De novo Sequencing and Analysis of Transcriptome from *Azadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis
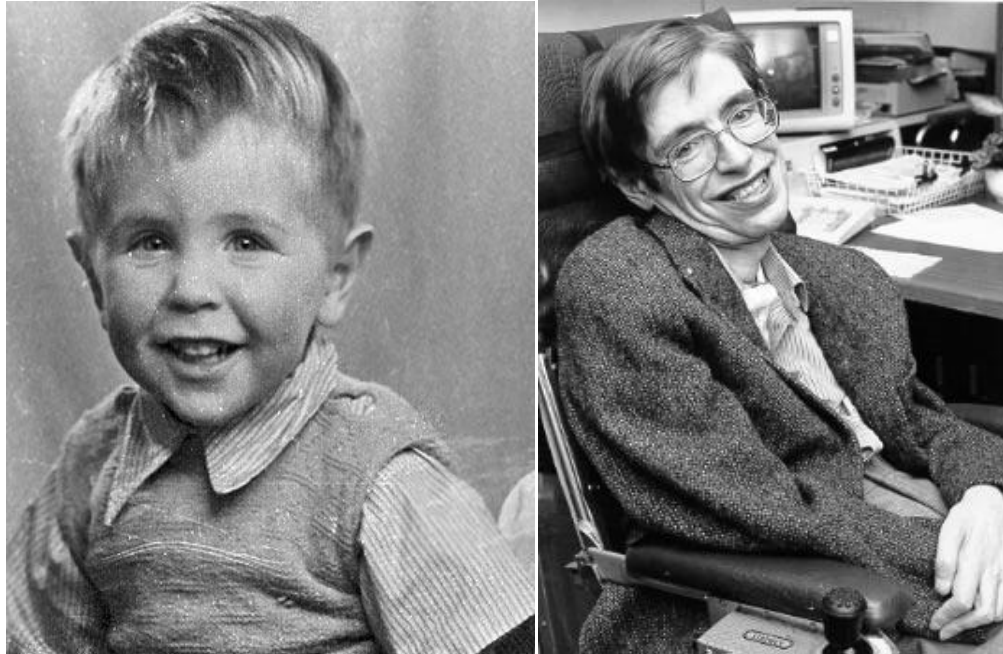
Thesis Submitted to **AcSIR**
for the Award of the Degree of
DOCTOR OF PHILOSOPHY
in **Biological Sciences**

By
Avinash Pandreka
10BB12A02015

Under the Guidance of
Dr. H. V. Thulasiram

CSIR-Institute of Genomics & Integrative Biology
New Delhi-110025, India
May 2018

"I realise the rare opportunity I've been given to live the life of the mind. But I know I need my body and that it will not last forever."

- Stephen William Hawking

CERTIFICATE

This is to certify that the work presented in this thesis entitled, "*De novo* Sequencing and Analysis of Transcriptome from *Azadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis" submitted by **Mr. Avinash Pandreka** to the Academy of Scientific and Industrial Research (**AcSIR**), for the degree of **Doctor of Philosophy** in **Biological Sciences**, embodies original research work under my supervision in the CSIR-Institute of Genomics & Integrative Biology, New Delhi – 110025, India. I further certify that this work has not been submitted to any other university or institution for the award of any degree or diploma. Any material that has been obtained from other sources has been duly acknowledged in this thesis.

Date: 14/05/2018

Place: New Delhi

Thulasiram. H.V.

Dr. H. V. Thulasiram

(Research Guide)

CSIR-Institute of Genomics & Integrative Biology

New Delhi, India

DECLARATION

I, Avinash Pandreka, hereby declare that the work incorporated in the thesis entitled "*De novo* **Sequencing and Analysis of Transcriptome from** *Azadirachta indica* **to Characterize the Genes Involved in Limonoid Biosynthesis**" submitted by me to the Academy of Scientific and Industrial Research (**AcSIR**), for the degree of **Doctor of Philosophy** in **Biological Sciences** has been carried out by me in the CSIR-Institute of Genomics & Integrative Biology under the guidance of Dr. H. V. Thulasiram. This work is original and has not been submitted to any other university or institution for the award of any degree or diploma. Such material, as has been obtained from other sources, has been duly acknowledged.

Avinash Pandreka

Dedicated to Family

Acknowledgement

It's a great pleasure to be surrounded by people who have supported, inspired and motivated me every single day during my PhD journey. Honestly, it would be another thesis if I begin to document each ones unselfish help rendered but I have to keep it short.

Firstly, I would like to express my in-depth gratitude to my research guide, Dr. H.V. Thulasiram, for giving me an opportunity to explore various dimensions in research; for his timely advice and for an exceptional research environment. I am extremely grateful to him for giving me the freedom to carry out research and for moulding me into a potential researcher. I sincerely acknowledge him for the support he extended during the ups and downs of my PhD journey.

Also, I express my heartfelt gratitude to Dr. Kausik Chakraborty, for his support right from my PhD joining to coursework and his valuable suggestions and inputs. A special mention and thanks to Dr. Dhanshekaran for guiding and teaching me about transcriptome analysis. I am highly obliged to Dr. Anand Bacchwat, Dr. Rakesh Sharma, Dr. Beena Pillai, Dr. Sureshkumar Ramasamy, Dr. Alok Sen, Dr. Nene, Dr. Subashchandrabose Chinnathambi and Dr. Mahesh J. Kulkarni for their support and valuable advice.

I would like to sincerely acknowledge, Shiva Shankar S, Kiran Kumar, Prashant G. Khanderao, Vignesh Kumar, Prabhakar Lal Srivastava, Krithika Ramakrishnan, Nagesh, Dipesh D. Jadhav, Santi, Saradhi, Atul Anand, Vaibhav Shah, Rincy Raju, Nilofer Jahan Siddiqui, Soniya Mantri, Priyadarshini B, Sharanbasappa Gangashetty, Dinesh, Joyti, Pankaj P. Daramwar, Asher Rajkumar, and Alaganuru Rajendraprasad for sharing their experiences and teaching me each and every step of research.

"Individually, we are one drop. Together, we are an ocean." I express my sincere gratitude to Chaya S. Patil, Vijayshree G. Shinde, Uttara S. Vairagkar, Nivedita Kolvekar, Jennifer Raphel, Aarthy Thiagarayaselvam, Fayaj A. Mulani, Bhagyashree D. Date, Ashish Kumar, Pooja Devi Sharma, Saikat Haldar, Shrikant Karegaonkar, Sudha Ramkumar, Singaravelansv G, Prasad Paphale, Devdutta Dandekar, Sharvani S. N, Balaji Kale, Kirti Gondkar and Rohil Sayed for their valuable work contribution towards the successful completion of my thesis.

I would like to extend a huge thanks to my labmates, Swati, Harshal, Ashish, Rajeshwari, Ajay, Shruti, Aaditi, Supriya, Sneha, Nikita, Jeevita, Ankita, Chitra,

Monica, Anirban, Pruthviraj, Rani, Radha, Manoj Kumar, Nilesh, Montu, Prabhu, Sridhar, Sonia, Anurag, Vijay, Yashpal, Nagendra, Debjyoti, Govinda, Datta, Joginder, Rahul and Mr. Khandekar for providing a healthy and pleasant environment for work.

I take this opportunity to acknowledge, The University Grants Commission (UGC), New Delhi, for providing the Research Fellowship. I would also like to thank, The Direction, CSIR-Institute of Genomics & Integrative Biology (IGIB), New Delhi and National Chemical Laboratory (NCL), Pune, for providing the infrastructure and support to accomplish the work. I am extremely obliged to AcSIR-IGIB, New Delhi team, for their support during the PhD documentation.

Each and every step in life becomes cheerful with friends. I would like to acknowledge my friends, Manna Ratan, Naveen, Sai Krishna, Raj, Koushik, Subba Rao, Suman Joshi, Arun Kumar, Prasad, Navya, Kalyan Kumar, Manikanta, Bhavani, Sammeswar Rao, Suresh, Hari, Shiva Ranjani, Kranti, Shravani, Sushma, Madan, Nagi Reddy, Madusudan, Jagadish, Manoj, Ashok, Niraj, Rajesh, Amitesh, Anil Giri, Priyanka, Shashank, Sindoori, Vibhav, Vishwa Bandu, Vishwanath, Reema, Arati, Tejal, Rubina, Shweta, Sandeep, Yashwant, Rashmi, Joyti, Rupa, Tanya, Deepanwita, Lakshmi and Chaitanya Kiran.

Words fall short to express my deep gratitude to my family for their constant love, support, blessings and sacrifices. I would like thank and express my love to the 'old gentlemen', Late Mr. Venkateswarlu and Late Mr. Peda Ramaya for their faith and affection on me. My parents have been my idols and my pillars of strength. The countless sacrifices and struggles they have been through, right from the beginning of time, to make my dreams come true, is something that I cannot express in words and will never be able to pay back. Thank you, Nanna and Amma (Mr. Veera Brahmamu and Mrs. Padmavathi). My sister, Deepthi, my childhood enemy but the one without my life is incomplete; thank you for the all the support and love you showered on me. My brother-in-law, Ram Prasad, for his constant faith, encouragement and understanding. I thank my uncles (Srinivas Rao, Raghu and Sambasiva Rao), my aunts (Padmavati, Eshwari Devi and Vijay Lakshmi) and cousins (Lalita, Yashwanth, Manikanta Srikant, and Sai Krishna) for their support and encouragement.

Last, but not least, I would like to thank each and every one for their help, support and as a source of inspiration.

Contents

List of Tables

List of Figures

Abbreviations

AA	Amino Acid
ADP	Adenosine Diphosphate
Amp	Ampicillin
ATP	Adenosine Triphosphate
BAS	β-Amyrin Synthase
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
CAS	Cycloartenol Synthase
C-B-C	Chair-Boat-Chair
C-C-C	Chair-Chair-Chair
C-C-C-B	Chair-Chair-Chair-Boat
C-C-C-C	Chair-Chair-Chair-Chair
cDNA	Complementary DNA
CDP-ME	4-(Cytidine-5'-Diphospho)-2-C-Methyl-D-Erythritol
cm	Centimetre
COSY	Correlation Spectroscopy
CPR	Cytochrome P450 Reductase
CSM	Complete Supplement Mixture
CSM-LEU	Complete Supplement Mixture Without LEU
CSM-URA	Complete Supplement Mixture Without URA
C-terminal	Carboxy-Terminal
CYP/CYP450	Cytochrome P450
DEPC	Diethyl Pyrocarbonate
DEPT	Distortionless Enhancement by Polarization Transfer
DMAPP	Dimethylallyl Diphosphate
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
DXP Pathway	1-Deoxy-D-Xylulose-5-Phosphate Pathway
DXP	1-Deoxy-D-Xylulose-5-Phosphate
DXR	DXP Reductoisomerase
DXS	1-Deoxy-D-Xylulose 5-Phosphate Synthase
ER	Endoplasmic Reticulum
FAD	Flavin Adenine Dinucleotide
FDS	Farnesyl Diphosphate Synthase
FMN	Flavin Mononucleotide
FPP	Farnesyl Diphosphate
Gal	Galactose
GAP	Glyceraldehyde-3-Phosphate
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GDS	Geranyl Diphosphate Synthase
GGDS	Geranylgeranyl Diphosphate Synthase
GGPP	Geranylgeranyl Diphosphate
GPP	Geranyl Diphosphate
h	Hour
HDR	HMBPP Reductase
HDS	HMBPP Synthase
HMBC	Heteronuclear Multiple Bond Correlation
HMBPP	4-Hydroxy-3-Methyl But-2-Enyl Diphosphate

HMG-CoA	Hydroxy-3-Methylglutaryl-CoA
HMGR	HMG-CoA Reductase
HMGS	HMG-CoA Synthase
HPL	Hydroperoxide Lyase
HRMS	High-Resolution Mass Spectrometry
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
INSIG-1	Insulin-Induced Gene 1
IP	Isopentenyl Phosphate
IPP	Isopentenyl Diphosphate
IPPI	IPP Isomerase
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
KAAS	KEGG Automatic Annotation Server
kb	Kilobase Pair
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
LA	Luria Agar
LB	Luria Broth
LS	Lupeol Synthase
Mb	Megabases
MDC	Mevalonate-5-Diphosphate Decarboxylase
MDS	2-C-Methyl-D-Erythritol-2,4-Cyclodiphosphate Synthase
MEcPP	2-C-Methyl-D-Erythritol-2,4-Cyclodiphosphate
MEP	2-C-Methyl-D-Erythritol-4-Phosphate Pathway
min	Minute
MJ	Methyl Jasmonate
MK	Mevalonate Kinase
mL	Millilitre
mg	Milligram
mRNA	Messenger RNA
MVA	Mevalonate Pathway
μg	Microgram
μM	Micromolar
μL	Microlitre
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Center for Biotechnology Information
ng	Nanogram
NGP	Neighbouring Group Participation
NGS	Next-Generation Sequencing
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Spectroscopy
N-terminal	Amino-Terminal
OD	Optical Density
OLC	Overlap Layout Consensus
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric Point
PMK	Phosphomevalonate Kinase

