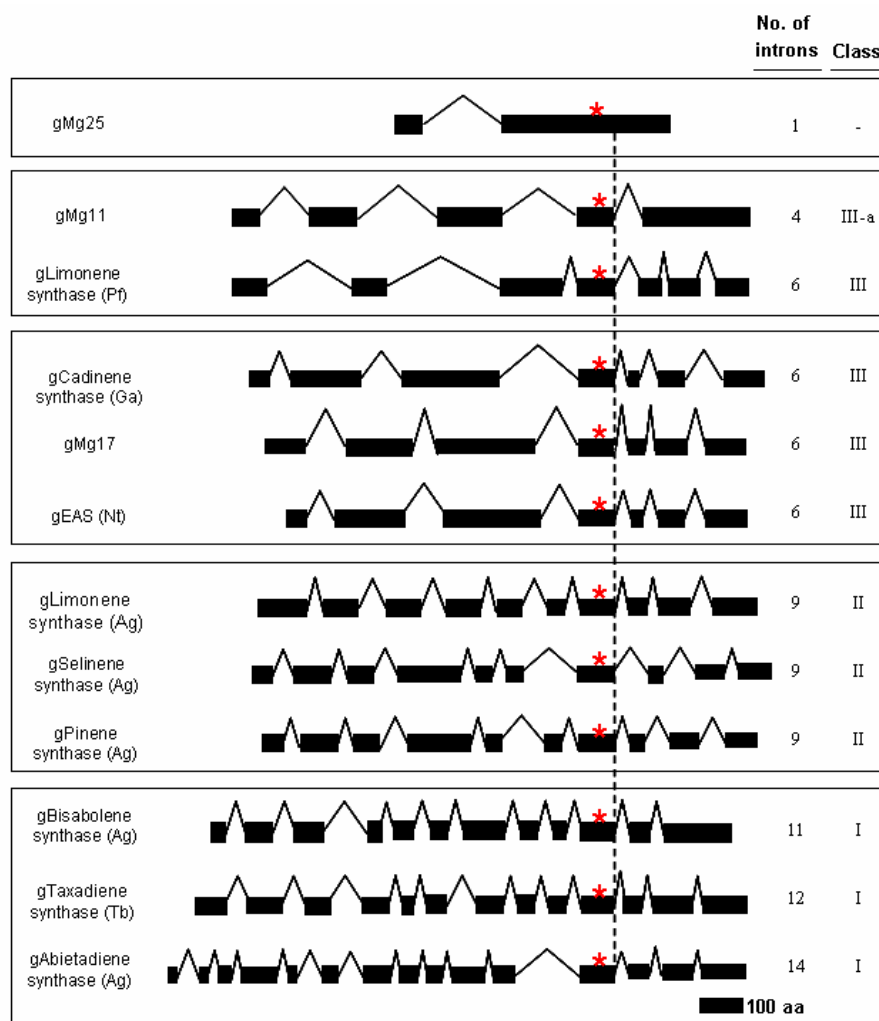


Figure 3.11. Comparison of the genomic organization of Magnolia terpene synthase genes to well known and characterized terpene synthase genes.

Mg17 was assigned to TPS Class III based on sequence alignments. However, Mg25 and Mg11 showed unique genomic structures in numbers and positions of exons/introns. Solid blocks (■) and opened arrowhead (Λ) represent exons and introns, respectively. Asterisks (*) represent the position of the DDXXD motif. Gene classification is based on the intron-exon organization according to Trapp and Croteau (2001). Ag represents *Abies grandis*; Nt, *Nicotiana tabacum*; Pf, *Perilla frutescens*; Ga, *Gossypium arboreum*; and Tb, *Taxus brevifolia*.

Table 3.1. A summary of terpene synthase genes used to infer structural and functional features of the *M. grandiflora* terpene synthase genes.

Product	Species	Class	GenBank accession no.	Genomic DNA (bp) ^a	Amino acid	No. of introns
β-Cubebene	<i>Magnolia grandiflora</i>	sesqui		2,339	551	1
-	<i>Magnolia grandiflora</i>	sesqui ^b		4,357	555	4
α-Terpineol	<i>Magnolia grandiflora</i>	mono		2,783	593	6
Trichodiene	<i>Myrothecium roridum</i>	sesqui	AF009416	1,215	385	1
(-)-Limonene	<i>Perilla frutescens</i>	mono	AB005744	3,996	603	6
5-epi-Aristolochene	<i>Nicotiana tabacum</i>	sesqui	LD4680	2,341	550	6
(-)-Limonene	<i>Abies grandis</i>	mono	AF326518	3,059	637	9
δ-Selinene	<i>Abies grandis</i>	sesqui	AF326513	3,351	582	9
(E)-α-Bisabolene	<i>Abies grandis</i>	sesqui	AF326515	4,647	814	11
Taxadiene	<i>Taxus brevifolia</i>	di	AF326519	3,999	862	12
Abietadiene	<i>Abies grandis</i>	di	AF326516	4,623	868	14

Genomic sequences for gMg25, gMg11 and gMg17 were verified by comparing these to their corresponding cDNA clones, deriving the number, position and size of introns.

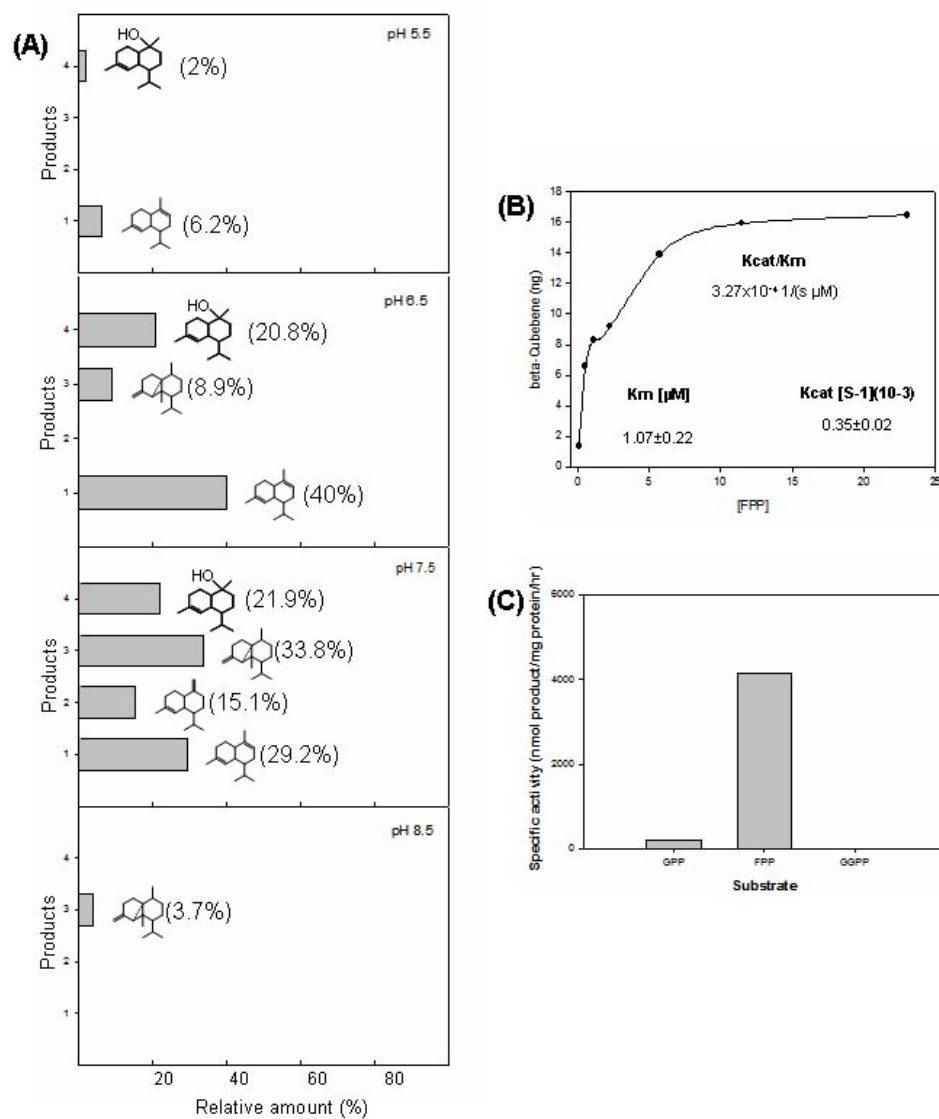
a, The size of each gene (genomic DNA) were calculated from the start codon to the termination codon.

b, Although an internal stop codon was substituted with a conserved amino acid tryptophan at this position, no enzyme activity was detected with a variety of substrates. Based on sequence homology and characteristic domains within the ORF, this gene was considered as a putative sesquiterpene synthase pseudogene.

Table 3.2. Phases of introns within Magnolia terpene synthase genes examined, and their comparison to those of known angiosperms and gymnosperms TPS genes.

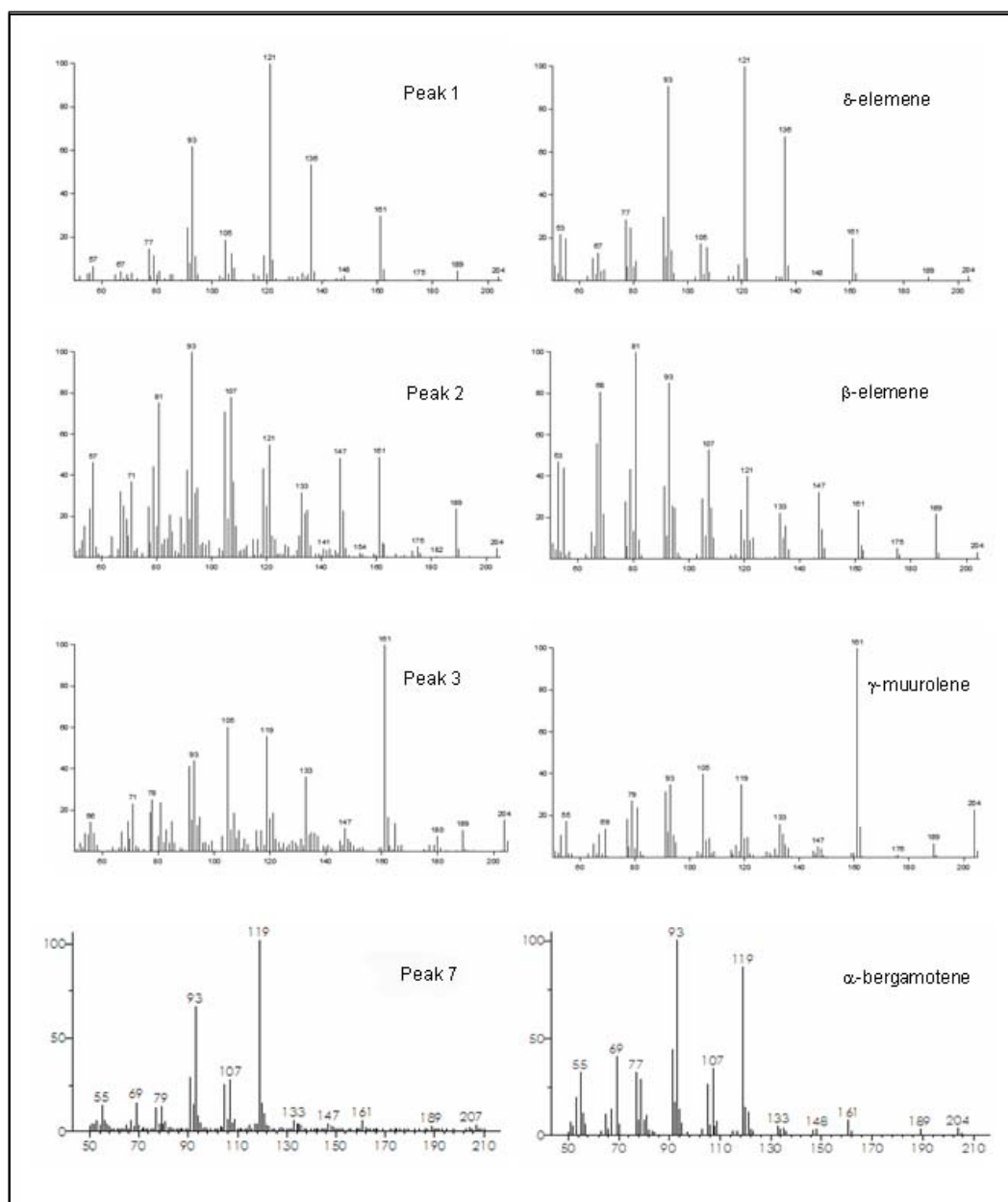
Roman numerals represent intron numbers. Intron phase number 0, 1 and 2 refers to the intron insertion before the first nucleotide or after the first or second nucleotide or a codon, respectively. A dash (-) indicates the absence of an intron. LS (Pf) refers to limonene synthase from *Perilla frutescens*; CS (Ga), δ -cadinene synthase from *Gossypium arboreum*; ES (Nt), 5-*epi*-aristolochene synthase from *Nicotiana tabacum*; LS (Ag), limonene synthase from *Abies grandis*; SS (Ag), selinene synthase from *Abies grandis*; PS (Ag), (-)-pinene synthase from *Abies grandis*; BS (Ag), α -bisabolene synthase from *Abies grandis*; TS (Tb), taxadiene synthase from *Taxus brevifolia*; and AS (Ag), abietadiene synthase from *Abies grandis*. All data except for of Magnolia genes were adapted from the literature (Trapp and Croteau 2001). CON represents conserved intron phasing in between Magnolia and the other *TPS* genes listed.

	Intron	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Basal Angiosperm	Mg25	-	-	0	-	-	-	-	-	-	-	-	-	-	-
	Mg11	-	-	0	-	-	-	-	1	-	-	2	2	-	-
	Mg17	-	-	0	-	-	-	-	1	-	-	2	2	0	0
Angiosperms	LS (Pf)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	CS (Ga)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	ES (Nt)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
Gymnosperms	LS (Ag)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	SS (Ag)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	PS (Ag)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	BS (Ag)	-	-	0	1	2	1	2	1	0	2	2	2	0	-
	TS (Tb)	-	-	0	1	2	1	2	1	1	2	2	2	0	0
	AS (Ag)	1	1	0	1	2	1	2	1	0	2	2	2	0	1
	CON	-	-	0	-	-	-	-	1	-	-	2	2	0	0

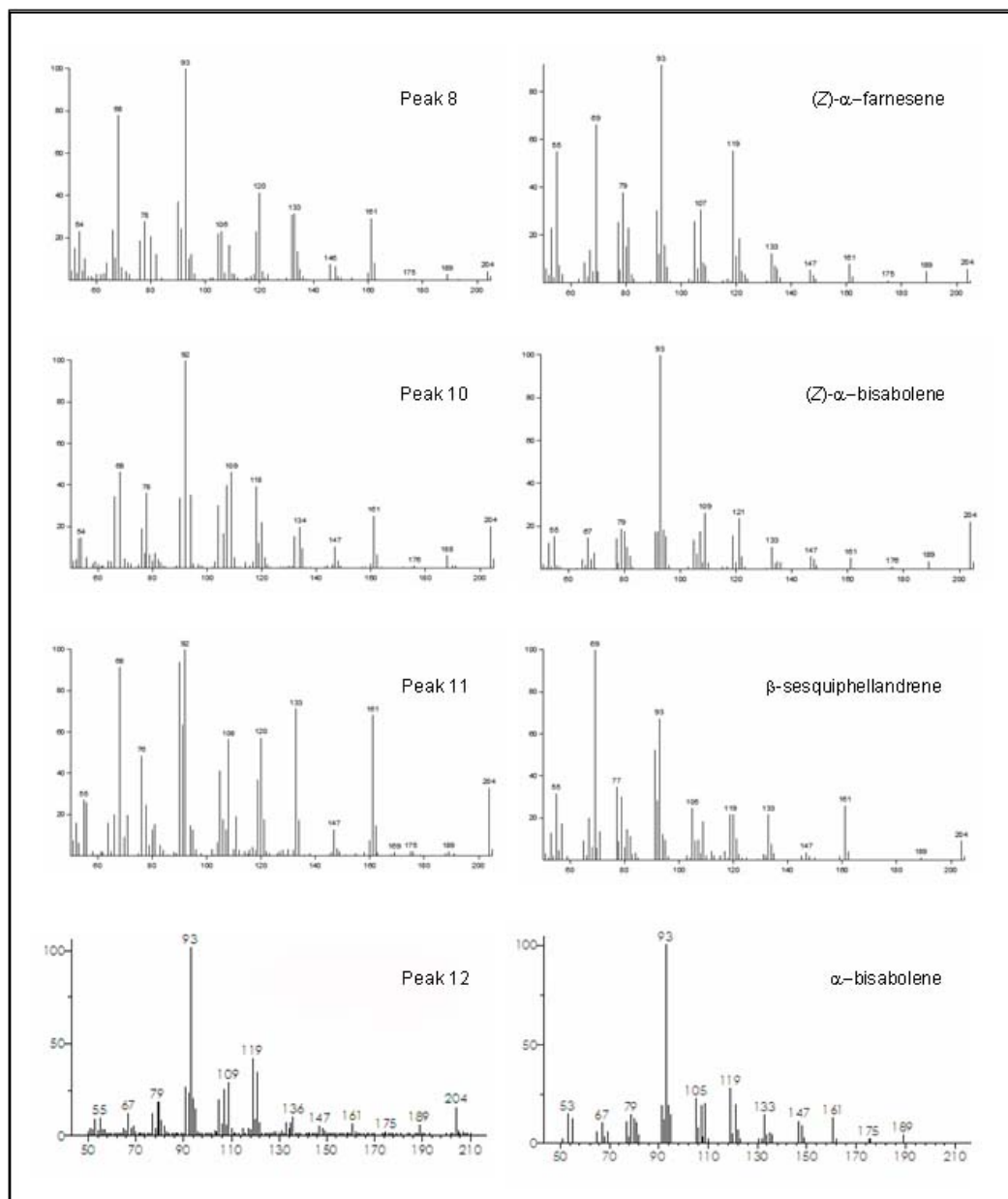


Supplementary Figure 3.1. Enzymatic properties of Mg25.

