INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI
800-521-0600
RICE UNIVERSITY

Cloning of Terpene Synthases from Plants

by

Ling Hua

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Seiichi Matsuda, Assistant Professor
Chemistry, Biochemistry and Cell Biology

Ronald J. Parry, Professor
Chemistry

Bonnie Bartel, Assistant Professor
Biochemistry and Cell Biology

Houston, Texas
February, 2000
ABSTRACT

Cloning of Terpene Synthases from Plants

By

Ling Hua

As part of a program to clone Artemisia annua sesquiterpene biosynthetic genes, a cDNA encoding epi-cedrol synthase was isolated from an A. annua leaf cDNA library by a homology-based strategy. The cDNA was functionally expressed in Escherichia coli and the encoded protein shown to produce (−) epicedrol and cedrol in a 96:4 ratio. The structure of epicedrol was determined by GC-MS and NMR and cedrol was identified by GC-MS and GC co-elution with the authentic sample. Neither cedrol nor epicedrol was detected in an extract of the plant from which the cDNA library was made. In addition, a sesquiterpene synthase like gene was cloned from A. thaliana ecotype Columbia by RT-PCR and the cDNA was expressed in E. coli.

The second part of this thesis, describes the study of triterpene biosynthesis. A yeast strain with squalene synthase (erg9) and lanosterol synthase (erg7) double mutation was made by homologous recombination. This strain has been used by other group members to express A. thaliana cycloartenol synthase mutants and S. cerevisiae lanosterol synthase mutants. More accurate enzyme product ratios were obtained for the mutants expressed in this erg9 and erg7 double mutant yeast strain compared to the erg7 single mutant SMY8.
In order to study the mechanism of oxidosqualene cyclases, *A. thaliana* was chosen as the model plant for cloning oxidosqualene cyclases. A triterpene synthase was isolated from the *A. thaliana* ecotype Landsberg young seedling cDNA library and the cloned gene had the second exon missing compared to cycloartenol synthase and lupeol synthase cloned from same organism. The cloned gene was expressed in yeast and no oxidosqualene cyclase activity was detected. An oxidosqualene cyclase like open reading frame ORF3, which is adjacent to lupeol synthase on the genomic DNA was cloned from *A. thaliana* ecotype Landsberg cDNA library. An open reading frame ORF1 found in the *A. thaliana* sequencing database, that has 71.4% identity to lupeol synthase was cloned from *A. thaliana* ecotype Columbia total RNA by RT-PCR. The cDNA of these two ORFs were expressed in yeast and no oxidosqualene cyclase activity was detected.
ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Seiichi P. T. Matsuda, his special interest in molecular biology made it possible for me to switch my career from an organic chemist to a molecular biologist. He not only teaches me science in both chemistry and molecular biology but also teaches me interpersonal skill as a young scientist. He found every possible way to help me improving my English, during my four years graduate study in his group.

I am also deeply indebted to Professor Bonnie Bartel, who helped me to set up the concept of molecular biology and taught me the basic biotechniques a good molecular biologist should have. I also want to thank my other committee member Professor Ronald Parry, for spending his precious time to read and improve my third year proposal and my thesis.

I would like to thank Dr. Jennifer Herrera, who helped me in constructing the cDNA library, and who is always there whenever I need help. I enjoyed very much working with her in the past four years. When she decided to graduate a few months ahead of me, I felt truly miserable because it is hard for me to imagine work in the lab without her. I would also like to thank other members in Dr. Matsuda’s lab, you guys are great team players and I enjoyed working with you. I would like to thank Rosie Tellez, for useful discussions and for supplying the A. thaliana total RNA.

Finally, I would like to thank my husband Dr. Dunming Zhu for his unconditional love. He is always there to comfort me whenever I feel frustrated or depressed. He is an amazing partner, an incredible friend and a great mentor. Without his sacrifice, it would be impossible for me to get this degree. I would like to thank my son James. He is so
understanding and is so sweet, and he helps me to improve my pronunciation while I spend time with him. He knows how to take care of himself because his mom is always busy. I would like to thank my parents and my grandparents for their love, support and encouragement. Thank God for giving me such a wonderful family and providing everything I need.
To my family
TABLE OF CONTENTS

ABSTRACT ii
ACKNOWLEDGEMENTS iv
LIST OF FIGURES xi
ABBREVIATIONS xii

PART I: Cloning of Sesquiterpene Synthases from Plants

Chapter 1. Background and Introduction 1
  1.1.1. Terpenoids 1
  1.1.2. Terpenoid biosynthesis 2
  1.1.3. Methods for studying terpene biosynthesis 6
  1.1.4. Artemisinin is an antimalarial drug 7
  1.1.5. The action of artemisinin 8
  1.1.6. Artemisinin biosynthesis 8
  1.1.7. Total synthesis of artemisinin 11
  1.1.8. Semisynthesis of artemisinin from artemisinic acid 12
  1.1.9. Plant breeding and cell culture of A. annua 13
  1.1.10. Genetic engineering the A. annua 14
  1.1.11. Homology-base approaches for cloning sesquiterpene synthase 14
  1.1.12. Putative sesquiterpene synthase from A. thaliana 15

Chapter 2. Experimental Procedures 17
  1.2.1. Artemisia annua mRNA isolation 17
  1.2.2. A. annua cDNA library construction 18
  1.2.3. Degenerate PCR amplification of the A. annua cDNA library 22
  1.2.4. Colony hybridization of epicedrol synthase cDNAs 23
  1.2.5. Cloning the full length epicedrol synthase cDNA 24
  1.2.6. Making epicedrol synthase expression constructs 25
  1.2.7. Epicedrol synthase protein analysis 26
1.2.8. Initial epicedrol synthase assay 28
1.2.9. Optimization of epicedrol synthase assay conditions 29
1.2.10. Large scale preparation of epicedrol synthase products 31
1.2.11. Structure identification of epicedrol synthase products 32
1.2.12. Analyzing an A. annua extract by GC-MS for epicedrol 33
1.2.13. Cloning a putative sesquiterpene synthase (LH6.6) from A. thaliana 33
1.2.14. Expression of LH6 in E. coli 35

Chapter 3. Results and Discussion 38
1.3.1. Cloning epicedrol synthase from A. annua 38
1.3.2. Epicedrol synthase expression and optimizing the in vitro enzymatic reaction 42
1.3.3. Epicedrol synthase product characterization 43
1.3.4. Sequence analysis of epicedrol synthase 43
1.3.5. Epicedrol synthase reaction mechanism 45
1.3.6. Relevance of epicedrol synthase to artemisinin biosynthesis 46
1.3.7. Cloning a sesquiterpene synthase-like gene from A. thaliana 48
1.3.8. Sequence analysis of LH6 51
1.3.9. Expression of LH6 in bacteria 54
1.3.10. Future work 55

Chapter 4. Conclusions 57

PART II Cloning of Triterpene synthases from Arabidopsis thaliana

Chapter 1. Background and Introduction 59
2.1.1. General introduction of triterpenes 59
2.1.2. Triterpene biosynthesis 60
2.1.3. Enzyme mechanism of oxidosqualene cyclase 63
2.1.4. Methods for studying triterpenes 66
2.1.5. Triterpenes from A. thaliana 67
2.1.6. Yeast as a model system 68
2.1.7. Engineering yeast strains 69

Chapter 2. Experimental Procedures 71
2.2.1. Yeast squalene synthase knockout 71
2.2.2. Yeast ERG9 and ERG7 double deletion 73
2.2.3. Recovery of LEU2 marker from LHY2 75
2.2.4. Construction of yeast expression vector pRS426GAL 76
2.2.5. Construction of pRS416GPD/Not I and pRS426GPD/Not I 76
2.2.6. Cloning LH1 from an A. thaliana cDNA library 77
2.2.7. LH1 sequencing subclones 78
2.2.8. Searching for a full-length LH1 cDNA 79
2.2.9. Expression of LH1 in yeast 80
2.2.10. Cloning ORF3 (LH2) from an A. thaliana cDNA library 81
2.2.11. LH2 sequencing subclones 83
2.2.12. Expression of LH2.14 in yeast 84
2.2.13. Cloning ORF1 (LH5) using RT-PCR 86
2.2.14. Expression of LH5.2 in yeast 87

Chapter 3. Results and Discussion 90
2.3.1. Construction of the squalene synthase deletion plasmid 90
2.3.2. Yeast homologous recombination 92
2.3.3. Yeast erg9 and erg7 double mutant 92
2.3.4. Making LHY2 derivatives 93
2.3.5. Application of yeast double knockout strains 95
2.3.6. Cloning LH1, a putative triterpene cyclase from A. thaliana 96
2.3.7. Expression of cloned LH1 in yeast 99
2.3.8. Searching for the missing exon 99
2.3.9. Cloning of ORF3 from A. thaliana cDNA library 103
2.3.10. Expression of ORF3 in yeast 105
2.3.11. Cloning of ORF1 by RT-PCR 106
Materials and Methods

1. Instruments

2. Materials

2.1. Chemicals and reagents

2.2. Enzymes and vectors

2.3. Substrates
   A. Synthesis of farnesyl pyrophosphate
   B. Synthesis of oxidosqualene

2.4. Medium
   A. Bacterial media
   B. Yeast media

3. Methods

3.1. E. coli transformation

3.2 Yeast transformation

3.3. Plasmid isolation

3.4. DNA analysis

3.5. Subcloning

3.6. Dephosphorylation of linearized DNA

3.7. PCR programs

3.8. Adding A-tails blunt-ended PCR product

3.9. Making pBluescript KS(+) T-vector

3.10. Labeling probe for hybridization

3.11. Colony hybridization

3.12. Site-directed mutagenesis

3.13. Protein analysis on SDS-PAGE gel

3.14. Oligonucleotides

3.15. DNA sequencing

3.16. Small scale assay for oxidosqualene cyclase activity
3.17. TMS-ether derivatize triterpene alcohol

References

Appendix

1. List of Plasmids
2. List of Strains
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Biosynthesis of IPP through mevalonic acid</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Biosynthesis of IPP through a mevalonate-independent pathway</td>
<td>4</td>
</tr>
<tr>
<td>3.</td>
<td>Terpene biosynthesis in plants</td>
<td>6</td>
</tr>
<tr>
<td>4.</td>
<td>Structures of antimalarial drugs</td>
<td>8</td>
</tr>
<tr>
<td>5.</td>
<td>Proposed artemisinin biosynthetic pathway</td>
<td>10</td>
</tr>
<tr>
<td>6.</td>
<td>Total synthesis of artemisinin</td>
<td>12</td>
</tr>
<tr>
<td>7.</td>
<td>Semisynthesis of artemisinin from artemisinic acid</td>
<td>13</td>
</tr>
<tr>
<td>8.</td>
<td>Alignment of sesquiterpene synthases</td>
<td>39</td>
</tr>
<tr>
<td>9.</td>
<td>DNA and protein sequences of epi-cedrol synthase</td>
<td>41</td>
</tr>
<tr>
<td>10.</td>
<td>Proposed mechanism of epi-cedrol synthase</td>
<td>46</td>
</tr>
<tr>
<td>11.</td>
<td>DNA and protein sequences of LH6.6</td>
<td>50</td>
</tr>
<tr>
<td>12.</td>
<td>Sequence alignment of monoterpene and sesquiterpene synthases</td>
<td>53</td>
</tr>
<tr>
<td>13.</td>
<td>Triterpene and sterol biosynthesis</td>
<td>62</td>
</tr>
<tr>
<td>14.</td>
<td>Mechanism of oxidosqualene cyclases</td>
<td>65</td>
</tr>
<tr>
<td>15.</td>
<td>Construction of plasmid LH7.3</td>
<td>91</td>
</tr>
<tr>
<td>16.</td>
<td>Sterol biosynthesis in LHY2</td>
<td>90</td>
</tr>
<tr>
<td>17.</td>
<td>Sequence alignment of two oxidosqualene cyclases from <em>A. thaliana</em></td>
<td>95</td>
</tr>
<tr>
<td>18.</td>
<td>cDNA and protein sequences of LH1.2</td>
<td>101</td>
</tr>
<tr>
<td>19.</td>
<td>Genomic DNA and amino acid sequences of LH1.18</td>
<td>102</td>
</tr>
<tr>
<td>20.</td>
<td>cDNA and protein sequences of LH2.14</td>
<td>104</td>
</tr>
<tr>
<td>21.</td>
<td>cDNA and protein sequences of ORF1</td>
<td>108</td>
</tr>
</tbody>
</table>
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amu</td>
<td>atom mass unit</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine-5’-triphosphate</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CIP</td>
<td>calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine-5’-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine-5’-triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine-5’-triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine-5’-triphosphate</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DMAPP</td>
<td>dimethylallyl pyrophosphate</td>
</tr>
<tr>
<td>DMF</td>
<td>$N, N$-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>mixture of dATP, dCTP, dGTP and dTTP</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>FOA</td>
<td>5-fluoro-orotic acid</td>
</tr>
<tr>
<td>FPP</td>
<td>farnesyl pyrophosphate</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione-S-transferase</td>
</tr>
<tr>
<td>HEPES</td>
<td>$N$-2-hydroxyethylpiperazine-$N’$-2-ethanesulfonic acid</td>
</tr>
</tbody>
</table>
his  histidine
IPP  isopentenyl pyrophosphate
IPTG  isopropyl-1-thio-β-D-galactoside
kbp  kilobase pair
leu  leucine
β-ME  2-mercaptoethanol
MOPS  morpholinopropanesulfonic acid
MS  mass spectroscopy
NMR  nuclear magnetic resonance
PCR  polymerase chain reaction
PEG  polyethylene glycol
SDS  sodium dodecyl sulfate
TCA  trichloroacetic acid
TEMED  N,N,N′,N′-tetramethylethylenediamine
THF  tetrahydrofuran
TLC  thin layer chromatography
TMS  trimethylsilane
Tris  tris(hydroxymethyl)aminomethane
ura  uracil
UV  ultraviolet spectroscopy
X-gal  3-bromo-4-chloro-3-indolyl-β-D-galactoside

Media:

LB: Luria broth, 1% tryptone, 0.5% yeast extract and 0.5% NaCl
YPD: 1% yeast extract, 2% peptone and 2% glucose

SCDHET-Ura: 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.4% amino acid mix lacking ura3, 2% glucose and supplied with 20 μg/mL of ergosterol, 13 μg/mL of heme and 5 μg/mL of Tween 80.

SCDHET-His: 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.4% amino acid mix lacking histidine, 2% glucose and supplied with 20 μg/mL of ergosterol, 13 μg/mL of heme and 5 μg/mL of Tween 80.

SCDHET-Leu: 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.4% amino acid mix lacking leucine, 2% glucose and supplied with 20 μg/mL of ergosterol, 13 μg/mL of heme and 5 μg/mL of Tween 80.

SCGHET-Ura: 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.4% amino acid mix lacking ura3, 2% galactose and supplied with 20 μg/mL of ergosterol, 13 μg/mL of heme and 5 μg/mL of Tween 80.

SCGHET-Leu: 0.34% yeast nitrogen base, 1% ammonium sulfate, 0.4% amino acid mix lacking leucine, 2% galactose and supplied with 20 μg/mL of ergosterol, 13 μg/mL of heme and 5 μg/mL of Tween 80.
Part I

Cloning of Sesquiterpene Synthases from Plants

Chapter 1: Background and Introduction

1.1.1. Terpenoids

Terpenoids constitute the largest family of natural products: over 15,000 known compounds of this class have been described. Terpenoid structures are diverse and range from relatively simple linear hydrocarbon chains to some of the most complex ring structures known. The greatest structural and functional diversity of terpenoids is in the plant kingdom, which can be divided into primary metabolites and secondary metabolites. Terpenoids that are primary metabolites include sterols, carotenoids, hormones, and polyrenol substituents of dolichols, quinones and proteins. These compounds are essential for membrane integrity, photoprotection, orchestration of developmental program and anchoring of essential biochemical function to specific membrane systems, respectively. Terpenoids that are secondary metabolites include monoterpenes, sesquiterpenes, diterpenes, sesterterpenes, triterpenes, and tetraterpenes. Monoterpenes, sesquiterpenes and diterpenes are not essential for viability, but mediate interactions between plants and their environment. Some sesquiterpenes and diterpenes serve as toxic defense compounds to ward off herbivores or fungal attacks, while
some monoterpenes serve to attract pollinators to floral tissues.\textsuperscript{8,9} Numerous plant terpenoids have significant roles in human health and commerce. For example, many monoterpenes and sesquiterpenes of essential oils are important flavoring and fragrance agents;\textsuperscript{10} and some sesquiterpenes and diterpenes have pharmacological significance including the anticancer drug taxol\textsuperscript{11} and the antimalarial drug artemisinin.\textsuperscript{12}

1.1.2. Terpenoid biosynthesis

The isoprenoid biosynthetic pathway is sometimes referred to as the mevalonate pathway.\textsuperscript{5} This pathway has been well characterized. The initial reaction of the mevalonate pathway involves the sequential condensation of three acetyl-CoA units to generated (3S)-3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). First, two acetyl-CoA moieties are fused by a Claisen condensation and the third acetyl-CoA reacts with acetoacetyl-CoA through an aldol-type condensation. HMG-CoA is converted to mevalonate by HMG-CoA reductase. The mevalonate is converted to the C\textsubscript{5} intermediate isopentenyl diphosphate (IPP) by a three-step biosynthetic sequence. First, mevalonate is pyrophosphorylated by two enzymes, mevalonate kinase and phosphomevalonate kinase. Then, the resulting mevalonate-5-pyrophosphate undergoes a decarboxylative elimination catalyzed by pyrophosphomevalonate decarboxylase to generate IPP. Figure 1 shows the biosynthesis of IPP through mevalonic acid. IPP serves as the basic building block in terpenoid construction.
Figure 1. Biosynthesis of IPP through mevalonic acid

Until recently, it was generally assumed that all isoprenoids were synthesized from acetyl-CoA via the classical mevalonate pathway to the central precursor IPP. However, in 1993, Rohmer and coworkers demonstrated a mevalonate-independent pathway for isoprenoids that was first found in bacteria and green alga\textsuperscript{13,14} and then was found in higher plants.\textsuperscript{15} This mevalonate-independent pathway\textsuperscript{15} also known as the deoxyxylulose-5-phosphate pathway. It starts with pyruvate and glyceraldehyde 3-phosphate. In higher plants, the gene involved in the initial step of the pathway has been cloned.\textsuperscript{15} It encodes a transketolase that catalyzes the condensation of pyruvate with D-glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylulose 5-phosphate (DXP). The
second enzyme of the pathway is a reductase, which rearranges and reduces DXP to give 2-C-methylerythritol 4-phosphate (MEP). The gene has been cloned from peppermint. The enzyme that catalyzes the ATP-dependent phosphorylation of isopentenyl monophosphate (IP) to IPP\textsuperscript{17} has also been cloned from peppermint. The isopentenyl monophosphate kinase is proposed to be the enzyme that catalyzes the last step of the mevalonate-independent pathway. Other enzymes on the pathway are still unidentified. Figure 2 shows the proposed mevalonate-independent pathway.

![Chemical structures](image)

Figure 2. Biosynthesis of IPP through a mevalonate-independent pathway (modified from Ref. 17)

Terpenoid precursors are synthesized by condensing IPP with its isomer DMAPP\textsuperscript{18}. IPP is first isomerized to dimethylallyl pyrophosphate (DMAPP) by IPP isomerase. Condensation of DMAPP with one unit of IPP in a head-to-tail fashion
generates the allylic C_{10} compound geranylgeranyl pyrophosphate (GPP), addition of a second IPP generates the C_{15} allylic pyrophosphate, farnesyl pyrophosphate (FPP), and a third IPP couples with FPP to generate the C_{20} precursor geranylgeranyl pyrophosphate (GGPP). These elongation reactions are catalyzed by prenyltransferases, enzymes that mediate the condensation of IPP with an allylic pyrophosphate forming a new allylic pyrophosphate containing five carbon atoms more than the original. Two molecules of FPP dimerize in a head-to-head fashion to form squalene (C_{30}), the triterpene precursor.

The enzymes that cyclize GPP, FPP, and GGPP are referred as monoterpene, sesquiterpene, and diterpene cyclases. In each category of the cyclases, the same substrate is cyclized to dramatically different enzyme products. Monoterpene synthases cyclize GPP to the various skeletal types of monoterpenes, such as limonene and pinene. Sesquiterpene synthases cyclize FPP to sesquiterpenes with even greater structural diversity than monoterpenes; over 7000 sesquiterpenes have been isolated from plants.\(^1\) Aristolochene and trichodiene are two examples of sesquiterpenes. Diterpene synthases cyclize GGPP to diterpenes, like casbane and abietadiene. In higher plants, enzymes that use the mevalonate pathway are localized to the cytosolic compartment, and produce the precursors of triterpenes (sterols) and certain sesquiterpenes;\(^4\) enzymes on the deoxyxylulose 5-phosphate pathway are localized in the plastids and operate to supply IPP for the synthesis of monoterpenes, diterpenes,\(^{19,20}\) tetraterpenes (carotenoids), and the prenyl side chains of chlorophyll and plastoquinone.\(^21\) These two pathways cooperate between cytosol and plastid to biosynthesize some terpenoids in a variety of plants.\(^{22,23}\) Figure 3 shows the organization of the two pathways for terpene biosynthesis in plants.
Figure 3. Terpene biosynthesis in plants (modified from Ref. 23)

1.1.3. Methods for studying terpene biosynthesis

Investigation of natural product biosynthesis starts with establishing the actual sequence of metabolic transformation. The study of terpene biosynthesis parallels that of other natural products, in that such information was once obtained by administering radiolabeled precursors to intact plant tissues, isolating the products, measuring their specific activities and determining the site of labeling by chemical degradation. Later, cell free plant extracts were used instead of intact tissues. However, the extremely low
level of terpene biosynthetic enzymes hinders using a cell free extract to do such study. Over the last few years, the methodology of biosynthetic research has been dramatically altered by the explosive advances in molecular biology. The application of recombinant DNA techniques has now made it possible to readily isolate and manipulate genes encoding terpene biosynthetic enzymes. The isolated gene can be expressed in prokaryotic and eukaryotic cells to generate large quantities of protein for detailed mechanistic and catalytic studies. In this thesis, recombinant DNA techniques are applied to study terpene biosynthesis in plants.

1.1.4. Artemisinin is an antimalarial drug

Malaria is the major health problem in many areas of the world, especially in tropical Africa. The World Health Organization reports that some 300 million people are believed to be infected with malaria and 85% of cases are caused by Plasmodium falciparum. The control of malaria is becoming more difficult due to the increased resistance of Plasmodium strains to commonly used drugs such as quinine and its derivatives. Artemisia annua is a fragrant annual herb widely distributed in Asia, Europe and North America. The use of this plant (Qinhaosu) in Chinese traditional medicine was recorded before 168 BC and the Chinese naturalist Li Shi-Zhen described the use of A. annua for treatment of fever and malaria in his 1596 book “Compendium of Materia Medica”. Extensive chemical analyses of the essential oil of A. annua have demonstrated the presence of artesiminin as the active principle. Artemisinin and its derivatives artemesunate and artemether have been approved as some of the most potent
chemotherapeutic agents against malarial resistant strain *Plasmodium*. Figure 4 shows the structures of these three antimalarial drug.

![Chemical structures](image)

Artemisinin  
Artemether  
Artesunate

Figure 4. Structures of antimalarial drugs

1.1.5. The action of artemisinin

Artemisinin is a sesquiterpene lactone with an endoperoxide bridge, which reacts with heme and iron to release free radicals. Artemisinin is a better antimalarial drug than others against parasites because of its 1, 2, 4-trioxane nature. Instead of reacting with oxygen and producing oxygen-containing free radicals, artemisinin itself becomes a free radical which effectively inhibits parasites. Artemisinin and its derivatives are toxic to malaria parasites at nanomolar concentrations, whereas micromolar concentration are required for toxicity to mammalian cells.

1.1.6. Artemisinin biosynthesis

Artemisinin is an endoperoxide sesquiterpene lactone derived from farnesyl diphosphate. Whereas the biosynthetic route from farnesyl diphosphate to artemisinin is
uncharacterized, the sesquiterpene artemisinic acid has been shown to be biosynthesized by *A. annua*\(^{31}\) and converted to artemisinin.\(^{32}\) Arteannuin B is an intermediate in the bioconversion of artemisinic acid to artemisinin.\(^{33}\) From the artemisinic acid structure, we predict that the precursor sesquiterpene hydrocarbon is 4,11 amorphadiene. The 4,11 amorphadiene has been isolated from a natural source\(^{34}\) but not from *A. annua*. The proposed artemisinin biosynthetic pathway is shown in Figure 5. Farnesyl pyrophosphate is cyclized to 4,11 amorphadiene, which undergoes oxidation to yield artemisinic acid. Artemisinic acid is then converted to arteannuin B, which is further biotransformed to the antimalarial principle artemisinin.
farnesyl pyrophosphate  

4,11-amorphadiene  

artemisinic acid

arteannuin B  

artemisinin

Figure 5. Proposed artemisinin biosynthetic pathway (modified from Ref. 32 and 33)

Currently, artemisinin is purified from plant extracts. However, a full ton of dried A. annua leaves is needed to produce about 6 kg of artemisinin. The yield of artemisinin from the dry herb varies between 0.01% - 0.5%, depending on the origin of the herb and the harvest time. The leaves before budding possess the highest artemisinin content. Since the artemisinin yield from the dry plant is very low, at present, artemisinin is available in large enough amounts for extensive clinical trials and treatment, only in China where this plant grows abundantly,
1.1.7. Total synthesis of artemisinin

Because of the low yield of artemisinin from natural sources, organic chemists have pursued total syntheses of artemisinin. The complicated structure, especially the peroxide ring, provides a challenge for the organic chemist to develop a total synthetic route to artemisinin. Several total syntheses have been achieved. The first total synthesis was reported in 1983 by Schmid and Hofheinz.\textsuperscript{38} It was a 13-step total synthesis. It started with (−)-isopulegol and the overall yield was 2.1%. The most efficient synthesis was accomplished by Avery \textit{et al.},\textsuperscript{39} starting from \((R)-(+)\)-pulegone and the overall yield was 3.6%. Figure 6 shows the total synthesis schemes. Each procedure for the total synthesis of artemisinin requires a final photo oxidative step. Commercial production of artemisinin by total synthesis is unlikely because the low overall yield and the high cost of \textit{de novo} synthesis.
13 steps, 2.1% overall yield

(-)-Isopulegol

10 steps, 3.6% overall yield

$R$-(+)-Pulegone

Figure 6. Total synthesis of artemisinin (modified from Ref. 38 and 39)

1.1.8. Semisynthesis of artemisinin from artemisinic acid

Semisynthesis of artemisinin from its native precursors is an alternative. Although the complete biosynthetic pathway for artemisinin has not been established, artemisinic acid is a biosynthetic precursor of artemisinin that can be biotransformed into artemisinin in vivo and in vitro. Artemisinic acid is a relatively abundant sesquiterpene in \textit{A. annua}, which occurs at a 8- to 10-fold higher concentration than artemisinin.\textsuperscript{40} Artemisinic acid was converted to artemisinin in three steps as shown in Figure 7, and the
overall yield from artemisinic acid was 30%.\textsuperscript{41} The low yield of artemisinic acid in the plant precludes using this route to make sufficient artemisinin for treatment. However, it would be a good alternative if metabolically engineered routes to artemisinic acid were available.