PMSF	Phenyl Methyl Sulphonyl Fluoride
qPCR	Quantitative Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
RNA	Ribonucleic Acid
RPKM	Reads per Kilobase of Transcript per Million Mapped Reads
Rt	Retention Time
SCAP	SREBP Cleavage-Activating Protein
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Sec	Second
SMART	Single Molecular Real Time
SQE	Squalene Epoxidase
SQS	Squalene Synthase
SREBP	Sterol Regulatory Element Binding Protein
TIC	Total Ion Chromatogram
TTS	Triterpene Synthases
UPLC	Ultra-Performance Liquid Chromatography

Thesis Abstract

De novo Sequencing and Analysis of Transcriptome from *Azadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis

Abstract

Neem (*Azadirachta indica*) is an evergreen tree, native to the Indian sub-continent. It has potential use in medicine, agriculture, environment protection and pest management. In neem, around 150 different limonoids were identified and characterized. These limonoids are synthesized by the triterpenoid biosynthetic pathway. In this study, we analyzed the neem transcriptome for identification of putative genes involved in limonoid biosynthesis by correlating transcriptome data with metabolite profiling. Metabolic fingerprinting data revealed that basic and C-seco limonoids were abundant in pericarp and kernel, respectively and low in flowers. Functional annotation of transcriptome predicted genes related to MVA and MEP pathways, prenyl diphosphate synthases, squalene epoxidases, CYP450 reductases and triterpene synthases. Expression profile data of genes related to terpenoid metabolism in neem shows that triterpene biosynthetic genes were highly expressed in seeds (kernel and pericarp), which is in line with metabolic fingerprinting data. Steroid biosynthetic pathway genes such as methylsterol monooxygenase, sterol 14-α-demethylase and 7-dehydrocholesterol reductase were highly expressed in flowers and diterpenoid biosynthetic genes were highly expressed in leaves. From differential gene expression studies, MVA pathway rate-limiting enzymes HMG-CoA synthase and HMG-CoA reductase, mevalonate kinase, farnesyl diphosphate synthase (AiFDS), squalene synthase (AiSQS), squalene epoxidase (AiSQE) and triterpene synthase (AiTTS1) were over-expressed in the kernel as compared to flowers. These gene expression profiles support the involvement of AiTTS1 in triterpenoid biosynthesis. Fifteen CYP450 genes were predicted to be involved in triterpenoid biosynthesis from functional annotation and expression analysis of transcriptome.

Homomeric geranyl diphosphate synthase (AiGDS), AiFDS and AiSQE1 were cloned and characterized in order to identify the genes involved in the formation of triterpene intermediate from the basic isoprene units, IPP and DMAPP. The triterpene cyclic product formed by AiTTS1 was purified and analyzed by NMR studies.

Heterologous expressed recombinant AiTTS1 produces tirucalla-7,24-dien-3β-ol and it is one of predicted tetracyclic product in limonoid biosynthesis such as euphol, tirucallol, butyrospermol and tirucallol-7,24-dien-3β-ol. AiTTS2 gene codes for only 704 amino acids with missing C terminal β-sheet which is essential for cyclization of 2,3-oxidosqualene as it provides key residues/base, hence the gene was found to be non-functional. Two CYP450 reductase genes, which transfer the electrons from NADPH to the downstream CYP450 enzymes, have been characterized. Co-expression of AiSQE1 and AiTTS1 resulted in an increase in production of tirucallol-7,24-dien-3β-ol by two-fold in yeast. AiCYP1 expression in neem reveals its involvement in limonoid biosynthesis. This is the first report regarding functional characterization of prenyltransferase, squalene epoxidase and triterpene synthases from neem. Further, this study will help in the systematic understanding of limonoid biosynthesis in neem and metabolic engineering in yeast.

Chapter 1

Introduction

Plants produce a diverse array of metabolites. Nearly 50000 compounds' structure have been elucidated, but hundreds of thousands are yet to be identified[1,2], the majority of which are secondary metabolites. These compounds play a critical role in plant interactions with the environment such as pathogens and herbivores, abiotic stress and attracting other organisms for pollination and seed dispersal[3]. Apart from their role in plants, secondary metabolites are important to humans in numerous ways. These compounds are used as medicines, drugs, dyes, fragrance and essential oils. Based on biosynthesis and structural similarities, these compounds were divided into alkaloids, phenolics and isoprenoids[3]. Alkaloids are nitrogen-containing plant secondary metabolites, having pharmacologically active properties. Phenolics contain aromatic rings with one or more hydroxyl groups, which has antioxidant properties[4]. Isoprenoids are structurally the most diverse group of metabolites, derived from C_5 isoprene units (IPP and DMAPP). Over 75,000 isoprenoids have been identified in different organisms. In plants, isoprenoids play a key role in plant-pathogen interactions, membrane fluidity, respiration, growth and development, photosynthesis, attraction of pollinators and seed-dispersing animals[5]. Triterpenoids are one of the classes of isoprenoids synthesized from isoprene units through C_{30} squalene intermediate[6]. Triterpenoids have various biological properties like anti-inflammatory,

antiviral, anti-cancer, insecticidal and for the treatment of metabolic and vascular diseases[7]. Limonoids are tetranortriterpenoids occurring in Meliaceae family. Total 300 limonoids were identified, majority of which were accounted to be in *Azadirachta indica* (Neem) and *Melia azedarach* (Chinaberry)[8,9]. Limonoids are abundant in neem seeds as compared to other tissues. Based on structural similarities, neem limonoids were divided into basic and C-seco limonoids. Gedunin, azadiradione, nimbin, salannin and azadirachtin are the most important limonoids from neem showing different biological activities[10,11]. The MVA and MEP pathways lead to the biosynthesis of 2,3-oxidosqualene, which serves as the precursor for limonoid biosynthesis. Further, 2,3-oxidosqualene is cyclized by triterpene cyclases. Based on oxygenated C_{30} compounds isolated from Meliaceae, the precursor cyclic molecule (protolimonoid skeleton) for limonoids biosynthesis have been predicted to be euphol or tirucallol derivative. When tritium labelled euphol, tirucallol, Δ^7-tirucallol and butyrospermol were fed to the leaves of neem all were incorporated into nimbolide[26-28]. However, euphol was more effectively incorporated into nimbolide as compared to others. Hence, the predicted protolimonoid skeleton for limonoid biosynthesis in neem was Δ^7-isomer (butyrospermol) of euphane or tirucallane. Further, this cyclic product undergoes modification like hydroxylation, dehydrogenation, epoxidation, acetylation and tigloylation to form diverse limonoids in neem. The main aim of this work is to identify the genes involved in limonoid biosynthesis and metabolic engineering for large-scale production.

Chapter 2:

Neem Transcriptome Analysis

Sequencing and functional annotation of transcriptome are the primary tools for the discovery of novel genes, especially in non-model plants for which full genome sequencing is not economically feasible[12,13]. Integration of transcriptomics and metabolic fingerprinting helps in the understanding of plant secondary metabolism.

Levels of total and fifteen major individual limonoids were quantified in various tissues of the neem plant. Tissue-specific variation in the abundance of limonoids has been observed. The mature seed kernel and pericarp of initial stages were found to contain the highest amount of triterpenoids. Furthermore, a wide

diversity of triterpenoids especially C-seco triterpenoids were observed in the kernel as compared to the other tissues, whereas pericarp, flower and leaf contain mainly ring-intact limonoids[14,15]. To extensively cover the transcriptome, RNA isolated from limonoid rich tissues such as fruit stage 4, leaves and flowers were pooled. A total of 79,079,412 (79.08 million) paired-end reads each of 72 bp length were generated by Illumina GA II platform. These reads were given as input for Oases to generate 41,140 transcripts. The average length of transcripts obtained was 1,331 bp and the N50 length was 1,953 bp. Blastx, virtual ribosome, Pfam and KAAS (KEGG Automatic Annotation Server) tools were used for functional annotation. Functional annotation results helped in identification of genes related to MVA pathway, MEP pathway, prenyltransferases, squalene epoxidases and triterpene cyclases.

To gain more insights into the secondary metabolism of neem, tissue-specific transcriptome analysis was done (pericarp, kernel, flower and leaf). Total of 127,815 transcripts were generated and functional annotation was done as described earlier. Based on these analyses, genes for each terpenoid metabolic pathways in neem were predicted. Expression profile (RPKM) of genes involved in each terpenoid metabolite pathway explains that MVA pathway and triterpenoid related genes were highly expressed in kernel and pericarp, which found a correlation with the metabolic fingerprinting dataset (Figure 1). Genes encoding for steroid biosynthetic pathway, short chain prenyltransferases and MEP pathway genes were highly expressed in flower, which may help in mono- and sesqui-terpene biosynthesis for attracting pollinators. MEP pathway, diterpenoid and short chain prenyltransferase genes were highly expressed in leaves too. To identify the downstream enzymes, mainly CYP450s, we applied differential gene expression. AiHMGS, AiHMGR (key genes from MVA pathway), AiFDS, AiSQS, AiSQE and AiTTS1 were highly expressed in the kernel as compared to flowers. Downstream enzymes of triterpene biosynthesis (CYP 450, methyltransferases, acyltransferases, desaturase and glycosyltransferases) were selected to be involved in limonoid biosynthesis based on gene expression analysis.
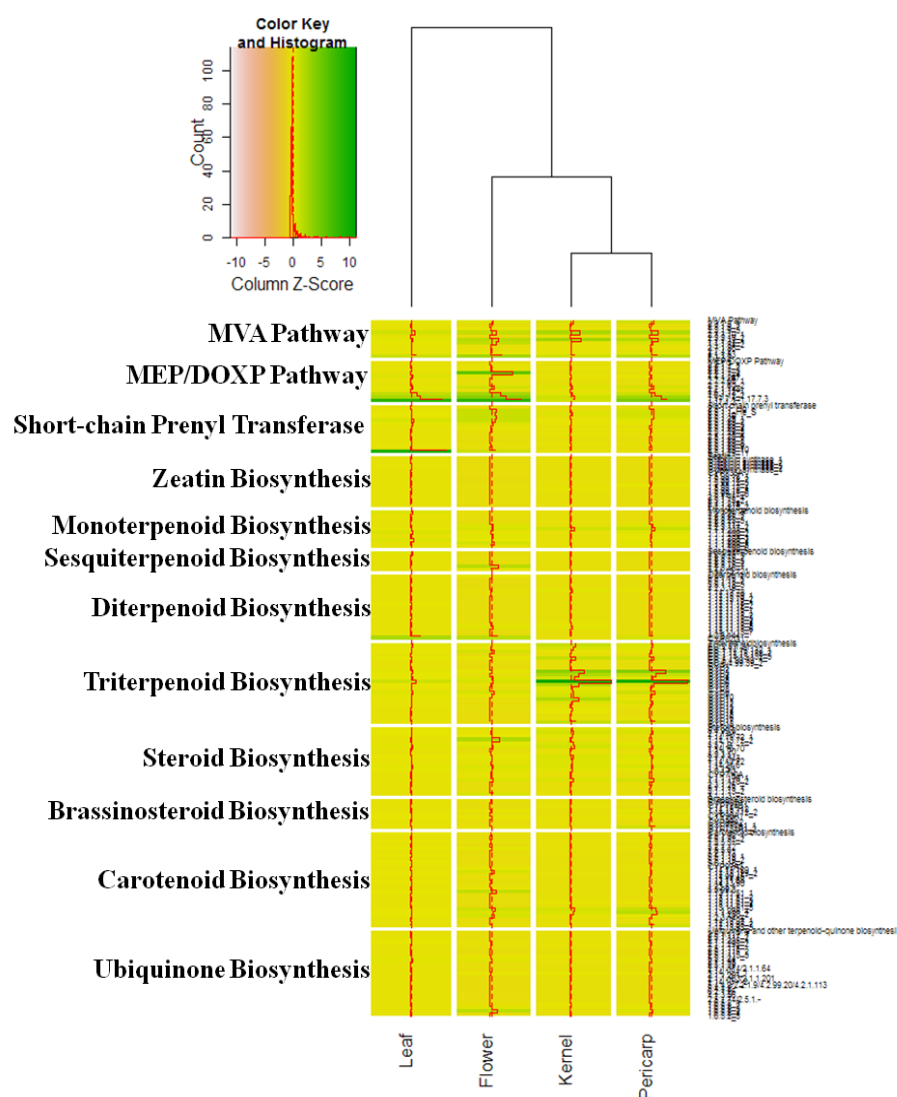
Figure 1: Comparison of RPKM Levels of Neem Terpenoid Metabolic Pathways.