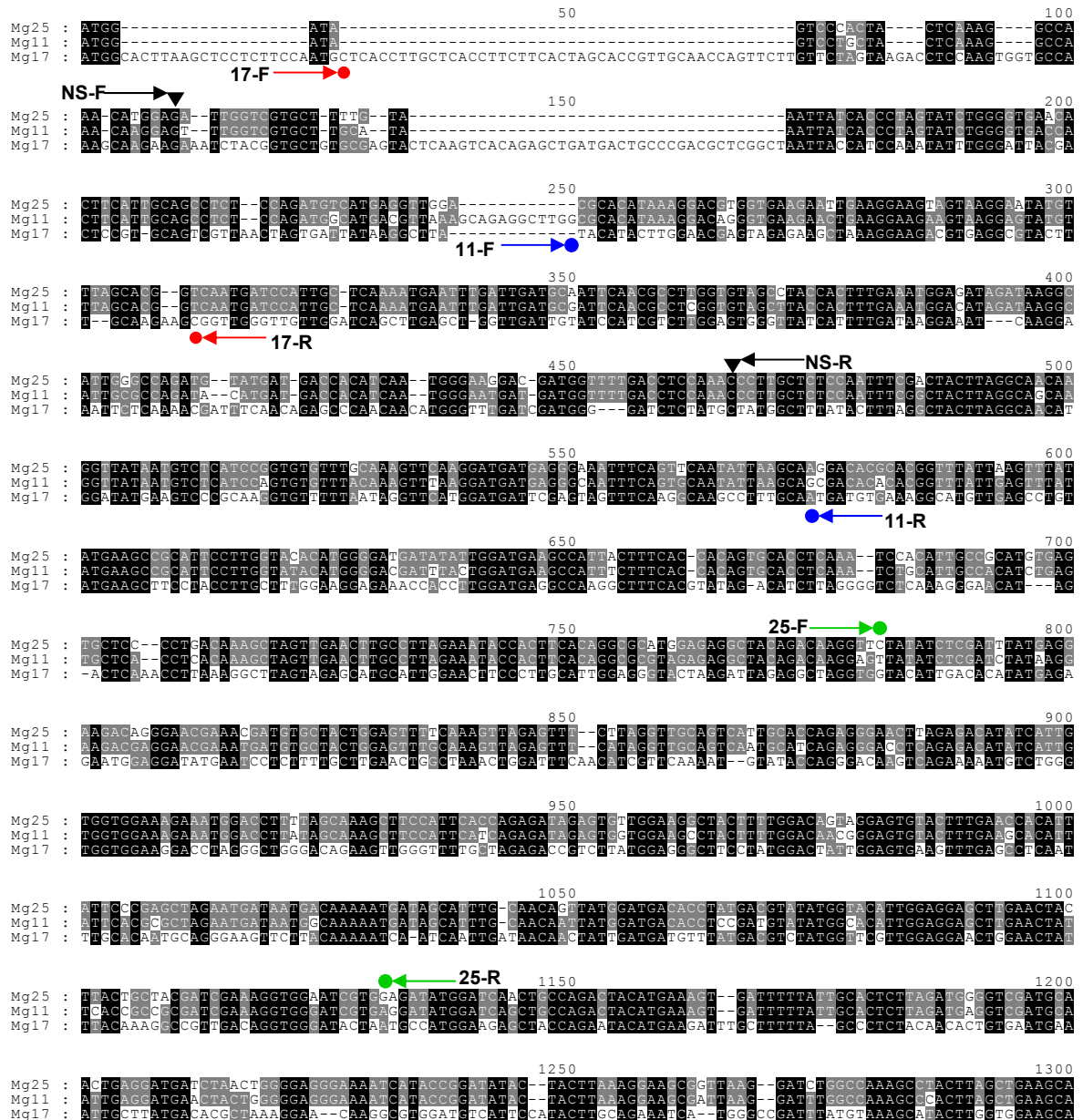
(A) pH-dependent reaction product profiles for major 4 hydrocarbon products: 1, α -muurolene; 2, τ -muurolene; 3, β -cubebene; 4, δ -cadinol, (B) enzyme kinetic analysis for FPP, and (C) allylic diphosphate substrate specificities.



Supplementary Figure 3.2. Additional MS spectra from GC peaks shown in Figs. 3.3 and 3.5 in comparison to NIST standards.



Supplementary Figure 3.2 (continued). Additional MS spectra from GC peaks shown in Figs. 3.3 and 3.5 in comparison to NIST standards.



Supplementary Figure 3.3. Nucleotide sequence alignments of *Mg25*, *Mg11*, and *Mg17* in order to design gene-specific PCR primers.

Primer names reflect their gene specific nature and orientation. For example, 17-F is a forward primer specific for *Mg17*. The NS primers were designed for non-specific amplification all three Magnolia *TPS* genes. The alignment was performed with the ClustalW server (<http://www.ch.embnet.org/software/ClustalW.html>) and graphically depicted with GENEDOC (<http://www.psc.edu/biomed/genedoc>).

CHAPTER IV FUNCTIONAL IDENTIFICATION OF A C-12 HYDROXYLASE FROM LETTUCE, THE FIRST COMMITTED MONOOXYGENASE IN COSTUNOLIDE BIOSYNTHESIS

INTRODUCTION

Lettuce plant (*Lactuca sativa*) accumulates a wide range of compounds that contribute to its flavor as a highly desired vegetable crop. In particular, the bitter components have been identified as sesquiterpene lactones that accumulate in specialized secretory cells, laticifers, which are distributed throughout and within the leaf and stem tissues (Olson et al., 1969). Less appreciated is the importance of these accumulating sesquiterpenes for the plant itself and their potential pharmaceutical applications. For example, lactucin (Figure. 4.1) which accumulates constitutively in the latex of lettuce and chicory (*Cichorium intybus*) possesses significant antimalarial activity (Leclercq, 1984; Bischoff et al., 2004). In contrast, the guaianolide lettucenin A (Figure. 4.1) accumulates in lettuce seedlings in response to microbial challenge and hence is considered a phytoalexin (Takasugi et al., 1985; Monde et al., 1990). Very interesting sesquiterpene lactone conjugates are also found in lettuce. Some *Lactuca* species accumulate oxalyl- and sulfate conjugated lactucin, deoxylactucin, and lactucopicrin (Figure. 4.1) in the latex of wounded tissues (Sessa et al., 2000). Costunolide (Figure. 4.1) has been widely observed in plants within the Asteraceae family including lettuce, chicory, artemisia, and costus (Rao et al., 1958; Herout and Sorm, 1959; Pyrek, 1977; de Kraker et al., 2002). Costunolide has drawn significant attention recently because of its potent and diverse pharmacological activities (Taniguchi et al., 1995; Kim et al., 2002; Choi et al., 2005).

Given the structural diversity observed within this class of sesquiterpenes found in lettuce and other related species and the documentation of their biological importance, efforts to define the pathways responsible for the biosynthesis of these interesting compounds have taken on greater importance. Progress in elucidating the biosynthetic pathway for costunolides in chicory root, for example, have recently been reported and enzyme extracts from chicory root shown to convert germacrene A to costunolide through several intermediates (Fig. 4.2) (de Kraker et al., 2001; de Kraker et al., 2002).

The pathway is thought to be initiated by the cyclization of farnesyl diphosphate (FPP) to the macrocyclic hydrocarbon germacrene A by the action of a sesquiterpene synthase (Fig. 4.2). In fact, several chicory germacrene A synthase cDNAs were recently characterized by Bouwmeester et al. (2002). Furthermore, de Kraker et al. (2001, 2002) observed that germacrene A could be hydroxylated to a germacra-1(10),4,11(13)-trien-12-ol in a reaction characterized as being mediated by a cytochrome P450 enzyme, and that the germacra-1(10),4,11(13)-trien-12-ol could be transformed into germacra-1(10),4,11(13)-trien-12-oic acid. These authors also suggested that an additional hydroxylation at C-6 could lead to lactone-ring formation either by a spontaneous or catalytic mechanism (de Kraker et al., 2001; de Kraker et al., 2002).

Given the recent successes of Ralston et al. (Ralston et al., 2001) and Takahashi et al. (Takahashi et al., 2005) in functionally characterizing cytochrome P450s (CYPs) that hydroxylate eremophilene sesquiterpene skeletons, we reasoned that similar efforts could facilitate the cloning and functional characterization of CYPs catalyzing similar reactions with germacrene A in lettuce. Such an effort was also encouraged by the recent release of a large EST sequence database (<http://cgpdb.ucdavis.edu>) for cDNAs prepared against lettuce and sunflower mRNA. That EST database was hence queried for sequences similar to previously characterized sesquiterpene hydroxylases, which lead to the isolation of several targeted, full-length cDNAs. Functional characterization of one of these cDNAs identified a germacrene A C-12 hydroxylase, the monooxygenase catalyzing the first committed step in costunolide biosynthesis.

MATERIALS AND METHODS

Biological and Chemical Materials

Chemical reagents for all the molecular manipulations and the media preparations were purchased from Invitrogen (Carlsbad, CA), Sigma-Aldrich Chemical Company (St. Louis, MO), Clontech (Mountain View, CA) or Promega (Madison, WI). All organic solvents were purchased from Fisher Scientific Inc. (Pittsburgh, PA) unless stated otherwise.

Lettuce (*Lactuca sativa* L.) was grown in a greenhouse maintained on a 16 hr photoperiod at 25°C to 28°C. After 2 weeks, the first true leaves were harvested and frozen at -80°C until used.

Maintenance, genetic engineering and propagation of yeast lines CALI7-1 and WAT11 were essentially as described by Takahashi et al. (Takahashi et al., 2005; Takahashi et al., 2007b). YPAD media for the production of sesquiterpenes by CALI7-1 cultures consisted of 1% (w/v) Bacto-yeast extract, 1% (w/v) Bacto-peptone, 2% (w/v) glucose, ergosterol (5 mg L⁻¹) from a stock of 10 mg L⁻¹ in 50% (v/v) Triton X-100, and 0.02% (w/v) adenine. SCE1 media (pH 5.5) for the selection of auxotrophic mutant CALI7-1 contained 0.6% (w/v) succinic acid, 0.54% (w/v) KOH, 0.67% (w/v) Bacto-yeast nitrogen base (without amino acids), 1.4 g L⁻¹ of Yeast synthetic drop-out media without histidine, leucine, tryptophan and uracil (Sigma, Product No. Y2001), 2% (w/v) glucose, ergosterol (5 mg L⁻¹) from a stock of 10 mg L⁻¹ in 50% (v/v) Triton X-100, and appropriate nitrogen base and amino acids (20 mg L⁻¹ uracil, 20 mg L⁻¹ tryptophan, 20 mg L⁻¹ histidine, and 30 mg L⁻¹ leucine) to supplement for auxotrophic mutants. SGI media for the selection of WAT11 transformants contained 0.1% (w/v) Bacto-casamino acids, 0.7% (w/v) Bacto-yeast nitrogen base (without amino acids), 2% (w/v) glucose, 0.02% (w/v) adenine, and select nitrogen base and amino acids (20 mg L⁻¹ uracil, 20 mg L⁻¹ tryptophan, 20 mg L⁻¹ histidine, and 30 mg L⁻¹ leucine). YPGE media for propagation of yeast for microsome preparation contained 1% (w/v) Bacto-peptone, 1% (w/v) Bacto-yeast extract, 0.5% (w/v) glucose, and 3% (v/v) ethanol.

Isolation of Lettuce CYP450 cDNAs

Putative *L. sativa* cytochrome P450s (LsCYPs) were initially identified via database searches at the web-based Compositae Genome Project Database (<http://cgpdb.ucdavis.edu>) site. Characterized plant sesquiterpene hydroxylases, CYP71AV1 (amorpha-4,11-diene monooxygenase from *Artemisia annua*, Genbank Accession No. DQ315671) (Teoh et al., 2006) and CYP71D20 (5-*epi*-aristolochene-1,3-dihydroxylase from *Nicotiana tabacum*, Genbank Accession No. AF368377) (Ralston et al., 2001) were used to query the EST library. Sequences for six partial cDNAs were recovered that showed the high probable scores with known plant P450 sequences. These

sequences were then used to obtain full-length cDNAs via 5' and 3' RACE methods. To isolate total RNA, leaves were collected, frozen immediately in liquid nitrogen, and then pulverized to a fine powder in liquid nitrogen. TRIzol RNA isolation reagent (Invitrogen, Carlsbad, CA) was added to the leaf powder and total RNA extracted from this material according to the manufacturer's recommendations. Small aliquots of mRNA (poly A+ RNA) were purified from 0.1 mg of total RNA using the Oligotex mRNA mini-preparation kit (Qiagen, Valencia, CA) according to the manufacturer's instruction manual. The purified mRNA was immediately reverse-transcribed into complementary DNA and employed in the amplification of double-stranded, cDNA ends. 5'- and 3'-RACE reactions were accomplished with the BD SMARTTM RACE cDNA Amplification Kit (Clontech, Palo Alto, CA). The adaptor sequence harboring a nested primer element was ligated onto the 5'-end of the double-strand cDNA, and a nested adaptor primer used together with the gene specific oligonucleotides for PCR amplifications. In some cases, a gene specific primer was used for the initial reverse transcription reaction to generate an initial gene specific, first strand, cDNA (GSD). The 3' and 5' RACE reaction products were amplified with the Advantage[®] 2 PCR Kit (Clontech, Palo Alto, CA). Fifty μ L reactions included 5 μ L of 10X PCR buffer, 200 nM dNTP mix, 1 μ L of Advantage[®] 2 Polymerase Mix, 200 μ M of adaptor-specific primer, 200 μ M of gene-specific primer, and 2.5 μ L of 100 to 300-fold diluted RACE-ready cDNA. The first amplification was completed with the following PCR program: 95°C for 3 min; 30 cycles of 30 sec at 95°C, 40 sec at 60 to 65°C, and 2 to 3 min at 72°C. The nested (secondary) PCR reaction was then performed with 5 μ L of a 200-fold diluted first amplification reaction aliquot, and nested oligonucleotide primer sets with a suitable PCR temperature program. The amplicons were cloned into the pGEM[®]-T Easy vector and ultimately subjected to automated DNA sequencing (BigDye Terminator Cycle sequencing Kit, Perkin-Elmer, Waltham, MA) with an ABI PRISM[®] 310 Genetic Analyzer, (Applied Biosystems, Foster City, CA). All custom oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA).

In order to obtain 3'-end amplicons for QGCA Contig2101 (*Ls2101*), the gene-specific forward primer used was 5'-AGA TCT AAA GGA AGA AAC AAG AGC GAT-3', and for QGG1C10 (*Ls1C10*), 5'-TTT CCG GAT TTT AAG TTT CTT TAC

GTG GTT-3'. To enhance gene specific amplification, nested gene-specific primers were designed for *Ls2101* (5'-GCA CAA GAA GAA GTG AGA TCG GTG GTG AAA GG-3') and for *Ls1C10* (5'-CAC AAA ACT CTC GAC AAA ATC TTT AAT GAC-3'). 3'-RACE for QGCA Contig7108 (*Ls7108*), QGCA Contig3597 (*Ls3597*), and QGCA Contig3399 (*Ls3399*) were not necessary as these sequences were reported in the database. However, for 5'-RACE, gene-specific reverse oligonucleotide primers for all clones were necessary (except for *Ls2101*); for *Ls7108*, 5'-TCG GAT CAC GAG GTT TAG GTA TGC GAG GTC-3'; for *Ls3597*, 5'-TGA CTC CAT GAC TTT CAG TCA TGT CAA TAT CTT TAG G-3'; for *Ls3399*, 5'-ACA CCA AAT AAA ATT CCT GGA CAG ATT CTC CTT CC-3'; and for *Ls1C10*, 5'-ATG AAA CGA TGA CTG CTG AAA TTT GGC CAA GC-3'. For nested gene-specific amplifications: for *Ls7108*, 5'-TTT GCT TGA GGT GCT AAC ATG GTG TTC AGC GAC-3'; for *Ls3597*, 5'-TCT AGC AAG AGT GAG CTC AAT AGA GAC CAA ACC-3'; for *Ls3399*, 5'-TCC CTT GAA ATC ATA ACA ATT GTC TTC GAA TCT TTC AGG-3'; and for *Ls1C10*, 5'-TCC CGC TAG CTG GTG AAT GTT TCC TAT CAG AGG-3' were used. The first 5'-end methionine found in-frame with the longest ORF was selected as the translation start site.