![Chemical reaction diagram]

Figure 7. Semisynthesis of artemisinin from artemisinic acid (modified from Ref. 40)

1.1.9. Plant breeding and cell culture of \textit{A. annua}

Because of the high cost of the chemical synthesis and the low yield of isolation from plant, plant cell and tissue cultures had been considered as a potential alternative to produce artemisinin. Disappointingly, the cell culture method does not appear to offer a viable production alternative,\textsuperscript{42} since the callus cells do not accumulated any artemisinin and the artemisinin leaches into the medium, where the amount obtained was low (ca. 8 \(\mu\)g/mL).\textsuperscript{36} Research has been shown that the hairy root cultures of \textit{A. annua} gives 550 mg/L of artemisinin.\textsuperscript{43} Attempts also have been made for selection and breeding of high-artemisinin-yielding strains of \textit{A. annua}. The best effort was a clone from China
that has the artemisinin concentration up to 1.1% of the dried leaves by weight. However, this number is still too low for commercial exploitation.

1.1.10. Genetic engineering of *A. annua*

In order to develop a genetically modified strain of *A. annua* with high artemisinin production, studying artemisinin biosynthesis and cloning the genes of the artemisinin biosynthetic pathway is necessary. When sesquiterpene overproduction is induced in this plant, squalene synthase activity is suppressed. Overexpressing the sesquiterpene synthase should increase production of sesquiterpene hydrocarbon precursor of artemisinin. This should increase artemisinin biosynthesis and decrease squalene biosynthesis. The other study showed that adding an artemisinin precursor increased artemisinin yields in tissue culture, confirming that overexpression of upstream genes in the artemisinin biosynthetic pathway should increase the artemisinin yield in the plant. If we can clone and express the genes in the artemisinin biosynthetic pathway, and transform the genes into the plant, we should be able to make a transgenic *A. annua* with increased artemisinin yields.

1.1.11. Homology-based approaches for cloning sesquiterpene synthase

More than a dozen sesquiterpene synthases are expressed in *A. annua*; over 300 sesquiterpenes belonging to different structural classes have been isolated from this plant. The synthases involved in the biosynthesis of these sesquiterpenes are most likely similar to other terpene synthases, particularly those sesquiterpene synthases from other angiosperms. Several sesquiterpene synthases have been cloned from a variety of
plants, and the primary sequences of these enzymes show a relatively high level of homology. The similarity-based cloning technique provides a method to use consensus sequence elements of the cloned sesquiterpene synthases to design degenerate primers for PCR amplification.

To investigate the biosynthesis of sesquiterpenes and in particular artemisinin in *A. annua*, I have begun to isolate sesquiterpene synthase clones from a cDNA library prepared from *A. annua* leaves before budding. Our eventual goal is to exhaustively characterize terpene cyclases, including the cyclase involved in artemisinin biosynthesis. We will transform the cyclase gene into *A. annua* to make a transgenic plant with a high level artemisinin production. One of the many advantages of this approach over protein purification, is that we can obtain a variety of new cyclase genes for structure-function studies. In the course of this work, I have identified a new sesquiterpene synthase from *A. annua*, which catalyzes the formation of a sesquiterpene alcohol epicedrol. Epicedrol has neither been previously reported as a natural product nor isolated from *A. annua*. This cyclase is also the first sesquiterpene alcohol synthase that has been cloned.

### 1.1.12. Putative sesquiterpene synthase from *A. thaliana*

The other project described in this part is the cloning of a sesquiterpene synthase from *Arabidopsis thaliana*. *A. thaliana* is a small plant in the mustard family, which has been used as a model system for research in plant molecular biology and cell biology. The entire life cycle of *Arabidopsis thaliana* is completed in six weeks, including seed germination, formation of a rosette plant, bolting of the main stem, flowering and maturation of the first seeds. *A. thaliana* has a highly compact genome with only an 120
megabases genome organized into five chromosomes. It has about 20,000 genes.\textsuperscript{53} The entire genome is scheduled to be sequenced by the end of year 2000. Many different ecotypes are available for experimental purposes; the ecotype Columbia was selected for sequencing. The secondary metabolites of \textit{Arabidopsis} have not been well studied except the genes on the gibberellin biosynthetic pathway.\textsuperscript{54} No sesquiterpene synthase has been cloned from this plant. A sesquiterpene synthase like sequence was obtained from the Genbank \textit{Arabidopsis} sequencing database and the cDNA was cloned by RT-PCR. The open reading frame of the gene was expressed in \textit{E. coli} and the encoded enzyme shows weak phosphatase activity to farnesyl pyrophosphate.
Chapter 2: Experimental Procedures

1.2.1. *Artemisia annua* mRNA isolation

All the containers were treated with DEPC water and then either autoclaved or baked at high temperature. DEPC treated water was made by mixing 700 μL of diethyl pyrocarbonate (DEPC) into 700 mL of Milli-Q H₂O in an RNase free bottle. After being vigorously shaken and standing at room temperature overnight, the solution was autoclaved on liquid cycle for 45 min.

Frozen *A. annua* leaves (1.5 g) were ground in prechilled mortars to a fine powder in a pool of liquid nitrogen. Acid guanidinium thiocyanate-phenol-chloroform (15 mL Trizol Reagent) was added to the homogenized sample. The homogenized sample was transferred to a tube and incubated for 5 min at room temperature. After adding 3.0 mL of chloroform, the tube was shaken vigorously and incubated at room temperature for 2 to 3 min. The tube was then centrifuged, and the RNA-containing aqueous phase was transferred to another tube, leaving the protein containing phenol-chloroform phase in the original tube. The RNA was precipitated from the aqueous phase by adding 18 mL of isopropyl alcohol. The precipitate was washed once with 15 mL of 75% DEPC-treated ethanol, and briefly air-dried. The resulting solid was then dissolved in RNase-free water by incubating at 55 to 60 °C for 10 min. The total RNA yield (25%) was quantified by UV absorbance at 280 nm. The mRNA was prepared from total RNA by selectively binding it to an oligo(dT) column (Gibco-BRL) as follows. Total RNA (585 μg) was dissolved in 2.0 mL of TNES buffer (10mM Tris-HCl, pH7.5, 1 mM EDTA, 0.5 M NaCl
and 0.1% SDS) and the resulting solution was heated at 65 °C for 5 min and then chilled on ice. To this solution, oligo(dT) cellulose (1 ml of 50% suspension) was added, and the solution was incubated at 37 °C for 10 min. The mRNA-bound cellulose was separated by filtration via a filter syringe and washed once with 3 ml of TNES buffer and once with 3 ml of DEPC-treated water to remove the unbound RNA. The mRNA was then eluted from cellulose with DEPC-treated water at 65 °C. The yield was 10 µg of mRNA.

1.2.2. A. annua cDNA library construction

The Superscript Plasmid System from Gibco-BRL was used to make the cDNA library. This system features directional construction of the library by introducing asymmetry at the ends of cDNA. This was achieved by a Not I primer-adapter, which was an oligo dT primer for first strand synthesis that also contains restriction sites. The primer-adapter used here had the following sequence 5’-pGACTAGTTCTAGATCGCGAGCGGCGCCGCCC(T)15-3’, which contains restriction sites for Not I, Nru I, Xba I and Spe I. The 8-base NotI recognition sequence next to the 15-bp Ts was extremely rare in most DNAs. First strand synthesis was catalyzed by SuperScript II reverse transcriptase. Second strand strain synthesis was catalyzed by E. coli DNA polymerase I in combination with E. coli RNase H and E. coli DNA ligase. After the second strand cDNA synthesis, Sal I adapters were ligated to both ends of the cDNA. The sequence of Sal I adapter was 5’- TCGACCCACGCAGTCGG-3’ for the upper strand and 3’-GGGTGCAGGCACP-5’ for the lower strand. The adapter contained a Mlu I recognition site and only one of the oligomers in the adapter was phosphorylated, which eliminated self-ligation of the adapters at the Sal I overhangs during ligation to the
cDNA. The cDNAs were then digested with Not I to release the Not I sites at the 3' ends of the cDNAs. Size fractionation of cDNA is a very important step that can eliminate a large molar excess of Sal I adapters and deplete Not I-Sal I fragments released from 3' ends of the cDNAs after the Not I digestion. This also reduced the representation of smaller (<500 bp) inserts in the cDNA library. Vector pSPORT1 cut with Not I and Sal I was ligated with size-fractionated cDNA inserts. E. coli host strain DH10B was transformed with the ligated plasmids by electroporation. The detailed protocols were as follows: Five micrograms of mRNA were dissolved in 5 μL of DEPC water and 2 μL of Not I primer-adapter (0.5 μg/μL) was added. The mixture was heated to 70 °C for 10 min and chilled on ice. The following reagents were added: 4 μL of 5 × first strand buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM MgCl₂), 2.0 μL of 0.1 M DTT, 1.0 μL of 10 mM dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP) and 1.0 μL of [α-32P]dCTP (1 μCi/μL), and incubated at 37 °C for 2 min to equilibrate the temperature. SUPERSCRIPT II reverse transcriptase (5 μL, 200 units/μL) was added to the above mixture and incubated at 37 °C for 1 h. The first strand synthesis reaction was terminated by chilling the tube on ice. A 2.0 μL aliquot was withdrawn and mixed with 43 μL of 20 mM of EDTA (pH 7.5) and 5.0 μL of yeast tRNA (1 μg/μL) to quantitate the first strand cDNA yield. Ten microliter aliquots were spotted onto two glass fiber filters. One filter was air dried and the other one was washed three times with 50 mL of ice cold 10% (w/v) TCA containing 1% (w/v) sodium pyrophosphate, once with 50 mL of 95% ethanol and then air dried. The first strand cDNA yield (14.5%) was obtained by scintillation counting these two filters.
The rest of the first strand cDNA was used as template to synthesize the second strand cDNA. The following reagents were added into first strand cDNA solution sequentially: 93 μL of DEPC treated water, 30 μL of 5 × second strand buffer (100 mM Tris-HCl, pH 6.9, 450 mM KCl, 23 mM MgCl₂, 0.75 mM β-NAD⁺, 50 mM ammonium sulfate), 3.0 μL of 10 mM of dNTP mix, 1.0 μL of *E. coli* DNA ligase (10 units/μL), 4.0 μL of *E. coli* DNA polymerase (10 units/μL) and 1.0 μL of *E. coli* RNase H (2 units/μL). After the mixture had been incubated at 16 °C for 2 h, 2.0 μL of T4 DNA polymerase (5 units/μL) was added and incubation was continued at 16 °C for 2 min. The second strand cDNA synthesis was quenched by adding 10 μL of 0.5 mM EDTA and chilled on ice. The enzymes were denatured with 150 μL of phenol : chloroform:isoamyl alcohol (25 : 24 : 1). The aqueous layer was removed to a fresh tube and the double-stranded cDNA was precipitated with 70 μL of 7.5 M NH₄OAc and 500 μL of 100% ethanol and washed with 500 μL of 70% cold ethanol. The DNA pellet was air-dried and the following reagents were added sequentially: 25 μL of DEPC treated water, 10 μL of 5 × T4 DNA ligase buffer (250 mM Tris-HCl, pH 7.6, 50 mM MgCl₂, 5 mM ATP, 5 mM DTT, 25% (w/v) PEG 8000), 10 μL of *Sal* I adapters and 5 μL of T4 DNA ligase (1 unit/μL). The reaction was incubated at 16 °C for at 16 h. The enzymes were denatured with 50 μL of phenol : chloroform:isoamyl alcohol (25 : 24 : 1). The aqueous layer was removed to a fresh tube and the cDNA was precipitated with 25 μL of 7.5 M NH₄OAc and 150 μL of 100% ethanol and washed with 500 μL of 70% cold ethanol. The resultant cDNA was digested with 4.0 μL of *Not* I (15 units/μL) in 41 μL of DEPC treated water and 5 μL REACT 3 buffer at 37 °C for 2 h. The unreacted *Not* I was extracted away with 50 μL of phenol:chloroform:isoamyl alcohol (25 : 24 : 1). The aqueous layer was removed to a
fresh tube and the cDNA was precipitated with 25 μL of 7.5 M NH₄OAc and 150 μL of 100% ethanol and washed with 500 μL of 70% cold ethanol. The cDNA was dissolved in 100 μL of TEN buffer (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 25 mM NaCl and loaded on a TEN-equilibrated size fractionation column (4 x 800 μL). The cDNA was eluted with TEN and collected 550 μL total in microcentrifuge tube (35 μL each). The cDNA (248 ng) yield was measured by scintillation counting. The cDNA was precipitated with 5 μg of yeast tRNA, 40 μL of 7.5 M NH₄OAc and 250 μL of 100% ethanol. Ten nanograms cDNA aliquots were transferred into a new tube and centrifuged. The cDNA pellet was washed with 500 μL of 75% of ethanol and dissolved in 10 μL of TEN. Each 10 ng of cDNA was ligated with 50 ng of pSPORT1 Not I and Sal I digested vector (50 ng/μL) at room temperature for 3 h in the following reaction, 4 μL of DEPC treated water, 4 μL of 5 x DNA ligase buffer (250 mM Tris-HCl, pH 7.6, 50 mM MgCl₂, 5 mM ATP, 5 mM DTT, 25% (w/v) PEG 8000), and 1 μL of T4 DNA ligase (1 unit/μL). The ligated DNA was precipitated with 5 μg of yeast tRNA, 12.5 μL of 7.5 M NH₄OAc and 70 μL of 100% ethanol. The DNA pellet was washed with 500 μL of 70% cold ethanol and dissolved in 3 μL of DEPC water. Each ligation was electroporated into 25 μL of ElectroMAX DH10B cells (Gibco-BRL), which were plated on LB plus 100 μg/mL ampicillin. Three transformations were done from three ligations. The resultant cDNA library had about 10⁶ independent transformants, and the average size of 24 inserts was about 1.3 kbp. These transformants were pooled in 10 mL LB, brought to 50% glycerol, and stored in aliquots at -80 °C. Plasmid DNA was prepared from a library aliquot following overnight growth in 100 mL LB plus 100 μg/mL ampicillin.
1.2.3. Degenerate PCR amplification of the *A. annua* cDNA library

Conserved segments of *Hyoscyamus muticus* vetispiradiene synthase,\(^50\) *Nicotiana tabacum* epiaristolochene synthase\(^49\) and two *Gossypium arboreum* cadinene synthase genes were identified,\(^51,52\) and 11 forward and 11 reverse degenerate PCR primers were designed based on these homologous sequences. All possible combinations of these primers were tested for their ability to amplify DNA from the *A. annua* cDNA library. Each PCR reaction included 200 ng library DNA, 5.0 µL 10 x PC2 buffer (500 mM Tris-HCl pH 9.1, 160 mM ammonium sulfate, 35 mM MgCl\(_2\)), 4.0 µL dNTPs (each at 2.5 mM), 5.0 µL of each degenerate primer (20 pmol/µL), and H\(_2\)O to 50 µL. The PCR program included a 4 min 95 °C hot start during which 0.5 µL *Taq* polymerase (Fisher, 5.0 U/µL) was added, and 40 cycles with 1 min annealing using a temperature gradient from 68 °C to 48 °C (- 0.5 °C/cycle), 3 min extension at 72 °C, and denaturation for 45 s at 95 °C. The program was terminated with a 5 min extension at 72 °C. Each PCR reaction (5 µL) was analyzed on a 1% agarose gel. Of all these combinations, only one pair of primers gave a clear band at the expected size (129 bp). They were forward primer SQF9: GAYMNGYNGTNGARKGTAYTTYTGG [corresponding to \(DR(A/V)VE(C/G)YFW\)] and reverse primer SQR3: SCRTANGMRTCRWANGTRTCRTC [corresponding to the reverse complement of DDT(F/Y)D(A/S)Y(G/A)]. The PCR fragment was gel purified and subcloned into pBluescript II KS(+) T-vector\(^55\) with T4 DNA ligase. The ligation was transformed into *E. coli* DH5\(\alpha\) chemically competent cells and plated on LB plus 100 µg/mL ampicillin with X-Gal (20 µg/mL) for the blue/white selection. Forty-eight colonies were screened...
by restriction digestion for plasmid with insert. Five that had an insert were sent to
sequence and three of them had the same DNA sequences, which were consistent with a
fragment of a sesquiterpene synthase. One of them was named LH4.0. The LH4.0 insert
was used as a probe to hybridize to the original cDNA library.

1.2.4. Colony hybridization of epcedrol synthase cDNAs

The LH4.0 EcoRI-Hind III fragment was used as template for making
radiolabeled probe. Plasmid LH4.0 (20 µg) was double digested with 5.0 µL of each
restriction enzymes in a 200 µL reaction using EcoRI unique buffer (NEB). The 129 bp
insert was gel-purified and 25 ng of this fragment (13.5 µL) was combined with 1 µg of
specific primer LH4F1: GACAGGGCGGTGGGAAGG in a 1.5 mL screw-cap tube. The
α-32P-dCTP labeled probe was made as described in the Materials and Methods section.
A total of 1.2 × 10⁶ colonies was screened. After two rounds of standard hybridization as
described in Materials and Methods, four positive clones were found according to the X-
ray film. The positive colonies were further mapped with the restriction enzymes Hind
III, Kpn I, and Ssp I. The insert sizes were determined by Not I and Sal I double
digestion. Two colonies with longer inserts (~ 1.7 kbp) and one colony with a shorter
insert (~ 1.2 kbp) were obtained from hybridization. The colony with short insert was
named LH4.1, and the two other colonies with longer inserts were named LH4.2 and
LH4.3. Extensive mapping of these three plasmids indicated that they are different
fragments of the same cDNA. LH4.2 was sequenced from T7, M13 forward primer and
sequencing primers (LH4.2S1: CGTTGCGCTACATACCTT, LH4.2S2: AAAGAGGCTGTCCAAAGC, LH4.2S3: GTTCGGTGATGTCCTTT). The insert
sequence appeared to encode a sesquiterpene synthase with ~200 bp missing from the 5' end.

1.2.5. Cloning the full length epicedrol synthase cDNA

In order to obtain the N-terminus of this gene, two nested primers were designed, LH4.2R1: CGTCCCTCTTTGTCTTTG and LH4.2R2: TGTCGCATGAGTCGAAAC. LH4.2R1 was paired with T7 sequentially to PCR amplify the cDNA library. The PCR program was CAPS (see details in the Materials and Methods) and the PCR conditions were as follows: 50 ng cDNA library, 5.0 μL 10 x PC2 buffer, 4.0 μL dNTPs (each at 2.5 mM), 3.0 μL of T7 (20 pmol/μL), 1.0 μL of LH4.2 R1 (20 pmol/μL), 0.5 μL Taq polymerase (Fisher, 5.0 U/μL) and H2O to 50 μL. Five microliters of the PCR reaction was analyzed on a 2% agarose gel, which showed that a faint band at 480 bp was obtained from this first round of PCR. The faint band was used as template to do a second round of PCR using nested primer LH4.2R2 and T7. A pipette tip was used to poke the agarose gel where the faint band was and the pipette was dipped into water in the PCR tube. The PCR program and the PCR conditions were same as first round amplification except 1.0 μL of LH4.2R2 was used instead of LH4.2R1, and only 1.0 μL of T7 primer was used. After the second round of PCR, a 450 bp fragment was obtained. After gel purification, 25 ng of this DNA fragment was subcloned into 30 ng of pBluescript KS(+) T vector with T4 DNA ligase. Twenty-four colonies were mapped by restriction digestion. One plasmid with a 450 bp insert was named LH4.6 and was sequenced with EXT-7. The sequence overlapped with LH4.2, and a specific primer LH4.6F1: CCGTACTAATTCCGATCT was designed upstream of the presumed initiator
codon of LH4.6 sequence. LH4.6F1 was paired with LH4.2R2 to PCR-amplify the cDNA library using the high-fidelity polymerase *Pfu* (Stratagene). Each PCR reaction included 150 ng cDNA, 5.0 μL 10 × *Pfu* buffer (100 mM KCl, 100 mM (NH₄)₂SO₄, 200 mM Tris-HCl (pH 8.75), 20 mM MgSO₄, 1% Triton X-100, 1mg/mL BSA), 4.0 μL dNTPs (each at 2.5 mM), 1.0 μL of each primer (20 pmol/μL), 1 μL *Pfu* DNA polymerase (2.5 units/μL) and H₂O to 50 μL. The PCR program was CAPS. The PCR fragment was gel purified and A-tailed as described in the Materials and Methods. The A-tailed PCR product was purified by Qiaex suspension according to the manufacturer’s instructions and subcloned into pBluescript KS (+) T-vector. Twenty-four colonies were screened and one colony with the same insert as LH4.6 was sequenced and named LH4.7. The LH4.7 *Nde* I and *Pst* I fragment was subcloned into the longest cDNA LH4.2 digested with same restriction enzymes. The resultant plasmid was named LH4.8, which has a 1.7 kbp insert (not including poly A tail). The *Pst* I and *Not* I fragment of LH4.8 was subcloned into pBluescript KS (+) digested with same restriction enzymes to give LH4.9.

1.2.6. Making epicedrol synthase expression constructs

A *Xho* I site was introduced immediately upstream of the initiator codon by PCR to facilitate expressing this gene as a fusion protein in expression vector pET15b⁵⁶ and pGEX-KTO.1.⁵⁷ The primer ATCTCGAGATGTCTTTATA (*Xho* I site was underlined) was paired with LH4.2R2 (see cloning the full length gene) to PCR amplify the 350 bp 5' end of LH4.9. The PCR conditions and program were same as those used to PCR amplify LH4.7. The PCR product was digested with *Xho* I and *Nde* I, and
subcloned into LH4.9 cut with the same two enzymes to give LH4.11. The insert was excised from LH4.11 with Xho I and Not I and was subcloned into a pET15b (Novagen) derivative in which a Not I site had been inserted in the polylinker. The resultant plasmid (LH4.13) encoded a protein tagged with six histidine residues at its amino terminus. The LH4.11 Xho I-Not I insert was also subcloned into pGEX-KTO.1 expression vector. The resultant plasmid (LH4.15) encoded a protein fused with glutathione-S-transferase.

1.2.7. Epicedrol synthase protein analysis

LH4.13 was transformed into E. coli BL21(DE3) (Novagen) for expression. Several colonies were picked from the transformation to check the protein expression level by SDS-PAGE gel. One milliliter of overnight culture was diluted into 10 mL of fresh LB amp media and inoculated at 30 °C until OD_{600} reached 0.7, 100 mM IPTG was added to 50 μM final concentration. The flask was shaken at 30 °C for another 3 h. Cells from a 3 mL aliquots were collected and the cell pellet was resuspended in 500 μL of lysis buffer (50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 100 mM NaCl). The cells were sonicated twice at 4 °C for 10 sec each, and the tube was chilled on ice between sonication. A Branson Sonifier 450 was used and the output control was set at 5. The lysed cells were centrifuged to separate the soluble and insoluble fraction. The insoluble fraction was resuspended in 500 μL of lysis buffer. Ten microliters of each soluble and insoluble fraction were mixed with an equal amount of 2 × SDS sample buffer (0.1 M Tris, pH 6.8, 20% glycerol, 4% SDS, 2% 2-mercaptoethanol, 0.001% bromophenol blue). The samples were heated at 100 °C for 5 min and chilled on ice before loading. The protein samples were loaded onto a 9% SDS-PAGE gel along side wide-range protein
markers (Sigma). Most of the LH4.13 protein was expressed in an insoluble form (inclusion body); only a very faint band appeared in the soluble fraction at 61 kDa and the background of *E.coli* proteins was very high.