Chapter 3:

Cloning and Characterization of Prenyltransferases

Terpenoid biosynthesis starts with building blocks such as isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are synthesized through the mevalonate (MVA) or methylerythritol phosphate (MEP) pathway[5,16]. Allylic diphosphate, DMAPP undergoes condensation with one or more IPP in head-to-tail fashion to produce linear diphosphates such as geranyl diphosphate (C_{10}, GPP), farnesyl diphosphate (C_{15}, FPP) and geranylgeranyl diphosphate (C_{20}, GGPP)[17,18]. Two molecules of FPP undergo 1-1' (head to head) condensation to form squalene via NADPH dependent reduction of presqualene diphosphate intermediate catalyzed by squalene synthase (SQS)[19]. Thus squalene is the first committed precursor for the

biosynthesis of triterpenoids. Squalene undergoes oxidation to form (*S*)2,3-oxidosqualene mediated by squalene epoxidase (SQE). Further triterpene cyclases catalyzes the formation basic triterpene skeletons from SQE[20,21]. Prenyltransferases such as GDS and FDS play a key regulatory role in triterpenoid and phytosterol biosynthesis[22,23]. Therefore, identification and functional characterization of prenyltransferases will assist in better understanding of triterpenoid biosynthesis.

Out of the 10 prenyltransferases identified from neem transcriptome, homomeric geranyl diphosphate synthase (AiGDS, Neem_transcript_10912) and farnesyl diphosphate synthase (AiFDS, Neem_transcript_25722) were selected for cloning and functional characterization (Figure 2). Real-time PCR analysis of AiGDS and AiFDS clearly indicates the involvement of AiFDS in limonoid biosynthesis.
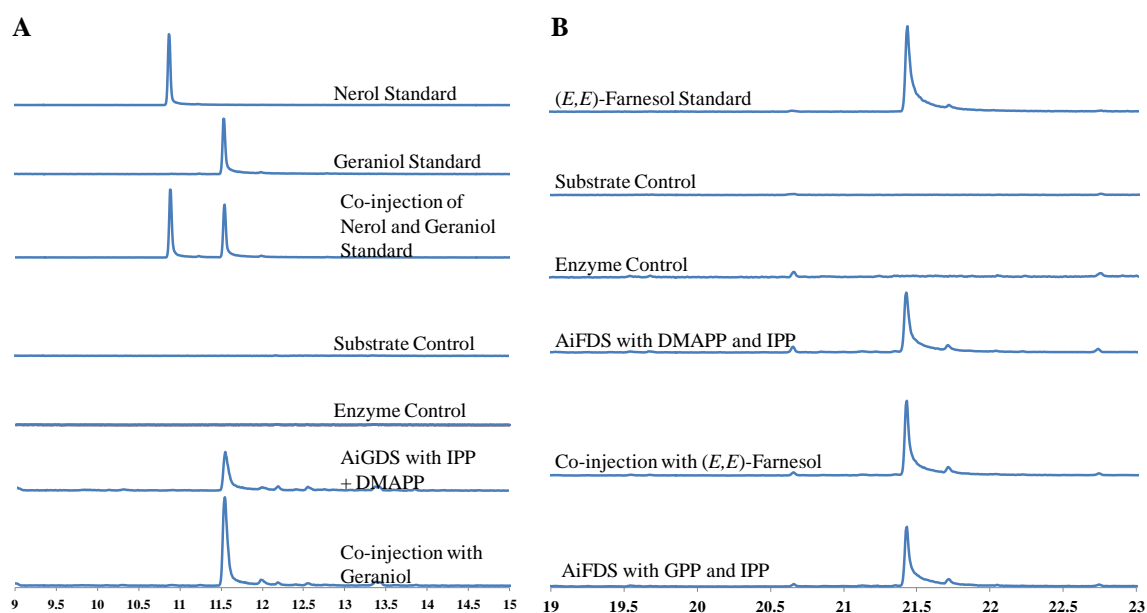
Figure 2: Total Ion Chromatograms (TICs) of AiGDS and AiFDS Assays. (A) TICs of AiGDS Assays and (B) TICs of AiFDS Assays.

Chapter 4:

Cloning and Functional Characterization of Squalene Epoxidase (SQE) and Triterpene Synthases (TTS)

From neem, more than 150 limonoids and their derivatives were isolated and characterized[24]. Seed kernel found to be rich source of limonoids as compared to other tissues. In seeds, basic limonoids and C-seco limonoids are distributed in pericarp and

kernel, respectively[15]. This suggests that limonoids production and distribution are highly regulated in neem. The first committed step for triterpenoid biosynthesis is the production of cyclic triterpene from 2,3-oxidosqualene by triterpene cyclases. Squalene epoxide is involved in the production of plant steroids and triterpenoids. Plants have more than one triterpene cyclases[20,25]. In steroid biosynthesis, cycloartenol is synthesized from 2,3-oxidosqualene by the action of cycloartenol synthase and further modified to phytosterol, while other triterpene cyclases convert 2,3-oxidosqualene into triterpene cyclic product, which is further modified by CYP450 enzymes into triterpenoids. According to triterpenoid biosynthesis prediction in neem, euphol or tirucallol or butyrospermol or tirucalla-7,24-dien-3β-ol may be the triterpene cyclic products formed by cyclization of 2,3-oxidosqualene[26-28]. Till date, there is no study reported on the involvement of these triterpene hydrocarbons in limonoid biosynthesis.

In neem, six triterpene cyclases were identified from transcriptome analysis. Neem_transcript_27436 (AiCAS) showed 92 % similarity with cycloartenol synthase [Q8W3Z3] from *Betula platyphylla*. Neem_transcript_28920 (AiTTS1) showed 86 % similarity, while Master_Control_74892 (AiTTS2) showed 93 % similarity with β-amyrin synthase [Q8W3Z1] from *Betula platyphylla*. These two triterpene cyclases have 76 % similarity at amino acid level. The phylogenetic analysis predicts that AiTTS1 is single- and AiTTS2 is multi-product forming enzyme. TTSs genes were cloned into a pYES2/CT vector and expressed in yeast INVSc cells. Saponification and n-hexane extraction of yeast cells was carried out to isolate the metabolites. Silver nitrate column was used to purify AiTTS1 metabolite. Based on GC-MS and NMR analyses, AiTTS1 metabolite is identified as tirucalla-7,24-dien-3β-ol, which is Δ^7-tirucallol. In AiTTS1, Phe 260 and Phe 729 stabilize the Markovnikov secondary cation created at C-14 then C-20 (dammarenyl cation). Lupenyl cation and oleanyl cation were not stabilized due to Val 550, Lue 553 and Tyr 125 of AiTTS1.

AiTTS2 transcript was not full length, around 1000 bp were missing from 3' end. 3' RACE was performed to get full-length AiTTS2. In AiTTS2, C terminal β-sheet, which play a key role in cyclization of 2,3-oxidosqualene was found to be missing. From these results, AiTTS2 can be predicted as an inactive enzyme. To further confirm, the full-length gene was cloned into a pYES2/CT vector and expressed in INVSc1. As we predicted, triterpene cyclic product was not observed

from yeast extracts. From the transcriptome, expression level studies as well as functional characterization, clearly indicates that AiTTS1 is involved in limonoid biosynthesis.

Total of three transcripts related to squalene epoxidase (AiSQE) are identified by transcriptome analysis. Out of these, Neem_transcript_11701 (AiSQE1) was selected and cloned into pRS315-TEF and pESC-LEU vector. Functional characterization of AiSQE was done by co-expressing with AiTTS1. A two-fold increase in the production of tirucalla-7,24-dien-3β-ol was observed when AiSQE1 was coexpressed with AiTTS1, which confirms that AiSQE1 involved in the production of 2,3-oxidosqualene (Figure 3B).

Figure 3: Characterization of AiTTS1 and AiSQE1. A) NMR analysis of AiTTS1 metabolite tirucalla-7,24-dien-3β-ol, **B**) Fold change in AiTTS1 metabolite in coexpression with AiSQE1 under gal and tef promoter. The retention time of individual metabolite given in parentheses.

Chapter 5:

Cloning and Functional Characterization of Cytochrome P450 Systems

Azadirachtin, nimbin and salannin are highly modified limonoids produced in neem. Azadirachtin contains 16 stereocenters, of which 7 are quaternary, and is nevertheless one of the most highly oxidized limonoids known[29]. The complexity of limonoids is due to the action of different cytochrome P450 systems present in neem. CYP450s are the largest gene family present in plants. These are involved in numerous biosynthetic and xenobiotic pathways. Most of the CYP450s act by taking electrons from CYP450 reductases for the oxidation of substrates. Analysing CYP450 systems in neem helps in understanding the biosynthesis and complexity of limonoids.

A total of 134 transcripts predicted as CYP450s and two transcripts as cytochrome P450 reductases were identified. Based on BLAST results, with reference to *Arabidopsis thaliana* CYP450 database, Neem CYP450s were classified into 39 families and 78 subfamilies, out of which most of the CYP450s belonged to the CYP71 family. Seven transcripts related to plant steroid biosynthesis and fifteen transcripts related to triterpenoid biosynthesis were predicted. Two CYP450 reductases were cloned and functionally characterized by cytochrome C and potassium ferricyanide reduction assays (Figure 4-A, B). To overexpress AiCYP1 in neem, the ORF was cloned into pRI101-AN. To silence AiCYP1 in neem, initial 300 bp fragment of sense and antisense strand of AiCYP1 were cloned into pRI101-AI harbouring wheat starching branching intron. *Agrobacterium*-based syringe infiltration method was used for transient transformation of AiCYP1 constructs into neem leaves. In AiCYP1 overexpression, azadirachtin A production was observed maximum in day 2 (3.5 folds) and Day 4 (2.09 folds) as compared to pRI101 vector control. In AiCYP1 silencing, Day 4 showed the highest effect i.e., 5 fold lesser was observed in AiCYP1 silenced neem leaves as compared to vector control. Based on metabolite extraction and HRMS analysis of different time intervals of neem transient transformed leaves, azadirachtin A level states that AiCYP1 is involved in limonoid biosynthesis.
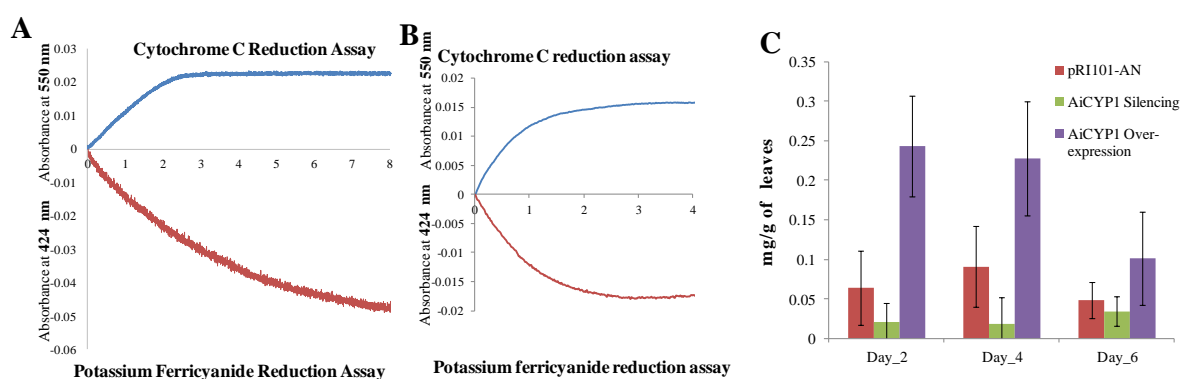
Figure 4: AiCPRs and AiCYP1 characterization. A) Cytochrome C and potassium ferricyanide reduction assays for AiCPR1, B) Cytochrome C and potassium ferricyanide reduction assays for AiCPR2, C) Fold change in production of azadirachtin A in neem when AiCYP1 was overexpressed and silenced.

References:

1.   Pichersky, E. & Gang, D.R. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in Plant Science* **5**, 439-445 (2000).

2.   Anarat-Cappillino, G. & Sattely, E.S. The chemical logic of plant natural product biosynthesis. *Current Opinion in Plant Biology* **19**, 51-58 (2014).

3.   Nan Zhao, G.W., Ayla Norris, Xinlu Chen & Feng Chen. Studying Plant Secondary Metabolism in the Age of Genomics. *Critical Reviews in Plant Sciences* **32**, 369-382 (2013).