Full-length of cDNAs were amplified with the primers based on the sequence information from 5'- and 3'-RACE clones. Polymerase chain reactions were then performed with *Pfu* Ultra High-fidelity DNA Polymerase using 3'-RACE Ready cDNA as a template. Forward and reverse primers were as follows: *Ls7108*-F-*NotI* (a Forward primer with a *NotI* restriction enzyme site), 5'-ATG GAG CTT TCA ATA ACC ACC TCC-3', *Ls7108*-R-*SpeI* (a Reverse primer with *SpeI* restriction enzyme site), 5'-CTA AAA ACT CGG TAC GAG TAA CAA CTC AGT C-3'; *Ls3597*-F-*NotI*, 5'-ATG GCG ATC GAT ATG ACC ATC ATC CTC ATG-3', *Ls3597*-R-*SpeI*, 5'-TCA TGA AAA CGG AAT AAA AAC AGT TGG GAT TAC C-3'; *Ls2101*-F-*NotI*, 5'-ATG GAG CCT CTC ACC ATC GTC TCC-3', *Ls2101*-R-*SpeI*, 5'-CTA GGA CTT GAG GAT CGG GAC GAT TTT CAA CTG-3'; and *Ls3399*-F-*NotI*, 5'-ATG TCC CCC TTG ATG AGC CTC ATA GTG-3', *Ls3399*-R-*SpeI*, 5'-TTA GGA GTC CCC GCA ATT GAG ATC ATA AGG-3'. PCR conditions were as follows: 3 min at 95°C, 30 cycles of 30 sec at 95°C, 30 sec at 52 to 58°C, 1 to 2 min at 72°C, and final extension for 5 min at 72°C. In the case of low abundant transcripts, seven additional cycles were performed at 2 to 3°C

higher annealing temperatures after the 30 cycles of standard amplifications. The amplified full-length of cDNAs were ligated into a pGEM-T[®]-Easy vector and all the clones verified by DNA sequencing.

Yeast Vector Construction

To construct yeast expression vectors, the full-length P450 cDNAs were first isolated from the respective recombinant plasmids by *NotI* and *SpeI* restriction digestions and then ligated into the pET-GW-2ADHR vector. This donor vector harbors an *A. thaliana* CYP reductase gene, a modified multi-cloning site with *NotI*/*SpeI* cloning sites, two ADH1 promoters installed in opposite directions all integrated within a gateway cloning cassette (RfB). One of the ADH1 promoters was oriented to drive expression of the P450 gene inserted into the *NotI*/*SpeI* cloning site, and the other was oriented to direct transcription of the Arabidopsis cytochrome P450 reductase gene. The final donor vectors containing different combinations of unique lettuce CYPs and the Arabidopsis CYP reductase were mobilized in recombination cloning reactions into the YEp-GW-LEU vector according to (Takahashi et al., 2007b). The lettuce germacrene A synthase gene used in this work was originally isolated and characterized by Dr. Shuiqin Wu working in the Chappell laboratory. Dr. Wu was also responsible for construction of the yeast vector YEp352-LsGS-URA which directs over-expression of the lettuce germacrene A synthase gene in yeast.

Yeast Transformation

For the generation of sesquiterpene products via yeast fermentation, the YEp-2ADH-P450R-LEU vectors containing full-length CYPs and a cytochrome P450 reductase were transformed into yeast strain CALI7-1 (*leu2*, *trp1*, *his3*, *ura3Δ*, *erg9Δ::HIS3*, *sue*, *dpp1*) according to the method described previously by (Takahashi et al., 2007b). For *in vitro* biochemical work requiring microsomal preparations, the pESC-Ls7108-URA vector was transformed into the WAT11 yeast strain (Urban et al., 1997) using the method described in Takahashi et al. (2007). For both transformation protocols, competent yeast cells were prepared from a yeast colony cultured in 5 mL of YPGE media for 24 hrs at 30°C. Cells were diluted in 100 mL of YPGE media making OD₆₀₀ of

0.05 to 0.08 and were grown for 4 to 5 hrs at 30°C. When OD₆₀₀ of the cell was 0.4 to 0.5, the cells were harvested by centrifugation at 3,000g for 5 min, washed with 25 mL of sterile water and then resuspended in 1 mL of sterile water. For transformations, 100 µL of the competent cells in a 1.5 mL eppendorf tube were mixed with 100 µg of denatured salmon sperm DNA, 0.5 to 1 µg of plasmid DNA, 0.1 M lithium acetate, and PEG 3500 (final to 16.7%, v/v), followed by vigorous vortex mixing for 30 sec. The mixtures were then incubated at 42°C for 15 min. After centrifugation, the pellet was resuspended in 600 µL of sterile water and plated onto selection plates for appropriate prototrophic growth (Takahashi et al., 2007b).

Co-expression of Germacrene A Synthase and Putative P450 in Yeast

The YEp-2ADH-P450R-LEU vectors and YEp352-LsGS-URA were transformed separately and in various combinations into CALI7-1 yeast cells. Five to twenty independent colonies grown on selection media were analyzed by a PCR method and positives selected for further growth. Transformed colonies were cultured in 10 mL of appropriate selection media. To extract hydrocarbons, a two-step extraction procedure was employed. First, cells were broken by mixing equal volumes of water-resuspended yeast cells with acetone. After 20 to 30 min incubation at room temperature, terpenes were extracted twice with two volumes of pentane:ether mixture (4:1, v/v). Terpenes in the pentane:ether extracts were identified by GC-MS analysis after concentrating the extracts to final volume of 10 to 30 µL under a gentle stream of nitrogen gas.

For a large-scale production, 1 L seeded cultures were grown for 7 to 8 days and acetone/pentane:ether extracts prepared as described above. After concentrating of pentane:ether extracts under nitrogen gas, the culture products were separated on the fluorescent silica gel TLC (Thin Layer Chromatography) plate using hexane:ethyl acetate (85:15, v/v) as the developing solvent. Desired compounds were isolated from the TLC plate by scraping the corresponding zones and eluting the TLC scrapings with ethyl acetate. The culture product(s) was analyzed further by GC-MS and NMR analyses. The NMR analysis was kindly provided by Dr. Juan Faraldos and Dr. Bob Coates, Chemistry Dept., University of Illinois, Urbana-Champaign, IL.

P450 Expression in Yeast Cells and Enzyme Assay

For microsome preparation, one colony of the WAT11 cell transformed with pESC-Ls7108-URA was grown in 10 mL of SGI selection media. After 24 hrs of incubation, the culture was diluted by adding 5 mL of the seed culture into 250 mL of the identical media. The cell cultures were incubated for up to 2 days until the media glucose was depleted. Once the glucose was depleted, the cultures were supplemented with 2% (w/v) galactose to induce transcription mediated by the *GAL10* promoter. After an additional 16 to 20 hrs of incubation, cells were collected by centrifugation at 7,000 g for 10 min. The media was discarded and the resulting pellet was dissolved in 100 mL of TES buffer (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 0.6 M sorbitol). The cell suspension was centrifuged as above, and the pellet was resuspended in TES-M buffer (TES buffer with 10 mM of β -mercaptoethanol). After 10 min incubation at room temperature, the cells were harvested by centrifugation as above and resuspended in 2 to 3 mL of fresh extraction buffer (1% (w/v) bovine serum albumin, 2 mM β -mercaptoethanol, and 1 mM phenylmethylsulfonyl fluoride). The cells were subsequently homogenized and the microsomal fraction isolated by high-speed centrifugation according to Greenhagen et al. (2003). The amount of properly folded CYP protein accumulating in the microsomes was verified by CO difference spectroscopy and quantified using the 450 nm extinction coefficient of $91 \text{ mM}^{-1} \text{ cm}^{-1}$ (Omura and Sato, 1964).

Hydroxylase assays were carried out in 200 μL of total reaction volume in 1 mL glass vials. The reaction contained 4% (v/v) DMSO, 100 mM Tris-HCl, pH 8.0, 2.4 mM NADPH, 10 to 200 pmol of microsomal proteins, and 2 to 100 μM of germacrene A as a substrate. Where indicated, microsomes were pre-incubated with germacrene A at 30°C for 5 min before adding 2.4 mM NADPH to initiate the reaction. The reaction was then terminated and extracted by adding five volumes of ethyl acetate. The extractions were performed twice and the combined extract was concentrated to 10 to 30 μL under the stream of nitrogen gas. Reaction products were identified by GC-MS analyses using a Restec Rtx-5 capillary column (30 m x 0.32 mm, 0.25- μm phase thickness). The injector port was maintained at 220°C in the splitless mode and the initial oven temperature of 40°C (0.5 min) was increased a $4^\circ\text{C}/\text{min}$ gradient to 200°C followed by a $20^\circ\text{C}/\text{min}$

gradient to 300°C. Mass spectra were recorded at 70 eV, scanning from 35 to 270 atomic mass units.

NMR Analysis

¹H NMR spectra were recorded on a Varian 500 spectrometer (500 MHz for ¹H and 125 MHz for ¹³C). Chemical shifts are given in parts per million using TMS as internal standard. CDCl₃ (Aldrich Chemical Co.) was purified by filtering through basic alumina (Brockmann I, standard grade, 150 mesh, 58 Å) and dried overnight over molecular sieves (4 Å) in the dark. Benzene-*d*₆ was used without purification. Preparative TLC was carried out with Merck 60F-254 plates with 0.25 mm thickness. Visualizations of the TLC spots were performed by spraying with a 0.1% solution of berberine-HCl in EtOH, and examination with a UV light. *n*-Pentane (HPLC grade, Fisher Scientific) was used without further purification. Molecular modeling was performed with the program CS Chem 3D Pro[®] (version 8.0) from *Cambridge Soft*. The start geometries were calculated by the molecular mechanical MM2 force field method (Allinger, 1977). All calculations were carried out on a standard PC equipped with a Pentium[R] 4 CPU 2.40 GHz processor and 512 MB RAM.

Molecular Modeling

As previously reported for germacrene A (Faraldos et al., 2007), a total of 24 start geometries were modeled using the MM2 force field method with relative steric energies ranging from 0.0 to 15.0 kcal/mol. After the minimization process, the four conformations (UU, UD, DU and DD) shown in Fig. 4.9 were judged to be the most likely to be populated in the gas phase. As before, the UU (20.09 kcal/mol) conformation was found to correspond to the global minimum, and the conformations UD, DD, and DU were interpreted as local minima with steric energies of 21.05, 21.13, and 22.58 kcal/mol, respectively.

RESULTS

Cloning Full-Length Lettuce Cytochrome P450s

In order to study the possible enzymes involved in the downstream processing of germacrene A in lettuce to more biologically active oxygenated species, the UC-Davis lettuce EST database was queried for cytochrome P450s, and especially those possibly associated with sesquiterpene metabolism. A large number of putative CYP ESTs are evident in this database with general queries for sequence domains highly conserved in all CYPs. However, fewer ESTs were identified when the database was searched using sequences for well characterized sesquiterpene oxygenases. Five such candidate EST candidates (*Ls7108*, *Ls3597*, *Ls2101*, *Ls3399*, and *Ls1C10*) were hence identified because of their similarity to biochemically defined P450s from the Asteraceae (CYP71AV1) (Teoh et al., 2006) and the Solanaceae (CYP71D20) (Ralston et al., 2001). All 5 ESTs represented partial cDNAs, thus necessitating experiments to obtain the complete sequence information for each of the corresponding genes. 5'- and 3'-end RACE experiments were successfully employed to determine the putative complete sequence information for 4 of the genes, and this information subsequently used to amplify full-length cDNA clones from mRNA isolated from lettuce seedlings. Three full-length cDNAs (*Ls7108* (1,467 bp), *Ls3597* (1,500 bp), *Ls2101* (1,473 bp)) were hence isolated. The three cDNAs were similar in size and showed strong similarities (60 to 65%) and identities (42 to 48%) between their protein and nucleotide sequences to one another (Fig. 4.3). These clones were also closely related to those CYPs used in the initial database query. *Ls7108* (489 aa) showed the highest amino acid sequence similarity to CYP71AV1 (87% identity, 95% similarity) (Teoh et al., 2006), while *Ls3597* (500 aa) was 51% identical and 70% similar to HPO (*Hyoscyamus muticus* cytochrome P450 monooxygenase) (Takahashi et al., 2007a) and *Ls2101* was 48% identical and 69% similar to HPO (*Hyoscyamus muticus* cytochrome P450 monooxygenase) (Takahashi et al., 2007a). All the three cDNAs contained highly conserved domains typical among the P450 families, such as the heme-thiolate-binding region (FXXGXRXCXG) and the oxygen-binding domain (A/G)GX(D/E)T(T/S).

Functional Verification of *Ls7108* as a Germacrene A Monooxygenase

To determine if any of these putative P450s might be involved in the costunolide biosynthesis, each of the P450s were expressed in yeast CALI7-1 cells, along with a

previously isolated germacrene A synthase cDNA (Wu and Chappell, unpublished data). CALI7-1 cells have been engineered for high level production of farnesyl diphosphate (FPP) (Takahashi et al., 2007) and co-expression of the lettuce germacrene A synthase cDNA was shown to shunt that FPP into high level production of germacrene A (Wu and Chappell, unpublished observations and Fig. 4.4B). Hence, if any of lettuce P450 enzymes were associated with costunolide metabolism and could be functionally coupled with the endogenous supply of germacrene A in the CALI7-1 cells, we would expect new, oxygenated terpene product(s) to accumulate in the cultures.

Each of the lettuce CYP and germacrene A synthase co-expression lines were compared with control yeast lines. The control lines included CALI7-1 without any transgenes, a line only harboring the germacrene A synthase (*LsGS*) gene (Fig. 4.4B) or lines harboring only the respective P450 gene (Fig. 4.4A). Among the yeast cells transformed with both the *LsGS* and the putative P450s, only the *LsGS* and *Ls7108*-combination line accumulated two new compounds (Fig. 4.4C, peaks c and d), as well as a high level of germacrene A (peak a) and small amount of farnesol (peak b). Farnesol accumulates in the CALI7-1 cells as a consequence of FPP de-phosphorylation or if the FPP is not readily converted to other terpene compounds (Takahashi et al., 2007). Neither of the other lines co-transformed lines with *Ls2101* or *Ls3597* in combination with *LsGS* showed any other compounds accumulating other than germacrene A and farnesol (data not shown), suggesting that these cDNA are either not properly expressed in the yeast or the encoded enzymes cannot utilize germacrene A as a substrate.

Both of the new compounds accumulating in the *LsGS*-*Ls7108* cell line appear to be primary reaction products from germacrene A, rather than successive reaction products. Both compounds showed greater polarity (greater retention time with increasing polarity) than germacrene A, and both compounds have MS ions of 220 amu (atomic mass units) (Fig. 4.5). The observed parent MS ion for germacrene A is 203 amu. The addition of a single hydroxyl substituent would thus add 17 amu, and the molecular mass for a mono-hydroxylated germacrene A would be 220 amu. Therefore, the two compounds (peaks c and d) accumulating in the *LsGS*-*Ls7108* CALI7-1 line could be different mono-hydroxylated compounds.

To fully identify the accumulating compounds, large cultures of the yeast were propagated in a nutrient-rich production media (SCE) and extractable terpene compounds isolated by preparative TLC. Significant amounts (approximately 0.4 mg L⁻¹) of the compound corresponding to peak (d) (Fig. 4.4C) were successfully isolated in this manner, but not for the compound corresponding to peak (c), and the isolated compound was verified by GC-MS. NMR analysis of the isolated compound was performed by Dr. Juan Faraldos in Dr. Robert Coates laboratory in the Chemistry Department at the University of Illinois, Urbana-Champaign, IL. The isolated compound was unequivocally identified as 12-acetoxy-germacrene A based on the observed NMR signals. To the best of our knowledge, NMR assignments for this compound have not been reported in the literature to date.