LH4.15 was a GST fusion protein, which can be easily purified by absorption with glutathione-agarose beads. LH4.15 was transformed into BL21(DE3) and cultured as described for LH4.13. The cells from a 3 mL aliquot of induced *E. coli* cells with LH4.15 were collected and resuspended in 500 μL of PBS solution (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄, 1.4 mM KH₂PO₄, pH 7.3), and the cells were lysed by sonication as described for LH4.13. The cell-free extract was mixed with 50 μL of 50% glutathione beads slurry which was made as follows: after preswelling S-linked glutathione agarose beads 1 h in 10 volumes of PBS, the beads were washed twice with PBS and stored as a 50% slurry in PBS. The protein was absorbed on the beads by rotating at 4 °C for 30 min. The beads were collected by centrifugation at 500 × g and the supernatant was discarded. An equal volume (25 μL) of 2 × SDS sample buffer was added to the beads and the samples were heated at 100 °C for 5 min before loading. The protein samples (20 μL) were loaded with 10 μL wide-range markers (Sigma) on a 9% SDS-PAGE protein gel. After the primary purification, a clear 87 kDa band appeared in the soluble fraction. In this experiment, the best IPTG concentration for expressing LH4.15 was also determined. Four cultures were induced with IPTG to final concentrations of 50 μM, 100 μM, 200 μM and 400 μM, respectively. The culture induced with 50 μM of IPTG gave the most soluble protein (still a tiny amount) on the gel.
1.2.8. Initial epicedrol synthase assay

LH4.13 was transformed into *E. coli* strain BL21(DE3)pLysS. One colony was inoculated into 2.5 mL of LB supplied with 100 µg/mL of carbenicillin and 25 µg/mL chloramphenicol for overnight incubation. One milliliter of the above overnight culture was diluted into 14 mL LB supplemented with 100 µg/mL of carbenicillin and 25 µg/mL chloramphenicol and shaken at 30 °C for 1.5 h (OD<sub>600</sub> = 0.7). IPTG (100 mM) was added to 50 µM final concentration and the flask was shaken at 30 °C for another 3 h. The cells were harvested by centrifugation, the pellet was resuspended in 750 µL of lysis buffer (50 mM Tris-HCl, pH8.0, 1 mM EDTA and 100 mM NaCl) and the cells were lysed by adding 50 µL of fresh made of lysozyme solution (10 mg/mL) and incubated at room temperature for 1 h. Thirty microliters of the cell lysate were added into 500 µL of assay buffer with 10.0 µL of 20 mM farnesyl pyrophosphate. Two kinds of buffer were used: buffer A (10 mM Tris-HCl, pH 8.0, 15% glycerol, 5 mM MgCl<sub>2</sub> and 5 mM β-ME)<sup>58</sup> and buffer C, which was buffer A with MnCl<sub>2</sub> instead of MgCl<sub>2</sub>. The enzymatic reaction was incubated at 30 °C overnight and extracted with pentane. The organic layer was dried over sodium sulfate and concentrated under a nitrogen stream. The enzymatic product was analyzed on GC, including a pET15b/Not I empty vector reaction (prepared exactly as LH4.13) as a negative control. GC analyses were performed on an HP series 6890, column: Rtx-5 (Restek Inc.), 30 m × 0.25 mm I.D., 0.10 mm df, 1/40 split ratio, helium carrier gas (15 psi), the oven temperature started at 70 °C for 1 min and increased at 15 °C/min to 250 °C, injector was 280 °C and detector was 280 °C. The GC data showed that enzyme was active in the presence of Mg<sup>2+</sup>; the retention time of the enzyme product was 10 min and Mn<sup>2+</sup> was not a suitable cofactor.
1.2.9. Optimization of epicedrol synthase assay conditions

**Type of reductant:** The cell lysate of LH4.13 was prepared as the initial enzyme assay. The enzyme was assayed in two different buffers: buffer A (above) and buffer A', which was buffer A with 2.5 mM DTT instead of β-ME. The GC data showed that 5 mM DTT gave better product yield than 5 mM β-ME.

**Induction time:** LH4.13 was inoculated as the initial enzyme assay. IPTG was added to the culture to 50 μM final concentration to induce the cells. After 3 h, half of the culture was harvested and assayed as the initial enzyme assay. The rest of the culture was induced for one more hour and harvested as before. The enzyme activity was assayed in buffer A' and the GC data showed that 3 h induction produced more enzymatic product than 4 h.

**E. coli. strain and expression vector:** LH4.13 was transformed into *E. coli* strain BL21(DE3) and BL21(DE3)pLysS and LH4.15 was transformed into BL21(DE3). LH4.13 and LH4.15 in BL21(DE3) were inoculated in LB amp and LH4.13 in BL21(DE3)pLysS was inoculated in LB supplemented with 100 μg/mL of carbenicillin and 25 μg/mL of chloramphenicol. The cells were induced and inoculated as the initial enzyme assay. After the cells were harvested, the cell pellets were suspended in lysis buffer to make the cell concentration at 100 mg/mL and the lysozyme was added to final concentration at 1 mg/mL. The enzyme was assayed same as before in buffer A' (10 mM Tris-HCl, 15% glycerol, 5 mM MgCl₂ and 2.5 mM DTT). The GC data showed that LH4.15 in BL21(DE3) made very little sesquiterpene compared to LH4.13 in
BL21(DE3). LH4.13 expressed in BL21(DE3)pLysS strain is the best system for this gene.

**Cell free extract vs. whole cell lysate:** LH4.13 in BL21(DE3)pLysS was inoculated and induced same as initial enzyme assay. After 3 h induction, the culture (30 mL) was harvested and resuspended in 1.5 mL of lysis buffer and 100 μL of fresh lysozyme solution (10 mg/mL). After incubated at room temperature for 10 min, half of the cell suspension was transferred to a new microfuge tube and sonicated twice for 10 s at setting 3. After sonicating, the cells were incubated for another 40 min. The rest of the sample was incubated at room temperature for 1 h. Half of each sample (sonicated and non-sonicated) was transferred to a fresh microfuge tube and centrifuged, and the cell-free extract of each sample was removed to a fresh microfuge tube. The enzymatic reactions were set up as the initial enzyme assay except different cell lysates were used. The following four lysates were used: 1) sonicated cell free extract, 2) sonicated whole cell lysate, 3) lysozyme cell free extract, 4) lysozyme whole cell lysate. The GC data showed that neither cell free extract gave any detectable enzyme activity and the cells that were only lysed by lysozyme had better enzyme activity than the one was sonicated.

**Substrate concentration:** When buffer A' was used to assay LH4.13 in BL21(DE3)pLysS, different amounts of 20 mM FPP solution (from 3.0 μL to 15 μL, 3 μL increase) was added into 500 μL of buffer A' with 80 μL (100 mg/mL) of cell lysate. The GC data of product/farnesol ratio showed that 9.0 μL of substrate gave the highest product/farnesol ratio.

**Mg²⁺ and DTT concentration of assay buffer:** Mg²⁺ and DTT concentration were optimized from buffer A'. The assay was same as the initial enzyme assay. When
10 μL of 20 mM FPP was used for assay, DTT concentration was varied from 0.5 mM to 2.5 mM (in 0.5 mM steps) while the MgCl₂ concentration was fixed at 5 mM and the MgCl₂ concentration was varied from 1 mM to 5 mM (in 1 mM steps) while the DTT concentration was fixed at 2.5 mM. The GC data showed that buffer H: 10 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT and 1 mM MgCl₂ gave the best yield among all these buffer conditions.

**Final epicedrol synthase assay buffer** After DTT and MgCl₂ concentration were decided, 10 mM Tris-HCl (pH 8.0) was replaced by 50 mM Tris-HCl to give buffer E: 50 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT and 1 mM MgCl₂, which gave slightly better yield for epicedrol synthase than buffer H. Buffer E was used for large scale enzyme product preparation.

### 1.2.10. Large scale preparation of epicedrol synthase products

Based on the small-scale assay data, LH4.13 in BL21(DE3)pLysS was chosen for the large-scale enzymatic reaction. Eighty milliliters of overnight culture was diluted into 1.2 L of LB supplied with 100 μg/mL of carbenicillin and 25 μg/mL chloramphenicol, and the resultant culture was incubated at 30 °C until the OD₆₀₀ reached 0.7 to 1.0. IPTG was added to 50 μM and the flask was shaken at 30 °C for another 3 h. The cells were harvested by centrifugation at 5000 rpm in a GS-3 rotor (Sorvall RC 5C Plus) for 20 min at 4 °C. The cell pellet was washed once with lysis buffer and then suspended in lysis buffer at 100 mg/mL cell concentration. The cells were lysed by French press at 7000 psi 3 times and the cell lysate was incubated overnight with 3 mL of 50 mM synthetic FPP in 400 mL of buffer E. The enzymatic reaction was quenched by
adding an equal volume of ethanol. The precipitate (cell debris) and denatured protein were removed by centrifugation and the supernatant was extracted with 3 volumes of hexane and dried over sodium sulfate. The crude product was chromatographed and eluted with a mixed solvent of hexane, CH₂Cl₂ and EtOAc at a ratio of 50:45:5 (Rf = 0.4) and 7 mg of pure sesquiterpene synthase products were obtained.

1.2.11. Structure identification of epicedrol synthase products

The GC-MS analysis was performed on VG ZAB-HF, EI⁺, 70 eV, 200 °C ion source temperature, scan range: 50-550 amu, scan speed: 1 sec, HP 5890A gas chromatography, column temperature started 70 °C for 1 min and increased 15 °C/min to 250 °C, injector temperature 250 °C, GC-MS transfer line: 270 °C, column: DB-5ms (J&W Scientific Inc.), 60 m x 0.25 mm I.D., 0.1 mm df. The GC spectrum showed two isomers (R₁ = 9.84 min and 9.96 min) in the sesquiterpene synthase product and their molecular weights were 222 instead of 204, indicating that the enzymatic products were sesquiterpene alcohols instead of hydrocarbons. The major product was identified as (-) epi-cedrol by spectroscopic comparison with authentic sample. m/z (relative intensity) 222 (6.75, [M]⁺), 204 (41.20, [M-H₂O]⁺), 150 (100, [M-H₂O-C₄H₆]⁺). ¹H NMR (500 MHz, Bruker AMX500 spectrometer, 25 °C, 5 mM in CDCl₃, referenced to tetramethylsilane) δ ppm 0.851 (3H, d, J = 7.0 Hz), 1.014 (3H, d, J = 0.6 Hz), 1.139 (3H, br d, J = 0.5 Hz), 1.322 (3H, s), 1.275 (1H, m), 1.339 (1H, m), 1.397 (1H, m), ~1.519 (1H, m), 1.527 (1H, m), 1.56 (1H, m), 1.57 (1H, m), 1.59 (1H, m), ~1.6 (1H, m), ~1.678 (1H, m), 1.733 (1H, m), 1.883 (1H td, J = 12, 6 Hz), 1.902 (1H, dd, J = 11.7, 1.6 Hz);

¹³C NMR (62.5 MHz, Bruker AC 250 spectrometer, 25 °C, 15 mM in CDCl₃, referenced
to tetramethylsilane) δ ppm 15.4 (C-12), 25.3 (C-4), 28.1 (C-14), 29.0 (C-13), 30.5 (C-10), 30.6 (C-15), 34.3 (C-9), 36.9 (C-3), 39.9 (C-11), 41.8 (C-2), 41.9 (C-6), 53.4 (C-1), 56.2 (C-5), 61.5 (C-7), 73.2 (C-8). The optical rotation was about –8.0 °(577 nm, CHCl₃). The minor product was identified as cedrol by comparison with the MS spectrum with MS library data and GC co-migration with authentic sample.

1.2.12. Analyzing an A. annua extract by GC-MS for epicedrol

Several branches of the A. annua specimen from which the cDNA library was prepared was extracted with ether in a Soxhlet extractor. A small amount (1 mL) of this essential oil was dissolved in ether and loaded on a preparative TLC plate along with epicedrol standard. After the TLC plate was developed in a mixed solvent of hexane, CH₂Cl₂, EtOAc at a ratio of 50:45:5, the band that comigrated with epicedrol was excised and extracted with ether. The resultant fraction was examined by GC-MS as described above for the product of the recombinant protein. Neither cedrol nor epicedrol was detected from plant extract by GC-MS, indicating that the amount of these two compounds in A. annua are lower than 0.001%.

1.2.13. Cloning a putative sesquiterpene synthase (LH6) from A. thaliana

A genomic sequence in the A. thaliana DNA database showed homology with sesquiterpene synthases. The genomic DNA sequence was downloaded and the introns were conceptually spliced out according to the prediction (http://www.cbs.dtu.dk/NetPlantGene.html). The resultant predicted cDNA is about 1.9 kbp long. A pair of primer was designed to flank the open reading frame of the cDNA,
AtSF2: TCGACAGAAGTTCAAAAAAGAAAATGGAGG and AtSR1: GCGGCCGAAGACTCTAAAGAGTGATG. The cDNA was amplified from *A. thaliana* by RT-PCR. Total RNAs extracted from dry seeds, two week old shoots, four week old stems, and four week old siliques of *A. thaliana* ecotype Columbia\textsuperscript{57} were used to synthesize first strand of cDNA for RT-PCR. Total RNA (8 \(\mu\)g, 2 \(\mu\)g each) was mixed with 0.5 \(\mu\)L (20 pmol/mL) oligo dT primer (18-mer), 81.5 \(\mu\)L DEPC H\(_2\)O, 10 \(\mu\)L 3 M NaOAc and 200 \(\mu\)L 100% ethanol, and precipitated at \(-20^\circ\)C overnight. After centrifugation, the RNA pellet was washed with 200 \(\mu\)L 70% DEPC-treated ethanol, air-dried and dissolved in 24 \(\mu\)L DEPC-treated H\(_2\)O. Half of the above total RNA and oligo dT mixture (12 \(\mu\)L) was used to synthesize the first strand cDNA. The mixture was heated at 70 \(^\circ\)C for 10 min and chilled on ice. The following reagents (from the Gibco-BRL cDNA library kit) were added to the total RNA solution: 4.0 \(\mu\)L 5 \(\times\) first strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl and 15 mM MgCl\(_2\)], 2.0 \(\mu\)L 0.1 M DTT, 1.0 \(\mu\)L 10 mM dNTP mix. After incubating at 42 \(^\circ\)C for 2 min, 1.0 \(\mu\)L SuperScript II reverse transcriptase II (200 units/\(\mu\)L) was added into first strand cDNA synthesis reaction, and the reaction was incubated at 42 \(^\circ\)C for 50 min. The reverse transcriptase was inactivated by heating at 70 \(^\circ\)C for 15 min. The RNA was degraded by incubating with 1.0 \(\mu\)L of *E. coli* RNase H at 37 \(^\circ\)C for 20 min. One microliter of this reaction mixture was used as template to PCR amplify this sesquiterpene synthase. The “hot start” PCR program was used with Fisher Taq polymerase. The PCR fragment was subcloned into pBluescript KS(+) T-vector. One colony was named as LH6.01 and the plasmid DNA was fully sequenced by primer walking (LH6S1: TACGCTTGTAGTCTGGG, LH6S2: CCACACATCACCACTC, LH6S3:
TTGAAAGGTGGGACTCCG, LH6S4: AATGGCTGAGGTCAAGAC, LH6S5: CCCAGACTAAACAAGCGTA, LH6S6: GAGCTTGGTGATGTGG, LH6S7: CGGAGTCCCACCTTCAA, LH6S8: GTCTTGACCTCAGCCATT). Ten DNA level mutations (two at amino acid level) were found by comparing the LH6.01 sequence with the predicted cDNA sequence from the Genbank. These two amino acid mutations on LH6.01 were removed by site directed mutagenesis with two primers in one site-directed mutagenesis reaction. The two mutagenesis primers were LH6M2: AACATTTCCATTTGATCATGCTCTAGTCTCGT (Bcl I site is underlined) and LH6M3: AGAGCTCATAATTTCCATACGCTCGT (Xho I site is underlined). Mutated plasmids were mapped with Bcl I and Xho I and the one with the correct sequence was named LH6.5. In order to subclone LH6.5 into an E. coli expression vector to make a fusion protein, a Nde I site was introduced in front of initial methionine by site-directed mutagenesis. The mutagenesis primer was CAAAAAGCATATGGAGGCCA (Nde I site is underlined), and the resultant plasmid was named as LH6.6.

1.2.14. Expression of LH6 in E. coli

The Nde I-Not I fragments of LH6.6 was subcloned into pGEX-KTO.157 to give LH6.8, which encoded the putative sesquiterpene synthase fused to GST. LH6.8 was transformed into DH5α and expressed on a small scale. One milliliter of overnight culture was diluted into 20 mL of fresh LB amp and incubated at 30 °C until the OD₆₀₀ reached 0.7 to 1.1. After IPTG (100 mM) was added to induce the expression of the gene, the culture was shaken for 3 h at 30 °C (250 rpm), and 1.5 mL of cells were harvested by centrifugation. The cell pellet was resuspended in 600 μL of PBS buffer
and the cells were lysed by sonication as described in protein analysis. The cell free extract was mixed with 30 µL of 50% PBS equilibrated glutathione beads slurry and rotated on at 4 °C overnight. The unbound fraction was removed by centrifugation at 500 × g and the glutathione beads were washed with 200 µL of PBS. The glutathione beads bound to protein was mixed with equal volume of 2 × SDS loading buffer and heated at 100 °C for 5 min. The denatured protein was loaded on a 9% SDS-PAGE gel. Different IPTG amounts (50 µM, 100 µM, 200 µM and 400 µM) was used to induced the cells, and the best one was 100 µM as showed on the protein gel. The gel showed that some LH6.8-encoded protein was expressed in a soluble form. The enzyme activity was assayed as epicedrol synthase and no activity was detected by GC from this construct.

The Nde I and Not I fragment of LH6.6 was subcloned into pET15b derivative in which a Not I site had been inserted in the polylinker to give LH.6.7 as a His-tag fusion. This construct was expressed in E. coli strain BL21(DE3). Two milliliters of overnight culture was diluted in 30 mL of LB amp media and shaken at 30 °C for 1.5 h (OD₆₀₀= 0.6), 100 mM IPTG was added to a final concentration of 100 µM and the culture was incubated for an additional 3 h at 30 °C. The culture was split into two 15 mL aliquots and centrifuged. The supernatants were removed, one cell pellet was resuspended in 600 µL of extraction buffer A (0.1% Triton, 50 mM Tris-HCl, pH 8.0, 1.0 mM EDTA, 100 mM NaCl) and the other cell pellet was resuspended in 600 µL of extraction buffer B (0.1% sarkosyl, 50 mM Tris-HCl, pH 8.0, 1.0 mM EDTA, 100 mM NaCl). Thirty five microliters of lysozyme solution (10 mg/mL) was added to both cell suspensions and the mixtures were incubated at room temperature for 1 h. The enzyme activity assay was as follows: cell lysate (60 µL) was added into 600 µL of assay buffer with 10 µL of 20 mM
FPP. For each extraction buffer, 10 different assay buffers were used. They were buffer A: 50 mM HEPES (pH 8.0), 15% glycerol, 2.5 mM DTT, 1.0 mM MgCl₂, buffer C: 10 mM HEPES (pH 8.0), 15% glycerol, 2.5 mM DTT, 1.0 mM MgCl₂, buffer E: 15% glycerol, 2.5 mM DTT, 1.0 mM MgCl₂, buffer G: 50 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT, 1.0 mM MgCl₂, buffer I: 10 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT, 1.0 mM MgCl₂, buffer B, D, F, H, J were paired with buffer A, C, E, G, I, except they used MnCl₂ instead of MgCl₂. The pET15b/Not I empty vector was assayed same way in extraction A and assay buffer G. After overnight incubation, the reactions were extracted with pentane and the organic layer was dried over sodium sulfate. The crude products were analyzed on GC, and same GC conditions were used as assay for LH4.13. The enzyme was active in buffer A, C, E, G and I. Extraction buffer A gave better cell lysate than extraction B. An empty pET15b/Not I vector was expressed in BL21(DE3) and prepared same as negative control. The GC retention time of enzyme product was 10.609 min and which co-eluted with farnesol, and the GC-MS data also showed as farnesol. The enzyme was not active with Mn²⁺ cofactor. Since FPP was partially hydrolyzed by the vector control lysates, purified enzyme is necessary to characterize this gene. A construct that should produce more soluble protein was made for future protein purification. The BamH I and EcoR I fragment of LH6.8 was subcloned into pET32a (Novagen) to give LH6.9. The LH6.9 will be further expressed in BL21(DE3) at 20 °C.
Chapter 3: Results and Discussion

1.3.1. Cloning epicedrol synthase from *A. annua*

Significant progress has been made toward elucidating the artemisinin biosynthetic pathway (Figure 5). Although the initial steps in the pathway (cyclization of FPP to 4,11-amorphadiene and subsequent oxidation to artemisinic acid) have not yet been experimentally verified, no other route seems plausible. It is generally accepted that sesquiterpenes are derived from FPP. A reasonable mechanism can be drawn for the cyclization to 4,11-amorphadiene, and enzymatic allylic oxidations such as that from 4,11-amorphadiene to artemisinic acid are common, and artemisinic acid has been shown to be an artemisinin precursor. Sesquiterpene synthases cyclized FPP to structurally related compounds, and these enzymes have similar sequences. Anticipating that the *A. annua* sesquiterpene synthase would resemble other dicot sesquiterpene synthases, we decided to clone the first sesquiterpene synthase from *A. annua* using a homology-based PCR approach.

According to a literature survey, the young leaves before budding have the highest amount of artemisinin. A cDNA library of *A. annua* was constructed from young leaves with triazol reagents. The library had $10^6$ individual clones and the average insert was about 1.3 kbp. At that time only a few sesquiterpene synthases had been cloned: vetispiradiene synthase from *Hyoscyamus muticus*, epiaristolochene synthase from *Nicotiana tabacum* and two cadinene synthases from *Gossypium arboreum*. The sequences alignment of these genes is shown in Figure 8.
Figure 8. Sequence alignment of sesquiterpene synthases. Sequences were aligned with megalign (DNAStar) using the Clustal method. Sequences shown are vetispiradiene synthase\textsuperscript{50} from \textit{Hyoscyamus muticus}, epiaristolochene synthase\textsuperscript{49} from \textit{Nicotiana tabacum}, two cadinene synthases\textsuperscript{51,52} from \textit{Gossypium arboreum}. Arrows depict positions of degenerate oligos, and asterisks indicate the successful primer pair.
The conserved the regions of these four genes were identified and 11 forward and 11 reverse degenerate primers were designed, the location of the primers are shown on the sequence alignment with arrows (Figure 8). All possible combinations of these primers were used to PCR amplify the cDNA library. Only one primer pair (SQF9 and SQR3) gave a PCR fragment with the anticipated size. The sequence of this 129-bp fragment was consistent with it being a sesquiterpene synthase. The forward primer SQF9 has $2^{15}$ fold degeneracy and the reverse primer SQR3 has $2^{12}$ fold degeneracy. This PCR fragment was used as a probe to hybridize to the *A. annua* cDNA library, and the three clones that were isolated represented the same gene. The longest one was named as LH4.2, and was only a partial cDNA compared to other dicot sesquiterpene synthases. The N-terminus of this cDNA was found by anchored PCR using two nested primers with the T7 primer on the vector. The 450 bp anchored PCR fragment was joined with LH4.2 to give a cDNA (LH4.8) with 1641 bp (547 amino acids) open reading frame. The sequence of this cloned sesquiterpene synthase is still a few amino acids shorter than the cadene, vetispiradiene and epiaristolochene synthase. Extensive anchored PCR amplification of the cDNA library did not provide any upstream sequence longer than 450 bp. Thus, the 1641 bp open reading frame is probably the full-length cDNA of the sesquiterpene synthase. The DNA and amino acid sequences of this gene are shown on Figure 9.
Figure 9. DNA and protein sequences of epidendrol synthase cDNA (LH4.8), Genbank accession No. AF157059
1.3.2. Epicedrol synthase expression and optimizing the in vitro enzymatic reaction

The reconstructed cDNA was expressed as a fusion to facilitate future purification. The gene (LH4.8) was subcloned into pET15b/Not I\textsuperscript{56} and pGEX-KTO.1\textsuperscript{57} respectively, and the expressed proteins were examined for sesquiterpene synthase activity. The recombinant protein converted FPP into a new compound evidenced by GC analysis. A parallel cell-free extract of \textit{E. coli} carrying empty vector did not produce this compound, suggesting that the gene-encoded protein catalyzed the reaction. The GC result shows that the pET15b/Not I construct produces more active protein than the pGEX-KTO.1 construct. Vector pGEX-KTO.1 has a 26-kDa glutathione-S-transferase fused to the N-terminus of the gene-encoded protein, which might interfere with protein folding to compromise enzyme activity. Expression host \textit{E. coli} BL21(DE3)pLysS (B2383) gave better enzyme activity than that in BL21(DE3). BL21(DE3)pLysS carries an extra pLysS plasmid which produces a low level of T7 lysozyme.\textsuperscript{61} T7 lysozyme is a natural inhibitor of T7 RNA polymerase. The low level of T7 lysozyme supplied by plasmid pLysS, which is compatible with the pET15b vector for expressing a gene from a T7 promoter, is sufficient to stabilize the foreign plasmid and allow high level protein to be produced upon induction of T7 RNA polymerase. The enzymatic reaction was optimized by varying divalent metal cofactors, glycerol, and reductant. Like other angiosperm terpene synthases, epicedrol synthase is activated by Mg\textsuperscript{2+} (1 mM MgCl\textsubscript{2} is optimal) but not by Mn\textsuperscript{2+}. Either DTT or β-mercaptoethanol can serve as reductant, but 2.5 mM DTT gave slightly better epi-cedrol yield. The optimal buffer used for the large-scale reaction was 50 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT, and 1 mM MgCl\textsubscript{2}. 
1.3.3. Epicedrol synthase product characterization

FPP (150 µmol) was fed to the lysate from a 1 L culture. The enzymatic product was pre-purified with silica gel chromatography and 7 mg of enzymatic product was obtained. GC analysis of the crude product with flame-ionization detection showed that the sesquiterpene product of the A. annua enzyme had two distinct isomers with retention times 9.84 and 9.96 min in a 4 : 96 ratio, respectively. GC-MS analysis showed that these two compounds have a molecular weight of 222 instead of 204, which corresponds to C_{13}H_{26}O, the formula of sesquiterpene alcohols. No C_{15}H_{24} sesquiterpene hydrocarbon products were found from the GC-MS analysis. The minor product is cedrol, as its EI fragmentation matched that of authentic cedrol, and it coeluted with cedrol in GC coinjection experiments. Although the EI mass spectrum of the major product also resembled that of cedrol from the MS library search, these compounds had different GC mobilities. The $^1$H NMR and $^{13}$C NMR data of the major product match with the authentic sample of (-) 8-epicedrol and are consistent with the published data.$^{62,63}$ 8-Epicedrol is numbered according to the tricyclo[5.3.1.0(1,5)]undecane skeleton.