4.   Dai, J. & Mumper, R.J. Plant Phenolics: Extraction, Analysis and Their Antioxidant and Anticancer Properties. *Molecules* **15**, 7313 (2010).

5.   Vranova, E., Coman, D. & Gruissem, W. Structure and Dynamics of the Isoprenoid Pathway Network. *Molecular Plant* **5**, 318-333 (2012).

6.   Xu, R., Fazio, G.C. & Matsuda, S.P.T. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**, 261-291 (2004).

7.   Connolly, J.D. & Hill, R.A. Triterpenoids. *Natural Product Reports* **18**, 131-147 (2001).

8.   Roy, A. & Saraf, S. Limonoids: Overview of Significant Bioactive Triterpenes Distributed in Plants Kingdom. *Biological and Pharmaceutical Bulletin* **29**, 191-201 (2006).

9.   Giovanni Benelli, A.C., Chiara Toniolo, Akon Higuchi, Kadarkarai & Murugan, R.P.M.N. Neem (*Azadirachta indica*): towards the ideal insecticide? *Natural Product Research* **31**, 369-386 (2016).

10.   Ponnusamy, S. *et al*. Gedunin and Azadiradione: Human Pancreatic Alpha-Amylase Inhibiting Limonoids from Neem (*Azadirachta indica*) as Anti-Diabetic Agents. *PLOS ONE* **10**, e0140113 (2015).

11.   Kushwaha, P. *et al*. *Azadirachta indica* triterpenoids promote osteoblast differentiation and mineralization *in vitro* and *in vivo*. *Bioorganic & Medicinal Chemistry Letters* **26**, 3719-3724.

12.   Goossens, A. *et al*. A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proceedings of the National Academy of Sciences* **100**, 8595-8600 (2003).

13.   Yonekura-Sakakibara, K. & Saito, K. Functional genomics for plant natural product biosynthesis. *Natural Product Reports* **26**, 1466-1487 (2009).

14.   Haldar, S. *et al*. Biocatalyst mediated functionalization of salannin, an insecticidal limonoid. *RSC Advances* **4**, 27661-27664 (2014).

15.   Haldar, S. Chemical finger-printing, metabolic profiling, and biotransformation of neem (*Azadirachta indica*) Limonoids. *Savitribai Phule Pune University* (2014).

16.   Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci Biotechnol Biochem* **66**, 1619-1627 (2002).

17.   Dewick, P.M. The biosynthesis of C5-C25 terpenoid compounds. *Natural Product Reports* **19**, 97-130 (2002).

18. Thulasiram, H.V., Erickson, H.K. & Poulter, C.D. A common mechanism for branching, cyclopropanation, and cyclobutanation reactions in the isoprenoid biosynthetic pathway. *Journal of the American Chemical Society* **130**, 1966-1971 (2008).

19. Hill, R.A. & Connolly, J.D. Triterpenoids. *Natural Product Reports* **29**, 780-818 (2012).

20. Phillips, D.R., Rasbery, J.M., Bartel, B. & Matsuda, S.P. Biosynthetic diversity in plant triterpene cyclization. *Current Opinion in Plant Biology* **9,** 305-314 (2006).

21. Xu, R., Fazio, G.C. & Matsuda, S.P.T. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65,** 261-291(2004).

22. Kim, O.T. *et al*. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. *Biologia Plantarum* **54,** 47-53 (2010).

23. Johnson, E.E., Jetter, R. & Wasteneys, G. Rapid induction of the triterpenoid pathway in *Arabidopsis thaliana* mesophyll protoplasts. *Biotechnology letters* **36,** 855-858 (2014).

24. Tan, Q.G. & Luo, X.D. Meliaceous limonoids: chemistry and biological activities. *Chemical Reviews* **111,** 7437-7522 (2011).

25. Yutaka Ebizuka, Y.K., T. Tsutsumi, Tetsuo Kushiro and M. Shibuya. Functional genomics approach to the study of triterpene biosynthesis. *Pure Appl. Chem.* **75**, 369-374 (2003).

26. Akhila, K.R.A. Biosynthetic Relationship Between Nemocinol and Nimocinolide in *Azadirachta indica*. *Natural Product Letters* **4**, 179-182 (1994).

27. E.U. Ekong, D. & A. Ibiyemi, S. Biosynthesis of nimbolide from [2-^{14}C,(4R)4-^{3}H1] mevalonic acid lactone in the leaves of *Azadirachta indica*. *Phytochemistry* **24**, 2259-2261 (1985).

28. Ekong, D.E.U., Ibiyemi, S.A. & Olagbemi, E.O. The meliacins (limonoids). Biosynthesis of nimbolide in the leaves of *Azadirachta indica*. *Journal of the Chemical Society D: Chemical Communications*, 1117-1118 (1971).

29. Veitch, G.E. *et al*. Synthesis of Azadirachtin: A Long but Successful Journey. *Angewandte Chemie International Edition* **46**, 7629-7632 (2007).

Secondary Metabolites

Translation: "We can designate as by-products of metabolism such compounds which are formed during metabolism but which are no longer used for the formation of new cells. ... Any importance of these compounds for the inner economy of the plant is so far unknown.   **Julius Sachs, 1873.**
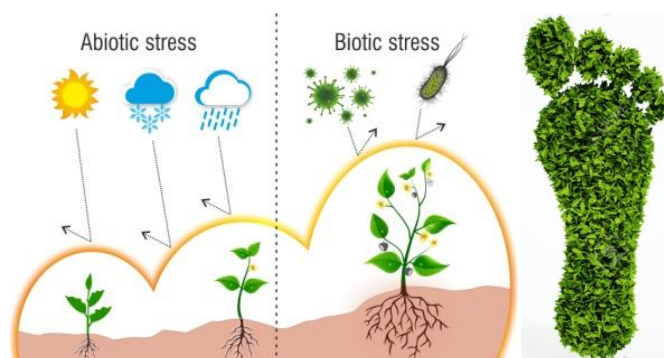
Chapter 1

Introduction

Secondary Metabolites

"A metabolic intermediate or product, found as a differentiation product in restricted taxonomic groups, not essential to growth and life of the producing organism, and biosynthesized from one or more general metabolites by a wider variety of pathways than is available in general metabolism."

J.W.Bennett, Ronald Bentley, 2008.

1.1 Introduction

Metabolites are the molecules present in the living system and are the substrates and products of enzymatic reactions. Based on their functional significance in the cell, they can be classified as primary and secondary metabolites. Primary metabolites, present in all living organisms are vital for life and includes carbohydrates, nucleic acids, amino acids, lipids and vitamins. Secondary metabolites are synthesized from the primary metabolites for enhancing particular biological functions in living systems, mainly to help in their interaction with the environment[1] (Figure 1.1). Investigation of the properties of secondary metabolites started around 200 years ago[2] (Figure 1.2). These secondary metabolites are majorly present in the plant kingdom and are involved in communication and defence in plants[3]. Plants synthesize very diverse, complex and most effective molecules as compared to other organisms. Till now over 200,000 secondary metabolites have been discovered from the plant kingdom and still hundreds of thousands yet to be identified and most of them are complex in nature[3,4]. Depending on physiological and developmental stages of the plant, secondary metabolites are synthesized and are of very low quantity (less than 1% of dry weight). Secondary metabolite biosynthesis is organized as complex biosynthetic pathways, under the action of enzymes which are encoded by genes. Approximately 20,000 - 60,000 genes are present in plant genome and among which nearly 25 % of the genes encode for enzymes involved in secondary metabolite biosynthesis.

Plants synthesize a wide range of secondary metabolites, in order to achieve fitness during evolution and to defend themselves from competing plants, herbivores[5] and pathogens[6,7] (Figure 1.1). For example, secondary metabolites in flowers act as fragrance and attract pollinators[8]. In fruits, they help in attracting animals and birds for seed dispersal. In roots, these metabolites are useful for better communication with soil microorganisms[7]. Some of these metabolites serve for a cellular function such as resistance to abiotic/biotic stress[9]. Each plant species have evolved a set of secondary metabolites for defence and communication. For example, to attract a specific class of pollinators, plants floral fragrances vary widely from species to species[3,10]. Plants synthesize these secondary metabolites and store them in vacuoles, resin ducts, trichomes, laticifers or cuticle, where they do not interfere with the

plant's own metabolism[7]. Plants not only synthesize and store specific class of secondary metabolites, but a complex array of diverse metabolites of different classes in order to obtain a synergistic effect as a response to environmental factors[11]. These secondary metabolites are produced at different concentrations varying from organ to organ, within a developmental cycle of plant[12,13]. In addition to their biological functions, plant secondary metabolites are very useful for mankind due to their pharmacological and toxicological properties[14].
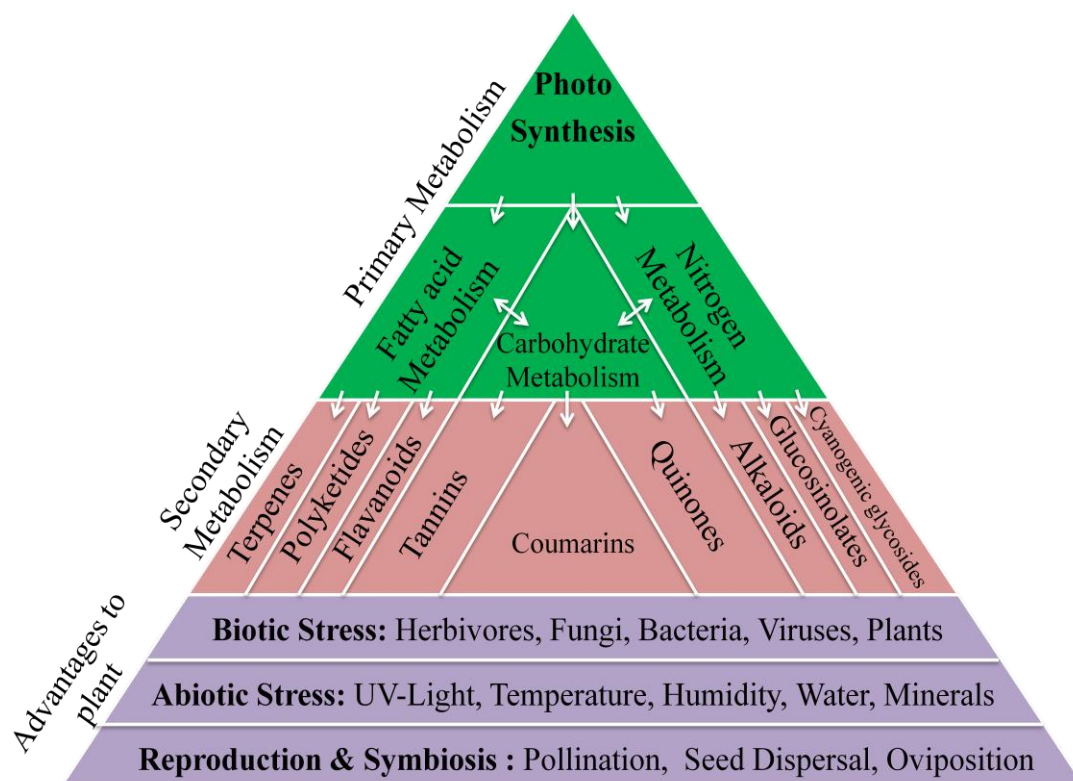
Figure 1. 1 Secondary Metabolites in Plants.