***In vitro* Characterization of the Lettuce Germacrene A Mono-Hydroxylase Activity**

To confirm and extend the functional identification of the *Ls7108* cDNA with *in vitro* biochemical investigations, microsomes from the WAT11 yeast cells over-expressing the *Ls7108* cDNA were prepared and incubated with germacrene A plus and minus reducing equivalents in the form of NADPH. In the presence of NADPH, the microsomes were able to convert germacrene A to a single monooxygenated product which migrated with the same retention time and had the same MS as the compound corresponding to peak (c) observed in the earlier *in vivo* experiments (Fig. 4.4C). The dominant sesquiterpene product detected in the earlier *in vivo* studies, 12-acetoxy-germacrene A, was not detected in the *in vitro* incubations (Fig. 4.6B). Hence, a variety of experiments were performed to optimize the reaction conditions for product formation and to determine if the germacrenol compound generated *in vitro* was a true catalytic product. Without NADPH, peak (c) was not evident in the GC-MS chromatograms (Fig. 4.6). Generation of this novel germacrenol compound was also dependent upon the amount of enzyme added (varying amounts of microsomes yielded variable amounts of the germacrenol compound, data not shown), pH and incubation time. Similar to the results reported by de Kraker et al. (2001), little if any mono-hydroxylated product was formed at pHs lower than 5.5 and the optimal pH appeared to be around 8.0 (Fig. 4.7). Equally important to note, mono-hydroxylated product formation was linear in

incubations up to 60 min without any other products, including the germacrene-12-ol (Fig. 4.8), being detected by GC-MS profiling.

NMR Data for 12-Acetoxy-Germacrene A

The compound corresponding to peak d in the GC chromatograms was unequivocally identified by NMR analysis. The pertinent data are as follows: TLC R_f 0.57 (10% AcOEt/hexane); ^1H NMR (500 MHz, CDCl_3 , 25° C); 5.21 (1H, broad s, H5 conformer DU, 18%), 5.09 (1H, broad s, H5 conformer UD, 33%), 5.03-4.95 (3H, broad m, H1 conformer DU and UD (51%), and H13 conformer DU, UD and UU (100%)), 4.93 (1H, s, H13' conformer DU, UD and UU (100%)), 4.78 (1H, d, $J = 11.7$ Hz, H1 conformer UU, 49%), 4.59 (2H, broad s, H12 conformers DU, UD and UU (100%)), 4.49 (1H, d, $J = 10.6$ Hz, H5 conformer UU, 49%), 2.24-1.74 (11 H, m, all conformers), 2.11 (3H, s, 12-OCOCH₃ conformer UU, 49%), 2.10 (6H, s, 12-OCOCH₃ conformer DU and UD, 51%), 1.68 (3H, s, H14 conformer DU), 1.60 (3H, s, H14 conformer UD), 1.53 (3H, s, H15 conformer UU), 1.49 (6H, broad s, H15 conformer DU and UD), 1.39 (3H, s, H14 conformer UU) ppm. The NMR data is in good agreement with those values previously reported for the parent sesquiterpene germacrene A (Faraldos et al., 2007), and similar esters (Zdero et al., 1988) in CDCl_3 at room temp.

^1H NMR (500 MHz, C_6D_6 , 25° C); 5.21 (1H, broad s, H5 conformer DU), 5.07 (1H, broad s, H5 conformer UD), 5.01 (1H, broad s, H1 conformer UD), 4.98 (1H, s, H13 conformer DU, UD and UU), 4.91 (1H, broad s, H1 conformer DU), 4.87 (1H, H13' all conformers), 4.74 (1H, broad d, $J = 11.6$ Hz, H1 conformer UU), 4.59 (2H, broad s, H12 all conformers), 4.44 (1H, broad d, $J = 10.6$ Hz, H5 conformer UU), 2.24-1.74 (11H, m, all conformers), 1.74-1.67 (1H, broad m, H8 $_{\alpha}$, all conformers), 1.71 (3H, s, 12-OCOCH₃ all conformers), 1.59-1.54 (1H, broad m, H8 $_{\beta}$, all conformers), 1.42 (3H, broad s, H15), 1.38 (3H, broad s), 1.28 (3H, broad s, H14). The NMR Data is in good agreement with those values previously reported for the parent sesquiterpene germacrene A (Faraldos et al., 2007) in C_6D_6 at room temp.

DISCUSSION

While de Kraker et al. (1998, 2001, 2002) were able to postulate a pathway for costunolide biosynthesis based on physiological experiments performed with chicory root extracts, a complete recapitulation of the pathway from FPP to costunolide has yet to be achieved. Obtaining the genes encoding for the various steps, over-expressing these in a suitable heterologous host in various combinations, or isolating the individually over-expressed enzymes from heterologous hosts and combining these *in vitro* assays should provide the tools to more conclusively define the costunolide pathway. Hence, the aim of the current work was directed at isolating a gene coding for the initial mono-hydroxylation of germacrene A. This effort was facilitated in large part because lettuce, like chicory is a member of Asteraceae and accumulates sesquiterpene lactones including costunolide, has been the subject of a very large EST sequencing project orchestrated out of the University of California-Davis. Our initial goals were to identify the genes coding for a germacrene A synthase, work completed independently by Shuiqin Wu in the Chappell laboratory (unpublished results), and a mono-hydroxylase responsible for the first and putative rate-limiting step in costunolide biosynthesis.

Interestingly, a single cytochrome P450 cDNA designated *Ls7108* was shown to convert germacrene A into two mono-hydroxylated forms in an *in vivo* format, but only a single mono-hydroxylated product when assayed *in vitro*. While we were unable to conclusively determine the structure of the *in vitro* mono-hydroxylated product (more specifically, the regio- and stereo-chemical positioning of the hydroxyl substituent), the second and dominate product generated *in vivo* was identified as 12-acetoxy-germacrene A. This compound was not an expected mono-hydroxylated germacrene A, but it was possible that this compound might be generated by adding an acyl group to the hydroxylated germacrene A catalyzed by endogenous yeast enzyme(s). To solve this problem, additional efforts will be required to demonstrate that the peak c product is germacra-12-ol. One way to identify this unique compound would be to purify the peak c product and analyze it by NMR spectrometry. Alternatively, germacra-12-ol could be recovered and identified from the chemical reactions such as deacylation or saponification, where the acyl group is cleaved from 12-acetoxy-germacrene A.

Several plant cytochrome P450s capable of catalyzing the hydroxylation of sesquiterpene scaffolds have recently been described (Ralston et al., 2001; Takahashi et al., 2005; Teoh et al., 2006). CYP71D20, for instance, was shown to catalyze the successive hydroxylation of an eremophilene scaffold (Takahashi et al. 2005), and particularly relevant, CYP71AV1 was demonstrated to successively hydroxylate a terminal methyl group of the isopropenyl substituent of amorphadiene, a cadinene-type sesquiterpene (Teoh et al., 2006). The description of these enzymes provides critical information to further our analysis of Ls7108. For example, CYP71AV1 has been shown to actually catalyze the successive oxidation of the methyl substituent of amorphadiene through an alcohol to an aldehyde, and finally onto an acid form. This was demonstrated *in vitro* with heterologously expressed enzyme (Teoh et al., 2006) as well as *in vivo* with coupled over-expression of the terpene synthase (amorphadiene synthase) and CYP71AV1 in yeast (Ro et al., 2006). A closer comparison between the methodologies employed by these other researchers working with CYP71AV1 and those used here are certainly warranted. Additionally, more rigorous structure-function comparisons between the two enzymes, like that recently described by Takahashi et al. (2007), might also provide additional insights into the specific mechanism catalyzed by these enzymes and hints as to how Ls7108 might be engineered for enhance activity(s).

Lastly, the identification of the *Ls7108* gene should be advantageous for finding the genes coding for the next steps in the costunolide pathway. Ls7108 should obviously be used as a search query of the lettuce EST library to isolate other homologs, perhaps finding more efficient germacrene A mono-hydroxylases or hydroxylases functioning downstream of Ls7108 in the costunolide pathway. But equally important will be the utility of Ls7108 in providing germacra-1(10),4,11(13)-trien-12-ol, the substrate for the next step in the putative costunolide pathway. This may be achieved by facilitating the biosynthesis of this mono-alcohol in sufficient amounts for *in vitro* screens with enzymes obtained for over-expression of candidate genes in suitable heterologous hosts, or providing for coupled, *in vivo* experiments wherein yeast over-expressing the germacrene A synthase and Ls7108 can be used to screen for other candidate genes, also over-expressed and resulting in further metabolism of germacra-1(10),4,11(13)-trien-12-ol.

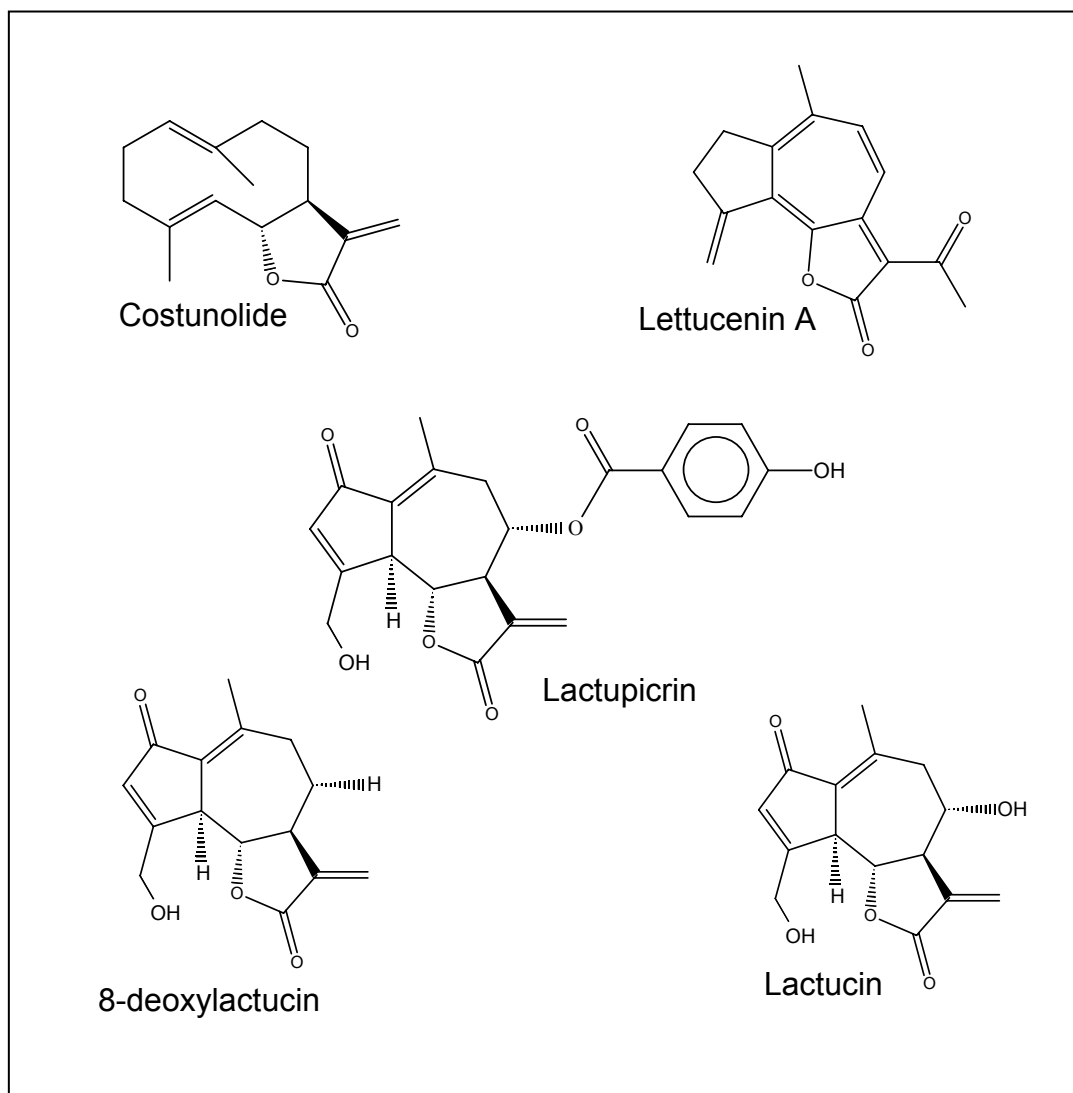


Figure 4.1. An illustration of the sesquiterpenes accumulating constitutively in the latex of lettuce and chicory (*Cichorium intybus*).

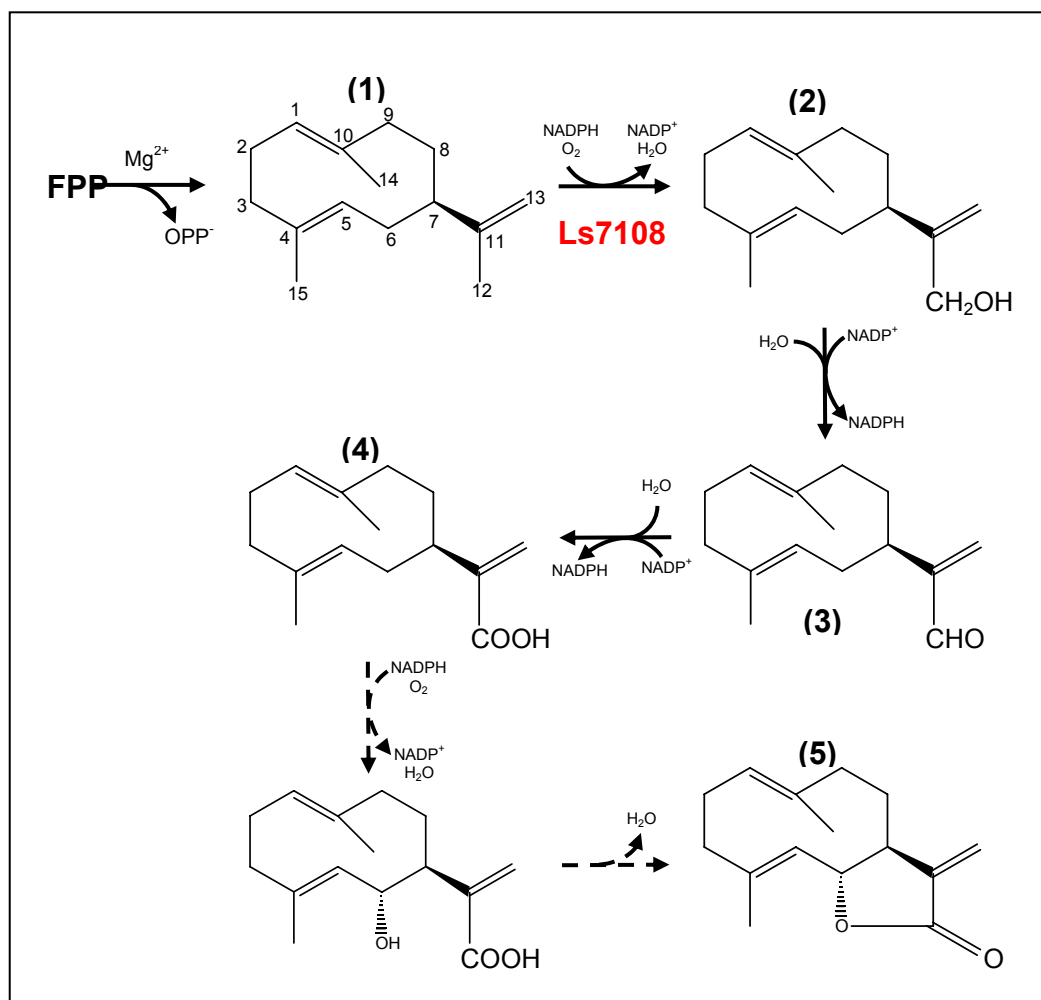


Figure 4.2. A proposed biosynthetic pathways for the sesquiterpene costunolide based on the work of de Kraker et al. (2001). The pathway begins with a conversion of FPP to (1) germacrene A catalyzed by a germacrene A synthase, followed by sequential and successive hydroxylation/oxidations of the hydrocarbon scaffold to (2) germacra-1(10),4,11(13)-trien-12-ol; (3) germacra-1(10),4,11(13)-trien-12-al; (4) germacra-1(10),4,11(13)-trien-12-oic acid; and ultimately formation of the lactone (+)-costunolide (5) by either spontaneous or enzymatic formation. Ls7108 represents a cytochrome P450 characterized in this work and shown to hydroxylate C12 of germacrene A. The dashed arrows represent proposed steps to the final lactone system found in costunolide.