1.3.4. Sequence analysis of epi-cedrol synthase

The deduced mass of the encoded protein is 61 kDa and the pI is calculated to be 4.94. The low pI and the molecular weight are similar to those reported for plant sesquiterpene cyclases.$^{49-51,64,65}$ The molecular mass of sesquiterpene synthases is smaller than that of mono- or diterpene synthases. Mono- and diterpene synthases are located in the plastids, which have plastidial targeting signal sequences located at their N-terminus, whereas sesquiterpene synthases are located in the cytosolic compartment.
The lack of an encoded N-terminal transit peptides in the epicedrol synthase cDNA indicates that it is located in the cytosol, and its precursor IPP is likely to be synthesized through the mevalonate pathway.

The deduced amino acid sequence of the *Artemisia* enzyme shows relatively high homology to other sesquiterpene synthase sequences and especially to those from angiosperms. A highly conserved aspartate-rich motif (DDXXD) is found in all terpenoid synthases, including those of microbial origin, as well as in prenyltransferases which operate by a related mechanism on the common prenyl diphosphate substrates. Based on directed mutagenesis and X-ray structural investigation, it has been shown that this motif is responsible for binding the divalent metal ions of the substrate diphosphate-metal ion complex.

The highest sequence homology is observed between the *Artemisia* cyclase and δ-cadinene synthase from *Gossypium* with similarities between 65 and 67% and identities between 41 and 43%. Vetispiradiene synthase from *Hyoscyamus muticus* and epi-aristolochene synthase from *Nicotiana tabacum* show similarities of 61 - 64% and identities of 40 - 42%. 8-Epicedrol synthase is more closely related to dicot sesquiterpene hydrocarbon synthases than to the monoterpen alcohol linalool synthase from the same organism (32%), but it is more distant from linalool synthase from other organisms (20% identical), indicating that the genes from the same species are more closely related those from other organisms. Sesquiterpene synthases (γ-humulene, δ-selinene and *E*-α-bisabolene synthase) from the gymnosperm *Abies grandis* exhibit 56-58.5% similarity and 29-34% identity to the *A. annua* clone, indicating that the
Sesquiterpene synthases from gymnosperms evolved independently from the angiosperm genes.

1.3.5. Epicedrol synthase reaction mechanism

Sesquiterpene synthases cyclize FPP to intermediates with different carbon skeletons, and the cationic reactions are terminated either by proton elimination to make C\textsubscript{15}H\textsubscript{24} hydrocarbons or by the equivalent of hydroxide addition to make C\textsubscript{15}H\textsubscript{26}O sesquiterpene alcohols. The best studied sesquiterpene synthases terminate the cyclization reaction by abstracting a proton from an intermediate cation.\textsuperscript{59} Only patchouli alcohol\textsuperscript{72} and epi-cubenol\textsuperscript{73} have been shown to be formed by quenching the intermediate carbocation with water. 8-Epicedrol synthase cDNA is the first cloned cyclic sesquiterpene alcohol synthase. A mechanism for the 8-epicedrol synthase-mediated cyclization is proposed in Figure 10. Farnesyl pyrophosphate is isomerized to nerolidyl diphosphate and then cyclized to generate the monocyclic bisabolyl cation. A hydride shift is followed by two cyclizations to yield the cedryl cation, and the reaction is terminated by quenching the tricyclic cation with a hydroxide equivalent (presumably by addition of water followed by proton loss) to form 8-epicedrol. The third cyclization is a syn addition, and therefore cannot be concerted with the second cyclization. The formation of cedrol as a minor product also suggests that the final carbocation has a significant lifetime. If water addition were concerted with the third ring closure, antiperiplanar addition would mandate quantitative 8-epicedrol formation.
1.3.6. Relevance of epicedrol synthase to artemisinin biosynthesis

An *A. annua* extract prepared from the same specimen as the cDNA was examined by GC-MS. Neither cedrol epimer was detected. This is in agreement with decades of research isolating and identifying the natural products found in this medicinal herb. Cedrane sesquiterpenes have not been reported from *A. annua*. In fact, epicedrol has not been reported as a natural product from any organism. The isolation of a gene encoding an epi-cedrol synthase implies that this product is made in plants, even
though this compound is undetectable in mature *A. annua* plants. It is possible that these alcohols are further metabolized to products not separable by GC or recognizable by MS. Cedrol has been reported from the related species *A. mongolic, A. feddei, A. vestita, A. dubia* and *A. Scoparia.*\(^7^4\) \(\alpha\)-Cedrene has been reported from *A. herba-alba.*\(^7^5\) The reason that cedrol exists in this plant is not clear.

epi-Cedrol synthase represents the first sesquiterpene cyclase to be described that produces the mechanistically complex cedrane structural family and the first that quenches the cationic reaction sequence with water. While of considerable interest in the study of artemisinin biosynthesis, the rationale for the presence of such an enzyme in *A. annua* is uncertain since the cedranes do not appear to be significant metabolites of this species. Converting epicedrol to artemisinic acid would require rearrangements that do not seem plausible. However, epi-cedrol synthase could be used as a probe to hybridize the *A. annua* cDNA library to clone other sesquiterpene synthases. After cloning the first cDNA, related genes from the same organism will be obtained more easily. For example, after first monoterpene synthase was cloned from *Abies grandis*\(^7^1\), four more monoterpene synthases were quickly cloned from the same organism by similarity-based cloning strategy.\(^2^3\)

This experiment has demonstrated that a homology-based gene cloning approach can be fruitful for isolating new sesquiterpene synthase genes from *A. annua*, and for identifying new natural products. The epi-cedrol cloning experiments illustrate the power of molecular biology to obtain catalysts that make natural products so rare that they cannot be detected by chromatographic methods as sensitive as GC-MS. This is currently an under-utilized approach in natural products chemistry. Others in our lab are
continuing to seek the 4,11-amorphadiene synthase gene by hybridization using epicedrol synthase as a probe or by PCR amplification using new primers based on epicedrol synthase sequence.

1.3.7. Cloning a sesquiterpene synthase-like gene from *A. thaliana*

The *A. thaliana* genome has been almost completely sequenced. At present, only the general class of terpene synthase (mono-, sesqui- and di-) can be predicted by sequence analysis, and this process is imprecise. Prediction of the exact terpene structural type is not currently possible by sequence analysis alone and awaits further understanding of synthase active-site structure. An *A. thaliana* genomic sequence encoding a predicted sesquiterpene synthase like protein was found in Genbank (gene F20P5.19, Accession number AC002062). Predicted splice sites and similarity to known sesquiterpene synthases were used to construct a hypothetical cDNA sequence. The cDNA was amplified from *A. thaliana* mRNA by RT-PCR using primers outside the open reading frame of the gene. The templates for the RT-PCR were total RNA isolated from dry seeds, two week old shoots, four week old stems and four week old siliques of *A. thaliana* ecotype Columbia. Since the error-prone *Taq* polymerase was used, two mutations at the amino acid level were found in the PCR amplified cDNA compared to the amino acid sequence from Genbank. These two mutations were removed by site-directed mutagenesis and the resultant DNA was sequenced. The 5'-terminal region flanking the coding DNA sequence was modified by site-directed mutagenesis to incorporate a unique *Nde* I site. The open reading frame of the amplified sesquiterpene synthase is 1833 kb, which corresponds to 633 amino acids. The deduced mass of the
encoded protein is 68 kDa and the pI is calculated to be 5.8. The DNA and protein sequences of LH6 are shown in Figure 11.
GTACCGGGCCCCCCTCGAGGTGAGCTGATCGTTACTGAGCTCTGATGATCTGACAGAGTCAAAAGCATATGAAGGCGC
75
ATACAAATATGGATTTTATTTTATGGTGGTTCTGTTGGCTATCAGATTGCTGCTTTCTATCGTACTTGGCTAC
150
ATKYGSGYFNVRSLRCWRLNLSS
TACCACTACCTCCTCTCTTAAATCATCTACTGCTATCTCTGATATTCTAATCTCCAAAAAGGTTGGCCTAGTGCA
225
YHYPPLKSSLSFSCRQSPKICKLVR
GCAAACAGCAATCTACAGATGTAACCTCGAGCAGCTCTTTAATTCTACACCTACACCACAAAGCTATGAGGGTCT
300
ATTNPTDDDNRSTTPHPSPSLWGH
CAGTTTCTCTCTACCTGCTGATCAAAAGCCAAATGAGTATCTTTTGGAGACACTGAGCACCCTAAACTATA
375
HFSLASVGVNTEOMDLWROIEALKPI
GTGAAATGCGGATGCTCCTCTTTGATACGGTGTGCAAGAGAATGATCTTGCCTCAGCCGCTATCAGCTTAC
450
YNAMLLPCNGADAKKICFCIFHTLVS
CTGAGGGTCTCTTACCTACAGTTTGGAGAAAGATGTTGAGTCTTGCTGAGAGAATATCGAGGACATG
525
LGVSYHFEKIFLVDFAFENIDM
ATCATCGGATTGCGAAGGAGATGATCCCTCAGCTGTTTTCATTATATCTCCGGTGTTTGCATAGATATTGCTCATTAC
600
I1DCLDKETLVSIIRFVFLYGHY
ATAACTCACCATATTTACCAACAGGTTTAAAGGQAAGCAGTGAATTTTAAAGAATATCGTCAATGATGATGAA
675
ITPDICITFRNRRFRQGDFGNNFKKLCNLON
GGTATGCTAAGTACTCAGAGCTAACACCTTGGCGAAGCAGAAGAATCTCCGTAAGAAGAACTGATTTTC
750
GMFLSFIYEAHFGTHFTDEILLEMSEF
ACACACAAAACACTTGAGGTATTTTTGTTAGTGGCGAGAAGGCAACATACTACACACACATAGCCTCATACAA
825
TkKHLEGEVKRLKHYPHITKLIO
GCTGCTCTTTTATACCTCAGAATTTCAATTTGAGAAATATTGTGCGAGAAGAATATTTGCTACCTACGACTA
900
AALYIPIQNFNLSEILVAREYIIOFVYE
GAGATCGATCAACATTAATGTTACTATAAGTGTGCTAAGAAGTACACTTAAAGGTCGTTCTGAGAGCTAATACCTA
975
ET2DHNEMLLKLAKLNLNFRKLILO
GACTGAAAAACACTAACAATGATGTTGGAAGAATCGTGCTATTTTACATTATCGAGCTCAGCAGCAGAGA
1050
DLKTLTTTWKELDLDVSKIPVFRER
TTGGCAGAGCCTTATTCTCTCGCGACCTGTTAATTTAGAGGCGACAAATATTACCTACGTGCGACATCAGCCTGCC
1125
LAEYPFWATGIIYEPOYSAAARMLALTACCATATTGGAATCTGAACTTCTGACAGTGAATGTTGTGCTGACAGA
1200
KSIIYDIVDONTAVFDYVTGEIDVESKL
GGTACAAGCAATTAAGGACGACTCCGGCGGAGATCTGCTACCTACACCCGTTAAGGAGCTACCTACAGAAACA
1275
VQAIERWDSADAVPVDLPOLYKLVVFR
ACHGGTGACCTTGTAAAAGATGAGCTCTGAGGAGACAGCTTTTACTATGCATATACGCCTAT
1350
TFDLFKELEYSEVSSEARSMOTYAY
GAGACGTCTAAAGGATCTGATTACATCTGCAAGACAGCTCTCGAGAAGCAGCGAATTCACATTAGGCGCAGGCCTCAAGCCT
1425
E0DLRLRIMKCYLLOEAEANWRSNHRGLPSH
GAGAAGACAGGCTCAGGCTCGGATGCGAACGGAGAGGAGAGCTCGCTAGAATGACCTTACCTACCCCTATGGGA
1500
EASYTEVYGAVSTAGEVILLAMTFIPMFGAGGCTCAGGGCTCAGGTATTTAGGCTGCGACAGCAAAACATGACCCACCTGCTTCTTGTGCAAATCT
1575
DAAGAEGYWEYELRSRPLKTALWFLVSKSCGTGCAGATATCTGCTACCTAATAGGAGAGATGGAAGAATGTTGGAATATTGAGGACTAACTGTTAT
1650
RLRDIATYKEEMKRGDRVNCINGICY
ACGAAGCAACCAAGAATAGGGAAGAGCTGTATAGGTGAGTTGGAAAAAGCATCACAATCAATGCGTAAAGCTT
1725
TKOHKYVEEAEIIIFEEKKTNHMKSKV
ATGAAACAGGATTGCGAAGGAGCTGTTTATTTACCTGCTTCTATATTGACCCGTGTTTTGTAATAGGCGACGCG
1800
MNEEFILKAAKFLIPHLIRPLPVLNYGR
TTGCGATGTCTCTAATAGAGATGAGGTGACCTCCTCGTGGAGGAATGCAAAATGATTATATACTCCCTTA
1875
LAADVICYKYDGTYFAEGKIKDYITSTGATTGCTTCATGCTCACTCTTAAAGGACGCTTCTCCGGCCCGCGAATTCGCTCAGCAGCGGCGGGGCAGGCTCAATTCC
1950
LYVDOLITT.
TTCTAGAGGCCGCAGCCACCCGGCTGAGGAGCTCCACATC
1987

Figure 11. cDNA and protein sequences of LH6.6, a putative sesquiterpene synthase, the bold sequence is the predicted chloroplast transit sequence.
1.3.8. Sequence analysis of LH6

The monoterpenes synthases are between 600 and 650 aa in length and sesquiterpene synthases are shorter by 50-70 aa. This difference largely is a result of the N-terminal transit peptides required for plastidial targeting of monoterpen synthases. The deduced amino acid sequence is 633 aa, which is the approximate length of a monoterpen synthase. However, it does not have a conserved tandem Arg in its N-terminus as the cleavage site for the signal of plastidal transit sequence of all the cloned monoterpen synthase to date. Most diterpene synthases are approximately 210 aa longer that monoterpen synthases, except casbene synthase from castor bean, which has only 601 aa. The N-terminus of the cloned gene possesses some expected characteristics of chloroplast transit peptides, which are often rich in serine and threonine and have a low abundance of acidic amino acids. The first 83 amino acids of this cloned gene are high in serine and threonine content, which includes 22 serine and threonine residues, but only 3 glutamate and aspartate residues. The computer program at http://genome.cbs.dtu.dk predicts that there is a chloroplast transit peptide sequence in the N-terminus of LH6.6 and the predicted cleavage site is between amino acid 54 (A) and 55 (T). So it is possible this gene is a monoterpen or diterpene synthase. The sequence alignment (Figure 12) with several other sesquiterpene synthases and a monoterpen synthase from angiosperms shows that the highest sequence homology is between this cloned gene and vistispiradiene synthase from Hyoscyamus muticus, which has 37.1% identity. Epi-aristolochene synthase from Nicotiana tabacum and two cadinene synthases from Gossypium arboreum show identities of 34.1-36.6%. It only shows 31.9% identity with acyclic sesquiterpene synthase (E)-β-farnesene synthase
from peppermint$^{64}$ and 27.8% identity with acyclic monoterpenoid alcohol synthase linalool synthase from *A. annua*.$^{69}$
Figure 12. Sequence alignment of monoterpene and sesquiterpene synthases. Sequences were aligned with megalign (DNASTar) using the clustal method. Sequences shown are LH6.6, farnesene synthase from peppermint (*Mentha x piperita*, L.), linalool synthase from *Artemisia annua*, epiaristolochene synthase from *Nicotiana tabacum*, vetispiradiene synthase from *Hyoscyamus muticus* and two cadinene synthases from *Gossypium arboriem*. 
One region containing the highly conserved aspartate-rich motif (DDXXD) in the C-terminal halves of the protein is found in all terpenoid synthases, including those of microbial origin. This DDXXD motif is involved in coordinating divalent metal ions for substrate binding, and possibly the ionization of the substrate allylic diphosphate to initiate the reaction. However, the copalyl diphosphate synthase from *Arabidopsis thaliana* lacks this DDXXD motif, which is believed to catalyze cyclization reactions by substrate double-bond protonation. The DDXXD motif is not in the cloned sequence or the genomic sequence in Genbank, which has a DNXXD instead. The nucleotides “GAC” encode Asp and the nucleotides “AAC” encode Asn, and they only have one nucleotide difference. Mutagenesis of any of the three aspartates of the DDXXD motif of limonene synthase to Glu reduced catalytic activity by 1,000-fold. It is unknown how the Asn in LH6.6 will affect the enzyme activity.

1.3.9. Expression of LH6 in bacteria

The cloned cDNA was cloned into *Nde* I and *Not* I digested pGEX-KTO for expression. The LH6.8 encoded protein having a 26-kDa glutathione-S-transferase fused to its N-terminus was expressed in *E. coli* DH5α. The expressed protein was purified using glutathione beads, and showed a protein at 94 kDa in the soluble fraction. The IPTG concentration was varied from 50 μM to 400 μM and the best concentration was found to be 100 μM from the protein gel result. I had previous experience that enzymes expressed as GST fusions did not give detectable enzyme activity. The gene was further cloned into pET15b/*Not* I for expression as a His-Tag form for future purification. The gene was expressed in *E. coli* BL21(DE3) strain carrying the inducible T7 RNA
polymerase in a prophage under the control of a LacUV5 promoter. The enzyme assay conditions were varied from lysis buffer to assay buffer. Two kinds of lysis buffer were used, one with Triton as detergent and the other with sarkosyl as detergent to help the cells to lyse. The 0.1% Triton combined with lysozyme gave better results for lysing the small amount of E. coli cells for the enzyme assay. The assay buffers were varied from HEPES to Tris-HCl and Mg\(^{2+}\) is the divalent cation. The enzyme is not active under Mn\(^{2+}\). The best assay condition is 50 mM Tris-HCl (pH 8.0), 15% glycerol, 2.5 mM DTT and 1.0 mM MgCl\(_2\). The enzyme activity was analyzed by GC and an enzymatic product showed retention time at 10.7 min. Since the enzyme product co-elutes with farnesol, it is possible that the hydrolase activity came from E. coli. A low level of farnesol was detected in extracts of E. coli harboring empty plasmid, and this is likely derived from the action of nonspecific phosphatase.

1.3.10. Future work

In order to exclude the non-specific phosphatase activity from the host cells, protein purification will be necessary. It has been reported that recombinant protein expressed at 20 °C can give better protein solubility.\(^{80}\) The gene was subcloned into a pET32a expression vector, which encodes a fusion protein thioredoxin gene and a heptahistidyl tag to simplify the protein purification. The fusion protein thioredoxin will increase the solubility of the recombinant protein expressed in E. coli. After large amount of soluble protein is obtained, the enzyme activity will be further assayed by feeding GPP, FPP, and GGPP to decide if the gene encodes a monoterpane, sesquiterpene or diterpene synthase. The other possibility on this project is to cleave some of the N-
terminal sequence, which is considered as the plastidal transit peptide and express the truncated protein in \textit{E. coli} and then test its activity against GPP, FPP and GGPP. A final consideration is that the enzyme may cyclize its substrate without cleaving the prenyl pyrophosphate (as does copalyl pyrophosphate synthase).
Chapter 4: Conclusions

In the process of studying artemisinin biosynthesis, a sesquiterpene synthase has been cloned from *A. annua* cDNA library by homology based PCR using a cDNA library made from the leaves before budding. The cDNA was expressed in *E. coli* using the into a pET15b vector. The enzyme activity was assayed by feeding the *E. coli* lysate farnesyl pyrophosphate and the products were analyzed on GC. The structure of the enzymatic product was determined by NMR and GC-MS. This sesquiterpene synthase cyclizes farnesyl pyrophosphate to epi-cedrol and cedrol at a ratio of 96 : 4. Neither cedrol or epi-cedrol was detected from an extract of the plant from which the cDNA library was made. epi-Cedrol synthase is the first sesquiterpene alcohol synthase that has been cloned to date. Progress has been made towards identifying catalysts in artemisinin biosynthesis because this gene can serve as a probe for further cloning of sesquiterpene synthases from *A. annua*. This experiment illustrates the power of molecular biology to obtain catalysts that make natural products so rare that they cannot be detected by chromatographic methods as sensitive as GC-MS.

Another project described in this part of the thesis is the cloning of a sesquiterpene synthase-like gene from *A. thaliana*. A sesquiterpene synthase-like open reading frame was found in the *A. thaliana* sequencing project and a cDNA was isolated from the *A. thaliana* ecotype Columbia total RNA by RT-PCR. The cDNA was expressed in *E. coli* using the pET15b. The expressed gene showed weak but detectable farnesyl pyrophosphate hydrolase activity compared to the *E. coli* harboring an empty vector. From the length of the open reading frame and the sequence characteristics of the
N-terminus, it is hard to conclude that the cloned gene is a sesquiterpene synthase, even though it most closely resembles sesquiterpene synthases. It will be necessary to purify the enzyme and then incubate the enzyme with different substrates including GPP, FPP, and GGPP under various assay conditions to determine what the enzyme activity is.
Part II

Cloning of Triterpene Synthases from *Arabidopsis thaliana*

Chapter 1: Background and Introduction

2.1.1. General introduction of triterpenes

Sterols and triterpenes are widely distributed isoprenoids found in various organisms and constitute one of the most important classes of natural products. Sterols serve as essential membrane constituents, growth regulating substances, and precursors of various hormones in mammals, plants, fungi and yeast. The most studied triterpenes are lanosterol, which is the sterol precursor for cholesterol in animals and for ergosterol in fungi. In higher plants, rather than lanosterol, cycloartenol is the precursor for phytosterol primary metabolites such as stigmasterol and stitosterol. Besides cycloartenol, a large number of nonsteroidal triterpene derivatives exist as secondary metabolites. Some of these secondary metabolites are stored as glycosides, saponins and steroidal alkaloids. In addition, some of them are stored in their latexes and resins, which are a chemical defense against pathogens and herbivores. All of these sterols and triterpenes are biosynthesized from a common precursor, oxidosqualene. In bacteria, hopene is the sterol precursor, which is derived from squalene instead of oxidosqualene. Its C₃₅ bacteriohopanepolyol derivatives stabilize the membrane and are believed to serve an analogous function as cholesterol, sitosterol and ergosterol in
eukaryotes. The study of sterol precursors has fascinated researchers over decades, because of their great biological and medicinal significance. However, the study of nonsteroidal triterpenes has not been well investigated until recently, when the bloom of molecular biology has accelerated the pace of triterpene research. More than 80 triterpene alcohols having the molecular formula \( (\text{C}_{30}\text{H}_{50}\text{O}) \) have been found in nature,\(^8\) they are apparently the enzymatic products of oxidosqualene cyclase. The chemistry behind the single step cyclization of oxidosqualene to the structural diversity of triterpenes fascinates researchers. Over the past decade, several oxidosqualene cyclases have been cloned from various plants.\(^8\)-\(^9\)

2.1.2. Triterpene biosynthesis

Triterpenes in eukaryotes are biosynthesized from oxidosqualene whereas triterpenes in bacteria are biosynthesized from squalene.\(^8\) \(^3\) \((3S)-2,3\)-oxidosqualene is biosynthesized via the mevalonate pathway.\(^4\) Farnesyl pyrophosphate is exclusively synthesized in the cytosol, where IPP is synthesized through the mevalonate pathway. The mevalonate-independent pathway uses pyruvate and D-glyceraldehyde-3-phosphate as precursors to synthesize IPP, the precursor for GPP and GGPP.\(^1\(^9\),\(^2\)\(^0\) This alternative isoprenoid pathway is localized in the plastid and is not involved in triterpene biosynthesis (Figure 3). Two molecules of FPP undergo a head-to-head condensation to give squalene. Squalene synthase is the critical branch-point for terpenoid biosynthesis in eukaryotes.\(^9\)\(^1\) Before squalene synthases, the biosynthesis of terpenoids branches to give the monoterpenes, sesquiterpenes, diterpenes and ubiquinones. After squalene synthase, the biosynthesis of triterpenoids branches to give the sterols and triterpenes.
Squalene is oxidized to oxidosqualene by a P450 type enzyme. The resulting oxidosqualene is then cyclized to the structurally diverse sterols and triterpenes. Eukaryotic oxidosqualene cyclases only accept the (3S)-enantiomer of oxidosqualene as a substrate. Figure 13 shows triterpene and sterol biosynthesis.
Figure 13. Triterpene and sterol biosynthesis
2.1.3. Enzyme mechanism of oxidosqualene cyclases

The cyclization of oxidosqualene into tetracyclic and pentacyclic carbon skeletons by oxidosqualene cyclases is one of the most complex and fascinating reactions found in nature. The most studied oxidosqualene cyclases are lanosterol and cycloartenol synthases. Lanosterol synthase has been purified from rat liver, and the gene has been cloned from *Saccharomyces cerevisiae*, *Candida albicans*, *Schizosaccharomyces pombe* and humans. Cycloartenol synthase has been cloned from *Arabidopsis thaliana*, *Pisum sativum*, *Panax ginseng*, and *Allium macrostemon*. The cyclization of oxidosqualene to lanosterol and cycloartenol proceeds via the “chair-boat-chair” conformation of (3S)-oxidosqualene, and the reaction is initiated by an acid-catalyzed oxirane ring opening with participation by a neighboring π-bond. The cyclization proceeds to give a protosteryl cation, which then undergoes a series of 1, 2-methyl and hydride shifts with proton elimination to yield either the lanosterol or cycloartenol skeleton. Unlike lanosterol and cycloartenol synthases, other nonsteroidal triterpene synthases that have been cloned so far proceed through another intermediate. Lupeol synthase from *A. thaliana*, *Olea europaea*, and *Taraxacum officinale* and β-amyrin synthase from *P. ginseng* cyclize the oxidosqualene via the “chair-chair-chair” conformation to give the tetracyclic dammarenyl cation. Ring expansion to the baccharenyl cation and a fifth ring closure give the tertiary lupenyl cation. Deprotonation of one of the gem-dimethyl groups of the lupenyl cation gives lupeol, while ring expansion of the lupenyl cation to the oleanyl cation followed by two 1,2-hydride shifts and proton abstraction from C-12 gives β-amyrin. It is believed that
tetracyclic and pentacyclic triterpenoids derived from these cationic intermediates are also widespread in nature. Figure 14 shows the mechanisms of oxidosqualene cyclases.
Figure 14. Mechanism of oxidosqualene cyclases
2.1.4. Methods for studying triterpenes

The structural diversity of triterpene skeletons in plants gives clues about the diversity of oxidosqualene cyclases and the size of this enzyme family. It is of great interest to determine how each triterpene synthase controls the product specificity during cyclization of oxidosqualene. The triterpene cyclases are membrane-bound, and therefore it is difficult to obtain a soluble protein in an active form by native enzyme purification.\textsuperscript{100,101} Recently, the use of recombinant DNA technology has begun to provide sources of protein for catalytic and physical studies, as well as the sequence information that serves as the blueprint for these enzymes. Most available triterpene synthases have been cloned following the strategy: natural product isolation and structural characterization, enzyme characterization, and gene cloning from that organism. Lupeol synthase was cloned by a different method. An \textit{Arabidopsis} DNA fragment was identified that was similar to the cycloartenol synthase from the same organism. Cloning and characterization of the full-length cDNA showed its encoded protein cyclizes oxidosqualene to lupeol. As more and more genomic sequences become available from the \textit{A. thaliana} sequencing project, cloning by PCR is another approach. Based on the sequence homology of cloned triterpene synthases, homology-based PCR allows the cloning of genes from various organisms. In order to obtain further insights in the molecular evolution of triterpene synthases and their catalytic mechanisms, cloning more triterpene synthases from the same source would be the first thing to do.
2.1.5. Triterpenes from *A. thaliana*

Cycloartenol, a primary metabolite, is the most abundant triterpene in plants, and cycloartenol synthase has been cloned from *A. thaliana* by a thin layer chromatography-based screen of a cDNA library. The nonsteroidal triterpene lupeol has not been isolated from *A. thaliana*, however, the first lupeol synthase has been cloned from *A. thaliana* by hybridization. We choose *A. thaliana* as a model plant for cloning of triterpene synthases because as the genome sequencing project progresses, more and more triterpene synthase-like sequences will be available for cloning and characterization. In addition, convenient access to an *A. thaliana* cDNA library and plants (from Dr. Bartel's lab) make cloning easier.