Humans are dependent on plants for their secondary metabolites which have a wide range of pharmacological property whereas, primary metabolites of plants act as essential amino acids and vitamins which cannot be synthesized in the human body. One-half of all the licensed drugs are natural products or their synthetic derivatives[15,16]. The earliest recording for use of the plant-derived substance was around 2600 B.C. in Mesopotamia. In India, the earliest documentation for the use of plant derivatives was in Atharva-veda around 1000 B.C.[17]. Some important examples are quinine (*Cinchona* species) and artemisinin (*Artemisia annua*) as antimalarial

drugs, reserpine (*Rauwolfia serpentina*) as an antihypertensive agent, ephedrine (*Epheda sinica, Ephedraceae* species) for the synthesis of antiasthma agents, vinblastine and vincristine (*Catharanthus roseus*) and paclitaxel (*Taxus* species) for treatment against cancer[17]. Atropine (*Solanaceae* species), caffeine (*Coffea arabica* and *Theobroma cacao*), cocaine (*Erythroxylum coca*), morphine (*Papaver somniferum*) and nicotine (*Nicotiana tabacum*) are known to have effects on the human brain[15]. Capsaicin (*Capsicum annuum*) and reserpine (*Rauwolfia serpentina*) have antiulcer activity[18]. Furthermore, in agriculture, triterpene saponins have anti-parasitic properties[19,20]. Azadirachtin and salannin (*Azadirachta indica*) are the most potent anti-insecticides[21]. Rotenone (*Milletia pachycarpa, Trpterygium forrestii* and *Rhododendron molle*), asimicin (*Asimina tribola*) and daphnane, orthoester skeletons (*Excoecaria agallocha*) are found to have insecticidal properties[20]. Apart from the medicinal uses, plant secondary metabolites are useful in flavours, dye, and fragrance industries. The wide range of properties that secondary metabolites have continuously being unravelled, showing great promises for present and future needs.
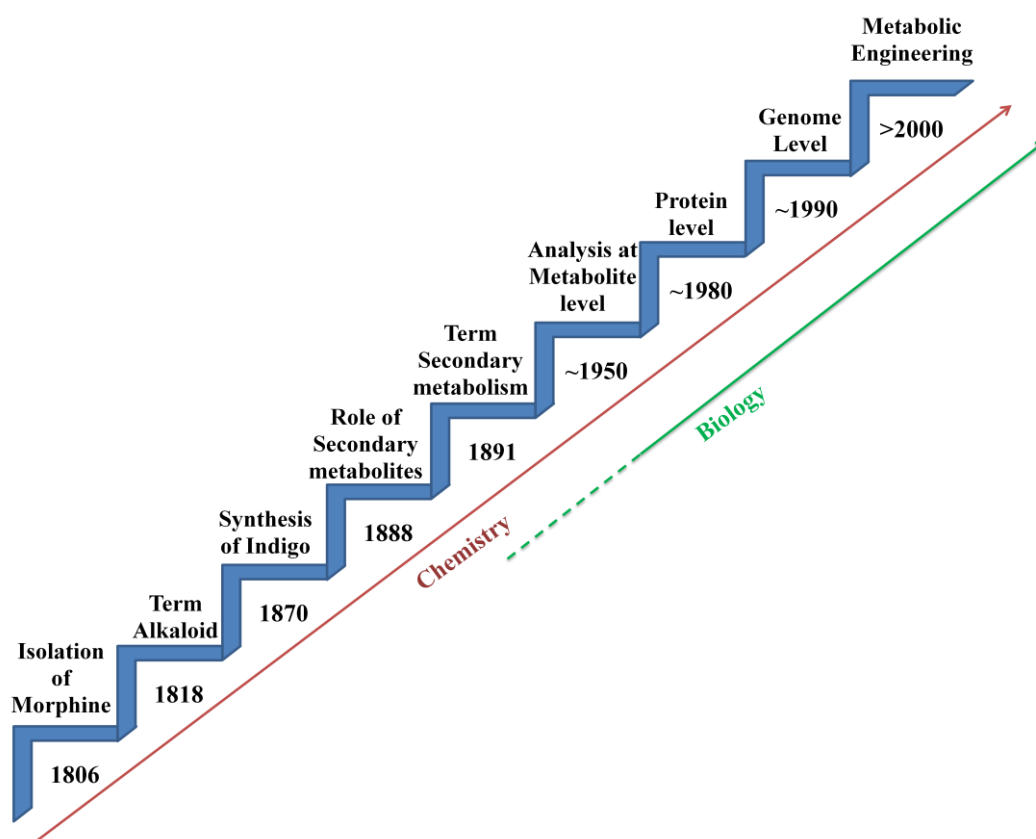
Figure 1. 2 Time Line of Plant Secondary Metabolites Investigation.

1.2 Classification

Secondary metabolites were classified into phenolics, alkaloids and isoprenoids, based on their structure and the pathway from which they are synthesized[22] (Figure 1.3).
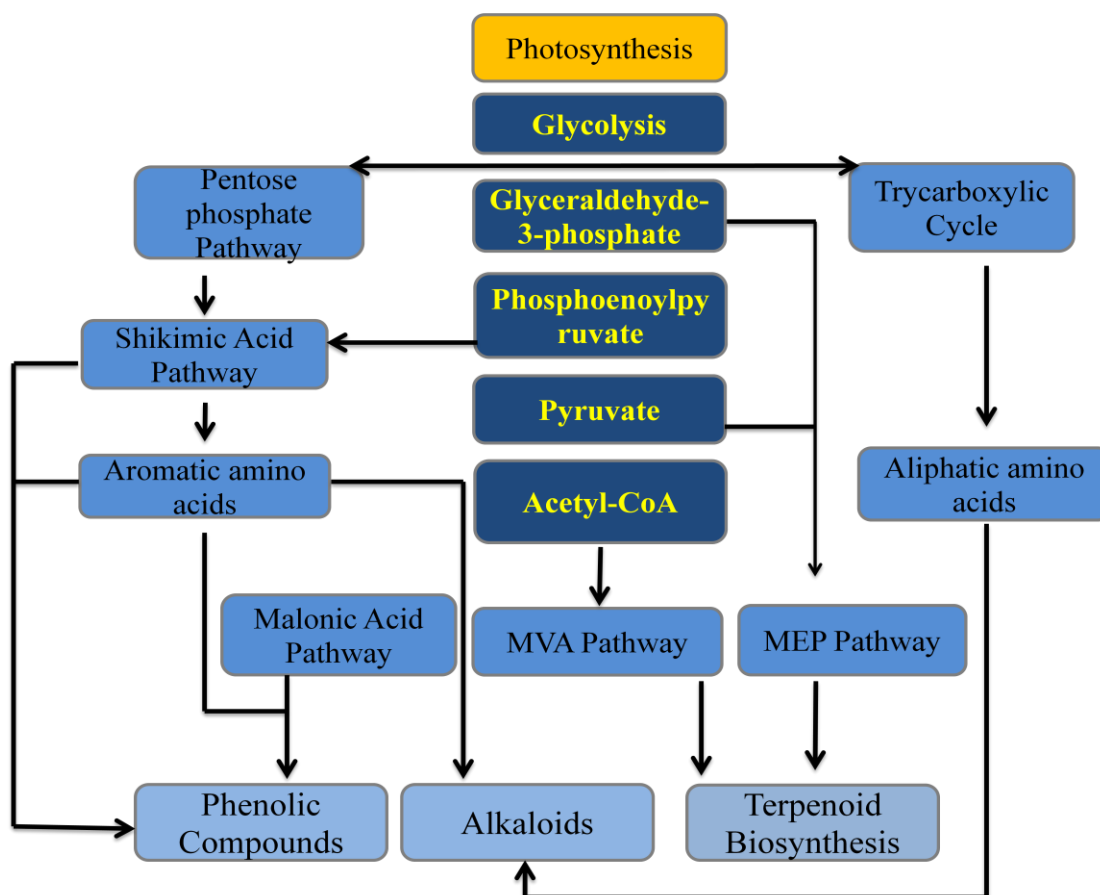
Figure 1. 3 Overview of Plant Secondary Metabolite Biosynthesis.

1.2.1 Phenolics

Phenolics are the compounds which contain one or more aromatic rings and hydroxyl groups. More than 8000 phenolic compounds are known until now. Phenolics are synthesized from phenylpropanoid pathway, and include flavonoids, phenolic acids, lignins and tannins. Flavonoids contain the flavan nucleus (15 carbons arranged in three rings as C6-C3-C6) and differ in the position of hydroxylation, methoxylation, prenylation or glycosylation. Phenolic acids are derivatives of benzoic acid (gallic acid) or cinnamic acid (coumaric and caffeic acid). Ferulic acid is a phenolic acid which is esterified to form hemicellulose in cereals. Tannins are

synthesized from the flavan-3-ol as basic units and bind to enzymes or proteins and act in defence against pathogens and herbivores[23]. Lignins are the polymer which acts as a matrix for cellulose microfibrils in the cell wall. Lignins are synthesized by phenolic oxidative coupling with monomers like hydroxyl cinnamoyl alcohol, coniferyl alcohol or sinapoyl alcohol (Figure 1.4)[24].

Phenolics are most abundant in dietary food and are also involved in pigmentation in plants, defence against ultraviolet radiation and pathogens. They have properties like antimicrobial, antioxidant, anti-inflammatory and anti-carcinogenic. Phenolics are widespread constituents of plant foods and beverages and are partially responsible for the overall organoleptic properties of plant foods[25].
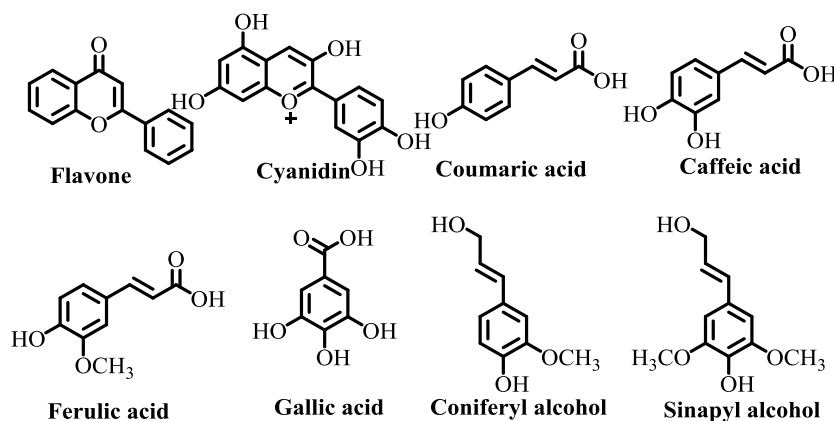
Figure 1. 4 Plant Phenolics.

1.2.2 Alkaloids

Alkaloids are basic or alkaline in nature and contain nitrogen group. More than 12,000 alkaloids structures were known and their roles in plants against herbivores and pathogens. The first alkaloid isolated was morphine in 1806[26,27]. Alkaloids include pure alkaloids, tropane alkaloids, benzylisoquinoline alkaloids and terpenoid indole alkaloids. Pure alkaloids are nitrogen-containing compounds which are originated from amino acids. Morphine, nicotine and caffeine are characteristic examples for pure alkaloids. Tropane alkaloids (cocaine) are synthesized from putrescine, an intermediate in nicotine biosynthesis[28]. Benzylisoquinoline alkaloids are synthesized from dopamine and 4-hydroxyphenylacetaldehyde which includes morphine (analgesic), colchicine (microtubule disrupter), sanguinarine (antimicrobial)

and (+)-tubocurarine (neuromuscular blocker)[29]. Terpene indole alkaloids contain indole moiety derived from amino acid tryptophan and terpenoid moiety derived from terpene biosynthetic pathway[28]. More than 3000 terpene indole alkaloids are known, including vinblastine (antineoplastic), quinine (antimalarial) and strychnine (rat poison). Alkaloids contain properties like anti-malarial, anti-asthma, anti-cancer, analgesic, herbivore deterrents and anti-hyperglycemic activities[30].
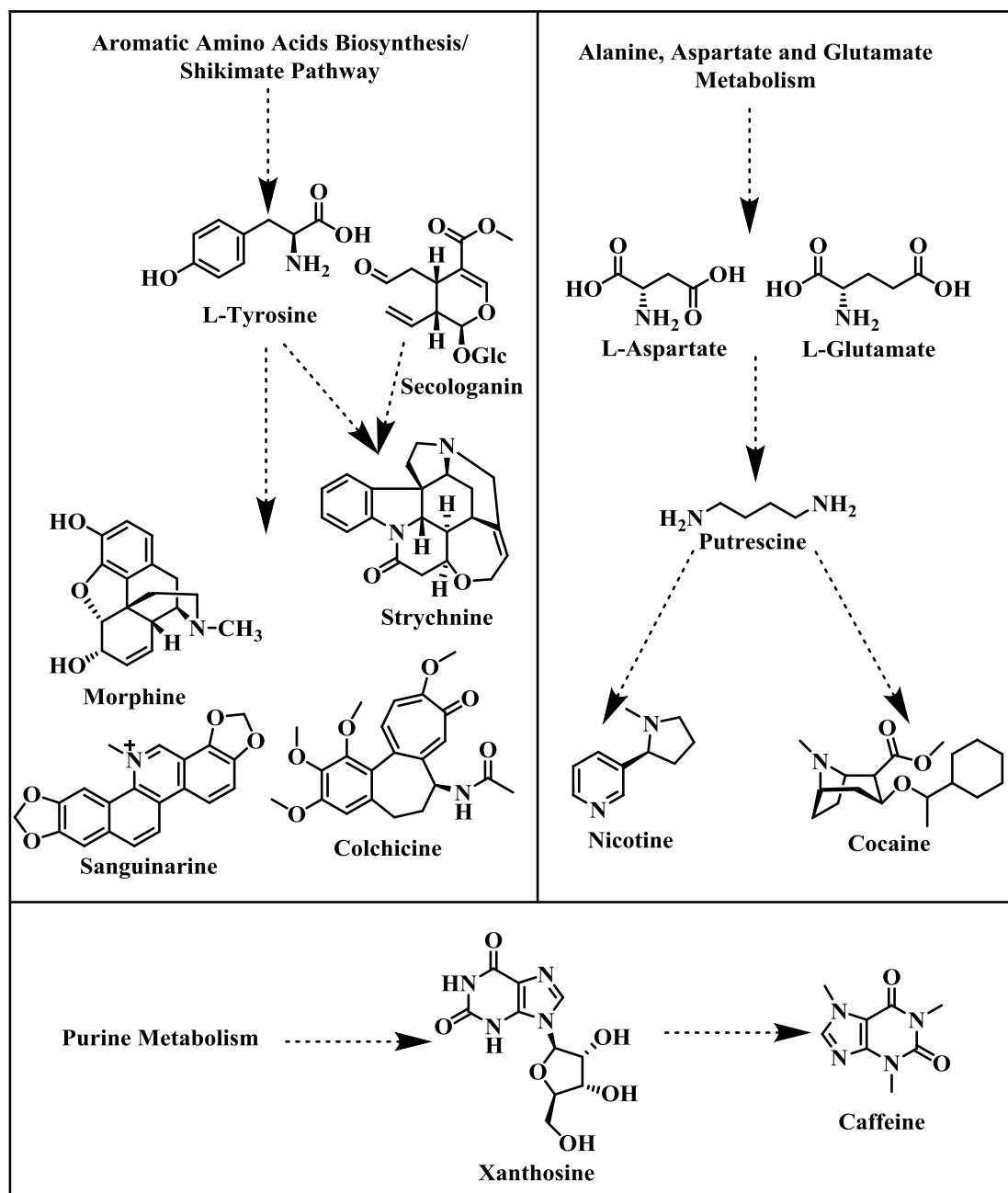
Figure 1. 5 Plant Alkaloids.