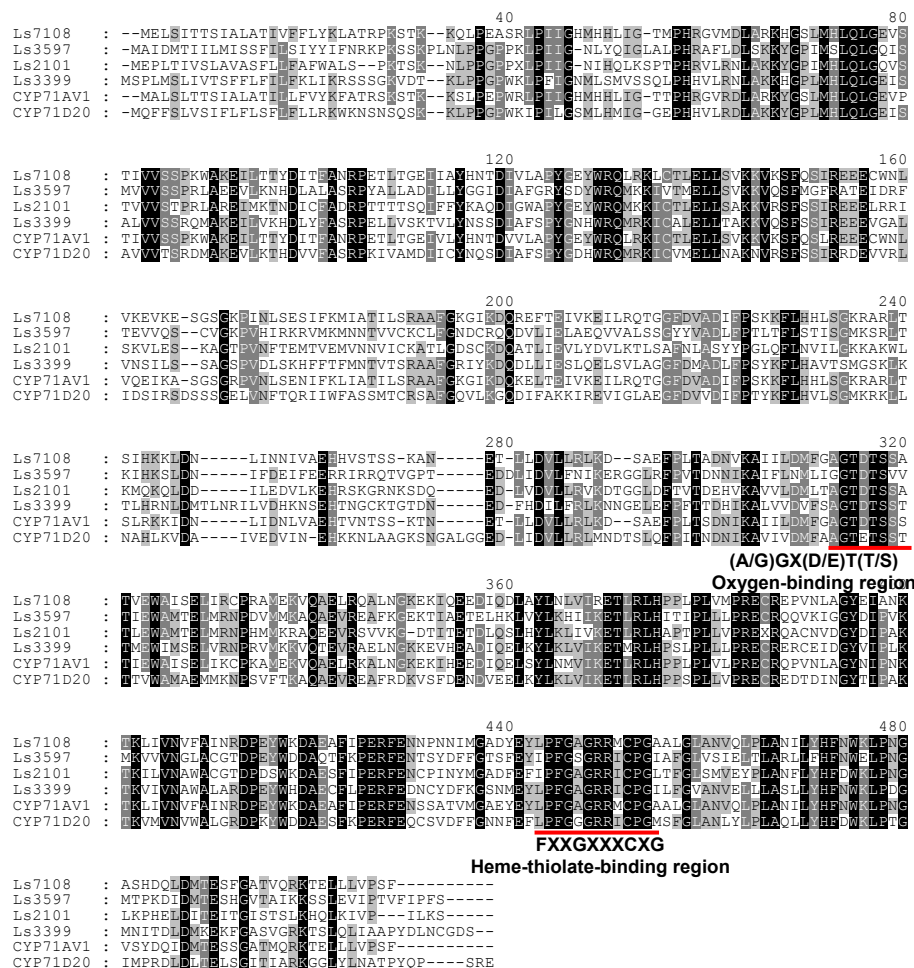


Figure 4.3. A deduced amino acid sequence alignment of 4 putative P450 genes isolated from *Lactuca sativa* in comparison to 2 well characterized sesquiterpene hydroxylases, 5-*epi*-arisolochene dihydroxylase (CYP71D20) and amorphadiene oxidase (CYP71AV1). All four genes from lettuce were initially identified from a public EST databases based on the presence of a characteristic heme-thiolate-binding motif, FXXGXXXCXG and an oxygen-binding domain, (A/G)GX(D/E)T(T/S), and full-length clones isolated using 5' and 3' RACE methodology. The black shades represent conserved residues among all six genes; the dark gray shades represent conserved residues among five genes; and the light gray shades represent conserved residues among four genes. CYP71AV1 (amorpha-4,11-diene monooxygenase) (Teoh et al., 2006) and CYP71D20 (5-*epi*-aristolochene-1,3-dihydroxylase) (Ralston et al., 2001) are known plant P450s from *Artemisia annua*, Genbank Accession No. DQ315671 and *Nicotiana tabacum*, Genbank Accession No. AF368377, respectively. The alignment was prepared using ClustalW and visualized with GENEDOC.

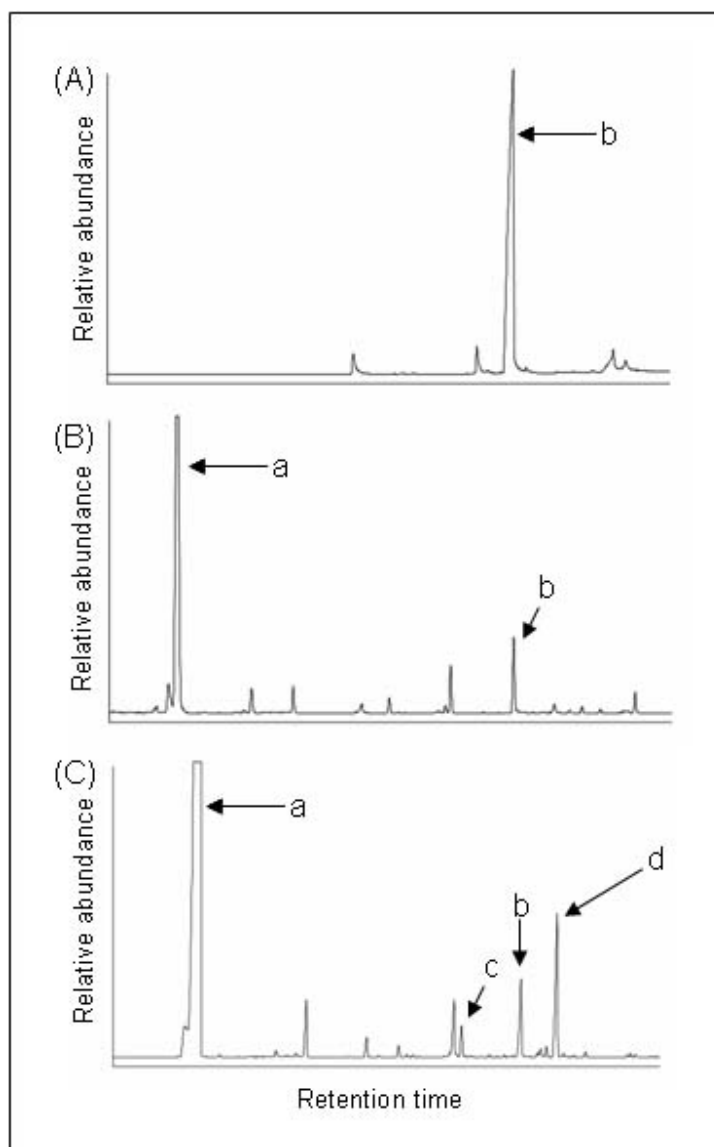


Figure 4.4. Functional identification of Ls7018 by its co-expression along with germacrene A synthase in yeast cells.

Yeast developed for high-level production of farnesyl diphosphate were transformed with a vector harboring only Ls7108 (A), or only germacrene A synthase (B), or both germacrene A synthase and Ls7018 (C). The terpene compounds synthesized by the yeast were extracted, profiled by GC-MS, and the peaks identified by comparisons to authentic standards and MS profiles in the NIST library except for peak (d), which was identified by NMR. Peak a is germacrene A; peak b, farnesol; peak c remains unknown; and peak d is 12-acetoxy-germacrene A.

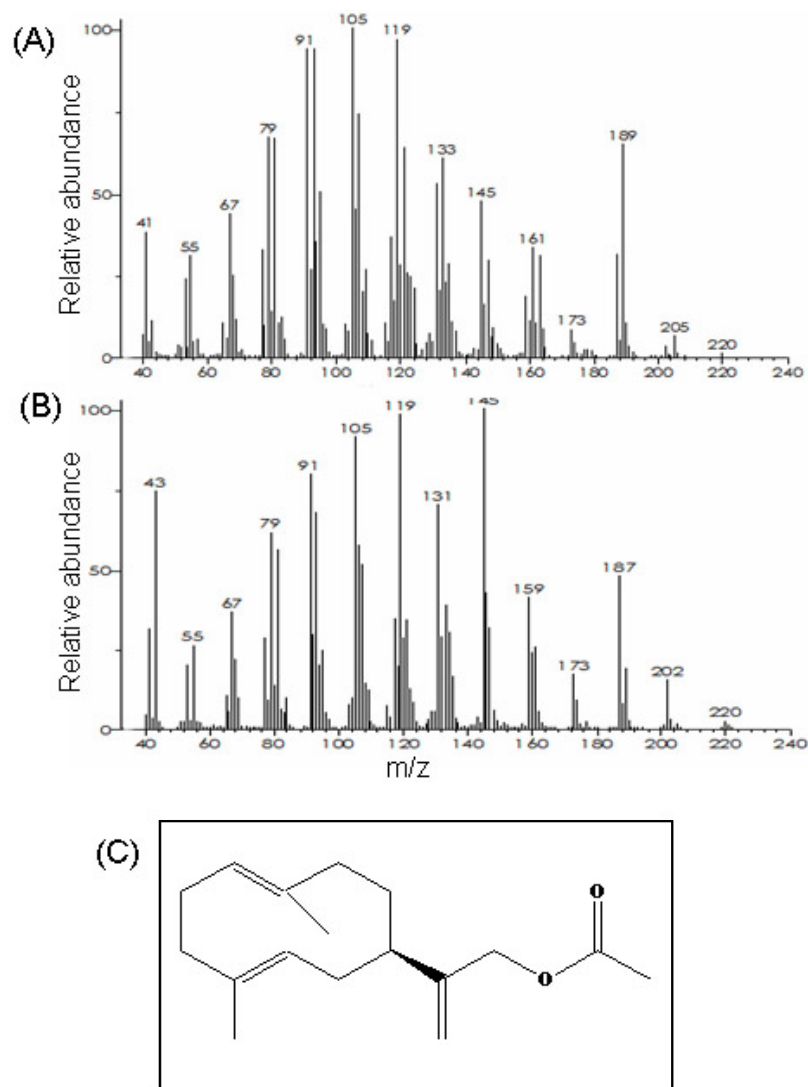


Figure 4.5. Mass spectral patterns for the compounds generated in yeast as a consequence of co-expressing Ls7018 along with germacrene A synthase.

The fragmentation pattern in panel A represents the mass spectra for the compound that migrated as peak (c) in the chromatographs shown in Fig. 4.4, and panel B represents the mass spectra for the compound that migrated as peak d. Final chemical identification of peak (d) as 12-acetoxy-germacrene A, whose structure is shown in panel C, was obtained by NMR analysis.

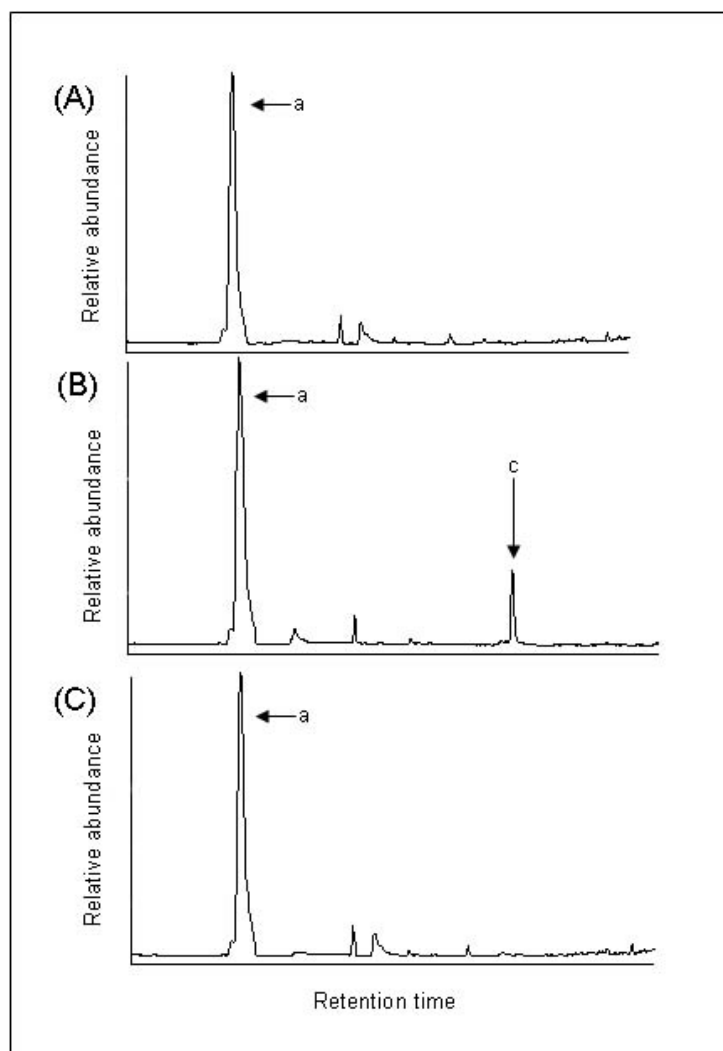


Figure 4.6. *In vitro* characterization of the cytochrome P450 encoded by Ls7018.

Microsomes were isolated from yeast over-expressing the Ls7018 cDNA, incubated *in vitro*, and the reaction products profiled by GC-MS. Incubations were performed without microsomes (A); with microsomes, 100 mM of germacrene A and 2.4 mM NADPH (B); and with microsomes and germacrene A (without NADPH) (C). Peak (a) corresponds to germacrene A and peak (c) is the same compound observed as an *in vivo* product generated by yeast over-expressing the Ls7018 cDNA (see Fig. 4.4).

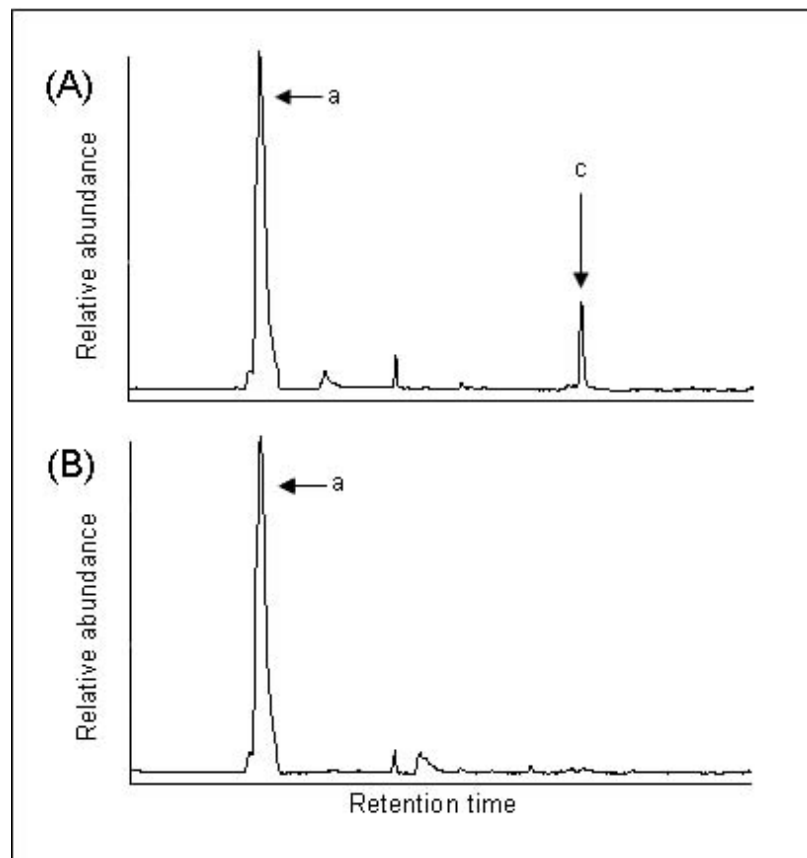


Figure 4.7. The pH-dependence of the enzyme activity encoded by the Ls7108 cDNA. Microsomes from yeast over-expressing the Ls7108 cDNA were incubated with 100 mM germacrene A and 2.4 mM of NADPH at pH 8.0 (A) or at pH 5.5 (B) and the reaction products profiled by GC-MS. Peak (a) corresponds to germacrene A and peak (c) to the same compound observed in Figs 4.4 and 4.6.

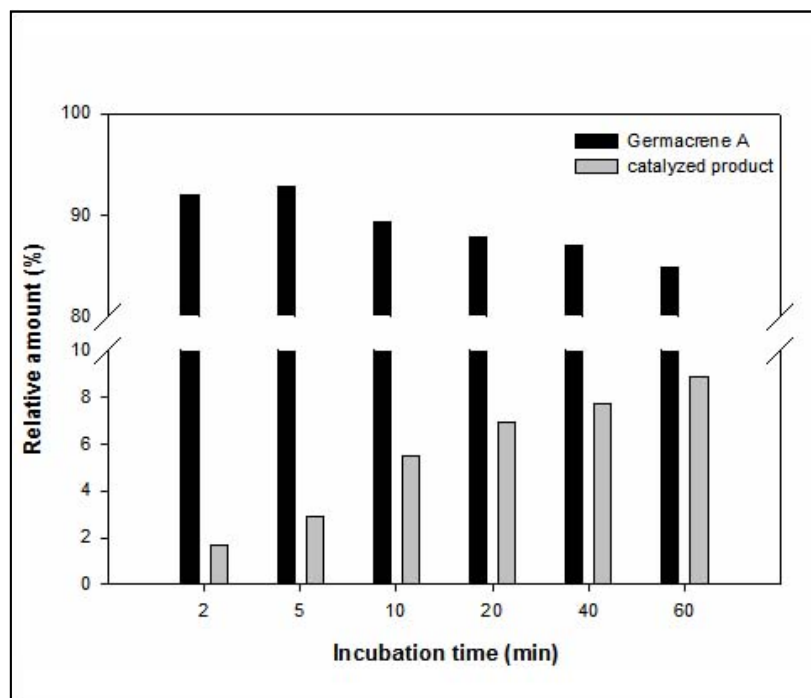


Figure 4.8. The time-dependence of the reaction catalyzed by the enzyme encoded by the Ls7108 cDNA.