When I started this project, only a few triterpene synthases had been cloned, including lanosterol synthase from *Candida albicans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, rat and human and cycloartenol synthase and lupeol synthase from *A. thaliana*. All of the known oxidosqualene cyclases have been shown to possess several highly conserved sequences including WGK and DGSW motifs. Speculating that these conserved motifs will be found in other oxidosqualene cyclases, homology-based PCR was used to amplify an *A. thaliana* cDNA library for the oxidosqualene cyclases. The cDNA was hybridized from the cDNA library, and found to have a sequence similar to the cycloartenol and lupeol synthase from the same organism, with the exception that the second exon is missing. Attempts at cloning and expressing this cDNA in yeast will be described. Two other triterpene synthase like sequences were found in the *Arabidopsis* genome. They are 70-80% identical to lupeol synthase from the
same organism. Efforts to clone and express these two cDNAs in yeast will also be described.

2.1.6. Yeast as a model system

The genetics and biochemistry of yeast have been well studied, especially its sterol biosynthesis. The genes of the yeast sterol biosynthetic pathway have all been cloned; a total of 20 enzymes and 19 steps from acetyl-CoA to ergosterol.\textsuperscript{105} Yeast (\textit{Saccharomyces cerevisiae}) cells are nonpathogenic, grow rapidly and have stable haploid and diploid states. Furthermore, it is a single celled eukaryote and has all the characteristics of eukaryotes. It is a superior organism for expression of genes cloned from other eukaryotes. A previous study showed that oxidosqualene cyclase cannot be expressed in \textit{E. coli}, perhaps because of folding and post-transcriptional differences between eukaryotic and prokaryotic genes.\textsuperscript{106}

The availability of a well-characterized genetic system and the ease of isolating mutants in a haploid organism have made yeast engineering feasible. Homologous recombination occurs frequently in yeast cells and this property helps target gene disruption.\textsuperscript{107} When sufficiently small pieces of yeast chromosomal DNA are conjugated with a foreign DNA and introduced into the original yeast strain, any chromosomal gene might be disrupted by homologous recombination between the introduced incomplete gene and the corresponding chromosomal gene.

Yeast has the capability to uptake sterols and also allows the usage of appropriate strains for complementation.\textsuperscript{108} Wild-type yeast grown under aerobic conditions cannot uptake sterols from the growth medium, because sterol uptake is dependent on heme
biosynthesis. Since oxygen is required for ergosterol biosynthesis, under anaerobic conditions, heme and sterol synthesis are precluded, and yeast imports sterols. However, when heme biosynthesis is disabled, yeast cells uptake sterols under aerobic conditions.

2.1.7. Engineering yeast strains

The advantage of yeast as a host for oxidosqualene cyclase expression is the common origin between sterols and triterpenes. As mentioned before, the biosynthesis of sterols and triterpenes is the same up to the point of oxidosqualene cyclase. In yeast, lanosterol synthase cyclizes the (3S)-2,3-oxidosqualene to lanosterol, precursor to the membrane sterol ergosterol. Other triterpene synthases expressed in yeast also use (3S)-2,3-oxidosqualene as substrate.

Several steps in the middle of the yeast sterol biosynthetic pathway are of particular interest for making yeast mutants. Squalene synthase (ERG9) is the branch point for yeast sterol biosynthesis. Two molecules of farnesyl pyrophosphate undergo a head-to-head condensation catalyzed by ERG9 to give squalene. Squalene epoxidase is responsible for the conversion of squalene to oxidosqualene, which is the immediate precursor to the initial sterol molecule in the ergosterol biosynthetic pathway. This enzyme functions only in the presence of molecular oxygen. Lanosterol synthase (ERG7) cyclizes oxidosqualene to lanosterol. The yeast strain SMY8, an erg7 mutant, was developed for cloning lanosterol synthase genes by complementation. This strain was constructed in a hem1’ background and therefore allows sterol import under aerobic conditions. This yeast strain has in vivo substrate oxidosqualene and was used to express cycloartenol synthase while uptaking ergosterol from the growth medium.
In the effort of cloning and expressing more oxidosqualene cyclases for the mechanistic study of this enzyme family, a yeast strain with no *in vivo* oxidosqualene is needed. Some triterpene synthases might produce compounds lethal to yeast with the *in vivo* oxidosqualene substrate, some oxidosqualene cyclase might produce more than one enzyme product, and some of the enzyme products might be further metabolized in yeast. In order to get an accurate ratio of enzyme products, a host expression system where oxidosqualene cyclases could not access *in vivo* oxidosqualene is required. With this goal in mind, I engineered a yeast strain deleted for both squalene synthase and lanosterol synthase. This mutated yeast strain can be used to express many kinds of oxidosqualene synthase and provides accurate quantitation of the product ratios.
Chapter 2: Experimental Procedures

2.2.1. Yeast squalene synthase (ERG9) knockout

EH7.1 (#84 in Dr. Matsuda lab’s plasmid library) is the yeast squalene synthase (ERG9) 5’-nontranslated region in pGEM-T vector; EH6.3 (#83 in Dr. Matsuda lab’s plasmid library) is the ERG9 coding sequence in pGEM-T vector; JR2.01 (#35 in Dr. Matsuda lab’s plasmid library) is the HIS3 gene in pBluescript KS(+).

ERG9F1 is a forward primer with an Acc I site (underlined) GTATACTCAGAGAGC TGCATCCAGAAT and ERG9R1 is a reverse primer with Sph I and Not I sites (underlined) GCATGC GCAGCCCCGTTATTTGGC. These primers were designed to PCR-amplify the 1.1 kbp C-terminus of squalene synthase. The PCR program was CAPS (see Materials and Methods section) and the DNA template was EH6.3. The PCR reaction included 40 ng DNA, 5.0 μL 10 × PC2 buffer, 4.0 μL dNTPs (each at 2.5 mM), 40 pmol of each primer, 2.5 units Taq polymerase (Fisher) and water to 50 μL. The 1.1 kbp PCR fragment was gel purified and subcloned into pGEM-T vector (Promega) using Fast-Link DNA ligation as described in the Materials and Methods. The ligation mixture (5.0 μL) was used to transform 100 μL of E. coli DH5α chemically competent cells. The positive clone was verified by restriction analysis and named LH7.0.

In order to gain a BamH I site upstream but not next to the Acc I site of the 1.1 kbp ERG9 fragment, the Acc I fragment of LH7.0 was subcloned into CIP-treated pBluescript KS(+) Acc I linearized vector using the Fast-Link DNA ligation kit. One out of 12 minipreps had the insert with the right direction, and this clone was named LH7.1.
LH7.1 was sequentially digested with *Sph* I and *BamH* I as follows: after LH7.1 (10 μg) was digested with 40 units of *Sph* I in NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9 at 25 °C) supplied with BSA at the DNA concentration of 100 ng/μL, 2.0 μL of 5 M NaCl was added to the digestion reaction to increase the salt concentration to match *BamH* I buffer (150 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9 at 25 °C), and *BamH* I was added. The resultant *Sph I/BamH* I fragment of LH7.1 was subcloned into EH7.1 digested with the same enzymes. The resultant plasmid LH7.2 contained the *ERG9* 5'-nontranslated region adjacent to 1.1 kbp of squalene synthase (*ERG9*) gene. The *HIS3* selective marker was inserted in the *BamH* I and *Acc* I sites between 5'-nontranslational region and the 3' end of the *ERG9* gene. Because *Acc* I cuts LH7.2 three times, (once in EH7.1, once in LH7.1 and once in the multiple cloning site of pGEM-T vector), a partial digestion was necessary. Twenty micrograms of LH7.2 plasmid was digested with 5 U *Acc* I at 37 °C for 1 h. The DNA cut only once with *Acc* I was gel-purified and then digested with *BamH* I. Since the target *Acc* I site was close to *BamH* I, the right *Acc I/BamH* I double digested LH7.2 should only give a large band and a very small band which should not be visible on a 1% agarose gel. The large DNA fragment of LH7.2 was gel-purified after *Acc* I and *BamH* I sequential double digestion. The *HIS3* gene from JR2.01 digested with the same restriction enzymes was subcloned into *Acc* I- and *BamH* I-linearized LH7.2 to give LH7.3.

LH7.3 was the construct used for the *ERG9* gene deletion in yeast. LH7.3 (18 μg) was digested with *Not* I and used to transform SMY498 using the lithium acetate yeast transformation method. Transformants were selected for the *HIS3* integrants using
synthetic media lacking histidine and supplied with heme (13 mg/L), ergosterol (20 mg/L) and Tween 80 (5 mg/L). One hundred original transformants were streaked on both SCD-His supplied with heme and SCD-His supplied with heme, ergosterol and Tween 80 plates to identify ergosterol auxotrophs. Only three colonies had a functional ERG9 gene and grew on SCD-His + heme plate, and ninety-seven colonies died on SCDH-His plate and grew on SCDHET-His plate. This indicated that the ERG9 gene was deleted by homologous recombination and the crossover efficiency was 97%. The positive clone was further verified by PCR amplification with LH7.3F2: AACAACTTACGACTTTACCATTCC and LH7.3R2: TAGTCGGCCATACCATCATTACC. This pair of primers gave a 600 bp fragment for wild type ERG9, and a 1.7 kbp fragment for erg9 interrupted by the HIS3 marker. A fresh single yeast colony was picked and heated at 95 °C for 10 min in 39 μL of sterile water to break the yeast cell wall. After the water cooled to room temperature, the other PCR reagents (5.0 μL 10 × PC2, 4.0 μL dNTPs (2.5 mM each), 1.0 μL of each primers (20 pmol/μL) and 0.5 μL of Taq polymerase) were added and CAPS program was used. Two yeast colonies were checked by PCR and they both gave 1.7 kbp PCR fragment. One of them was named LHY1.

2.2.2. Yeast ERG9 and ERG7 double deletion

JR3.1 (#35 in Dr. Matsuda lab’s plasmid library) was the yeast lanosterol synthase coding sequence (ERG7) disrupted with the LEU2 marker in pBluescript KS(+) . The PstI fragment of JR3.1 can undergo homologous recombination with yeast lanosterol synthase.
JR3.1 (10 μg) was digested with Pst I and used to transform 5 mL of LHY1 culture grown in rich media (YPDHET). The transformation was selected for LEU2 prototrophy on complete synthetic media lacking leucine and supplied with heme (13 mg/L), ergosterol (20 mg/L) and Tween 80 (5 mg/L). The positive clone with the erg7 gene deletion was further confirmed by enzyme activity assay. ERG7 (lanosterol synthase) cyclizes oxidosqualene to lanosterol, which can be detected by thin layer chromatography. SMY898 with the erg7 deletion can be used as a negative control. LHY1 has the wild type ERG7, but a mutant erg9 gene. The parent strain SMY4 has both ERG7 and ERG9 genes intact. The saturated yeast cultures (5 mL, YPDHET) were harvested by centrifugation. The cell pellets were resuspended in 60 μL of assay buffer (990 μL of 0.5 M pH 6.2 sodium phosphate, 10 μL of Triton X-100) and lysed with 100 μL of acid washed glass beads by vortexing for 4 min. The assays were set up with 5.0 μL of 20 x OS (20 mg/mL) or without substrate and the enzymatic reactions were incubated at room temperature for 3 h. A 5.0 μL aliquot of the reaction mixture was spotted on TLC plate. The TLC plate was air-dried and developed twice in ether to the baseline and then developed in 100% CH₂Cl₂ for identification of lanosterol. The LHY1 strain converted oxidosqualene to lanosterol in vitro, and a control without oxidosqualene did not produce lanosterol, indicating that oxidosqualene was not produced in vivo. The SMY8 and LHY2 strains did not produce lanosterol either in the presence or absence of oxidosqualene, indicating that lanosterol synthase was not active. The SMY4 strain produced lanosterol in vivo because of its intact sterol biosynthetic pathway. The TLC assay result elucidated that LHY2 had squalene synthase and lanosterol synthase double deletion.
2.2.3. Recovery of LEU2 marker from LHY2

Because the LEU2 marker was used to delete ERG7, the LHY2 strain can biosynthesize leucine, precluding using the useful pRS305GAL integrating vector. The LEU2 blaster, a yeast URA3 gene flanked by 1.1 kbp direct repeats of a bacterial sequence (hisG) inserted into LEU2 gene was used to delete the chromosomal LEU2 marker in the LHY2. The 5.0 kbp Bgl II fragment of pNKY85 (from Dr. Bartel) was transformed into LHY2 and the transformed yeast were selected for growth in the absence of uracil. The positive clones were further confirmed by streaking the transformants on both SCDHET-Ura and SCDHET-Leu plate. Two out of twenty-four colonies survived on SCD-Ura HET plate and died on SCD-Leu HET plate, indicating the LEU2 gene was not functional in these two colonies. One of these yeast colonies was named LHY3.

The URA3 marker in yeast was removed by selecting for 5-fluoro-orotic acid (5-FOA) resistance. An overnight culture of LHY3 in complete synthetic medium was plated on SCD plus heme (13 mg/L), ergosterol (20 mg/L), Tween 80 (5 mg/L) and FOA (1 mg/mL). The colonies that survived this plate lost the URA3 gene and left the LEU2 marker disrupted with one copy of hisG sequence. The resultant yeast strain LHY4 is MATa erg7::leu2::hisG erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal".
2.2.4. Construction of yeast expression vector pRS426GAL

Plasmid pRS426 was purchased from ATCC. The 800 bp GAL promoter was obtained from pRS316GAL\textsuperscript{113} using \textit{Kpn} I and \textit{Sal} I. Ten micrograms of pRS316GAL was digested with 40 units of \textit{Sal} I in NEB \textit{Sal} I buffer at the DNA concentration of 100 ng/\mu L at 37 °C for 2 h. The \textit{Sal} I linearized DNA was precipitated with 1/10 volume of 3 M pH 5.5 NaOAc and 2.5 volumes of 100% ethanol. The DNA was washed with 75% of ethanol and air-dried. The DNA pellet was dissolved in 86 \mu L of water and was further digested with 40 units of \textit{Kpn} I in NEB \textit{I} buffer supplied with BSA (100 \mu g/\mu L). The 800 bp GAL promoter with \textit{Sal} I and \textit{Kpn} I on each end gel purified and subcloned into \textit{Sal} I and \textit{Kpn} I digested pRS426\textsuperscript{114} to give the yeast expression vector pRS426GAL. The newly constructed pRS426GAL was tested using cycloartenol synthase from \textit{Arabidopsis thaliana} (AtCAS1). The AtCAS1 \textit{Xho} I-\textit{Not} I fragment was subcloned into \textit{Sal} I and \textit{Not} I double digested pRS426GAL and the resultant construct transformed into yeast strain SMY8 showed cycloartenol synthase activity by TLC as described in the Materials and Methods.

2.2.5. Construction of pRS416GPD/\textit{Not} I and pRS426GPD/\textit{Not} I

Vector pRS416GPD and pRS426GPD were obtained from Dr. Gustin (Biochemistry and Cell Biology Department, Rice University). These two vectors have a constitutive promoter from glyceraldehyde-phosphate dehydrogenase and a terminator from cytochrome-c oxidase.\textsuperscript{115} The multiple cloning sites of pRS416GPD and pRS426GPD were not compatible with the restriction sites we used in our lab for subcloning. A polylinker with four existing restriction sites (\textit{Hind} III, \textit{Cla} I, \textit{Sal} I and
Xho I) and three new restriction sites (Bgl II, Nde I and Not I) was inserted into the multiple cloning site of these two yeast expression vectors. Two primers encoding the top strand (5'-AAGCTTGGTATCGATTTACGAAGATCTCTGTCGACAAAAATACATATGCACGCACGCGCCGACCATCTCGAG) and bottom strand of this polylinker were annealed in a heating block (removed from heater) from 100 °C to room temperature in the reaction included 2.0 µL of each primer (50 pmol/µL), 10 µL of 10 x anneal buffer (200 mM Tris-HCl, pH 7.5, 20 mM MgCl₂ and 500 mM NaCl) and 86 µL of H₂O. After annealing, the polylinker was subcloned into Hind III- and Xho I-digested pRS416GPD and pRS426GPD to give pRS416GPD/Not I and pRS426GPD/Not I, respectively. These two yeast expression vectors were used to express triterpene cyclases LH2 and LH5.

2.2.6. Cloning LH1 from an A. thaliana cDNA library

One pair of primers OS13: ATHCCNWSNTGGGNAARWTNTGGYT (24576 fold degeneracy) and OS10: CCRTANGTRAARCANAYNSCCCA (2048 fold degeneracy) was used to PCR-amplify the A. thaliana young seedling cDNA library. The same PCR conditions and program were used as PCR amplification of Artemisia annua cDNA library for sesquiterpene synthase. A clear band was obtain at the expect size (1.2 kbp). This band was subcloned into pT7Blue vector (Novagen) and four out of twenty four clones had insert and these four plasmids were further mapped with Dra I, Hind III, Nde I and Pvu II to check for the cycloartenol synthase (AtCAS1) or lupeol synthase (AtLUP1) cDNA. One plasmid matched neither AtLUP1 nor AtCAS1 restriction mapping pattern. This plasmid was named LH1.0, and was sequenced from T7 and U-19
primer from the vector. The sequence was neither AtLUP1 nor AtCAS1, having 65% similarity with AtCAS1 and 45% similarity with AtLUP1.

*Kpn I* cuts once in the gene and once just 15 bp outside the gene in the T vector multiple cloning site. The 700 bp *Kpn I* fragment was excised from the LH1.0 plasmid and used as a template to synthesize $^{32}$P radiolabeled probe with random 12-mers, (see Materials and Methods). A total of $3 \times 10^5$ colonies from *A. thaliana* young seedling cDNA library$^{116}$ were screened. After two rounds of standard hybridization (details described in the Materials and Methods), four positive colonies were found and were further checked by PCR. The degenerate primers OS13 and OS10 were used and the PCR conditions and program were same as PCR amplification of the cDNA library. The clone gave the right size PCR fragment and had a 2.4 kbp insert by *Not I* digestion, and was named LH1.1. Because the cDNA library was constructed in the pFL61 yeast expression vector, LH1.1 was subcloned into pBluescript KS (+) for making sequencing constructs.

2.2.7. LH1 sequencing subclones

The *Not I* fragment of LH1.1 was subcloned into CIP-treated *Not I* linearized pBluescript KS (+) to give LH1.2 and LH1.2R (reverse direction insert of LH1.2). A series of subclones were made from LH1.2 and LH1.2R for sequencing the top and bottom strands. LH1.2 (10 μg) was digested in parallel with *EcoR V*, *Kpn I*, *Sac I*, *Hind III* and *EcoR I* respectively to give a large and a small fragment. The large fragment was gel-purified and recircularized to give LH1.3, LH1.4, LH1.5, LH1.7, and LH1.8 for sequencing. The small fragment of *Sac I* digestion was subcloned into CIP-treated *Sac I*
linearized pBluescript KS (+) to give LH1.6. LH1.2R (10 μg) was digested with EcoRI, HindIII and EcoRI respectively to give large and a small fragments. The EcoRI and EcoRI V large fragments of LH1.2R were reclosed to give LH1.14 and LH1.15. Gel purified LH1.12R EcoRI, EcoRI V and HindIII small fragment were subcloned into CIP-treated and corresponding restriction enzyme linearized pBluescript KS (+) to give LH1.11, LH1.12 and LH1.13 for sequencing. By making all these constructs, the LH1.2 was fully sequenced on both strands. The sequence data showed that there were two extra methionine sites before the predicted start site. SphI cut once in front of the predicted start site and SmaI cut once on the vector, and SphI and SmaI double digestion removed the two extra methionines. LH1.2 (4.5 μg) was digested with SphI and SmaI in NEB 4 buffer supplied with BSA (100 μg/mL) in 50 μL of total volume. After incubating at 37 °C for 2 h, 6.0 μL of 10 mM ZnSO4, 1.0 μL of 10 × NEB 4 with BSA, 1.5 μL of water and 1.5 μL (3 units/μL) Mung Bean Nuclease (NEB) were added to the above reaction and the mixture was incubated at 30 °C for another hour. The Mung Bean Nuclease was used to blunt the SphI sticky end and was deactivated by adding 0.6 μL of 1% SDS solution. The DNA was gel purified and the intramolecular blunt end ligation gave LH1.16.

2.2.8. Searching for a full-length LH1 cDNA

Sequence alignment of LH1.2 with AtCAS1 and AtLUP1 cDNAs showed that the second exon of LH1.2 was missing. A pair of primers LH1F1: CCATAATTTCGTCATGC and LH1R1: GGAATAGTGATAGAGCTCT, which covered 800 bp beginning part of LH1.2, including the missing exon region, was used to
PCR amplify *A. thaliana* cDNA library to search for the gene with missing exon. The PCR program was CAPS and the PCR conditions were as follows: 200 ng of cDNA library, 5.0 μL of 10 × PC2, 4.0 μL dNTPs (2.5 mM each), 1.0 μL of each primers, 0.5 μL of *Taq* polymers and water to 50 μL. No PCR fragment larger than 800 bp was found. The same pair of primers was used to amplify genomic DNA using hot start PCR program and the PCR conditions were same as before, except 2.0 μL of genomic DNA was used. A PCR fragment (2.3 kbp) was obtained and subcloned into pBluescript KS(+) T-vector. The resultant plasmid was named LH1.18 and sequenced by primer walking (LH1.18S1: CGCATATTGGTACAGCAG, LH1.18S2: CTGTAAGGCAGGATTAA, LH1.18S3: ACAAGATGGGTTTTGGCC. This genomic DNA did show some degree of homology with AtCAS1 and AtLUP1 second exon, however there is no obvious splicing site. Another pair of primer was designed flanking the open reading frame of LH1.2 to PCR amplify the *A. thaliana* cDNA library. The primers were LH1.2F1: ACGGCATTGAATATCACA and LH1.2R1: GGAGACGCGAAAAAGATT. The PCR program was CAPS, the PCR conditions were same as using LH1F1 and LH1R1 to amplify the cDNA library. No band longer than LH1.2 was amplified.

### 2.2.9. Expression of LH1 in yeast

LH1.1 and LH1.16 derivatives were expressed in yeast strain SMY8.98 LH1.1 was in yeast expression vector pFL61116 and under a constitutive phosphoglycerate kinase (PGK) promoter. LH1.1 was transformed into SMY8 and grown in 5 mL of rich media (YPD) supplied with heme (13 mg/L), ergosterol (20 mg/L) and Tween 80 (5 mg/L). After the cells were saturated, the cells were harvest and the cell pellet was
resuspended in 2 volume of 0.5 M pH 6.2 sodium phosphate buffer and 20 x oxidosqualene (details see Materials and Methods). The cells were lysed with acid washed glass beads and the enzymatic reaction was incubated overnight at room temperature. The product was checked by TLC. Nothing was found from this enzyme assay. LH1.16 was constructed in pRS316GAL and pRS426GAL galactose inducible yeast expression vector to give LH1.19 and LH1.20. LH1.19 and LH1.20 were transformed into SMY8 and assayed as before. No enzymatic product was detected by TLC from this assay. In order to get rid of the extra stop codon before initial methionine, LH1.16 was mutated into LH1.21 by inserting a Sal I site before initial methionine by site directed mutagenesis. The mutagenesis primer was TAACCTTTAACCTCCACATGTGACCTGTATCTATCTCTCCCT (underlined was Sal I). LH1.21 Sal I and Not I fragment was subcloned into the same sites of pRS426GAL to give LH1.22. LH1.22 was expressed in LH2 and assayed in different pHs. Potassium acetate buffer (0.1 M) covered the pH range from 3.6 to 5.6, and 0.1 M potassium phosphate buffer covered the pH range from 5.8 to 7.8. The assay was set up at 0.4 pH range apart and no enzyme activity was detected from TLC in any of the assays.