1.2.3 Isoprenoids/Terpenoids

Isoprenoids/terpenoids are the compounds which are synthesized from five-carbon isoprene units. Terpenoids are structurally and functionally most diverse group of metabolites and are classified into hemi-, mono-, sesqui- di-, ses-, tri-, tetraterpenes based on the number of isoprene units. They are of more than 75,000 compounds which are found in all domains of life and play important biological functions. They are usually linear hydrocarbons or chiral carbocyclic skeletons with different modifications like hydroxyl, ketone, aldehyde and peroxide groups. In plants, terpenoids are involved in primary metabolic processes such as membrane fluidity, respiration, photosynthesis, and regulation of growth and development. Terpenoids formed as a result of secondary metabolic processes are involved in the plant-pathogen interaction, protect plants against herbivores and pathogens, attract pollinators and seed dispersal animals[31]. Natural rubber, formed by the *cis*-polymerization of isoprene units have enormous application in polymer and medical industries[32,33]. Terpenoids such as mono- and sesquiterpenes are used as flavours and fragrances. These terpenoids have various biological properties such as anti-microbial, anti-fungal, anti-viral, chemotherapeutic, anti-hyperglycemic, anti-allergic, insecticidal, anti-parasitic, anti-inflammatory, immunomodulatory and anti-spasmodic. Recently, terpenoids have emerged as a biofuel because of branches and rings found in their hydrocarbon chain. Isopentanol and farnesane are potential alternatives for gasoline. Advanced biofuels such as ring structure containing bisabolene and pinene are used as an alternative source for jet-fuels[34].

1.3 Isoprenoids/Terpenoids Biosynthesis

Isoprenoid biosynthesis starts with condensation of isoprene units i.e. isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These isoprene units are synthesized by very well-known mevalonate pathway (MVA) or HMG-CoA reductase pathway and the recently characterized 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway, also known as the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway. These isoprene units, by condensation in the head-to-tail fashion form short or long chain diphosphates by the action of diphosphate synthases.

These diphosphates undergo cyclization or modification to respective terpenes. By the action of downstream tailoring enzymes, terpenes undergo further modifications like skeletal rearrangement, addition of hydroxyl, ketone, aldehyde and peroxide groups, acetylation, tigloylation, glycosylation to form terpenoids[35,36].

1.3.1 Mevalonate Pathway

Bloch and Lynen established the MVA pathway in the 1960s, for which they received the Nobel Prize in Physiology in 1964. MVA pathway is present in all eukaryotes (mammals, the cytosol of plants, fungi), archaea and some eubacteria. MVA pathway operates in the cytosol and the synthesized isoprene units are used for sesquiterpenoid, triterpenoid, sterols and side chains of ubiquinone[31].

In this pathway, three molecules of acetyl-CoA are utilized to form isoprene units in seven steps. MVA pathway begins with condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA by the action of acetyl-CoA acyltransferase or acetoacetyl-CoA thiolase (AACT; EC 2.3.1.9, Yeast; ERG10). AACT belongs to class II thiolase and having two conserved cysteine residues. *Arabidopsis thaliana* AACT2 mutant is lethal, indicating that AtAACT2 has an essential role in isoprenoid biosynthesis. Acetoacetyl-CoA condenses with one more acetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by the action of HMG-CoA synthase (HMGS; EC 2.3.3.10, Yeast; ERG13). In plants, HMGS shows correlation with rapid cell division and induction with methyl jasmonate, ozone, salicylic acid, which suggest that it is one of the regulatory enzymes of MVA pathway. The most important committed step is the reduction of HMG-CoA to mevalonic acid in the presence of NADPH (Nicotinamide adenine dinucleotide phosphate) as a cofactor. This reaction is catalyzed by HMG-CoA reductase (HMGR; EC 1.1.1.34, Yeast; HMG1 and HMG2), which is encoded by several paralogous genes in most plants. Plant HMGR protein is bound to endoplasmic reticulum, with catalytic domain towards cytosol. In plants, HMGR determines the flux of isoprenoid biosynthesis, regulated by different developmental and environmental signals. Mevalonate is converted into mevalonate-5-diphosphate by two-step phosphorylation catalyzed by mevalonate kinase (MK; EC 2.7.1.36, Yeast: ERG12) and phosphomevalonate kinase (PMK; EC 2.7.4.2, Yeast; ERG8) respectively. Both these kinases belong to GHMP kinase ATP-binding protein

family. The last step in MVA pathway is ATP-dependent decarboxylation of mevalonate-5-diphosphate into IPP by the action of mevalonate-5-diphosphate decarboxylase (MDC; EC 4.1.1.33, Yeast; MVD1), also named as diphospho-mevalonate decarboxylase. IPP isomerase (IPPI; EC 5.3.3.2, Yeast; IDI1) converts IPP into its isomer DMAPP and maintains equilibrium between both isoprene units and are of two types. Type I is a Zn^{2+}-dependent metalloprotein and catalyzes through carbocation intermediate[37]. Whereas type II requires Mg^{2+}, redox coenzymes FMN and NAD(P)H for its activity (Figure 1.6)[38 39].

An alternate MVA pathway was observed in most of archaea and some bacteria like *Chloroflexi* bacterium and *Roseiflexus castenholzii*. In these organisms, mevalonate-5-diphosphate is decarboxylated to isopentenyl phosphate (IP) by phosphomevalonate decarboxylase. Then IPP is synthesized by the action of isopentenyl phosphate kinase on IP[40,41]. The third alternate MVA pathway was observed in *Thermoplasma acidophilum*, in which mevalonate-3-phosphate and mevalonate-3,5-diphosphate are intermediates. The homolog of MDC (mevalonate-5-diphosphate decarboxylase) i.e, mevalonate-3-kinase converts MVA into mevalonate-3-phosphate[42]. Mevalonate-3,5-diphosphate is synthesized from mevalonate-3-phosphate by the action of mevalonate-3-phosphate-5-kinase[43]. Mevalonate-3,5-diphosphate is converted into IP by an unknown enzyme. By the action of isopentenyl phosphate kinase, IP is converted to IPP (Figure 1.6)[44].
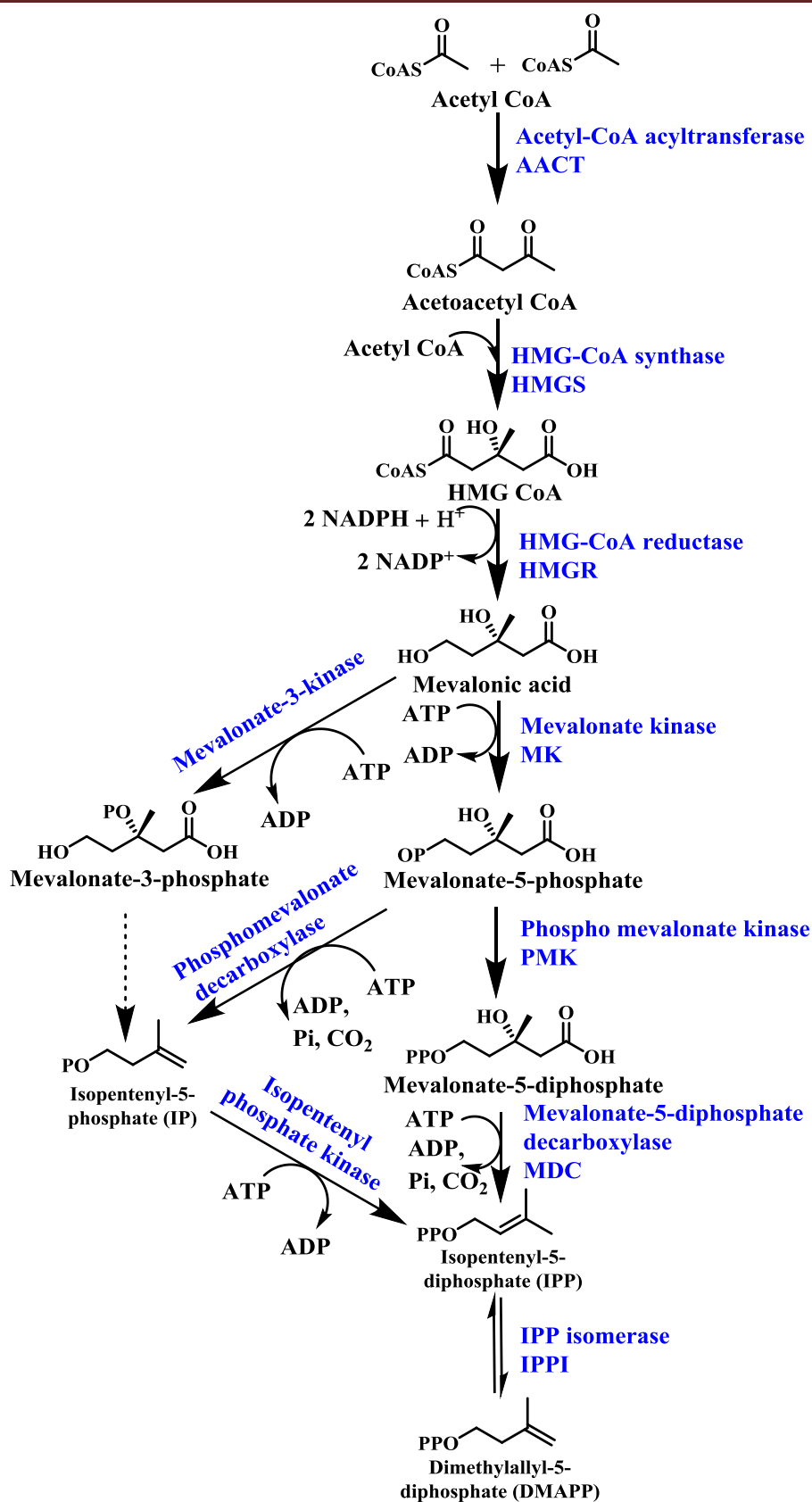
Figure 1. 6 Mevalonate (MVA) Pathway.