Microsomes isolated from yeast over-expressing the Ls7108 cDNA were incubated with 100 mM of germacrene A and 2.4 mM of NADPH at pH 8.0 for the indicated times before extracting and profiling the reaction products by GC-MS. Relative amounts of germacrene A and the only observed reaction product (peak c) were calculated from total ion chromatographs.

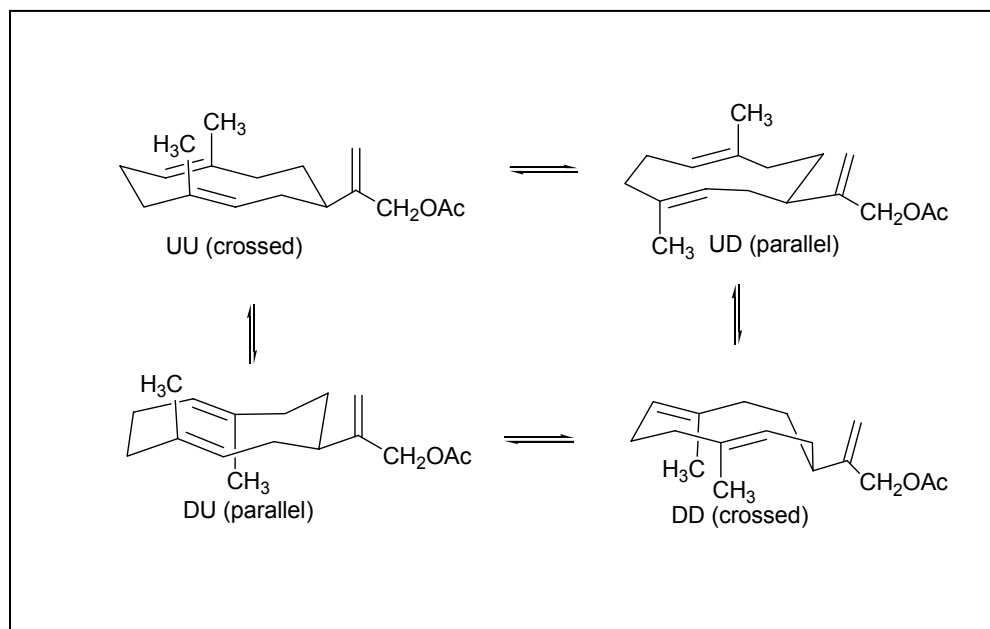


Figure 4.9. Possible conformation of 12-acetoxy-germacrene A.

The conformers are denoted as UU, UD, DU, and DD in reference to the U (Up) and D (down) orientations of C10 and C14 methyl groups with respect to the 10-membered ring (Allinger, 1977).

CHAPTER V SUMMARY AND PERSPECTIVES

Terpenes play significant roles in plants by contributing to secondary metabolism as well as primary metabolism throughout the entire life of plants. In order to fully understand terpene metabolism, efforts to define the biochemical mechanisms responsible for the biosynthesis of these structurally diverse molecules, and to elucidate the molecular mechanisms controlling terpene metabolism are being pursued by numerous laboratories worldwide. Although many terpene genes have been isolated from various plant sources and functionally characterized, the mysteries of the biosynthetic mechanisms and the origin of such a wide array of terpene genes are still poorly understood.

This thesis has attempted to define unknown and redundant features of terpene biosynthetic genes by assessing the terpene synthase genes from an ancestral angiosperm *Magnolia grandiflora*. Three putative *Magnolia* terpene synthase genes were isolated, their biochemical functions identified, and the genomic organization of these genes characterized. In particular, genes encoding for β -cubebene synthase and α -terpineol synthase were described. Further structure-function relationship studies of these enzymes would be interesting. For instance, the *Magnolia* β -cubebene synthase could be evaluated relative to the cadinene synthase from cotton. Both enzymes utilize *cis*, *trans* configuration of farnesyl cation, yet the structural features of terpene synthases responsible for the *trans* to *cis* conversion have not been identified. Because β -cubebene and cadinene synthases catalyze the biosynthesis of distinctly different reaction products, one could compare the enzymes for structural features in common, features that would potentially be involved with the isomerization reaction. Furthermore, β -cubebene synthase has relatively strict substrate specificity for FPP, while α -terpineol synthase appears able to utilize GPP as well as FPP as substrates. This observation suggests that there is some special feature of the *Magnolia* α -terpineol synthase that allows it to accommodate substrates of differing sizes, structural features which might be identifiable by comparisons of monoterpene synthases having strict substrate specificity to the *Magnolia* α -terpineol synthase.

A more provocative notion is that the Magnolia α -terpineol synthase arose from a sesquiterpene synthase gene during molecular evolution by the re-direction of an ancient sesquiterpene synthase enzyme from the cytoplasm to the chloroplast and mitochondria. The α -terpineol synthase characterized here thus might represent an enzyme in evolutionary transition from its original sesquiterpene characteristics to its monoterpene characteristics. Current molecular evolution theory suggests that diversification of gene function arose by the duplication of genes which allows for one of the duplicated genes to evolve new functional attributes. If this were the case for the α -terpineol synthase described here, then I would predict the existence of yet another α -terpineol synthase like gene, but one that encodes for a cytoplasmic targeted enzyme and one that would have the ability to utilize both FPP and GPP as substrate.

The three different genomic structures, intron-exon organizations, for the Magnolia *TPS* genes described here are not wholly consistent with the current model for the molecular evolution of the TPS gene families. Two of the gene structures were not easily categorized into current model proposed by Trapp and Croteau (2001). All characterized angiosperm *TPS* genes; including all the *TPS* genes found in the Arabidopsis genome, show a conserved intron-exon organization of seven exons interrupted by six introns, with the relative insertion site of introns within the genes very well conserved. One of the Magnolia TPS genes described here fit into this model of organization, but one contained a single intron and the other 4 introns. According to the model of Trapp and Croteau (2001), these TPS genes with fewer introns might have arose for the loss of introns over evolutionary time. An alternative explanation for these variants TPS genes is that these genes with fewer introns represent progenitor genes, while the genes with six introns arose from the progenitor genes by intron insertion. Discern which of these possibilities is correct will not be easy, but future studies of other basal angiosperm and gymnosperm TPS genes could shed light onto how unique or general the gene structure found in the Magnolia TPS is.

Recent success in engineering terpene biosynthesis into microbial and plant host systems have drawn attention to the biosynthetic modifications of the terpene cyclic scaffolds or templates. Terpenes usually obtain biological activities after modifications such as hydroxylations, acetylation, lactonizations, and epoxidations. For example,

costunolide is a complex sesquiterpene found in lettuce harboring several hydroxyl substituents and a lactone ring, and is a potential anti-cancer compound. However, the biosynthetic pathway for costunolide has not been established. Using an available lettuce EST database, this thesis attempted to isolate putative P450s responsible for the hydroxylation of germacrene A, which is likely the first committed intermediate in the costunolide biosynthetic pathway. Three full-length P450 cDNAs were hence co-expressed in yeast along with a germacrene A synthase to provide the appropriate substrate. One of the P450 cDNAs, Ls7108, when co-expressed with germacrene A synthase in yeast, produced two novel, *de novo* synthesized sesquiterpene products. The major product was 12-acetoxy-germacrene A, which was confirmed by NMR spectrometry. This major product, however, was not detected *in vitro* enzyme assay, where only the minor reaction product was detected instead. These results together suggest that Ls7108 encodes for a cytochrome P450 hydroxylase that converts germacrene A into germacra-1(10),4,11(13)-trien-12-ol, then an endogenous enzyme(s) in yeast, an acyltransferase for example, catalyzes the acetylation at the C12 alcohol function. These results provided the tools to screen for other P450s that might be involved in catalyzing various intermediates in the costunolide biosynthesis, P450s that might exhibit substrate specificity for the mono-hydroxylated form of germacrene A, for example. Such experiments would entail the co-expression of germacrene A synthase, Ls7108 and the other lettuce P450s with hopes of identifying the enzymes responsible for converting the mono-hydroxylated sesquiterpene to additional hydroxylated forms and formation of the sesquiterpene lactone.

REFERENCES

- Abdin, M.Z., Israr, M., Rehman, R.U., and Jain, S.K.** (2003). Artemisinin, a Novel Antimalarial Drug: Biochemical and Molecular Approaches for Enhanced Production. *Planta Medica*, 289-299.
- Akashi, K., Grandjean, O., and Small, I.** (1998). Potential dual targeting of an Arabidopsis archaeobacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS Letters* **431**, 39-44.
- Allinger, N.L.** (1977). Conformational-Analysis .130. Mm2 - Hydrocarbon Force-Field Utilizing V1 and V2 Torsional Terms. *Journal of the American Chemical Society* **99**, 8127-8134.
- Arimura, G.-i., Huber, D.P.W., and Bohlmann, J.** (2004a). Forest tent caterpillars (*Malacosoma disstria*) induce local and systemic diurnal emissions of terpenoid volatiles in hybrid poplar (*Populus trichocarpa* × *deltoides*): cDNA cloning, functional characterization, and patterns of gene expression of (-)-germacrene D synthase, PtdTPS1. *The Plant Journal* **37**, 603-616.
- Arimura, G.-i., Ozawa, R., Kugimiya, S., Takabayashi, J., and Bohlmann, J.** (2004b). Herbivore-Induced Defense Response in a Model Legume. Two-Spotted Spider Mites Induce Emission of (E)-{beta}-Ocimene and Transcript Accumulation of (E)-{beta}-Ocimene Synthase in *Lotus japonicus*. *Plant Physiology* **135**, 1976-1983.
- Aubourg, S., Lecharny, A., and Bohlmann, J.** (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Molecular Genetics and Genomics* **267**, 730-745.
- Azuma, H., Garcia-Franco, J.G., Rico-Gray, V., and Thien, L.B.** (2001). Molecular phylogeny of the Magnoliaceae: The biogeography of tropical and temperate disjunctions. *American Journal of Botany* **88**, 2275-2285.
- Azuma, H., Thien, L.B., Toyota, M., Asakawa, Y., and Kawano, S.** (1997a). Distribution and differential expression of (E)-4,8-dimethyl-1,3,7-nonatriene in leaf and floral volatiles of *Magnolia* and *Liriodendron* taxa. *Journal of Chemical Ecology* **23**, 2467-2478.
- Azuma, H., Toyota, M., Asakawa, Y., Yamaoka, R., Garcia-Franco, J.G., Dieringer, G., Thien, L.B., and Kawano, S.** (1997b). Chemical Divergence in Floral Scents of *Magnolia* and Allied Genera (Magnoliaceae). *Plant Species Biology* **12**, 69-83.
- Back, K., and Chappell, J.** (1995). Cloning and Bacterial Expression of a Sesquiterpene Cyclase from *Hyoscyamus-Muticus* and Its Molecular Comparison to Related Terpene Cyclases. *Journal of Biological Chemistry* **270**, 7375-7381.
- Back, K., and Chappell, J.** (1996). Identifying functional domains within terpene cyclases using a domain-swapping strategy. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6841-6845.
- Back, K., Nah, J., Lee, S.B., Song, J.H., Shin, D.H., and Kim, H.Y.** (2000). Cloning of a sesquiterpene cyclase and its functional expression by domain swapping strategy. *Molecules and Cells* **10**, 220-225.
- Bak, S., Kahn, R.A., Olsen, C.E., and Halkier, B.A.** (1997). Cloning and expression in *Escherichia coli* of the obtusifolios 14 alpha-demethylase of *Sorghum bicolor* (L) Moench, a cytochrome P450 orthologous to the sterol 14 alpha-demethylases (CYP51) from fungi and mammals. *Plant Journal* **11**, 191-201.

- Beale, M.H., Birkett, M.A., Bruce, T.J.A., Chamberlain, K., Field, L.M., Huttly, A.K., Martin, J.L., Parker, R., Phillips, A.L., Pickett, J.A., Prosser, I.M., Shewry, P.R., Smart, L.E., Wadhams, L.J., Woodcock, C.M., and Zhang, Y.** (2006). Aphid alarm pheromone produced by transgenic plants affects aphid and parasitoid behavior. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10509-10513.
- Bennett, M., Mansfield, J., Lewis, M., and Beale, M.** (2002a). Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). *Phytochemistry* **60**, 255-261.
- Bennett, M.H., Mansfield, J.W., Lewis, M.J., and Beale, M.H.** (2002b). Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). *Phytochemistry* **60**, 255-261.
- Berry, C.** (2003). The molecular clock. *Qjm-an International Journal of Medicine* **96**, 545-546.
- Bertea, C.M., Voster, A., Verstappen, F.W.A., Maffei, M., Beekwilder, J., and Bouwmeester, H.J.** (2006). Isoprenoid biosynthesis in *Artemisia annua*: Cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. *Archives of Biochemistry and Biophysics* **448**, 3-12.
- Betts, M.J., Guigo, R., Agarwal, P., and Russell, R.B.** (2001). Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? *Embo Journal* **20**, 5354-5360.
- Binns, A.N.** (1994). Cytokinin Accumulation and Action: Biochemical, Genetic, and Molecular Approaches. *Annual Review of Plant Physiology and Plant Molecular Biology* **45**, 173-196.
- Bischoff, T.A., Kelley, C.J., Karchesy, Y., Laurantos, M., Nguyen-Dinh, P., and Arefi, A.G.** (2004). Antimalarial activity of Lactucin and Lactucopicrin: sesquiterpene lactones isolated from *Cichorium intybus* L. *Journal of Ethnopharmacology* **95**, 455-457.
- Blake, C.C.F.** (1978). Do Genes-in-Pieces Imply Proteins-in-Pieces. *Nature* **273**, 267-267.
- Bohlmann, J., Meyer-Gauen, G., and Croteau, R.** (1998a). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4126-4133.
- Bohlmann, J., Crock, J., Jetter, R., and Croteau, R.** (1998b). Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-alpha-bisabolene synthase from grand fir (*Abies grandis*). *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6756-6761.
- Bouwmeester, H.J., Kodde, J., Verstappen, F.W.A., Altug, I.G., de Kraker, J.W., and Wallaart, T.E.** (2002). Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiology* **129**, 134-144.
- Bremer, B., Bremer, K., Chase, M.W., Reveal, J.L., Soltis, D.E., Soltis, P.S., Stevens, P.F., Anderberg, A.A., Fay, M.F., Goldblatt, P., Judd, W.S., Kallersjo, M., Karehed, J., Kron, K.A., Lundberg, J., Nickrent, D.L., Olmstead, R.G., Oxelman, B., Pires, J.C., Rodman, J.E., Rudall, P.J., Savolainen, V., Sytsma, K.J., van der Bank, M., Wurdack, K., Xiang, J.Q.Y., and Zmarzty, S.** (2003).

- An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**, 399-436.
- Brown, J.W.S., and Simpson, C.G.** (1998). Splice site selection in plant pre-mRNA splicing. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**, 77-95.
- Cane, D.E.** (1990). Enzymatic Formation of Sesquiterpenes. *Chemical Reviews* **90**, 1089-1103.
- Cane, D.E.** (1999). *Comprehensive natural products chemistry*, Vol 2, Isoprenoids including carotenoids and steroids. (Amsterdam: Pergamon/Elsevier).
- Carroll, K.K., Guthrie, N., and Ravi, K.** (1992). Dolichol - Function, Metabolism, and Accumulation in Human Tissues. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **70**, 382-384.
- Castaneda-Acosta, J., Fischer, N.H., and Vargas, D.** (1993). Biomimetic Transformations of Parthenolide. *Journal of Natural Products* **56**, 90-98.
- Castanedaacosta, J., Cain, A.W., Fischer, N.H., and Knopf, F.C.** (1995). Extraction of Bioactive Sesquiterpene Lactones from *Magnolia-Grandiflora* Using Supercritical Carbon-Dioxide and near-Critical Propane. *Journal of Agricultural and Food Chemistry* **43**, 63-68.
- Cavalier-smith, T.** (1991). Intron Phylogeny - a New Hypothesis. *Trends in Genetics* **7**, 145-148.
- Chabregas, S.M., Luche, D.D., Farias, L.P., Ribeiro, A.F., van Sluys, M.A., Menck, C.F.M., and Silva, M.C.** (2001). Dual targeting properties of the N-terminal signal sequence of *Arabidopsis thaliana* THI1 protein to mitochondria and chloroplasts. *Plant Molecular Biology* **46**, 639-650.
- Chang, M.C.Y., Eachus, R.A., Trieu, W., Ro, D.K., and Keasling, J.D.** (2007). Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nature Chemical Biology* **3**, 274-277.
- Chappell, J.** (1995). Biochemistry and Molecular-Biology of the Isoprenoid Biosynthetic-Pathway in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **46**, 521-547.
- Chappell, J., Wolf, F., Proulx, J., Cuellar, R., and Saunders, C.** (1995). Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme-a Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants. *Plant Physiology* **109**, 1337-1343.
- Charles, D.J., Simon, J.E., Wood, K.V., and Heinstein, P.** (1990). Germplasm Variation in Artemisinin Content of *Artemisia-Annua* Using an Alternative Method of Artemisinin Analysis from Crude Plant-Extracts. *Journal of Natural Products* **53**, 157-160.
- Chen, X.Y., Chen, Y., Heinstein, P., and Davisson, V.J.** (1995). Cloning, expression, and characterization of (+)-delta-cadinene synthase: A catalyst for cotton phytoalexin biosynthesis. *Archives of Biochemistry and Biophysics* **324**, 255-266.
- Choi, D., Ward, B.L., and Bostock, R.M.** (1992). Differential Induction and Suppression of Potato 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Genes in Response to *Phytophthora infestans* and to Its Elicitor Arachidonic Acid. *Plant Cell* **4**, 1333-1344.