2.2.10. Cloning ORF3 (LH2) from an A. thaliana cDNA library

The genomic sequence of gene clusters ORF3/LUP1 was obtained from Genbank Arabidopsis thaliana sequencing project. Introns were identified by comparing the AtLUP1 cDNA sequence with its genomic DNA, and the ORF3 cDNA sequence was obtained by conceptually splicing the introns. One pair of specific primer with one
restriction enzyme on each primer was designed flanking the open reading frame of ORF3 cDNA. ORF3F1: GTCGACTACATAGCTTAGACCATCGTTATG (Sal I underlined), ORF3R1: GCGGCGCGTATTAAGAAAGAGCTAAAGATCTTG (Not I underlined). PCR was used to amplify Not I linearized A. thaliana young seedling cDNA library using these two specific primers. The PCR program was CAPS and the conditions were as follows: 200 ng of cDNA library, 5.0 µL of 10 x PC2, 4.0 µL dNTPs (2.5 mM each), 1.0 µL of each primer, 0.5 µL of Taq polymerase and water to 50 µL. A clear band was obtained at the expected size (~ 2.4 kbp). Mapping the PCR fragment with EcoR I, Hind III, Nde I, Pvu II and Xho I showed that was the ORF3 gene. This PCR fragment was gel purified and subcloned into pBluescript KS(+) T-vector. One out of twelve minipreps had an insert and was named LH2.0. The 700 bp Sph I fragment of LH2.0 was used as a template to synthesize the radiolabeled probe for hybridization. A total of 3 x 10^5 colonies was screened by hybridization, one positive clone (LH2.1) was obtained by two rounds of standard hybridization as described in the Materials and Methods. The LH2.1 insert was conformed as ORF3 by restriction mapping and sequencing. The 1.1 kbp Not I fragment of LH2.1 was subcloned into CIP treated Not I linearized pBluescript KS(+) to give LH2.2.

A more accurate enzyme was used to PCR amplify the cDNA library. An Expand High Fidelity PCR System (Boehringer Mannheim) was used according to the manufacturer's instructions as follows: Mixture one [200 ng cDNA library, 4.0 µL of dNTPs (2.5 mM each), 40 pmol of each primer (ORF3F1 & ORF3R1) and water up to 25 µL] was added to mixture two [5.0 µL of 10 x buffer with 15 mM MgCl2, 0.75 µL of polymerases (3.5 units/µL) and up to 25 µL of water] before the PCR started. The PCR
program was CAPS. A 2.3 kbp band was obtained from this PCR. The polymerase was a mixture of Taq and Pwo, and Pwo had 3’ to 5’ proofreading activity, the DNA amplified from this enzyme was a mixture of 3’ single A overhang products and blunt ended products. The PCR fragment was gel purified and A-tailed as described in the Materials and Methods. The resultant DNA was subcloned into pBluescript KS(+) T-vector. Eight out of twenty four miniprep DNA had inserts and two of them were named LH2.01 and LH2.02.

2.2.11. LH2 sequencing subclones

LH2.01 was digested with EcoR I, Xho I and Sph I and gel purified the both fragments of each digestion. The big fragment of EcoR I digestion was reclosed to give LH2.3 and the small EcoR I fragment was subcloned into pBluescript KS(+) to give LH2.4. The LH2.01 Sph I fragment was subcloned into CIP-treated Sph I linearized pUC-19 to give LH2.5. The big fragment of Xho I digestion was reclosed to give LH2.6. The EcoR I fragment of LH2.02 was subcloned into pBluescript KS (+) to give LH2.7 and the Xho I fragment was dropped out from LH2.7 to give LH2.8. The big EcoR I fragment of LH2.02 was reclosed to give LH2.9. All the LH2.01 and LH2.02 subclones were sequenced using vector-derived primers. Comparison of the LH2.01 sequence with ORF3 from Genbank showed that three mistakes at the amino acid level and one deletion from PCR. The two mistakes and one deletion region of LH2.01 was replaced by good sequence from LH2.02 (Ndel and EcoR I fragment). The LH2.3 Nde I and EcoR I fragment was replaced with LH2.8 Nde I and EcoR I (from LH2.02) fragment to give LH2.10. The EcoR I fragment of LH2.01 was inserted into EcoR I linearized LH2.10 to
give LH2.11. The last mistake was removed by site directed mutagenesis. The primer CTGTACAATCTGAGACTTGCCATCCATGGTCTCGATCGGACAGAGT (underlined was Nco I) was used to mutate LH2.11 to give LH2.14, which matched perfectly with ORF3 sequence from Genbank except three subspecies differences (comparing LH2.2 with the known genomic sequence from Genbank).

2.2.12. Expression of LH2.14 in yeast

LH2.14 Not I and Sal I fragment was subcloned into pRS426GAL to give LH2.16. LH2.16 was transformed into SMY8 and LHY2. The yeast cells were prepared as described as in the Materials and Methods section for small scale enzyme assay. The enzyme activity was assayed at different pHs. Potassium acetate buffer (0.1 M) covered the pH range from 3.6 to 5.6, and 0.1 M potassium phosphate buffer covered the pH range from 5.8 to 7.8 (all were 0.4 pH units apart). No enzyme activity was detected from TLC in any of the assays. The LH2.14 Not I-Sal I fragment was subcloned into pRS416GPD/Not I and pRS426GPD/Not I with Fast-Link ligation kit to give LH2.17 and LH2.18 and these two plasmids were transformed into LHY2 for expression. The enzyme activity was assayed as before under different pH condition and no enzyme activity was detected by TLC. The gene in a high copy vector gave smaller transformants than the gene in the low copy vector. After restreaking the transformants onto SCDHET-Ura plate, the yeast cells with high copy empty vector pRS426GPD showed as slow of growth as LH2.18. LH2.14 (10 μg) was sequentially digested with Sal I and Spe I and subcloned into the Xho I and Nhe I sites of the pESC-LEU (Stratagene) yeast expression vector. The plasmid with insert was identified by PCR amplifying the E. coli colonies
with primers ORF3S1: GGCAGAGGCATAACTTAC and ORF3S4: CGGCATGAATCAAACCCA. The PCR program was CAPS and the PCR conditions were as follows: an *E. coli* colony was picked up using a toothpick, which was dipped into 39 μL of water, and the other PCR reagents including 5 μL of 10 × PC2 buffer, 4.0 μL of dNTPs (2.5 mM each), 1.0 μL of each primers and 0.5 μL of *Taq* polymerase were added. The plasmid that gave an appropriately sized PCR fragment at 1720 bp was named LH2.19, which was driven by the *GAL1* promoter and followed by the *CYC1* terminator. LH2.19 and pESC-LEU were transformed into yeast strain LHY3 and LHY3[pSM60.21] was the positive control. One colony from each construct was inoculated into 5 mL of SCDHET-Leu media. The culture was grown to saturation, then 2.5 mL of the culture was collected and the pellet was resuspended into 31 mL of SCGHET-Leu. The cells were induced for two days at 30 °C and the cells were harvested to give about 400 mg of pellet each. The cell pellets were resuspended into 2 volumes of 0.1 M pH 6.4 sodium phosphate buffer and 20 × oxidosqualene stock (see the Materials and Methods). An equal volume of glass beads was added and the mixture was vortexed twice on high for 2 min and chilled on ice in between. The enzymatic reactions were incubated at room temperature overnight and then extracted with a mixture of ethanol, hexane, and ether (2:4:4). The organic layer was dried over sodium sulfate and concentrated under a nitrogen stream. The crude enzyme product was derivatized as the TMS-ether (see Materials and Methods section). The TMS-ether derivative was analyzed by GC using the following program: HP series 6890; column: Rtx-5, 30 m × 0.25 mm I.D., 0.1 mm df; oven temperature was 280 °C, isothermal; injector was at 280 °C; detector was at 290 °C; 1/40 split ratio; helium carrier gas; column head pressure: 13 psi.
The positive control LHY3[pSM60.21] showed a peak at retention time of 12.90 min, and nothing was detected from LHY3[LH2.19] by comparing its GC spectrum with negative control LHY3[pESC-LEU], indicating that LH2.19 encoded enzyme was not active on the substrate oidosqualene.

2.2.13. Cloning ORF1 (LH5) using RT-PCR

ORF1 was found in the Genbank A. thaliana sequencing project using a BLAST search with AtCAS1. A pair of primers with Sal I and Not I on each end was designed to amplify this gene from a cDNA library. ORF1F1: GTCGACTACGTAGCTACATTTTTGTAATG (Sal I underlined) and ORF1R2: GCGGCCGCAGTTATTAGAAAAGTGTCGTA (Not I underlined). Nothing was amplified from the cDNA library using this pair of primers in various conditions. The cDNA made from total RNA (see details in the cloning sesquiterpene synthase from A. thaliana) was therefore used as a template to amplify ORF1. The expand high fidelity PCR system was used with the same PCR conditions and program as for PCR-amplifying ORF3 from the cDNA library. One clear band was obtained at the expected size (~ 2.4 kbp). The PCR fragment was gel purified, A-tailed, and the resultant DNA was subcloned into pBluescript KS(+) T-vector. Five colonies having ORF1 (LH5) insert were identified by digestion miniprep DNA with Pvu II. LH5.02 was fully sequenced by primer walking (ORF1S1: CCGGTGAAAGTTGAAGAC, ORF1S2: AGCCTTTCTTGACGTT, ORF1S3: GCCATCCACAAGTAAATCG, ORF1S4: AGAGAGGTAGCTTTCTCC, ORF1S5: CTTCTTTCTTCCGTCG, ORF1S6: TTAAGTGCTGCCTGCTG, ORF1S7: GACTGTGAGGCAATACGC, ORF1S8:
AAACACCAGTGGCGGAAG, ORF1S9: GCAAGTAAACCATCGGC, ORF1S10: TCATACTCCGTTTTTGC). There were three mistakes at the amino acid level (four at the DNA level). LH5.03 was partially sequenced and the three mistakes on LH5.02 were repaired using BstEI-II-NdeI fragment and the EcoR V-NotI fragment of LH5.03. Both LH5.02 and LH5.03 were digested with BstEI II and NdeI I, and the big fragment of LH5.02 and the small fragment of LH5.03 were gel purified and ligated together to give LH5.1. LH5.1 and LH5.03 were digested with EcoRI I and PstI I, the big fragment of LH5.1 and the small fragment of LH5.03 were gel purified and ligated together to give LH5.2, which encodes the same protein as the conceptually spliced ORF1 genomic sequence.

2.2.14. Expression of LH5.2 in yeast

The LH5.2 NotI-SalI fragment was subcloned into pRS316GAL, pRS326GAL, pRS426GAL to give LH5.4, LH5.6, and LH5.7, respectively. Chris Neumann transformed LH5.4 into SMY8 and assayed the enzyme activity in a small scale as described in the Materials and Methods section and no enzyme activity was detected by TLC. LH5.6 was transformed into SMY8 and plated on SCDHET-Ura. Several colonies were grown on an SCDHET-Ura plate and restreaked onto SCGHET-Ura plate to induce the gene. After six days incubation at 30 °C, SMY8[LH5.6] did not grow on the galactose plate, indicating that LH5.6 produced a toxic compound in vivo that killed yeast cells. After the yeast erg9 and erg7 double mutant strain was made, LH5.7 was transformed into SMY8 and LHY2. The yeast cells were prepared as described in the Materials and Methods section for small scale enzyme assay. The enzyme activity was
assayed at different pHs. Potassium acetate buffer (0.1 M) covered the pH from 3.6 to 5.6, and 0.1 M potassium phosphate buffer covered the pH from 5.8 to 7.8. Assays were set up at 0.4 pH units apart and no enzyme activity was detected by TLC in any of the assays. The LH5.2 Not I-Sal I fragment was subcloned into pRS416GPD/Not I and pRS426GPD/Not I with Fast-Link ligation kit to give LH5.8 and LH5.9. The enzyme activity was assayed as before at different pHs and no enzyme activity was detected by TLC. The gene in a high copy vector gave smaller transformants than gene in low copy vector. After restreaking the transformants onto SCDHET-Ura plate, the yeast cells with high copy empty vector pRS426GPD showed the same slow growth as LH5.9. LH5.2 (10 μg) was sequentially digested with Sal I and Spe I and subcloned into the Xho I and Nhe I sites of pESC-LEU yeast expression vector. The plasmid with insert was identified by PCR amplifying the E. coli. colonies with sequencing primers ORF1S1 and ORF1S4 using the PCR program and conditions used for confirming LH2.19. The colony gave a appropriately sized PCR fragment at 1750 bp was named LH5.10. The LH5 insert in pESC-LEU was driven by GAL1 promoter and followed by the CYC1 terminator. LH5.10 and pESC-LEU were transformed into yeast strain LHY3 and LHY3[pSM60.21] was the positive control. One colony from each construct was inoculated into 5 mL of SCDHET-Leu media to saturation, and then 2.5 mL of the culture was collected and the pellet was resuspended into 31 mL of SCGHET-Leu. The cells were induced for two days at 30 °C and the cells were harvested to give about 400 mg of pellet. The cell pellets were suspended into 2 volumes of 0.1 M pH 6.4 sodium phosphate buffer and 20 × oxidosqualene stock (see Materials and Methods). Glass beads were added and the mixture were vortexed twice on high for 2 min and chilled on ice in between. The
enzymatic reactions were incubated at room temperature overnight and then extracted
with a mix of ethanol, hexane, and ether (2 : 4 : 4). The organic layer was dried over
sodium sulfate and concentrated under a nitrogen stream. The crude enzyme product was
derivatized as the TMS-ether as described in the Materials and Methods section. The
TMS-ether derivative was analyzed by GC using same program as analyzing of LH2.19.
The positive control LHY3[SM60.21] showed a peak at retention time of 12.90 min and
nothing was detected from LHY3[LH5.10] by comparing its GC spectrum with negative
control LHY3[pESC-LEU], indicating that LH5.10 encoded enzyme was not active on
the substrate oxidosqualene. Squalene was fed to both LHY2[LH5.8] and LHY2[LH5.9]
cell lysates and no enzyme activity was detected by TLC.
Chapter 3: Results and Discussion

2.3.1. Construction of the squalene synthase deletion plasmid

In order to abolish \textit{in vivo} oxidosqualene biosynthesis, the yeast squalene synthase 
gene was deleted by homologous recombination.\textsuperscript{117} The 200 bp N-terminus of squalene 
synthase was deleted with the \textit{HIS3} gene. The yeast squalene synthase (\textit{ERG9}) 
gene\textsuperscript{118,119} is 1.3 kbp, and the 1.1 kbp C-terminus of this gene was PCR-amplified with 
a forward primer flanked with an \textit{Acc} I site and a reverse primer flanked with \textit{Sph} I and 
\textit{Not} I sites. These restriction sites are intentionally inserted for the downstream 
subcloning. The 1.1 kbp PCR fragment was subcloned into pGEM-T vector to give 
LH7.0. pGEM-T (Promega) is a T-vector that has no \textit{BamH} I site on its multiple cloning 
site. In order to gain the \textit{BamH} I site, the LH7.0 \textit{Acc} I fragment was subcloned into CIP-
treated pBluescript KS(+) digested with the same enzyme, and a clone with \textit{BamH} I site 
upstream of the gene was found by restriction-mapping and named LH7.1. EH7.1 (from 
Elizabeth Hart) is the \textit{ERG9} 5’-nontranslated region subcloned in pGEM-T vector. The 
\textit{BamH} I/\textit{Sph} I fragment of LH7.1 was subcloned into EH7.1 digested with the same 
enzymes to give LH7.2, which is the \textit{ERG9} 5’-nontranslated region adjacent with the 1.1 
kbp C-terminus of the \textit{ERG9} gene. The selective marker \textit{HIS3} was subcloned into LH7.2 
between the 5’ nontranslated region and the \textit{ERG9} 3’ end using \textit{BamH} I and \textit{Acc} I 
restriction sites to give LH7.3. The detailed cloning of LH7.3 is illustrated in Figure 15.
Figure 15. Construction of plasmid LH7.3, the highlighted restriction enzymes are the ones used for subcloning.
2.3.2. Yeast homologous recombination

The Not I fragment of LH7.3 containing ERG9 5’-nontranslated region and 3’ coding sequence interrupted by the HIS3 selective marker was released from LH7.3 by digestion. The yeast strain SMY4 (MATa hem1::TRP1 ura3-52 trpl-Δ63 leu2-3,112 his3-Δ200 ade2 Gal′), a hem1 mutant strain was transformed with the LH7.3 Not I fragment. The transformation was selected for HIS3 integrants, and one hundred original transformants were further screened for ergosterol auxotrophy. Only three out of one hundred transformants remained capable of growing without ergosterol, indicating that homologous recombination between the deleted squalene synthase and yeast wild-type squalene synthase happened at a 97% rate. This yeast strain, named LHY1 has to import lanosterol or ergosterol for survival.

2.3.3. Yeast erg9 and erg7 double mutant

Even though sterol biosynthesis is disrupted by a squalene synthase mutation, the lanosterol synthase (ERG7) still functions in this yeast strain. The ERG7 gene was knocked out by homologous recombination. LHY1 was transformed with Pst I digested JR3.1, which is the ERG7 gene disrupted with a LEU2 marker. The transformants were selected for LEU2 integrants and colonies with the LEU2 gene were further checked for lanosterol synthase activity by TLC. Some yeast lysates were fed oxidosqualene in vitro and no lanosterol synthase activity was detected, indicating that yeast clone had both ERG9 and ERG7 genes deleted. The resultant yeast strain LHY2 with no in vivo substrate oxidosqualene for triterpene synthases and no lanosterol synthase, is an ideal
yeast strain for the expression of recombinant triterpene synthases. Figure 16 shows the sterol biosynthesis in engineered yeast strain LHY2.

\[
\begin{align*}
2 \text{ FPP} & \xrightarrow{-} \text{squalene} \xrightarrow{ERG1} \text{oxidosqualene} \xrightarrow{-} \text{lanosterol} \xrightarrow{-} \text{ergosterol} \\
\text{erg9}\Delta & \\
\text{erg7}\Delta \\
\end{align*}
\]

Figure 16. Sterol biosynthesis in LHY2

2.3.4. Making LHY2 derivatives

The yeast strains used in this work are derived from JBY575,\textsuperscript{98} which has four selectable markers: \textit{his3}, \textit{leu2}, \textit{trp1}, and \textit{ura3}. In the LHY2 strain, \textit{hem1} was deleted with \textit{TRP1}, \textit{erg9} was deleted with \textit{HIS3} and \textit{erg7} was deleted with \textit{LEU2}. Consequently LHY2 only can be used to express plasmids with the \textit{URA3} selectable marker. Any plasmid bearing \textit{HIS3}, \textit{LEU2}, or \textit{TRP1} selectable markers cannot be expressed in LHY2, because the host strain carries these markers. The lack of a \textit{LEU2} marker is particularly unfortunate because the \textit{LEU2}-marked pRS305GAL is convenient for integrating genes for expression. The \textit{LEU2} marker in LHY2 was consequently recovered with the \textit{LEU2} blaster.\textsuperscript{111} The \textit{LEU2} blaster is a yeast \textit{URA3} gene flanked by 1.1 kbp direct repeats of a bacterial sequence \textit{hisG} replacing partial \textit{LEU2} sequence. Homologous recombination occurs between \textit{LEU2} (at the \textit{erg7} locus) and the \textit{LEU2} blaster (\textit{Bgl II} fragment of pNKY85) to replace the \textit{LEU2} marker with \textit{URA3} marker. LHY2 transformed with the \textit{LEU2} blaster was selected for growth in the absence of uracil. Twenty four of the
original transformants were restreaked on the synthetic media supplied with heme and ergosterol in the absence of uracil or leucine. Two out of twenty-four colonies survived on the SCD-Ura HET plate and died on the SCD-Leu HET plate, indicating that only about 10% of the LEU2 blaster integrated at the LEU2 marker. One of these yeast colonies was named LHY3. LHY2 has two LEU2-derived sequences: the *leu2-3,112* allele at the LEU2 locus, and the functional LEU2 gene used to disrupt *ERG7*. The LEU2 blaster probably integrated most frequently at the *leu2-3,112* locus.

The *URA3* marker in LHY3 was removed by selection for 5-fluoroorotic acid (5-FOA) resistance. The URA3 enzyme converts 5-FOA to 5-fluoro-UMP, which is toxic to yeast. The fragment used to disrupt LEU2 gene consists of URA3 gene flanked by two repeats of hisG sequence. When the yeast grows in the presence of FOA, the two repeats of hisG sequence undergo mitotic recombination to eliminate the URA3 gene and leave behind a single copy of the hisG disrupting the *leu2* gene. LHY3 was inoculated in the SCDHET liquid media and plated on SCDHET plate supplied with 1 mg/mL FOA. The colonies that survived from this plate lost the URA3 gene and left the LEU2 marker disrupted with one copy of hisG sequence. The recombination frequency between hisG is $10^4$. The resultant yeast strain LHY4 (*MATα erg7::leu2::hisG erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal°") can be used to express recombinant DNA constructed in yeast vector with URA3 and/or LEU2 as selective markers.
2.3.5. Application of yeast double knockout strains

Bridget M. Joubert used LHY3 to express the *S. cerevisiae* lanosterol synthase mutants Val454X,\(^{120}\) where X represents phenylalanine, leucine, isoleucine, alanine and glycine. While the Val454Phe mutant inactivates the enzyme, the Val454Leu, and Val454Ile mutants act like wild-type lanosterol synthase and make 100% lanosterol. The Val454Ala and Val454Gly mutants make lanosterol and the monocyclic triterpene achilleol A. When Val454Ala and Val454Gly mutants were expressed in SMY8, before adding substrate, the yeast only accumulates achilleol A. Lanosterol made from the *in vivo* substrate oxidosqualene was metabolized to ergosterol by yeast. In order to accurately quantitate the enzyme products ratio of these two lanosterol synthase mutants, Ms. Joubert expressed these two mutant enzymes in a yeast strain lacking the *in vivo* substrate oxidosqualene, and the yeast imported sterol from the medium instead of making its own sterol from lanosterol. The Val454Ala mutant expressed in LHY3 gave a 95:5 ratio of lanosterol to achilleol A instead of 91:9 as in SMY8, and the Val454Gly expressed in LHY3 gave a ratio of 83:17 of lanosterol to achilleol A. The different ratios observed from SMY8 and LHY3 incubations show that yeast metabolizes lanosterol more rapidly than achilleol A.

The yeast strain LHY4 was been used to express a cycloartenol synthase (CAS1) mutant by Elizabeth A. Hart.\(^{121}\) The CAS1Ile481Val mutant converts oxidosqualene into a mixture of cycloartenol, lanosterol and parkeol. Lanosterol is the sterol precursor in yeast, and yeast consequently has enzymes dedicated to its metabolism. However, cycloartenol and parkeol are not normal yeast components, and might be metabolized less efficiently by yeast. Because of this complication, this CAS1 mutant had to be expressed
in a yeast strain without \textit{in vivo} oxidosqualene and no lanosterol synthase activity. The CAS1lle481Val mutant expressed in LHY4 produces cycloartenol, lanosterol and parkeol in relative ratios of 54:25:21 instead of 52:26:22 in SMY8. The enzyme product ratio difference between the CAS1lle481Val mutant expressed in SMY8 and LHY4 is negligible, indicating that not only lanosterol but also cycloartenol and parkeol are metabolized by yeast.