1.3.2 Methyl Erythritol Phosphate (MEP) Pathway/ 1-deoxy-D-xylulose-5-phosphate (DXP) Pathway/ Non-mevalonate Pathway

For several decades, MVA pathway was believed to be the unique source for isoprene units. In the 1990s, Rohmer and Arigoni's research lead to the proposition of an alternate 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose-5-phosphate (MEP/DXP) pathway, in which glyceraldehyde-3-phosphate (GAP) and pyruvate are the starting metabolites to synthesize IPP and DMAPP. MEP pathway is present in most of the bacteria, plastids of plants and green algae. In plants, MEP pathway genes have N-terminal plastid targeting sequence[45].
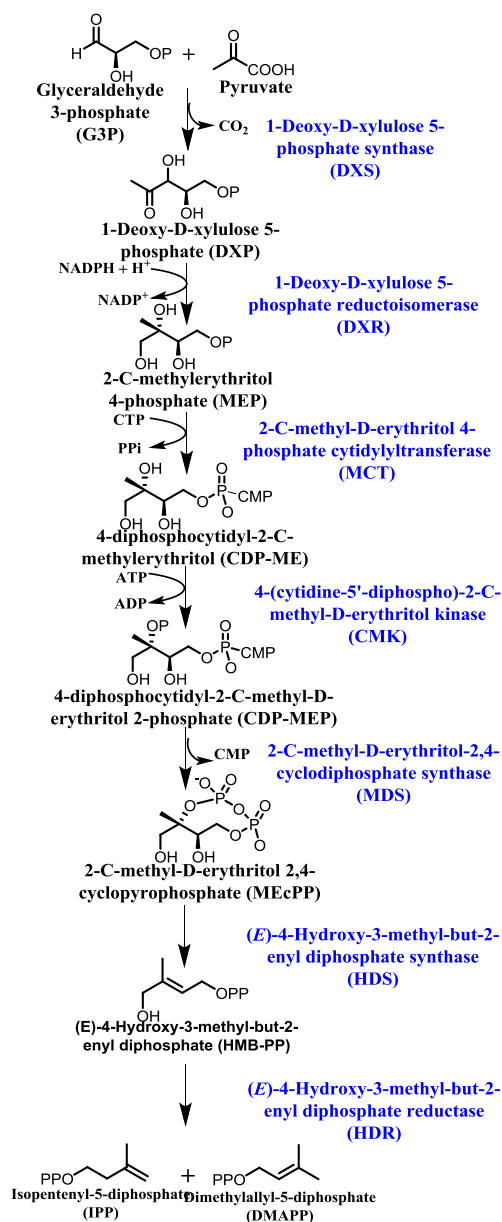
Figure 1. 7 Methyl Erythritol Phosphate (MEP) Pathway.

The first step in MEP pathway is condensation of pyruvate and GAP via transketolase-like decarboxylation reaction to form 1-deoxy-D-xylulose-5-phosphate (DXP), catalyzed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS; EC 2.2.1.7). Synthesis of DXP is irreversible and committed step for MEP pathway. DXS controls the flow of carbon in MEP pathway by negative feedback inhibition from IPP and DMAPP[46,47]. Herbicide, ketoclomazone inhibits the activity of DXS whereas, fosmidomycin blocks the activity of next step enzyme, DXP reductoisomerase and hence can be used as an anti-bacterial or anti-malarial agent. Further, intermolecular rearrangement and reduction of DXS into 2-C-methyl-D-erythritol-4-phosphate (MEP) is catalyzed by DXP reductoisomerase (DXR; EC 1.1.1.267), in the presence of NADPH. MEP is converted to 4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME) in a CTP dependent reaction catalyzed by 2-C-methyl-D-erythritol-4-phosphate cytidylyltransferase (MCT; EC 2.7.7.60), which requires Mg^{2+} for its catalytic activity. In bacteria, DXR and MCT have phosphorylation sites which when mutated to aspartate or glutamate (to mimic phosphothreonine) lead to the reduction of their activity. 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK; EC 2.7.1.148) belonging to GHMP kinase family, catalyzes the phosphorylation of 2-hydroxyl group in CDP-ME into 2-phospho-4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol (CDP-MEP). CMK requires ATP and Mg^{2+} for its activity. This mechanism might be involved in regulation of carbon flow in MEP pathway. Cyclic intermediate, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (MEcPP) formed by elimination of CMP from CDP-MEP and is catalyzed by 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MDS; EC 4.6.1.12)[48]. In bacteria, the accumulation of MEcPP is observed during oxidative stress caused by benzyl viologen or other redox mediators, suggesting it is an important antistressor[49]. In plants, during stresses (high light or wounding), MEcPP in the leaves increase, which triggers stress-inducible plastidial protein such as hydroperoxide lyase (HPL)[50]. MEcPP is then reduced to 4-hydroxy-3-methyl but-2-enyl diphosphate (HMBPP) by the action of HMBPP synthase (HDS; EC1.17.7.1). HDS belongs to GCPE protein family and requires flavodoxin/flavodoxin reductase/NADPH system. In photosynthetic tissues, HDS takes electrons directly instead of NADPH to reduce MEcPP. HMBPP has converted into IPP and DMAPP in a 5:1 ratio by HMBPP reductase (HDR; EC 1.17.1.2). HDR

belongs to LytB protein family and requires FAD, NADPH and divalent cation for its activity. Both HDS and HDR contains redox-active $[4Fe-4S]^{2+}$ cluster suggesting the involvement of radical intermediates (Figure 1.7)[51].

1.3.3 The crosstalk between MVA and MEP Pathways

The MVA and MEP pathways are compartmentalized in different cellular locations in plants. The MVA pathway takes place predominantly in cytoplasm and mitochondria synthesizing triterpenoids, sterols, certain sesquiterpenes and ubiquinones. The MEP pathway localized in plastid to synthesize hemiterpenes, monoterpenes, diterpenes (Figure 1.8). Systematic analysis of 130 isoprenoid biosynthesis from 86 plants species revealed that triterpenoids, sterols, carotenoids (tetraterpenoids) and phytol chain (diterpenoids) synthesis is strictly compartmentalized[52]. But monoterpenes, diterpenes, hemiterpenes, sesquiterpenes and polyterpenes may be synthesized from both pathways under specific environmental or ecological conditions[53]. Some studies show that cross-talk between the two pathways depends on the supply of precursor, end product and cultivation conditions[54, 55]. Plastid membranes possess Ca^{2+} gated unidirectional proton symport systems present to export isoprenoid intermediates involved in the crosstalk[56].
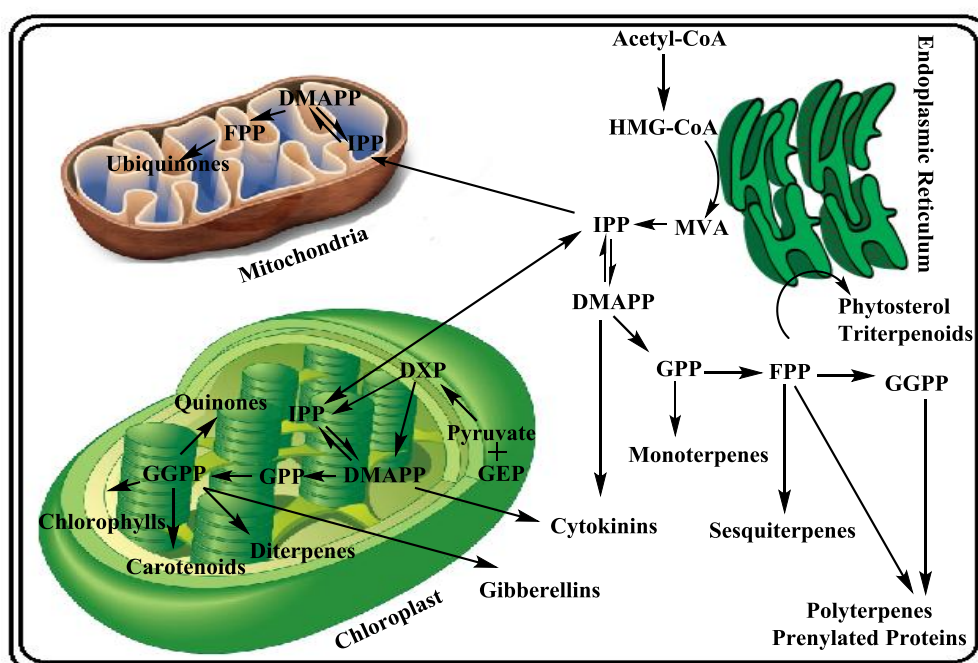
Figure 1. 8 Crosstalk between MVA and MEP pathways.

1.3.4 Regulation of MVA and MEP Pathways

Isoprene biosynthetic pathways are regulated at different levels such as transcription, translation, posttranslational modification and protein degradation. In bacteria, MEP pathway is regulated at transcription level as operon model. In yeast (*Saccharomyces cerevisiae*), during aerobic growth heme accumulation activates the Hap1P transcription factor, which further increases *HMGR1* transcription. Sterol and GGPP trigger endoplasmic reticulum (ER) associated degradation pathway to recognize and ubiquitinate HMGR2, thereby controlling protein turnover[57]. In mammals, when sterol levels are low, Insulin-induced gene 1 (INSIG-1) is released from SCAP (SREBP cleavage-activating protein) on ER membrane. Then SCAP forms a complex with SREBP (sterol regulatory element binding protein) and moves to the Golgi apparatus. The SREBP is then activated by cleaving off amino-terminal domain which releases SREBP from the membrane and allows it to enter the nucleus and activates *HMGR* transcription. When sterol levels are high, INSIG forms complex with SCAP and prevents its transportation to Golgi apparatus and proteolytic cleavage-dependent activation of SREBP. Along with this, INSIG binds to an amino-terminal domain of HMGR and recruits the enzymes which ubiquitinate HMGR, triggering its degradation (Figure 1.9A)[57,58].

Regulation of isoprene unit biosynthesis in plants is highly complex since both MVA and MEP pathways occur in a cell, and also due to other factors such as several pathway enzymes contain isozymes and a wide range of environmental stimuli are involved in plants. In *Arabidopsis thaliana*, both MVA and MEP pathway genes are expressed in all the tissues and developmental stages. In seedling and mature plants, MEP pathway genes are active primarily in photosynthetic tissues. Whereas, in the case of MVA pathway genes, they are most strongly expressed in radical, hypocotyl, roots, flowers and seeds[59]. In plants, isozymes of MVA and MEP pathways are evolved by gene duplication and may facilitate adaptation in dynamic environments[60,61]. *Cis*-regulatory motifs in the promoters of MVA pathway genes helps for high expression in flowers and oxidative stress conditions. MEP pathway genes are having *cis*-regulatory motifs related to light and circadian clock[31]. In *A. thaliana*, *DXS* and *HDR* expression follow a circadian pattern, reaching high mRNA levels during the day and decreasing levels during the night time (Figure 1.9B)[62,63].

ORCA3 (AP2 family) and AGL12 (MADS-box) transcription factors control tissue and cell-specific expression of *DXS* in *Catharanthus roseus*[64]. In tobacco, MEK2-SIPK/WIPK cascade activates *HMGR* expression during pathogen defence[65]. Sucrose nonfermenting 1 (SNF1)–related kinase 1 (SnRK1) phosphorylates a conserved Ser (Ser-577 in AtHMGR1S) and inactivates the catalytic domain of AtHMGR1 *in vitro*. PRL1 is a conserved WD protein and functions as a global regulator of sugar, stress, and hormone responses, by inhibiting SnRK1 activity[66,67]. A mitochondrial pentatricopeptide repeat protein LOI1 (Lovastatin insensitive 1) has been identified as a negative regulator of HMGR protein level[68].
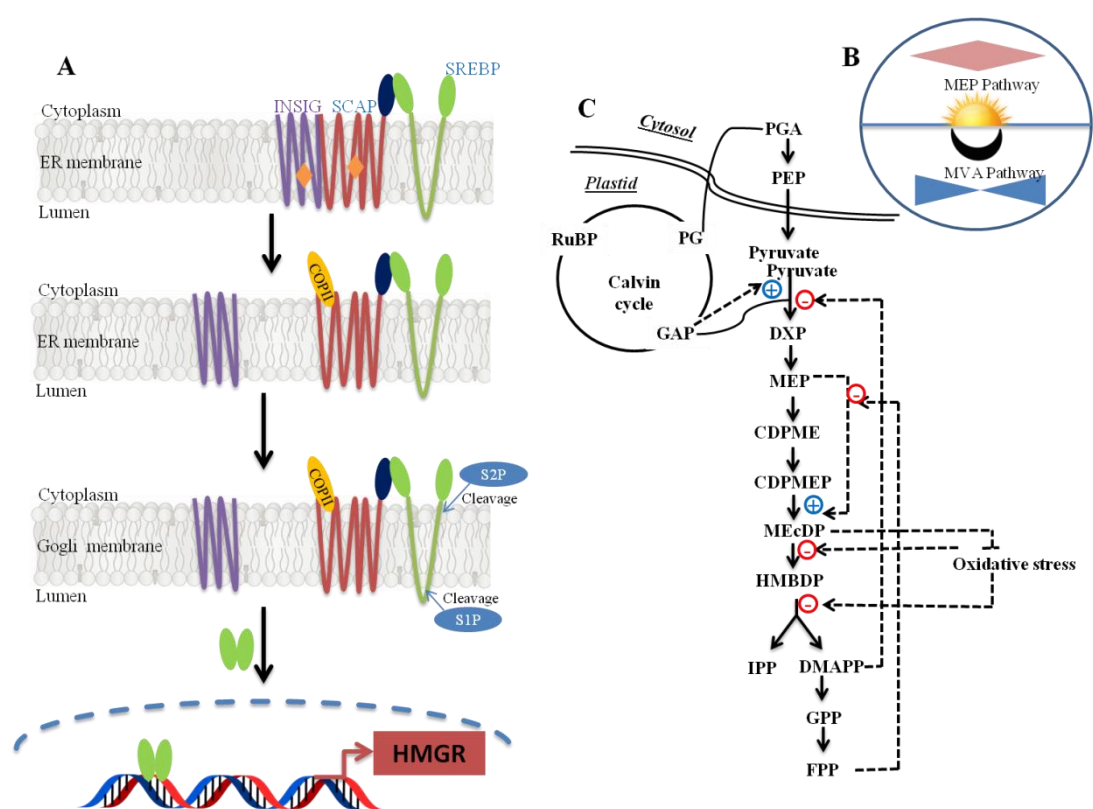
Figure 1. 9 Regulation of MVA and MEP Pathways in Mammals and Plants.
A) MVA Pathway regulation in mammals. B) Circadian pattern of MVA and MEP pathways in plants. C) MEP pathway regulation in plants.