- Choi, S., Im, E., Kang, H., Lee, J., Kwak, H., Bae, Y., Park, H., and Kim, N.** (2005). Inhibitory effects of costunolide on the telomerase activity in human breast carcinoma cells *Cancer Letters* **227**, 153-162.
- Christianson, D.W.** (2006). Structural biology and chemistry of the terpenoid cyclases. *Chemical reviews* **106**, 3412-3442.
- Clouse, S.** (2000). Plant development: A role for sterols in embryogenesis. *Current Biology* **10**, R601-R604.
- Clouse, S.D.** (1998). BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**, 427-451.
- Colby, S.M., Crock, J., Dowdle-Rizzo, B., Lemaux, P.G., and Croteau, R.** (1998). Germacrene C synthase from *Lycopersicon esculentum* cv. VFNT Cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 2216-2221.
- Crane, P.R., Friis, E.M., and Pedersen, K.R.** (1995). The origin and early diversification of angiosperms. *Nature* **374**, 27-33.
- Cronquist, A.** (1988). The evolution and classification of flowering plants. (New York: The New York Botanical Garden).
- Croteau, R.** (1987). Biosynthesis and Catabolism of Monoterpenoids. *Chemical reviews* **87**, 929-954.
- Croteau, R., Alonso, W.R., Koepp, A.E., and Johnson, M.A.** (1994). Biosynthesis of Monoterpenes - Partial-Purification, Characterization, and Mechanism of Action of 1,8-Cineole Synthase. *Archives of Biochemistry and Biophysics* **309**, 184-192.
- Cseke, L., Dudareva, N., and Pichersky, E.** (1998). Structure and evolution of linalool synthase. *Molecular Biology and Evolution* **15**, 1491-1498.
- Cunillera, N., Boronat, A., and Ferrer, A.** (1997). The *Arabidopsis thaliana* FPS1 gene generates a novel mRNA that encodes a mitochondrial farnesyl-diphosphate synthase isoform. *Journal of Biological Chemistry* **272**, 15381-15388.
- Cunningham, F.X., and Gantt, E.** (1998). Genes and enzymes of carotenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**, 557-583.
- Cunningham, S.A.** (1991). Experimental-Evidence for Pollination of Banksia Spp by Nonflying Mammals. *Oecologia* **87**, 86-90.
- Davis, E.M., and Croteau, R.** (2000). Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids* **209**, 53-95.
- de Kraker, J.-W., Franssen, M.C.R., de Groot, A., Konig, W.A., and Bouwmeester, H.J.** (1998). (+)-Germacrene A biosynthesis . The committed step in the biosynthesis of bitter sesquiterpene lactones in chicory. *Plant Physiology* **117**, 1381-1392.
- de Kraker, J.-W., Franssen, M.C.R., Dalm, M.C.F., de Groot, A., and Bouwmeester, H.J.** (2001). Biosynthesis of Germacrene A Carboxylic Acid in Chicory Roots. Demonstration of a Cytochrome P450 (+)-Germacrene A Hydroxylase and NADP+-Dependent Sesquiterpenoid Dehydrogenase(s) Involved in Sesquiterpene Lactone Biosynthesis. *Plant Physiology* **125**, 1930-1940.

- de Kraker, J., Schurink, M., Franssen, M., König, W., de Groot, A., and Bouwmeester, H.** (2003). Hydroxylation of sesquiterpenes by enzymes from chicory (*Cichorium intybus* L.) roots. *Tetrahedron* **59**, 409-418.
- de Kraker, J.W., Franssen, M.C.R., Joerink, M., de Groot, A., and Bouwmeester, H.J.** (2002). Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory. *Plant Physiology* **129**, 257-268.
- de Souza, S.J., Long, M.Y., and Gilbert, W.** (1996). Introns and gene evolution. *Genes to Cells* **1**, 493-505.
- DeJong, J.M., Liu, Y.L., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D., and Croteau, R.B.** (2006). Genetic engineering of Taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* **93**, 212-224.
- Devon, T.K., and Scott, A.I.** (1972). *Handbook of naturally occurring compounds*. (New York: Academic Press.).
- Dieringer, G., Cabrera, L., Lara, M., Loya, L., and Reyes-Castillo, P.** (1999). Beetle Pollination and Floral Thermogenicity in *Magnolia tamaulipana* (Magnoliaceae). *International Journal of Plant Sciences* **160**, 64-71.
- Dinkins, R.D., Conn, H.M., Dirk, L.M.A., Williams, M.A., and Houtz, R.L.** (2003). The *Arabidopsis thaliana* peptide deformylase 1 protein is localized to both mitochondria and chloroplasts. *Plant Science* **165**, 751-758.
- Disch, A., Hemmerlin, A., Bach, T.J., and Rohmer, M.** (1998). Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. *Biochemical Journal* **331**, 615-621.
- Dudareva, N., Cseke, L., Blanc, V.M., and Pichersky, E.** (1996). Evolution of floral scent in *Clarkia*: Novel patterns of S-linalool synthase gene expression in the *C. breweri* flower. *Plant Cell* **8**, 1137-1148.
- Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Faldt, J., Miller, B., and Bohlmann, J.** (2003). (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: Function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* **15**, 1227-1241.
- Elferaly, F.S., and Li, W.S.** (1978). Isolation and Characterization of Magnolol, Honokiol and Honokiol Monomethyl Ether from *Magnolia grandiflora* L Seeds. *Lloydia-the Journal of Natural Products* **41**, 653-653.
- Elferaly, F.S., and Chan, Y.M.** (1978). Isolation and Characterization of Sesquiterpene Lactones Costunolide, Parthenolide, Costunolide Diepoxide, Santamarine, and Reynosin from *Magnolia grandiflora*. *Journal of Pharmaceutical Sciences* **67**, 347-350.
- Elferaly, F.S., Chan, Y.M., and Benigni, D.A.** (1979a). Magnolialide - novel eudesmanolide from the root bark of *Magnolia grandiflora*. *Phytochemistry* **18**, 881-882.
- Elferaly, F.S., Chan, Y.M., Capiton, G.A., Doskotch, R.W., and Fairchild, E.H.** (1979b). Isolation and Characterization of Peroxycostunolide (Verlоторin) and Peroxyparthenolide from *Magnolia grandiflora* - C-13 Nuclear Magnetic-Resonance Spectroscopy of Costunolide and Related Compounds. *Journal of Organic Chemistry* **44**, 3952-3955.

- Emanuelsson, O., Nielsen, H., and Von Heijne, G.** (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**, 978-984.
- Facchini, P.J., and Chappell, J.** (1992). Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 11088-11092.
- Faraldos, J.A., Wu, S., Chappell, J., and Coates, R.M.** (2007). Conformational analysis of (+)-germacrene A by variable-temperature NMR and NOE spectroscopy. *Tetrahedron* **63**, 7733-7742.
- Feltenstein, M.W., Schuhly, W., Warnick, J.E., Fischer, N.H., and Sufka, K.J.** (2004). Anti-inflammatory and anti-hyperalgesic effects of sesquiterpene lactones from *Magnolia* and Bear's foot. *Pharmacology Biochemistry and Behavior* **79**, 299-302.
- Frugoli, J.A., McPeck, M.A., Thomas, T.L., and McClung, C.R.** (1998). Intron loss and gain during evolution of the catalase gene family in Angiosperms. *Genetics* **149**, 355-365.
- Gambliel, H., and Croteau, R.** (1984). Pinene Cyclase-I and Cyclase-II - 2 Enzymes from Sage (*Salvia-Officinalis*) Which Catalyze Stereospecific Cyclizations of Geranyl Pyrophosphate to Monoterpene Olefins of Opposite Configuration. *Journal of Biological Chemistry* **259**, 740-748.
- Gilbert, W.** (1978). Why Genes in Pieces. *Nature* **271**, 501-501.
- Giuliano, G., Bartley, G.E., and Scolnik, P.A.** (1993). Regulation of Carotenoid Biosynthesis during Tomato Development. *Plant Cell* **5**, 379-387.
- Goldstein, J.L., and Brown, M.S.** (1990). Regulation of the Mevalonate Pathway. *Nature* **343**, 425-430.
- Golenberg, E.M., Giannasi, D.E., Clegg, M.T., Smiley, C.J., Durbin, M., Henderson, D., and Zurawski, G.** (1990). Chloroplast DNA-sequence from a Miocene *Magnolia* species. *Nature* **344**, 656-658.
- Greenhagen, B.T., O'Maille, P.E., Noel, J.P., and Chappell, J.** (2006). Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9826-9831.
- Halim, A.F., Mansour, E.S., Badria, F.A., Ziesche, J., and Bohlmann, F.** (1984). A Guaianolide from *Magnolia-Grandiflora*. *Phytochemistry* **23**, 914-915.
- Harborne, J., and Tomas-Barberan, F.** (1991). *Ecological chemistry and biochemistry of plant terpenoids.* (Clarendon Press Oxford).
- Harrewijn, P., van Oosten, A.M., and Piron, P.G.M.** (2001). *Natural terpenoids as messengers : a multidisciplinary study of their production, biological functions, and practical applications* (Boston: Kluwer Academic Publishers).
- Hedden, P., and Kamiya, Y.** (1997). Gibberellin biosynthesis: Enzymes, genes and their regulation. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**, 431-460.
- Herout, V., and Sorm, F.** (1959). Isolation and Structure of Costunolide from *Artemisia-Balchanorum*. *Chemistry & Industry*, 1067-1068.
- Hoffmann-Benning, S., and Kende, H.** (1992). On the Role of Abscissic Acid and Gibberellin in the Regulation of Growth in Rice. *Plant Physiology* **99**, 1156-1161.

- Iijima, Y., Davidovich-Rikanati, R., Fridman, E., Gang, D.R., Bar, E., Lewinsohn, E., and Pichersky, E.** (2004). The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. *Plant Physiology* **136**, 3724-3736.
- Jones, S.B., and Luchsinger, A.E.** (1979). *Plant systematics*. (New York: McGraw-Hill).
- Kappers, I., Aharoni, A., van Herpen, T., Luckerhoff, L., Dicke, M., and Bouwmeester, H.** (2005). Genetic Engineering of Terpenoid Metabolism Attracts Bodyguards to Arabidopsis. *Science* **309**, 2070-2072.
- Karniely, S., and Pines, O.** (2005). Single translation-dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Reports* **6**, 420-425.
- Karp, F., Mihaliak, C.A., Harris, J.L., and Croteau, R.** (1990). Monoterpene Biosynthesis - Specificity of the Hydroxylations of (-)-Limonene by Enzyme Preparations from Peppermint (*Mentha-Piperita*), Spearmint (*Mentha-Spicata*), and Perilla (*Perilla-Frutescens*) Leaves. *Archives of Biochemistry and Biophysics* **276**, 219-226.
- Katoh, S., Hyatt, D., and Croteau, R.** (2004). Altering product outcome in *Abies grandis* (-)-limonene synthase and (-)-limonene/(-)-alpha-pinene synthase by domain swapping and directed mutagenesis. *Archives of Biochemistry and Biophysics* **425**, 65-76.
- Kim, S., Kang, S., Kim, H., and Kim, T.** (2002). Potentiation of 1,25-dihydroxyvitamin D3-induced differentiation of human promyelocytic leukemia cells into monocytes by costunolide, a germacranolide sesquiterpene lactone *Biochemical Pharmacology* **64**, 1233-1242.
- Klayman, D.L.** (1985). Qinghaosu (Artemisinin) - an Antimalarial Drug from China. *Science* **228**, 1049-1055.
- Koo, T.H., Lee, J.H., Park, Y.J., Hong, Y.S., Kim, H.S., Kim, K.W., and Lee, J.J.** (2001). A sesquiterpene lactone, costunolide, from *Magnolia grandiflora* inhibits NF-kappa B by targeting I kappa B phosphorylation. *Planta Medica* **67**, 103-107.
- Kumar, S., Gupta, S.K., Singh, P., Bajpai, P., Gupta, M.M., Singh, D., Gupta, A.K., Ram, G., Shasany, A.K., and Sharma, S.** (2004). High yields of artemisinin by multi-harvest of *Artemisia annua* crops. *Industrial Crops and Products* **19**, 77-90.
- Ladd, T.** (1980). Japanese beetle: enhancement of lures by eugenol and caproic acid. *Journal of Economic Entomology* **73**, 718-720.
- Lafever, R.E., and Croteau, R.** (1993). Hydride Shifts in the Biosynthesis of the P-Menthane Monoterpenes Alpha-Terpinene, Gamma-Terpinene, and Beta-Phellandrene. *Archives of Biochemistry and Biophysics* **301**, 361-366.
- Lambers, D.H.R., and Schepers, A.** (1978). Effect of Trans-Beta-Farnesene, Used as a Repellant against Landing Aphid *Alatae* in Seed Potato Growing. *Potato Research* **21**, 23-26.
- Lapkin, A.A., Plucinski, P.K., and Cutler, M.** (2006). Comparative assessment of technologies for extraction of artemisinin. *Journal of Natural Products* **69**, 1653-1664.
- Leclercq, E.** (1984). Determination of Lactucin in Roots of Chicory (*Cichorium-Intybus* L) by High-Performance Liquid-Chromatography. *Journal of Chromatography* **283**, 441-444.

- Lewin, B.** (2000). Genes VII. (Oxford, New York: Oxford University Press).
- Li, J., Nagpal, P., Vitart, V., McMorris, T.C., and Chory, J.** (1996). A role for brassinosteroids in light-dependent development of Arabidopsis. *Science* **272**, 398-401.
- Li, Y., Su, Z., and Chen, F.** (2002). Rapid Extraction of Genomic DNA From Leaves and Bracts of Dove Tree (*Davidia involucrata*). *Plant Molecular Biology Reporter* **20**, 185a-185e.
- Lichtenthaler, H.K.** (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **50**, 47-65.
- Liu, C.Z., Zhao, Y., and Wang, Y.C.** (2006). Artemisinin: current state and perspectives for biotechnological production of an antimalarial drug. *Applied Microbiology and Biotechnology* **72**, 11-20.
- Long, M.Y., Rosenberg, C., and Gilbert, W.** (1995). Intron phase correlations and the evolution of the intron exon structure of genes. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 12495-12499.
- Loughrin, J.H., Potter, D.A., Hamilton-Kemp, T.R., and Byers, M.E.** (1996). Volatile compounds from crabapple (*Malus* spp.) cultivars differing in susceptibility to the Japanese beetle (*Popillia japonica* Newman). *Journal of Chemical Ecology* **22**, 1295-1305.
- Lucker, J., Bowen, P., and Bohlmann, J.** (2004). *Vitis vinifera* terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. *Phytochemistry* **65**, 2649-2659.
- Lucker, J., El Tamer, M.K., Schwab, W., Verstappen, F.W.A., van der Plas, L.H.W., Bouwmeester, H.J., and Verhoeven, H.A.** (2002). Monoterpene biosynthesis in lemon (*Citrus limon*) - cDNA isolation and functional analysis of four monoterpene synthases. *European Journal of Biochemistry* **269**, 3160-3171.
- Lupien, S., Karp, F., Wildung, M., and Croteau, R.** (1999). Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. *Archives of Biochemistry and Biophysics* **368**, 181-192.
- Martin, D., Piulachs, M.D., Cunillera, N., Ferrer, A., and Belles, X.** (2007). Mitochondrial targeting of farnesyl diphosphate synthase is a widespread phenomenon in eukaryotes. *Biochimica et Biophysica Acta-Molecular Cell Research* **1773**, 419-426.
- Martin, D.M., and Bohlmann, J.** (2004). Identification of *Vitis vinifera* (-)-alpha-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry* **65**, 1223-1229.
- Martin, D.M., Faldt, J., and Bohlmann, J.** (2004). Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. *Plant Physiology* **135**, 1908-1927.

- Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D.** (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature biotechnology* **21**, 796-802.
- Mau, C.J.D., and West, C.A.** (1994). Cloning of Casbene Synthase cDNA - Evidence for Conserved Structural Features among Terpenoid Cyclases in Plants. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8497-8501.
- McCourt, P., Lumba, S., Tsuchiya, Y., and Gazzarrini, S.** (2005). Crosstalk and abscisic acid: the roles of terpenoid hormones in coordinating development. *Physiologia Plantarum* **123**, 147-152.
- Mcgarvey, D.J., and Croteau, R.** (1995). Terpenoid Metabolism. *Plant Cell* **7**, 1015-1026.
- Mellado, V., Soto, M.A.C., and Lozoya, X.** (1980). Pharmacological Screening of the Aqueous Extracts of *Magnolia-Grandiflora* L. *Archivos De Investigacion Medica* **11**, 335-346.
- Menand, B., Marechal-Drouard, L., Sakamoto, W., Dietrich, A., and Wintz, H.** (1998). A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11014-11019.
- Monde, K., Takasugi, M., Katsui, N., and Masamune, T.** (1990). Studies on Stress Metabolites .11. Total Synthesis of Lettucenin-a, a Guaianolide Phytoalexin from *Lactuca-Sativa* Var *Capitata*. *Chemistry Letters*, 1283-1286.
- Morant, M., Bak, S., Moller, B.L., and Werck-Reichhart, D.** (2003). Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current Opinion in Biotechnology* **14**, 151-162.
- Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W.B., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., Page, C.C., and Dutton, P.L.** (2002). P450BM3: the very model of a modern flavocytochrome. *Trends in Biochemical Sciences* **27**, 250-257.
- Nakai, K., and Horton, P.** (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends in Biochemical Sciences* **24**, 34-35.
- Naora, H., and Deacon, N.J.** (1982). Relationship between the Total Size of Exons and Introns in Protein-Coding Genes of Higher Eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**, 6196-6200.
- Olson, K.C., Tibbitts, T.W., and Struckme.Be.** (1969). Leaf Histogenesis in *Lactuca Sativa* with Emphasis Upon Laticifer Ontogeny. *American Journal of Botany* **56**, 1212-1216.
- Omura, T., and Sato, R.** (1964). Carbon Monoxide-Binding Pigment of Liver Microsomes .2. Solubilization Purification + Properties. *Journal of Biological Chemistry* **239**, 2379-2385.
- Page, R.D.M.** (1996). TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-358.

- Palmer, J.D., and Logsdon, J.M.** (1991). The recent origins of introns. *Current Opinion in Genetics and Development* **1**, 470-477.
- Park, S.Y., Shimizu, H., Adachi, S., Nakagawa, A., Tanaka, I., Nakahara, K., Shoun, H., Obayashi, E., Nakamura, H., Iizuka, T., and Shiro, Y.** (1997). Crystal structure of nitric oxide reductase from denitrifying fungus *Fusarium oxysporum*. *Nature Structural Biology* **4**, 827-832.
- Pechurskaya, T.A., Harnastai, I.N., Grabovec, I.P., Gilep, A.A., and Usanov, S.A.** (2007). Adrenodoxin supports reactions catalyzed by microsomal steroidogenic cytochrome P450s. *Biochemical and Biophysical Research Communications* **353**, 598-604.
- Peigler, R.** (1988). A review of pollination of *Magnolia* by beetles, with a collecting survey made in the Carolinas. *Magnolia (Journal of the Magnolia Society)* **45**, 1-8.
- Phillips, M.A., Wildung, M.R., Williams, D.C., Hyatt, D.C., and Croteau, R.** (2003). cDNA isolation, functional expression, and characterization of (+)-alpha-pinene synthase and (-)-alpha-pinene synthase from loblolly pine (*Pinus taeda*): Stereocontrol in pinene biosynthesis. *Archives of Biochemistry and Biophysics* **411**, 267-276.
- Picaud, S., Olsson, M.E., Brodelius, M., and Brodelius, P.E.** (2006). Cloning, expression, purification and characterization of recombinant (+)-germacrene D synthase from *Zingiber officinale*. *Archives of Biochemistry and Biophysics* **452**, 17-28.
- Pichersky, E., Raguso, R.A., Lewinsohn, E., and Croteau, R.** (1994). Floral Scent Production in *Clarkia* (Onagraceae) .1. Localization and Developmental Modulation of Monoterpene Emission and Linalool Synthase Activity. *Plant Physiology* **106**, 1533-1540.
- Podust, L.M., Poulos, T.L., and Waterman, M.R.** (2001). Crystal structure of cytochrome P450 14 alpha-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3068-3073.
- Postek, M.T., and Tucker, S.C.** (1983). Ontogeny and Ultrastructure of Secretory Oil Cells in *Magnolia-Grandiflora* L. *Botanical Gazette* **144**, 501-512.
- Prosser, I., Altug, I.G., Phillips, A.L., Konig, W.A., Bouwmeester, H.J., and Beale, M.H.** (2004). Enantiospecific (+)- and (-)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif. *Archives of Biochemistry and Biophysics* **432**, 136-144.
- Prosser, I., Phillips, A.L., Gittings, S., Lewis, M.J., Hooper, A.M., Pickett, J.A., and Beale, M.H.** (2002). (+)-(10R)-germacrene A synthase from goldenrod, *Solidago canadensis*; cDNA isolation, bacterial expression and functional analysis. *Phytochemistry* **60**, 691-702.
- Pyrek, J.S.** (1977). Terpenes of Compositae Plants. 5. Sesquiterpene lactones of *Lactuca serriola* L. The structure of 8-deoxylactucin and the site of esterification of lactupicrin. *Roczniki Chemii* **51**, 2165-2169.
- Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R.** (1992). Characterization and Mechanism of (4s)-Limonene Synthase, a Monoterpene Cyclase from the

- Glandular Trichomes of Peppermint (*Mentha X Piperita*). *Archives of Biochemistry and Biophysics* **296**, 49-57.
- Ralston, L., Kwon, S.T., Schoenbeck, M., Ralston, J., Schenk, D.J., Coates, R.M., and Chappell, J.** (2001). Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from tobacco (*Nicotiana tabacum*). *Archives of Biochemistry and Biophysics* **393**, 222-235.
- Rao, A.S., Kelkar, G.R., and Bhattacharyya, S.C.** (1958). Costunolide, a New Sesquiterpene Lactone from Costus Root Oil. *Chemistry & Industry*, 1359-1360.
- Richards, A.J.** (1986). *Plant breeding systems*. (London ; Boston: Allen & Unwin).
- Ridley, R.G.** (1999). Perspectives: Biomedicine - Planting the seeds of new antimalarial drugs. *Science* **285**, 1502-1503.
- Rising, K.A., Starks, C.M., Noel, J.P., and Chappell, J.** (2000). Demonstration of germacrene A as an intermediate in 5-epi-aristolochene synthase catalysis. *Journal of the American Chemical Society* **122**, 1861-1866.
- Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R., and Keasling, J.D.** (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943.
- Rohmer, M.** (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* **16**, 565-574.
- Roise, D., Theiler, F., Horvath, S.J., Tomich, J.M., Richards, J.H., Allison, D.S., and Schatz, G.** (1988). Amphiphilicity is essential for mitochondrial presequence function. *EMBO Journal* **7**, 649-653.
- Rost, B., Yachdav, G., and Liu, J.F.** (2004). The PredictProtein server. *Nucleic Acids Research* **32**, W321-W326.
- Runyon, J.B., Mescher, M.C., and De Moraes, C.M.** (2006). Volatile chemical cues guide host location and host selection by parasitic plants. *Science* **313**, 1964-1967.
- Ruzicka, L.** (1953). The Isoprene Rule and the Biogenesis of Terpenic Compounds. *Experientia* **9**, 357-367.
- Schnee, C., Kollner, T.G., Held, M., Turlings, T.C.J., Gershenzon, J., and Degenhardt, J.** (2006). The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *PNAS* **103**, 1129-1134.
- Schuler, M.A.** (1996). Plant cytochrome P450 monooxygenases. *Critical Reviews in Plant Sciences* **15**, 235-284.
- Schuler, M.A., and Werck-Reichhart, D.** (2003). Functional genomics of P450s. *Annual Review of Plant Biology* **54**, 629-667.
- Sessa, R.A., Bennett, M.H., Lewis, M.J., Mansfield, J.W., and Beale, M.H.** (2000). Metabolite profiling of sesquiterpene lactones from *Lactuca* species - Major latex components are novel oxalate and sulfate conjugates of lactucin and its derivatives. *Journal of Biological Chemistry* **275**, 26877-26884.
- Singh, G., Gavrieli, J., Oakey, J.S., and Curtis, W.R.** (1998). Interaction of methyl jasmonate, wounding and fungal elicitation during sesquiterpene induction in *Hyoscyamus muticus* in root cultures. *Plant Cell Reports* **17**, 391-395.

- Small, I., Peeters, N., Legeai, F., and Lurin, C.** (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581-1590.
- Song, Q., Gomez-Barrios, M.L., Fronczek, F.R., Vargas, D., Thien, L.B., and Fischer, N.H.** (1998). Sesquiterpenes from southern *Magnolia virginiana*. *Phytochemistry* **47**, 221-226.
- Song, W.C., and Brash, A.R.** (1991). Purification of an Allene Oxide Synthase and Identification of the Enzyme as a Cytochrome-P-450. *Science* **253**, 781-784.
- Sparrow, J., and Hanly, G.** (2002). *Subtropical Plants: A Practical Gardening Guide* (Timber Press).
- Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R.** (1998). Sesquiterpene synthases from grand fir (*Abies grandis*) - Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *Journal of Biological Chemistry* **273**, 2078-2089.
- Sun, G., Ji, Q., Dilcher, D.L., Zheng, S., Nixon, K.C., and Wang, X.** (2002). Archaeofractaceae, a new basal angiosperm family. *Science* **296**, 899-904.
- Suzuki, Y., Hibino, T., Kawazu, T., Wada, T., Kihara, T., and Koyama, H.** (2003). Extraction of total RNA from leaves of eucalyptus and other woody and herbaceous plants using sodium isoascorbate. *Biotechniques* **34**, 988-993.
- Takahashi, S., Zhao, Y.X., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M., and Chappell, J.** (2005). Kinetic and molecular analysis of 5-epiaristolochene 1,3-dihydroxylase, a cytochrome P450 enzyme catalyzing successive hydroxylations of sesquiterpenes. *Journal of Biological Chemistry* **280**, 3686-3696.
- Takahashi, S., Yeo, Y.-S., Zhao, Y., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M., and Chappell, J.** (2007a). Functional Characterization of Premnaspirodiene Oxygenase, a Cytochrome P450 Catalyzing Regio- and Stereo-specific Hydroxylations of Diverse Sesquiterpene Substrates. *Journal of Biological Chemistry* **282**, 31744-31754.
- Takahashi, S., Yeo, Y., Greenhagen, B.T., McMullin, T., Song, L., Maurina-Brunker, J., Rosson, R., Noel, J.P., and Chappell, J.** (2007b). Metabolic engineering of sesquiterpene metabolism in yeast. *Biotechnology and Bioengineering* **97**, 170-181.
- Takasugi, M., Okinaka, S., Katsui, N., Masamune, T., Shirata, A., and Ohuchi, M.** (1985). Isolation and Structure of Lettucenin-a, a Novel Guaianolide Phytoalexin from *Lactuca-Sativa* Var *Capitata* (Compositae). *Journal of the Chemical Society-Chemical Communications*, 621-622.
- Takeda, K.** (1974). Stereospecific cope rearrangement of the germacrene-type sesquiterpenes. *Tetrahedron* **30**, 1525-1534.
- Taniguchi, M., Kataoka, T., Suzuki, H., Uramoto, M., Ando, M., Arao, K., Magae, J., Nishimura, T., Otake, N., and Nagai, K.** (1995). Costunolide and dehydrocostus lactone as inhibitors of killing function of cytotoxic T lymphocytes. *Bioscience, Biotechnology, and Biochemistry* **59**, 2064-2067.
- Taylor, M., Lamb, D.C., Cannell, R., Dawson, M., and Kelly, S.L.** (1999). Cytochrome P450105D1 (CYP105D1) from *Streptomyces griseus*: Heterologous

- expression, activity, and activation effects of multiple xenobiotics. *Biochemical and Biophysical Research Communications* **263**, 838-842.
- Teoh, K.H., Polichuk, D.R., Reed, D.W., Nowak, G., and Covello, P.S.** (2006). *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *Febs Letters* **580**, 1411-1416.
- Tomita, M., and Nakano, T.** (1952). Studies on the Alkaloids of Magnoliaceous Plants .7. Syntheses of Salicifoline Chloride and Analogous Compounds. *Yakugaku Zasshi-Journal of the Pharmaceutical Society of Japan* **72**, 1256-1260.
- Trapp, S.C., and Croteau, R.B.** (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* **158**, 811-832.
- Trapp, S.C., Hohn, T.M., McCormick, S., and Jarvis, B.B.** (1998). Characterization of the gene cluster for biosynthesis of macrocyclic trichothecenes in *Myrothecium roridum*. *Molecular and General Genetics* **257**, 421-432.
- Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., and Pompon, D.** (1997). Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related *Arabidopsis thaliana* NADPH-Cytochrome P450 Reductases with P450 CYP73A5. *Journal of Biological Chemistry* **272**, 19176-19186.
- van der Hoeven, R.S., Monforte, A.J., Breeden, D., Tanksley, S.D., and Steffens, J.C.** (2000). Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. *Plant Cell* **12**, 2283-2294.
- Wallach, O.** (1914). Terpene und Camphor: Zusammenfassung eigener Untersuchungen auf dem Gebiet der alicyclischen Kohlenstoffverbindungen. (Leipzig: Veit).
- Wenzel, S.C., and Muller, R.** (2005). Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Current opinion in biotechnology* **16**, 594-606.
- Westerkamp, C.** (1990). Bird-Flowers - Hovering Versus Perching Exploitation. *Botanica Acta* **103**, 366-371.
- Wiedhopf, R.M., Young, M., Bianchi, E., and Cole, J.R.** (1973). Tumor Inhibitory Agent from *Magnolia-Grandiflora* (Magnoliaceae) .1. Parthenolide. *Journal of Pharmaceutical Sciences* **62**, 345-345.
- Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R.** (1998). Truncation of limonene synthase preprotein provides a fully active 'Pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry* **37**, 12213-12220.
- Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E.** (2000a). Microsomal cytochrome P450C5: comparison to microbial P450s and unique features. *Journal of Inorganic Biochemistry* **81**, 183-190.
- Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., and McRee, D.E.** (2000b). Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Molecular Cell* **5**, 121-131.
- Williamson, B.** (1977). DNA Insertions and Gene Structure. *Nature* **270**, 295-297.
- Wu, S.Q., Schalk, M., Clark, A., Miles, R.B., Coates, R., and Chappell, J.** (2006). Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nature Biotechnology* **24**, 1441-1447.

- Yang, M.H., Blunden, G., Patel, A.V., Oneill, M.J., and Lewis, J.A.** (1994). Coumarins and Sesquiterpene Lactones from *Magnolia-Grandiflora* Leaves. *Planta Medica* **60**, 390-390.
- Yoon, S.H., Lee, Y.M., Kim, J.E., Lee, S.H., Lee, J.H., Kim, J.Y., Jung, K.H., Shin, Y.C., Keasling, J.D., and Kim, S.W.** (2006). Enhanced lycopene production in *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. *Biotechnol Bioeng* **94**, 1025-1032.
- Yoshikuni, Y., Ferrin, T.E., and Keasling, J.D.** (2006). Designed divergent evolution of enzyme function. *Nature* **440**, 1078-1082.
- Zdero, C., Bohlmann, F., Solomon, J., and Dominguez, X.A.** (1988). Further isocedrene derivatives and other constituents from *Perezia* species. *Phytochemistry* **27**, 849-853.
- Zook, M.N., Chappell, J., and Kuc, J.A.** (1992). Characterization of elicitor-induction of sesquiterpene cyclase activity in potato tuber tissue. *Phytochemistry* **31**, 3441-3445.

VITA

Sungbeom Lee was born on February 9th, 1972 in Gwangju, Republic of Korea. He graduated from Soong-il High School in Gwangju, in 1990. In 1991, he attended Chonnam National University in Gwangju majoring in Agronomy. He met his obligatory military service in the Republic of Korea Air Force from November, 1993 to April, 1996. After receiving a Bachelor of Science degree in Agronomy in February, 1998, Mr. Lee continued his studies in the same institute and received a Master of Science in Agronomy in February, 2000. He was employed in the Department of Plant Genetic Engineering for two years as an intern researcher, supported by the Korea Science and Engineering Foundation. In July of 2002, he began his doctoral studies in the Plant Physiology program under the supervision of Dr. Joe Chappell at the University of Kentucky.