2.3.6. Cloning LH1, a putative triterpene cyclase from \textit{A. thaliana}

Our goal was to clone more triterpene cyclase genes from plants for mechanistic study of this gene family. \textit{A. thaliana} is a small plant and its genome will be completely sequenced by the end of this year. Since we had easy access to an \textit{A. thaliana} cDNA library and plant tissues, and two oxidosqualene cyclases have been cloned from this organism,\textsuperscript{85,86} we chose \textit{A. thaliana} as the model plant to clone more triterpene synthases. At the time I started this project, only five lanosterol synthases, one cycloartenol synthase and one lupeol synthase had been cloned from various organisms. The genes from plant origin were the \textit{A. thaliana} cycloartenol synthase\textsuperscript{85} and lupeol synthase\textsuperscript{86}. Conserved residues were identified from sequence alignment of lanosterol synthase from \textit{Candida albicans}\textsuperscript{103,122} \textit{Saccharomyces cerevisiae},\textsuperscript{96,123} \textit{Schizosaccharomyces pombe},\textsuperscript{98} \textit{rat}\textsuperscript{104,124} and human\textsuperscript{99,125} cycloartenol synthase\textsuperscript{85} and lupeol synthase\textsuperscript{86} from \textit{A. thaliana}. Several degenerate primers were designed and an \textit{A. thaliana} ecotype Landsberg young seedling cDNA library\textsuperscript{116} was chosen as the template for PCR amplification. One pair of primers, OS13 and OS10, gave a PCR fragment of the expected size and the PCR product was subcloned into pT7Blue T vector.
Using these two primers, CAS1 and LUPI were also amplified from the cDNA library. Extensive mapping of the miniprep DNA showed that one plasmid bore a insert which encoded neither cycloartenol synthase nor lupeol synthase. The insert of that plasmid was used as a template to hybridize to the cDNA library\textsuperscript{116} and one positive clone (LH1.1) was obtained. The library was constructed in the pFL61 yeast expression vector, and Not I digestion of LH1.1 gave a 2.4 kbp insert. The Not I fragment of LH1.1 was subcloned into pBluescript KS(+) to give LH1.2 and LH1.2R. LH1.2 was sequenced from both strands by making a series subclones of LH1.2 and LH1.2R. A comparison of the sequence of LH1.2 with cycloartenol synthase and lupeol synthase from same organism (Figure 17) shows that LH1.2 aligns well with cycloartenol synthase and lupeol synthase except that it lacks 62 amino acids corresponding to the second exon. It has 60.3% identity to cycloartenol synthase and 48.8% identity to lupeol synthase.
Figure 17. Sequence alignment of two oxidosqualene cyclases from *A. thaliana* with LH1.2, a putative cyclase. Sequences were aligned with megalign (DNASTAR) using the clustal method. The dash line shows the second missing exon of LH1.2.
2.3.7. Expression of cloned LH1 in yeast

There are two methionines followed by two stop codon before the initial methionine of the open reading frame. The extra methionines were removed and the insert was subcloned into pRS316GAL\textsuperscript{113} and pRS426GAL, which are galactose-inducible yeast expression vectors. The resultant plasmids were expressed in yeast strain SMY8 and assayed for oxidosqualene cyclase activity. No triterpene alcohol was detected by TLC. The extra two stop codons before the initial methionine were removed from the 5' of the open reading frame of LH1.16 by site-directed mutagenesis, and a Sal I site was inserted before the initial methionine. The resultant plasmid was named LH1.21 and its Sal I and Not I fragment was subcloned into yeast expression vector pRS426GAL and expressed in yeast strain LHY2. The oxidosqualene cyclase activity was assayed over a pH ranges, from acidic to basic, and triterpene was detected by TLC. Both pRS426GAL and pRS316GAL have been used to express cycloartenol synthase and lupeol synthase as the active form.\textsuperscript{85,86}

2.3.8. Searching for the missing exon

It is possible this cloned cDNA is not an oxidosqualene cyclase because the RNA was spliced alternatively for this particular cDNA. It is possible that the cloned cDNA is derived from a nonfunctional gene, the second exon was spliced by evolution. It is also possible that the cloned cDNA accidentally lost its second exon by missplicing. In \textit{Arabidopsis}, introns usually start with GTXX and ends with XXAG, if either G is mutated, the splicesome could miss one intron splicing site and go to the next splicing site to give a mRNA template which has the second exon missing. An effort to search the
missing exon was made by PCR amplification of the same cDNA library and the genomic DNA. A pair of primers flanking the missing exon region was designed. These would give a 800 bp of PCR fragment if LH1.2 was the template. If the library has the cDNA with missing exon, it will give about a 1.0 kbp PCR fragment. Extensive PCR amplification of the same cDNA library,\textsuperscript{116} did not find a fragment bigger than 800 bp. This result indicates that no full length LH1 cDNA exists in this particular cDNA library. It is possible that a cDNA library made from older plants might have the full length LH1 gene. The same pair of primers was also used for PCR amplification of the genomic DNA and a 2.3 kbp fragment was obtained. This PCR fragment was subcloned into a T vector and sequenced by primer walking. The sequence analysis of the genomic DNA (LH1.18) showed homology with the second exon of cycloartenol synthase and lupeol synthase. However, no obvious splicing sites before and after the putative second exon were found in the genomic DNA. The DNA and the amino sequences of LH1.2 cDNA and the genomic DNA are shown in Figure 18 and 19, respectively.
Figure 18: cDNA and protein sequences of LH1.2, the underline are the sequence of the primer pair used to PCR amplify the cDNA library for searching the missing exon
Figure 19. Genomic DNA and amino acid sequences of LH1.18. The bold amino acid sequence indicates the putative second missing exon of LH1.2
2.3.9. Cloning of ORF3 from an *A. thaliana* cDNA library

The ORF3 sequence was found in the Genbank *A. thaliana* ecotype Columbia sequencing project (gene YuP8H12R.43 on Accession No. AC002986) and the cDNA was PCR amplified from the *A. thaliana* ecotype Landsberg young seeding library. ORF3 is the middle gene of a three gene cluster on chromosome 1 that also includes lupeol synthase and another apparent oxidosqualene cyclase gene. Potential errors introduced by amplifying with Taq polymerase precluded expressing the PCR product directly. When the PCR fragment was subcloned into pBluescript T-vector and the 700 bp Sph I fragment from the 3' end of the gene was used as a probe to hybridize the cDNA library, only a 1.1 kbp gene fragment was obtained from the cDNA library. Using a probe close to the N-terminus instead of the C-terminus would have ensured that more complete clones would have been obtained. Cloning only a short fragment suggests that the cDNA is rare in tissue from which the cDNA was prepared. High fidelity polymerase was consequently used to obtain a PCR product with fewer errors. The PCR product was subcloned into a T-vector and the cDNA from two colonies was sequenced by making a series of subclones. Comparing the cDNA sequence from Genbank with the cDNA sequence obtained from one clone, shows one deletion and three mutations at the amino acid level. The two mutations and one deletion were removed by making chimerae with the other clone, and the last mutation was removed by site-directed mutagenesis. The resultant cDNA was sequenced and still has three amino acids different from the sequence from Genbank, however it matches the 1.1 kbp hybridized gene fragment, indicating that these three mismatches are ecotype differences. The DNA and protein sequences of LH2.14 is shown in Figure 20.
2.3.10. Expression of ORF3 in yeast

The plasmid LH2.14 was subcloned into pRS326GAL and pRS426GAL galactose-inducible yeast expression vectors and the resultant plasmids were expressed in yeast strains SMY8 and LHY2. The oxidosqualene cyclase activity was assayed between a pH range of 3.6 to 7.8 and no enzyme activity was detected.

Neither pRS326GAL nor pRS426GAL has a terminator and ORF3 is an open reading frame that doesn’t have a terminator or a poly-A tail. A terminator is required for efficient transcript termination in yeast.126 pRS416GPD and pRS426GPD115 are yeast expression vectors ("6" stands for the selective marker URA3, "1" stands for the CEN/ABS low copy number plasmid and "2" stands for the 2-micron high copy number plasmid). These two vectors are constitutive expression vectors with glyceraldehyde-3-phosphate dehydrogenase (GPD) promoters and cytochrome-c oxidase (CYC1) terminators. The multiple cloning sites of pRS416GPD and pRS426GPD were not compatible with the restriction sites we used for subcloning. A polylinker with three new sites (Bgl II, Nde I and Not I) was inserted into the existing multiple cloning site to give pRS416GPD/Not I and pRS426GPD/Not I. A LH2.14 Not I and Sal I fragment was subcloned into pRS416GPD/Not I and pRS426GPD/Not I to give LH2.17 and LH2.18 and the resultant plasmids were transformed into yeast strain LHY2. Oxidosqualene cyclase activity was assayed by feeding the cell lysate with substrate oxidosqualene, no triterpene alcohol was detected by TLC. The gene expressed in a high copy plasmid gave smaller transformants than when expressed in the low copy plasmid, but the yeast cells harboring empty plasmid pRS426GPD/Not I also show slow growth.
The LH2.14 Sal I and Spe I fragment was subcloned into yeast expression vector pESC-LEU Xho I and Nhe I sites to give LH2.19. pESC-LEU is a yeast epitope tagging vector. This vector contains the GAL1 and GAL10 yeast promoters in opposing orientation, two multiple cloning sites and the CYC1 terminator downstream of GAL1 and alcohol dehydrogenase (ADH) terminator downstream of GAL10. The epitope tag FLAG (DYKDDDDK) is located in the multiple cloning site downstream of the GAL10 promoter and the c-myc epitope (EQKLISEEDL) is located in the multiple cloning site of GAL1 promoter. The LH2.14 was subcloned under GAL1 promoter and the c-myc epitope is flanked with the N-terminus of the gene. The LH2.19 was expressed in LHY3 and the oxidosqualene cyclase activity was assayed and no triterpene peak was detected from GC by comparing the yeast harboring a empty vector and yeast harboring the CASI cDNA. Although GC is very sensitive, no triterpene alcohol was detected from GC indicating the cloned enzyme is not active or accepts a substrate different from oxidosqualene. Since the gene is on the same cluster as lupeol synthase and is 79.4% identity to lupeol synthase, it is possible that this gene is a lupeol synthase duplication and then evolved to a new function or lost its function as an oxidosqualene cyclase.

2.3.11. Cloning of ORF1 by RT-PCR

Another triterpene synthase-like open reading frame ORF1 was found in the Genbank A. thaliana ecotype Columbia sequencing project. ORF1 has 71.7% identity to ORF3, 71.4% identity to lupeol synthase and 57.8% identity to cycloartenol synthase. A pair of specific primers was designed to PCR amplify the A. thaliana ecotype Landsberg young seedling cDNA library, but no PCR product was obtained. It is possible that
this gene is only expressed in older plants or only exists in the ecotype Columbia. Total RNA from ecotype Columbia was used as template for RT-PCR with high fidelity polymerase. A PCR product of 2.4 kbp was obtained and was subcloned into pBluescript KS(+) T-vector. The cDNA of two clones was sequenced by primer walking. The cDNA sequence of one clone obtained from PCR has three mutations at amino acid level, which were replaced with the same region from the other clone. The resultant plasmid LH5.2 has the same sequence as found by the sequencing project. The cDNA and protein sequence of ORF1 are shown in Figure 21.
Figure 21. cDNA and protein sequences of ORF1.
2.3.12. Expression of ORF1 in yeast

The *Not* I and *Sal* I fragment of LH5.2 was subcloned into pRS316GAL, pRS326GAL, and pRS426GAL galactose-inducible expression vector and the resultant plasmids were expressed in yeast strain SMY8 and LHY2. The oxidosqualene cyclase activity was assayed from pH range 3.6 to 7.8, but no enzyme activity was detected. Since neither pRS326GAL nor pRS426GAL has a terminator, the LH5.2 *Not* I and *Sal* I fragment was also subcloned into the same sites of pRS416GPD/*Not* I and pRS426GPD/*Not* I constitutive yeast expression vectors, which have a *CYC1* terminator. The resultant two constructs were expressed in LHY2 and no oxidosqualene cyclase activity was assayed. No triterpene alcohol was detected by TLC. Similar results were obtained as LH2: when the gene was expressed in the high copy plasmid, it gave smaller transformants than when expressed in the low copy plasmid, indicating that the high copy plasmid is a burden to the yeast cells. The LH5.2 *Sal* I and *Spe* I fragment was also subcloned into *Xho* I and *Nhe* I sites of the pESC-LEU yeast expression vector and the resultant plasmid was expressed in LHY3. The oxidosqualene activity was assayed. No enzyme activity was detected by comparing to the yeast bearing the empty vector and yeast bearing a *CAS1* cDNA. The cloned enzyme is not active or does not accept oxidosqualene as a substrate. Another in frame methionine 9 amino acids upstream from the methionine we chose is located on the ORF1 genomic DNA. We picked the second methionine as the start site, because it is followed by a tryptophan like other oxidosqualene cyclases (Figure 16). It is possible that we picked the wrong start site for this cloned ORF1.
The *A. thaliana* sequencing project is almost complete. From sequence comparisons, we can sort the sequences with homology to families with known function, like oxidosqualene cyclases or sesquiterpene synthases. However, the limited information about cloned genes is insufficient to predict the enzyme product based on the sequence. Although genomic sequencing projects have uncovered more than dozen *A. thaliana* open reading frames with high homology to known oxidosqualene cyclases, the difficulty in finding activity upon expressing these genes raises the possibility that some may encode pseudogenes. Alternatively, though these ORFs have sequence homology with known oxidosqualene cyclases, it is possible that they have evolved to perform similar chemistry on different substrates.
Chapter 4: Conclusions

A yeast strain with an \textit{erg9} and \textit{erg7} double mutation was constructed by homologous recombination. This strain is very useful for expressing oxidosqualene cyclases. This strain has been used to express \textit{A. thaliana} cycloartenol synthase mutants and \textit{S. cerevisiae} lanosterol synthase mutants\textsuperscript{120,121} by other group members. The plasmids expressed in this strain gave more accurate product ratios than when expressed in a lanosterol synthase yeast mutant with a functional squalene synthase.

A cDNA was cloned from \textit{A. thaliana} ecotype Landsberg young seeding cDNA library\textsuperscript{116} The sequence was found to be 60\% identical to cycloartenol synthase and 49\% identical to lupeol synthase. The second exon of this cloned gene is missing compared to the cycloartenol synthase and lupeol synthase from the same organism. Although efforts have been made to express this gene in yeast, no oxidosqualene synthase activity was found. It is possible that this cloned gene is a pseudogene, a mis-spliced cDNA, or a functional gene that uses a different substrate.

An open reading frame (ORF3) adjacent to lupeol synthase on the genomic sequence was cloned from an \textit{A. thaliana} ecotype Landsberg young seedling cDNA library\textsuperscript{116} and expressed in yeast with various expression vectors. No oxidosqualene cyclase activity was detected either by TLC or GC. It is possible that the ORF3 is actually a duplicate of lupeol synthase that lost its function as a lupeol synthase by mutation during evolution. The other possibility is that the cloned gene is active towards a substrate other than oxidosqualene.
Another oxidosqualene cyclase like open reading frame (ORF1) was also found in the *A. thaliana* ecotype Columbia sequencing project. The gene was cloned by RT-PCR from the total RNA prepared from *A. thaliana* ecotype Columbia. The ORF1 was expressed in various yeast expression vector, and the oxidosqualene cyclase activity was assayed. No triterpene alcohol was detected by either TLC or GC.
Materials and Methods

1. Instruments

Nuclear Magnetic Resonance (NMR) spectra were recorded on a Bruker AC250 MHz or 400 MHz NMR spectrometer. Chemical shifts were reported in parts per million downfield from tetramethylsilane. Microcentrifugation was performed using an Eppendorf Centrifuge Model 5415C variable speed microcentrifuge. Centrifugations were performed using a Beckman Model TJ-6 Centrifuge, a Beckman GS-6R Centrifuge or an IEC Centra MP4R Centrifuge. DNA gel electrophoresis was performed using an Owl Scientific EASY-CAST Electrophoresis System. Gels were illuminated using a FisherBiotech Transilluminator FBDL T-88. Protein gel electrophoresis was performed using The Penguin™ water-cooled dual-Gel Electrophoresis system, model P8DS and the gel was cast using the Joey Gel Casting System, model JGC 4 and 0.8 mm thick, 10 well comb. Ultraviolet absorption spectra were measured on a Hewlett Packard 8452A Diode Array Spectrophotometer. Gas Chromatography spectra were obtained on a Hewlett Packard 6890 Series GC System using an Rtx-5 column from Retek Inc. GC-MS were measured on a VG ZAB-HF spectrometer using a DB-5ms column from J&W Scientific Inc. Analytical thin layer chromatography was performed on Merck 60 F254 precoated (0.25 mm) silica gel plates. *E. coli* and yeast were incubated in an incubator shaker series 25 from New Brunswick Scientific Co., Inc. Branson Sonifier Model 450 from Branson Sonic Power Company was used for sonication. A french Pressure Cell Press from SLM Instruments, Inc. was used to break bacteria and yeast cells for large-scale assays.
2. Materials

2.1. Chemicals and reagents

Ammonium molybdate, ascorbic acid, Dowex 50X4-400 resin, tetrabutyl ammonium hydroxide, mesyl chloride, triethylamine, lithium bromide, citric acid, sodium pyrophosphate decahydrate, squalene were from ACROS. Anthranilic acid and ammonium bicarbonate were from Aldrich. Cellulose powder CF11 and cellulose F$_{254}$ TLC plates from Whatman and Merck, respectively. Yeast nitrogen base, peptone, glucose, galactose, and all the amino acids were from FisherBiotech. IPTG was from US Biology and the other reagents were from Sigma. The cDNA library construction kit was from Gibco-BRL. The Qiagen plasmid preparation kit and Qiaex suspension were from Qiagen Inc.

2.2. Enzymes and vectors

Restriction enzymes, mung bean nuclase, exo’ Klenow fragment, T4 DNA kinase and T4 DNA polymerase were from New England BioLabs, Fast-Link DNA ligation kit was purchased from Epicentre Technologies, Taq polymerase was from FisherBiotech, Pfu polymerase was from Stratagene, and expand high fidelity PCR system was from Boehringer Mannheim.

pGEM-T vector was from Promega, pT7Blue T-vector was from Novagen, pESC-LEU was purchased from Stratagene, pET15/Not I was from Dr. Bartel’s plasmid library collection #11 and pGEX-KTO.1 was from Dr. Bartel’s plasmid library collection #250.
2.3. Substrates

A. Synthesis of farnesyl pyrophosphate

**Phosphate ester stain.** Nitric acid (2 mL, 70%) was added to 18 mL methanol containing 200 mg ammonium molybdate, and the mixture was stirred vigorously until the molybdate was completely dissolved. Ascorbic acid (200 mg) and anthranilic acid (200 mg) were added to the above solution and stirred until a clear yellow solution was obtained. This spray reagent was stable only for several hours. The acid molybdate treatment hydrolyzes the phosphate ester to free inorganic phosphate, which then forms the phosphomolybdate heteroanion. Iodine vapor was also used to stain double bonds.

**Synthesis of tris(tetra-n-butyl)ammonium hydrogen pyrophosphate.** The cation exchange column was slurry-packed by mixing Dowex 50X4-400 (36 meq) cation exchange resin (hydrogen form) with deionized water. The column was washed with deionized water until the pH was neutral. Sodium pyrophosphate decahydrate (2 g) was dissolved in 25 mL of deionized water and applied to the column. The pH of the eluent was monitored, and the eluent was collected when the pH became acidic (a few drops after loading the sodium pyrophosphate decahydrate on the resin). Collection was halted when the pH returned to that of deionized water. The solution of pyrophosphoric acid was then immediately titrated to pH 7.3 with tetra-n-butylammonium hydroxide. The salt was dried by lyophilization to yield 4.0 g of a hygroscopic, white solid. The salt was stored under nitrogen in a desiccator.

**Synthesis of farnesyl pyrophosphates.** The reaction was carried out under nitrogen in an oven-dried 50-mL round bottom flask with a magnetic stirring bar. A
solution of tris(tetra-n-butyl)ammonium hydrogen pyrophosphate (4.0 g, 4.5 mmol) in dried acetonitrile (10 mL) was added into a solution of farnesyl chloride (580 mg) in 1.0 mL of dried acetonitrile. The resulting mixture was stirred at room temperature, and the reaction was followed by TLC (developed in CH₂Cl₂) to monitor the diminishing of farnesyl chloride. After 2 h, the solvent was removed by rotary evaporation, and the resulting opaque residue (pale yellow) was dissolved in 6 mL of buffer consisting of isopropanol and 25 mM ammonium bicarbonate in a ratio of 1 : 49 (v/v). Half of the resulting solution was passed through a column of 72 meq Dowex 50X4-400 cation exchange resin. The ammonium form Dowex 50X4-400 cation exchange resin was made by passing 5 volumes of concentrated ammonium hydroxide through hydrogen form resin and equilibrating with two column volumes of 1 : 49 (v/v) isopropanol: 25 mM ammonium bicarbonate ion exchange buffer. The column was eluted with 90 mL of the same buffer. The milky eluent (slightly cloudy) was lyophilized to dryness to yield 0.8 g white solid. The white solid was suspended in 0.1 M ammonium bicarbonate (5 mL) and transferred to a 50 mL Falcon tube. The suspension was treated with 16 mL of a 1 : 1 (v/v) mixture of isopropanol and acetonitrile. A pale yellow sticky precipitate formed after rapid vortexing for 2 min and the clear extract was decanted into a new Falcon tube. The precipitate was suspended in 3 mL of 0.1 M ammonium bicarbonate, and extracted with 11 mL of 1 : 1 isopropanol: acetonitrile. The remaining solids were dissolved in 2.0 mL of 0.1 M ammonium bicarbonate and extracted with 9 mL of 1 : 1 isopropanol: acetonitrile. After extracting three times, the extracts were combined and clarified by centrifugation. The solvent was reduced by rotary evaporation to give crude farnesyl pyrophosphate.
Purifying farnesyl pyrophosphate with a cellulose column. The cellulose powder (60 g, about 180 mL) was pretreated by gentle washing for 30 min with two volumes of 0.2 N hydrochloric acid followed by three rinses with deionized water. The cellulose was then washed for 30 min with two volumes of 0.2 N sodium hydroxide solution rinsed three times with deionized water. This material was stored in a 1:1 (v/v) mixture of isopropanol and water. Columns were packed using a slurry technique. The bottom was plugged with glass wool and covered by a bed of clean sand 1 cm thick. The sand was covered with acetonitrile before adding a slurry consisting of one volume of cellulose in two volumes of the 1:1 (v/v) mixture of acetonitrile and 0.1 M ammonium bicarbonate. The cellulose was allowed to settle by gravity flow and packed under nitrogen pressure. Three column volumes of acetonitrile were passed through the column to remove any air pockets. The column was then equilibrated with four column volumes of a solvent mixture composed of isopropanol, chloroform, acetonitrile, and 0.1 M ammonium bicarbonate in a ratio of 5:3:1:1 (v/v/v/v), which was the solvent used to elute the farnesyl pyrophosphate from the column. The above crude material was dissolved in 2 mL of elution solvent and loaded onto column. Three column volumes of solvent were used to elute the column and 30 fractions (20-mL) were collected. Five fractions (7-11) that showed a single spot on cellulose TLC plate were combined and concentrated. The remaining aqueous solution was buffered with 25 mM ammonium bicarbonate solution to pH 7.2 and lyophilized to dryness. The yield was about 20% and the product identity was verified by proton NMR in deuterium oxide. The FPP (205 mg) was dissolved in 9.5 mL of 7:3 methanol/10 mM NH₄OH buffer to make 50 mM stock solution.
Regenerating the cellulose column Following chromatography, the column was regenerated by washing with a 1:1 (v/v) mixture of acetonitrile and 0.1 M ammonium bicarbonate (about 3 volumes) and washing with 2 volumes of isopropanol and acetonitrile in the ratio of 1:1 (v/v). The regenerated column was equilibrated with four volumes of isopropanol, chloroform, acetonitrile, and 0.1 M ammonium bicarbonate in a ratio of 5.5:2:1:1.5 (v/v/v/v). The column was ready to use for separating FPP.

Both column-purified and crude FPP were used as substrate for sesquiterpene synthase enzyme assay and worked well, indicating that the column purification of synthesized FPP is not necessary for this enzyme.

B. Synthesis of oxidosqualene

Oxidosqualene was synthesized according to the literature as follows: squalene (6.2 g, 0.015 mole) was dissolved in 410 mL of THF. The solution was stirred under nitrogen and cooled to 0 °C. The water was added to the solution until it became cloudy. Then a small amount of THF was added to clear the solution. N-bromosuccinimide (4.5 g, 0.025 mole) was added in small portions over 10 min. The reaction was stirred at 0 °C for 30 min, and the crude product was isolated by evaporation of about half the solvent, addition of water and extraction with hexane. Concentration of the organic layer gave 8.6 g of crude reaction mixture. The crude bromohydrin (8.6 g) was dissolved in 70 mL of methanol with 3.5 g of K₂CO₃, and the mixture was stirred at room temperature for 3 h. The crude oxide (6.3 g) was isolated by addition of water and extraction with hexane. The crude product was loaded on an automated column (dry packed) and eluted with 4% of ether in hexane. One hundred fractions were collected
and each fraction was about 20 mL. Fraction 35 to 44 contained oxidosqualene, which were combined and concentrated to give 1.12 g of pure oxidosqualene. Fraction 33 and 34 contained a little impurity and were combined to give 0.84 g of mixture. The overall yield, based on squalene was about 25%. The 20 × oxidosqualene stock solution was made as follows: (±)-2,3-oxidosqualene (200 mg) and Triton (2 g) were dissolved into a 50 mL round bottom flask with 2 mL of methylene chloride. The solution was concentrated by rotary evaporation and dl H$_2$O (8 mL) was added to the flask, and the solution was stirred in a warmed water bath for 1 h and stored at 4 °C.

2.4. Medium

A. Bacterial media

*E. coli* were propagated in Luria broth (LB). LB was prepared by dissolving 10 g of US Biological LB Broth Lennox (5 g tryptone, 2.5 g yeast extract and 2.5 g NaCl) in 500 mL dl water in a 1 L autoclave bottle, and sterilizing for 20 min in an autoclave. Plates were prepared identically, except that 7.5 g agar was added before sterilization. To make LB-ampicillin selective media, 25 mg/mL of ampicillin stock was made and filter sterilized. Two mL of this stock was added to 500 mL sterile LB liquid or LB agar media that had cooled below ~ 50 °C. For the plate used for blue/white selection, 20 μg/mL of 5-bromo-4-chloro-3-indoly-β-D-galactoside (X-Gal) added in LB-amp plate. X-Gal stock was made by dissolving 200 mg X-Gal in 50 mL DMF. The 2 × YT was prepared by dissolving 8 g tryptone, 5 g yeast extract and 2.5 g NaCl in 500 mL dl water in a 1 L autoclave bottle, and sterilizing for 20 min in an autoclave.
B. Yeast media

*S. cerevisiae* strains SMY series and their derivatives were propagated at 30 °C either in synthetic media (SCD) or rich media (YPD) supplemented with 20 μg/mL of ergosterol, 13 μg/mL of heme, and 5 μg/mL of Tween 80 (HET). YPD was made by mixing 5 g of yeast extract, 10 g of peptone, 10 g of glucose and 7.5 g of agar in 500 mL autoclave bottle and sterilizing in an autoclave for 20 min. A 100 × heme stock (1.3 mg/mL) was made by dissolving 13 mg of hemin in 10 mL 50% ethanol, 5 mM aqueous NaOH solution. A 100 × ergosterol stock (2.0 mg/mL) was made by dissolving 20 mg ergosterol in 10 mL 50% ethanol and 50% Tween 80 solution.

To make synthetic complete medium (SC), 1.7 g of yeast nitrogen base, 5.0 g of ammonium sulfate and 2.0 g of amino acids mix (10 g leucine, 2 g adenine, 2 g uracil and 2 g each of the following amino acid or its salt: alanine, arginine•HCl, asparagine, aspartic acid, cysteine•HCl, glutamic acid, glutamine, glycine, histidine•HCl, isoleucine, lysine•HCl, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine) were dissolved to 500 mL water in a 1 L autoclave bottle. Twenty grams of glucose (D) or galactose (G) and 15 g agar were added to 500 mL water in a separate 1 L autoclave bottle. Both bottles were sterilized in an autoclave for 20 min. After the medium had cooled to ~50 °C, equal volumes of sugar and synthetic media were mixed and the heme and erg stocks were added to the desired concentration (SCDHET or SCGHET). The plates were poured immediately. When selective plates were needed, uracil (Ura), histidine (His) or leucine (Leu) was omitted, as appropriate. Liquid synthetic complete media were prepared identically, except that the agar was omitted.
3. Methods

3.1. *E. coli* transformation

Transformation competent *E. coli* were prepared according to Okayama as follows. Frozen stock DH5α cells were thawed, streaked on an LB agar plate and incubated at 37 °C for overnight. About ten colonies with diameters 2-3 mm were picked and inoculated to 250 mL SOB medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂ and 10 mM MgSO₄; sterilized in an autoclave for 20 min) in a 2 L flask, and grown to an A₆₀₀ of 0.6 at 18 °C with vigorous shaking. Cells were harvested by centrifugation at 2500 × g for 10 min at 4 °C in a sterile tubes and resuspended by gentle shaking on ice in 80 mL of ice-cold TB (10 mM Pipes, 15 mM CaCl₂, and 250 mM KCl were mixed and the pH were adjusted to 6.7 with KOH, then MnCl₂ was added to 55 mM; filter sterilized) and incubated on ice for 10 min. The cells were spun down as above and the pellet was gentle resuspended in 20 mL TB, and DMSO was added with gently swirling to a final concentration of 7%. Aliquots were frozen in liquid nitrogen and stored at -70 °C. For transformation, a 100 μL aliquot was thawed on ice and 10-50 ng of DNA was added. The cells were incubated on ice for 20 min, then warmed to 37 °C for 5 min. Cells were spread directly on LB-amp plates and incubated at 37 °C for overnight. Typically, 1~3 × 10⁹ transformations/μg plasmid DNA were obtained. For the transformation of BL21(DE3), 40 ng of plasmid DNA was mixed with 25 μL of BL21(DE3) competent cells (Novagen), which were allowed to sit on ice for 5 min. The cells were heat shocked at 42 °C for 30 s and then spread directly on LB-amp plates. The plates were incubated at 37 °C overnight. For the transformation of BL21(DE3)pLysS, 100 ng of plasmid DNA was mixed with 100 μL of competent cells
from Dr. Bartel’s lab and transformed as for DH5α. The transformation was plated on LB supplied with 100 µg/mL ampicillin and 25 µg/mL of chloramphenicol and about 30 transformants were obtained from each transformation.

3.2. Yeast transformation

Carrier DNA was made by vigorously mixing 200 mg of high molecular weight DNA (deoxyribonucleic acid sodium salt Type III, Salmon Testes, Sigma D162U) with 100 mL of TE buffer (10 mM Tris-HCl, pH 8.0, 1.0 mM EDTA) on a magnetic stirrer for 2-3 h. The resultant carrier DNA was aliquoted into 200 µL fractions and stored at -20 °C. The carrier DNA was heated at 100 °C for 5 min and chilled on ice before use. Yeast cells were grown to an A600 of ~0.6 in 5 mL YPD. The cells were harvested by centrifugation at 1000 × g for 5 min and the cell pellet was resuspended in 100 µL of water. The 50 µg of carrier DNA, 1.0 µg of plasmid DNA and 2 mL of yeast transformation buffer (40% polyethylene glycol 3350, 0.1 M lithium acetate, 1 mM EDTA, 10 mM Tris pH 7.5, 0.1 M DTT) were added into the cell suspension. The transformation mixture were incubated at room temperature for at least 8 h and then centrifuged at 1000 × g for 5 min. The cell pellet was resuspended in 200 µL water and spread on two selective plates. Plates were incubated at 30 °C for colony formation (2-3 days). Typically, 10³ to 10⁴ transformants/µg plasmid DNA were obtained.

3.3. Plasmid isolation

Plasmid DNA was isolated from E. coli by using the Qiagen plasmid preparation Kit according to the manufacturer’s instructions.
3.4. DNA analysis

Restriction enzymes were used according to the manufacturer’s directions for analytical work as follows: DNA (200 ng) was added to 1 μL of the supplied 10 × buffer concentrate, diluted with water to 9.5 μL, and 0.5 μL of restriction enzyme was added. The tube with restriction digest mixture was incubated at 37 °C or other temperature according to the manufactures’ instructions. After 1 h incubation, 1.0 μL 10 × loading buffer (20% Ficoll 400, 0.1 M EDTA pH 8.0, 0.25% bromphenol blue, and 0.25% xylene cyanol) was added to the digestion reaction for sizing analysis. Tris-Acetic acid (TAE) agarose gels were used to determine molecular weights of DNA fragments and to preparatively fragments. A 50 × TAE stock was prepared by dissolving 242 g of Tris base, 57.1 g of acetic acid, 18.6 g of Na₂EDTA into up to 1 L of water. Gels were prepared by suspending 5 g of agarose in 500 mL of TAE buffer and heating in a microwave until the agarose dissolved. Upon cooling to ~ 50 °C, 5 μL ethidium bromide was added to 100 mL of 1% agarose and the solution was poured into a Owl EASY-CAST electrophoresis system. After the gel had solidified, it was overlaid with TAE buffer, and the restriction-digested DNA was electrophoresed alongside 250 ng of BstE II digested lamda DNA as molecular weight standards. DNA was visualized with a Fisher Scientific FBTI-88 transilluminator. Molecular weights were interpolated according to the standard DNA marker.
3.5. Subcloning

DNA fragments for subcloning were purified from agarose gels using Qiaex suspension from Qiagen according to the manufacturer’s instructions. Fragments were ligated together using Fast-Link DNA ligase according to the manufacturer’s instructions.

3.6. Dephosphorylation of linearized DNA

Calf intestinal alkaline phosphatase (CIP) was used to hydrolyze the 5’-phosphate group of the linearized vector to avoid vector self-ligation. Ten micrograms of pBluescript KS(+) plasmid DNA was digested with restriction enzyme in NEB buffer 2, 3 or 4 at the DNA concentration of 100 ng/μL. After incubation at 37 °C for 2 h, the digestion mixture was diluted into 50 ng/μL of DNA concentration with the same NEB buffer. The 1.0 μL of CIP (NEB) was added to hydrolyze the phosphate group at 37 °C for another hour. When the DNA was digested in NEB1 buffer, the DNA was precipitated and resuspended in the CIP buffer to the DNA concentration of 50 ng/μL after digestion. The CIP was added and phosphorylated as described before.

3.7. PCR programs

The CAPS program was 40 cycles with 30 s annealing at 56 °C, 3 min extension at 72 °C, and denaturation for 30 s at 95 °C. The “hot start” program included 4 min 95 °C hot start during which DNA polymerase was added, and 40 cycles with 1 min annealing using a temperature gradient from 68 °C to 48 °C (- 0.5 °C/cycle), 3 min extension at 72 °C, and denaturation for 45 s at 95 °C. The program was terminated with a 5 min extension at 72 °C.
3.8. Adding A-tails blunt-ended PCR products

After gel purification, the PCR product was mixed with 1.0 µL of 100 mM dATP, 5.0 µL of 10 × PC2 (500 mM Tris-HCl pH 9.1, 160 mM ammonium sulfate, 35 mM MgCl₂), 0.5 µL of Taq polymerase (Fisher) and up to 50 µL of water and incubated at 72 °C for 2 h. The A-tailed PCR product was purified with Qiaex suspension for subcloning according to the manufacture’s instruction as follows: The QX1 buffer (150 µL) and 20 µL of Qiaex suspension were added to above A-tailed reaction. The mixture was incubated at room temperature for 10 min and centrifuged at 1000 × g for 30 s. The pellet was washed twice with 500 µL of buffer PE and air-dried. The DNA was eluted from the Qiaex suspension with 20 µL of water twice.

3.9. Making pBluescript KS (+) T-vector

pBluescript KS (+) T-vector was made according to the literature:55 Ten microgram of pBluescript KS(+) vector was digested with 4 µL of EcoR V (NEB) in 10 µL of 10 × NEB 2 buffer with 100 µg/mL of BSA in 100 µL of total volume. The reaction was incubated at 37 °C for 1 h and linearized DNA was precipitated with 1/10 volume of 3 M of NaOAc (pH 5.5) and 2.5 volumes of 100% ethanol. The DNA pellet was washed with 500 µL of 70% ethanol and air dried. The DNA was dissolved in 43.5 µL of water and T-tailed with 5.0 µL of 10 × PC2 (500 mM Tris-HCl pH 9.1, 160 mM ammonium sulfate, 35 mM MgCl₂), 1 µL of dTTP (100 mM) and 0.5 µL of Taq polymerase (5 units/µL, Fisher) at 72 °C for 2 h. The T-tailed DNA was purified as follows. The QX1 buffer (150 µL) and 20 µL of Qiaex suspension were added to above
T-tail reaction. The mixture was incubated at room temperature for 10 min and centrifuged at 1000 x g for 30 s. The pellet was washed twice with 500 μL of buffer PE and air-dried. The DNA was eluted from the Qiaex suspension with 20 μL of water twice. The blunt ended DNA was removed by self-ligation and gel purification of the linearized DNA. A 50 μL of ligation reaction was set up as follows, 40 μL of DNA, 4.5 μL of water, 5 μL of T4 DNA ligase buffer (NEB), and 0.5 μL of T4 DNA ligase (400 units/μL). The reaction was incubated at room temperature overnight and the linearized T-tailed vector was isolated from gel purification. The concentration of T-tailed pBluescript KS(+) was adjusted to 25 ng/μL.

3.10. Labeling probe for hybridization

The DNA probe was combined with 1 μg random 12-mer oligonucleotides in a 1.5 mL screw cap tube in a final volume of 16.5 μL. The mixture was incubated at 100 °C for 3 min to denature the DNA and then placed on ice immediately. Klenow buffer (5.0 μL) with dNTPs (no C) and 5 units of NEB exo- Klenow fragment (1.0 μL) were added on ice to the above mixture. The tube was then moved to shielded rack and 2.5 μL (25 μCi) of α-32P-dCTP was added and mixed by pipetting up and down. The reaction was incubated at room temperature for 1 to 4 h. TE 8 (200 μL) was added to quench the reaction. Before adding to the hybridization solution, the probe was denatured by incubating at 100 °C for 3 min.
3.11. Colony hybridization

A small chump of cDNA library glycerol stock cells (≈ 20 μL) were dissolved into 11.0 mL LB amp, 1.0 mL of this mixture was used to titer library. After titering the library, about $10^6$ colonies were plated on six big LB amp plates. The plates were incubated at 37 °C for overnight and chilled at 4 °C for at least 30 min. Six nylon membranes were placed onto the colonies, pierced with a needle at four point asymmetrically for orientation, and removed onto paper towel to dry. Colonies were lysed by placing the filters colony side up on a filter paper saturated with 0.2 M NaOH/1.5M NaCl for 2 min. Filters were neutralized by placing on a filter paper saturated with 1 M Tris pH 7.5 for 2 min and then transferred into 500 mL of 2 × SSC (20 × stock: 3 M NaCl, 0.3 M sodium citrate, pH 7.0) solution to remove the cell debris. The filters were air dried and the DNA was fixed on filters by UV-crosslinking with the Stratalinker. The filters were prehybridized in 20 mL of Church buffer (5 g BSA, 185 mg Na₂EDTA, 9.66 g NaH₂PO₄•H₂O, 25.55 g Na₂HPO₄, 35 g SDS in 500 mL H₂O) at 65 °C while making the probe. The prehybridization solution was removed, and 7 mL of Church buffer and half of the $^{32}$P-labeled probe was added. Hybridization was carried out at 65 °C overnight in a roller drum. After hybridization, the solution was decanted and the filters were washed several times with 0.2 × SSC/0.2% SDS at room temperature. The filters were wrapped in Saran wrap and exposed to X-ray film with an intensifying screen. Dark spots on each film indicated colonies that hybridized to the labeled probe. These colonies were identified and further dilutions of these colonies were plated on small LB amp plates (50 to 100 colonies/plate). The hybridization procedure was repeated until a single clone was isolated from each positive colony.
3.12. Site-directed mutagenesis

The selected plasmid was transformed into RZ1032 (dutI ungI) and plated on an LB amp plate. A single colony was inoculated into 1 mL 2 × YT supplied with 100 μg/mL ampicillin and grown at 37 °C for 3 h. The helper phage M13K07 (5 μL) was added to above culture. After incubating at 37 °C for another 1.5 h, the culture was diluted into 20 mL of 2 × YT supplied with 100 μg/mL ampicillin and 50 μg/mL kanamycin. The culture was inoculated at 37 °C for 12 to 15 h and the cells were centrifuged at 3000 rpm for 20 min at 4 °C. The supernatant was carefully transferred to a new tube and precipitated with 5 mL of 20% PEG 8000 in 2.5 M NaCl. After at least 1 h of incubation on ice, the solution was centrifuged at 3000 rpm for 15 min at 4 °C. The pellet was air-dried and suspended in 1.2 mL of TE (pH 8.0). The suspension was transferred to a microfuge tube and incubated on ice for 15 min. The solution was centrifuged at room temperature for 10 min and the supernatant was transferred to a fresh tube. The phage particles were precipitated with 300 μL of 20% PEG 8000 in 2.5 M NaCl. After incubating 15 min on ice, the solution was removed from precipitates by centrifugation and the phage particles were dissolved in 200 μL of TE (pH 8.0). The uracil-containing single stranded DNA was purified from phage as follows: 4 μL of 5 M NaCl and 200 μL of 1:1 phenol/chloroform mixture were added to the above phage solution. The solution was vortexed and centrifuged. The aqueous phase was transferred to a new tube and the phenol/chloroform extraction was repeated. Two volumes of ethanol was added to the aqueous solution to precipitate the single stranded DNA. The mixture was incubated at -20 °C for 1.5 h and centrifuged. The DNA pellet was
dissolved in 50 μL of water. The DNA concentration was quantitated by UV and adjusted to 0.5 μg/μL.

The mutagenesis oligo was phosphorylated with T4 DNA kinase as follows: To 18 μL of 20 pmol/μL oligo was added 2 μL 10 × kinase buffer (700 mM Tris-HCl (pH 7.6), 100 mM MgCl₂, 50 mM dithiothreitol), 1 μL 10 mM ATP and 0.5 μL NEB T4 DNA kinase (10 units/μL). The mixture was incubated at 37 °C for 1 h and the enzyme was inactivated by adding EDTA and heating at 70 °C for 10 min. The concentration of phosphorylated oligo was adjusted to 10 pmol/μL by adding 14 μL of milli-Q water. The oligo was annealed to single stranded DNA under following conditions: 2.0 μL of single stranded DNA (0.5 μg/μL) was mixed with 2.5 μL of phosphorylated oligo (10 pmol/μL), 1.0 μL 20 × SSC (3 M NaCl, 0. 3 M sodium citrate, pH 7.0) and 14.5 μL of water. The reaction was placed in a 70 °C heat block, which was removed from the heater and cooled to lower than 35 °C. After the primer was annealed to the template, the second strand was synthesized by adding the following reagents in order: 49 μL of mill-Q H₂O, 10 μL 10 × T4 DNA ligase buffer, 20 μL dNTPs (2.5 mM each), 1.0 μL single-strand binding protein (1 μg/μL, Stratagene), 1.0 μL of T4 DNA polymerase (3 units/μL, NEB) and 0.5 μL of T4 DNA ligase (400 units/μL, NEB). The mixture was incubated on ice for 5 min, room temperature for 5 min and 37 °C for 1.5 h. The DNA was precipitated with 2.0 μL of 5 M NaCl and 200 μL of ethanol. The pellet was air-dried and dissolved in 10 μL of milli-Q water. The DNA (4 μL) was used to transformed to DH5α chemically competent cells and plated on LB amp plates. The right clone was chosen by mapping the miniprep DNA with the restriction enzyme on the mutagenesis oligo.
3.13. Protein analysis on SDS-PAGE gel

A 9% SDS-PAGE gel was made as follow: In a 25 mL Erlenmeyer flask, 4 mL of 1.5 M Tris pH 8.8, 7.1 mL of DI water, 80 µL of 20% SDS, 4.8 mL of 30% acrylamide/0.8% bisacrylamide, and 50 µL of 10% ammonium persulfate were mixed and degassed for 5 min. TEMED (10 µL) was added to polymerize the gel. The gel was poured into an OWL protein gel apparatus and overlaid with water-saturated n-butanol. After the gel solidified, the n-butanol was removed by rinsing with DI water. A stacking gel was prepared same way as separating gel and poured on top of the separating gel (650 µL 30% acrylamide/0.8% bisacrylamide, 1.25 mL 0.5 M Tris pH 6.8, 3.0 mL DI water, 25 µL 10% ammonium persulfate, 25 µL 20% SDS and 5 µL TEMED). Equal volumes of protein samples were loaded on the gel and along with a wide range marker (Sigma). The gel was run in SDS electrophoresis buffer (5 × stock: 15.1 g Tris base, 72 g glycine, 5 g SDS and water to 1 L) at 20 mA constant current. After running the protein was stained for at least 1 h in staining solution (50% methanol, 0.05% Coomassie brilliant blue R-250 (first dissolved in MeOH), 10% acetic acid and 40% of DI water) and destained overnight in 5% methanol, 7% acetic acid and 88% of DI water.

3.14. Oligonucleotides

Custom oligonucleotides were synthesized by Integrated DNA Technologies, Inc. The oligos longer than 100-mers were SDS-PAGE purified.
3.15. DNA sequencing

DNA sequencing was done on an automated Applied Biosystems International sequencer using dideoxy method by either Texas Health Science Center at Houston or at Lone Star Labs Inc., in Houston.

3.16. Small scale assay for oxidosqualene cyclase activity

The yeast strain (5 mL) expressed with oxidosqualene cyclase was grown to saturation in YP-dextrose media supplied with heme (13 mg/L), ergosterol (20 mg/L) and Tween (5 mg/L). An one milliliter aliquot of above culture was diluted into 5 mL of YP-galactose media supplied with heme (13 mg/L), ergosterol (20 mg/L) and Tween 80 (5 mg/L), the cells were induced for two days. The cells were harvested by centrifugation and the pellet was washed with water once. Two volumes of pH 6.2 sodium phosphate buffer and 20 × oxidosqualene solution were added to the cells. Glass beads (a little below the aqueous surface) were added and the solution was vortexed twice on high for 2 min. The tube was chilled on ice between vortexing. The enzymatic reaction was incubated at room temperature overnight. A 5 μL aliquot of the reaction was analyzed on TLC by two repeated partial elutions with diethyl ether and developed with the mixed solvent of hexane and ether at 1:1 ratio. The enzymatic product was visualized with p-anisaldehyde stain (350 mL of ethanol, 10 mL of concentrated sulfuric acid, 10 mL of glacial acid and 15 mL of anisaldehyde) with mild heat.
3.17. TMS-ether derivatize triterpene alcohol

A solution of 1 : 1 bis(trimethylsilyl)-trifluoracetamide and pyridine (500 μL) was added to the vial with triterpene alcohol. The air was removed by blowing nitrogen over the solution. The vial was capped and heated at 40 °C for 2 h. The samples were concentrated under a nitrogen stream and analyzed by GC.
References


(56) Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W. Meth. Enzymol. 1990, 185, 60-89.


(76) Williams, D. C.; McGarvey, D. J.; Katahira, E. J.; Croteau, R. Biochemistry 1998, 37, 12213-12220.


(80) Huang, K.-X.; Huang, Q.-L.; Scott, A. I. Arch. Biochem. Biophys. 1998, 352, 144-152.


Appendix 1: List of Plasmids

(The first number after each plasmid is the plasmid library number in Dr. Matsuda’s Lab
and the second number is the page number where the plasmid is described in this thesis)

Possible triterpene synthase from *A. thaliana ecotype Landsberg erecta*

LH1.0: PCR fragment in pT7Blue vector, #86, p77

LH1.1: LH1 gene in pFL61 vector, #87, p78

LH1.2: Not I of LH1.1 subcloned in pBluescript KS(+), #88, p78

LH1.2R: LH1.2 in the reverse direction of LH1.2, #89, p78

LH1.16: LH1.2 digested with Sph I/Sma I, mung Bean nuclease treated and reclosed, #90, p79

LH1.18: Genomic fragment of missing exon in pBluescript KS (+), #92, p80

LH1.19: LH1.16 in pRS316 GAL, #91, p81

LH1.20: LH1.16 in pRS426 GAL, #140, p81

LH1.21: LH1.16 with an extra Sal I site right before ATG, #167, p81

LH1.22: LH1.21 Sal I/Not I fragment in pRS426GAL, #169, p81

Possible triterpene synthase from *A. thaliana ecotype Landsberg erecta*

LH2.0: PCR ORF3 fragment subcloned in pT7Blue, #258, p82

LH2.1: cDNA fragment from *A. thaliana* cDNA library in pFL61 vector, #259, p82

LH2.01: PCR ORF3 (HF polymerase) in pBluescript KS(+) T vector (different clone), #93, p83

LH2.02: PCR ORF3 with HF polymerase in pBluescript KS(+) T vector, #261, p83
LH2.11: Recombinant DNA from LH2.01 and LH2.02, #95, p84

LH2.14: Site mutagenesis of LH2.11 R483Q, #97, p84

LH2.16: LH2.14 Not I/Sal I fragment subcloned in pRS426 GAL, #138, p84

LH2.17: LH2.14 Not I/Sal I fragment subcloned in p416GPD/Not I, #187, p84

LH2.18: LH2.14 Not I/Sal I fragment subcloned in p426GPD/Not I, #188, p84

LH2.19: LH2.14 Sal I/Spe I fragment in pESC-LEU Xho I/Nhe I site, #244, p85

Possible triterpene synthase from A. thaliana ecotype Columbia

LH5.02: RT-PCR of ORF1 in pBluescript KS(+) T-vector, #104, p87

LH5.03: RT-PCR of ORF1 subcloned into pBluescript KS(+) T-vector (different clone than LH5.02), #262, p87

LH5.1: LH5.02 BstE II/Nde I fragment replaced by LH5.03, #263, p87

LH5.2: LH5.1 EcoR V/Pst I fragment replaced by LH5.03, #103, p87

LH5.4: LH5.2 Not I/Sal I fragment subcloned into pRS316GAL, #102, p87

LH5.6: LH5.2 Not I/Sal I fragment subcloned into pRS326GAL, #101, p87

LH5.7: LH5.2 Not I/Sal I fragment subcloned into pRS426GAL, #139, p87

LH5.8: LH5.5 Not I/Sal I fragment subcloned into p416GPD/Not I, #189, p88

LH5.9: LH5.5 Not I/Sal I fragment subcloned into p426GPD/Not I, #200, p88

LH5.10: LH5.2 Sal I/Spe I fragment in pESC-LEU Xho I & Nhe I site, #245, p88

Epicedrol synthase from Artemisia Annua

LH4.0: PCR product (SQF9 & SQR3) in pBluescript KS(+) T-vector, #265, p23

LH4.2: the longest clone of A.annua cDNA from hybridization, #100, p23
LH4.8: full length epicedrol synthase in pSPORT1, #98, p25
LH4.9: full length epicedrol synthase in PBS KS(+) Pst I & Not I sites, #114, p25
LH4.11: LH4.9 with Xho I site right in front of ATG, #118, p26
LH4.15: LH4.11 Xho I/Not I subcloned in pGEX-KTO.1, #191, p26

Sesquiterpene synthase from *A. thaliana* ecotype *Columbia*

LH6.01: PCR fragment subcloned in pBluescript KS (+) T-vector, #120, p34
LH6.5: Site-directed mutagenized LH6.01, #145, p35
LH6.6: LH6.5 with a Nde I site right at the ATG site, #144, p35
LH6.7: LH6.6 Nde I/Not I fragment subcloned in pET15b/Not I, #172, p36
LH6.8: LH6.6 Nde I/Not I fragment subcloned in pGEX-KTO.1, #190, p35
LH6.9: LH6.8 BamH I/EcoR I fragment in pET32a, #266, p37

**ERG9 knock out plasmids and yeast expression vectors**

LH7.0: 1.1 kbp PCR fragment of EH6.3 in pGEM-T vector, #135, p71
LH7.1: Acc I fragment of LH7.0 in pBluescript KS(+) Acc I site, #142, p71
LH7.2: LH7.1 Sph I/BamH I fragment in the same sites of EH7.1, #143, p72
LH7.3: JR2.01 Acc I/BamH I fragment in the Acc I/BamH I linearized LH7.2, #146, p72
pRS426: purchased from ATCC, #65
pRS426GAL: GAL promoter (Sal I and Kpn I) in pRS426, #66, p76
pRS416GPD: yeast expression vector from Dr. Gustin, #170, p76
pRS426GPD: yeast expression vector from Dr. Gustin, #171, p76
pRS416GPD/Not I: Not I was inserted into the pRS416GPD Hind III/Xho I site, #186, p77

pRS426GPD/Not I: Not I was inserted into the pRS426GPD Hind III/Xho I site, #184, p77
Appendix 2: List of Strains

Yeast Strains

SMY4: MATa hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

SMY8: MATa erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

LHY1: MATa erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

LHY2: MATa erg7::LEU2 erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

LHY3: MATa erg7::leu2::hisG-URA3-hisG erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

LHY4: MATa erg7::leu2::hisG erg9::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 Gal^-

MLY2: MATa erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 LEU2, pGAL-CPI::leu2-3,112 his3-Δ200 ade2 Gal^-

E. coli Strains

DH5α: F^- Φ80dlacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (rK^-, mK^+)

phoA supE44λ^- thi-1 gyrA96 relA1

DH10B: F^- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 deoR recA1 endA1 araD139 Δ(ara, leu)7697 galU galKλ^- rpsL nupG

BL21(DE3): F''ompT hsdSB (rB^- mB^-)gal dcm (DE3)

BL21(DE3)pLysS: F''ompT hsdSB (rB^- mB^-)gal dcm (DE3) pLysS