1.3.5 Prenyl Diphosphates Synthases

Prenyl diphosphate synthases catalyze head-to-tail condensation and chain elongation reactions of DMAPP and IPP to produce prenyl diphosphates like geranyl diphosphate (GPP; C_{10}), farnesyl diphosphate (FPP; C_{15}), geranylgeranyl diphosphate

(GGPP; C_{15}) and other diphosphates. The discovery of IPP (biologically active isoprene unit) by two individual groups Lynen and of Bloch in 1958 lead to several enzymological studies[69,70]. Cornforth and Popják established the stereochemical details of the biosynthetic pathway from mevalonic acid to squalene[71,72]. Poulter and Rilling have established ionization-condensation-elimination mechanism by using fluorinated substrate analogues[73]. Molecular biology work related to prenyl diphosphate synthases started with the isolation of cDNA for FPP synthase from rat liver[74]. First prenyl diphosphate synthases crystal structure (Avian FPP synthase) is reported in 1994 by Poulter and Sacchettini[75]. Based on subunit composition, chain length and stereochemistry of final product, prenyl diphosphate synthases are classified into four groups; i) short-chain prenyl diphosphate synthases, ii) medium-chain prenyl diphosphate synthases, iii) long-chain (*E*)-prenyl diphosphate synthases and iv) (*Z*)-polyprenyl diphosphate synthases. First three groups belong to (*E*)-prenyl diphosphate synthases (Figure 1.10).

1.3.5.1 Short-Chain Prenyl Diphosphate Synthases

Short-chain prenyl diphosphate synthases include geranyl diphosphate synthase (GDS), farnesyl diphosphate synthase (FDS) and geranylgeranyl diphosphate synthase (GGDS). GDS catalyzes head to tail condensation of DMAPP and IPP to form GDP (C_{10}). GPP acts as a universal precursor for monoterpenes. GDS exists as homodimeric and heteromeric structures. Heteromeric GDS contains smaller subunit (SSU) and lager subunit (LSU). SSU is catalytically inactive in nature but the catalytic activity of LSU produces GGPP. When SSU interacts with LSU only GPP results as the sole product. FDS catalyze the condensation of allylic diphosphates (DMAPP, GPP) with IPP to form FPP, which act as a precursor for sesquiterpenes, steroids and triterpenoids. GGDS catalyzes the condensation of allylic diphosphates (DMAPP, GPP and FPP) with IPP to form GGPP, which then act as a precursor for the synthesis of diterpenoids, vitamin E and carotenoids[76].

1.3.5.2 Medium-Chain Prenyl Diphosphate Synthases

This class is composed of enzymes which are involved in the synthesis of prenyl diphosphates with length ranges from C_{25} to C_{35}. Bacteria is a good source for prenyl diphosphate synthases, mainly for the synthesis of medium and long-chain

diphosphates which are involved in the synthesis of ubiquinones. Geranylfarnesyl diphosphate synthase (EC 2.5.1.81) catalyzes the condensation of allylic diphosphates (DMAPP, GPP, FPP and GGPP) with IPP to form geranylfarnesyl diphosphate C_{25}, which acts as a precursor for archaeal C_{25}-C_{25} ether lipids and methanophenazine (an electron carrier utilized for methanogenesis)[77,78]. Hexaprenyl diphosphate (EC 2.5.1.33) synthase (*trans*-pentaprenyl*trans*transferase) isolated from bacteria (*Micrococcus luteus*) is capable of synthesizing (*all-E*)-hexaprenyl diphosphate C_{30}. Similarly, heptaprenyl diphosphate synthase (EC 2.5.1.30) from *Bacillus subtilis* synthesize (*all-E*)-heptaprenyl diphosphate C_{35}. Both hexaprenyl diphosphate synthase and heptaprenyl diphosphate synthase are two protein subunits; both subunits are required for the catalytic activity[79]. Both these C_{30} and C_{35} prenyl diphosphates are involved in the synthesis of coenzyme Q, menaquinone having polyprenyl side chain length of C_{30} and C_{35}[80].

1.3.5.3 Long-Chain Prenyl Diphosphate Synthases

The enzymes which synthesize prenyl diphosphate with length more than C_{35} belong to long-chain prenyl diphosphate synthases. (*all-E*)-octaprenyl diphosphate (C_{40}) synthase (EC 2.5.1.90), (*all-E*)-solanesyl diphosphate (C_{45}) synthase (EC 2.5.1.84) and (*all-E*)-decaprenyl diphosphate (C_{50}) synthase (EC 2.5.1.91) are characterized from *Escherichia coli*, *Micrococcus luteus* and *Paracoccus denitrificans*, respectively. All long-chain prenyl diphosphates are involved in ubiquinone biosynthesis. The long-chain prenyl diphosphate synthases are homodimers but require a protein factor to maintain efficient catalytic turnover[81,82].

1.3.5.4 (*Z*)-Prenyl Diphosphate Synthases

(Z)-prenyl diphosphate synthases or *cis*-prenyltransferases catalyze the condensation of IPP with allylic diphosphate to form long-chain prenyl diphosphates, which can occur in the membrane fraction of bacteria and a microsomal fraction of eukaryotic cells. *Cis*-prenyl diphosphates are the precursors of polyprenyl lipids required as carbohydrate carriers in the biosynthesis of the bacterial cell wall and of eukaryotic glycoproteins. In *H. brasiliensis*, it helps in the biosynthesis of rubber which contains around 5000 isoprene units. *Mycobacterium tuberculosis* contains *E,Z*-farnesyl diphosphate synthase (EC 2.5.1.68) to produce *E,Z*-farnesyl diphosphate

from the condensation of allylic diphosphates (DMAPP, *E*-GPP) with IPP. Decaprenyl diphosphate synthase catalyzes the chain elongation of *E,Z*-farnesyl diphosphate with seven IPP units to synthesize *E,Z*-decaprenyl diphosphate C_{50}[83,84]. Undecaprenyl diphosphate synthase (EC 2.5.1.31) catalyzes the condensation of eight IPP units with *E*-farnesyl diphosphate to form undecaprenyl diphosphate[85]. The (*Z*)-prenyl diphosphate synthase present in eukaryotic cells was dehydrodolichyl diphosphate synthase (EC 2.5.1.87) which catalyzes the chain elongation of *E*-farnesyl diphosphate with 11 to 20 IPP units in the *cis* conformation. Dehydrodolichyl diphosphate undergoes single dephosphorylation to form dolichyl phosphate which acts as sugar carrier lipid in the biosynthesis of N-glycosylated proteins and GPI-anchored proteins[86]. Rubber *cis*-polyprenyl*cis*transferase synthase (EC 2.5.1.20) has been found and characterized from *Hevea brasiliensis* and *Parthenium argentatum* which catalyzes the transfer of (*Z*)-polyprenyl diphosphates to IPP with the elimination of diphosphate[87]. In general, (*Z*)-polyprenyl diphosphates prefer the synthesis of long-chain diphosphates[88].
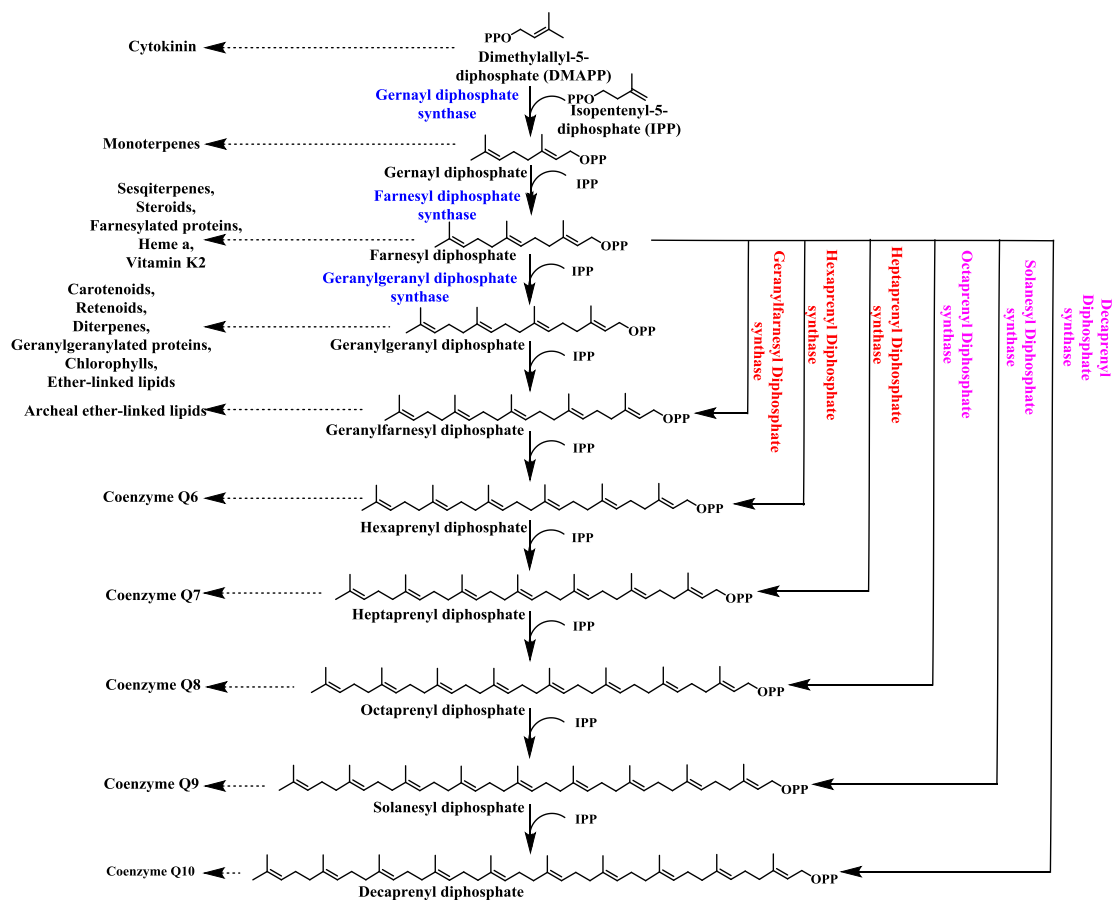
Figure 1. 10 Prenyl Diphosphate Synthases.

1.3.5.5 Structure and Mechanism of Prenyl Diphosphate Synthases

(*E*)-prenyl diphosphate synthases contain two conserved aspartate-rich motifs (DDxxD motif). The crystal structure of FDS clearly shows that first DDxxD motif binds with allylic diphosphate and the second motif binds with IPP. FDS contain 13 α-helices, out of which 10 are surrounding the large central catalytic cavity. Fourth and fifth position to first DDxxD motif determines which type of prenyl diphosphate has to be synthesized. GGDS contain aromatic amino acid in the fifth position but FDS contains aromatic residue in fourth and fifth positions[75,89]. The conserved regions of (*Z*)-prenyl diphosphate synthases differ from (*E*)-prenyl diphosphate synthases (Figure 1.11). Crystal structure of undecaprenyl diphosphate synthase from *Micrococcus luteus* B-P 26 and *E.coli* shows that (*Z*)-prenyl diphosphate synthases contain 20-23 α-helices and 12 parallel β-sheets. Two α-helices and 2-4 β-sheets surrounding an active site which is a large elongated hydrophobic cleft[90,91].

Z/Cis-Prenyl Diphosphate Synthase

Region 1	Region 2	Region 3
HxxxMDGNxRxA	GHxxG	TxxAFSxENxxRPAQEVxxLMxL

Region 4	Region 5
AxxYGGRxEIxxA	DLxIRTxGExRxSNFKKWQxxYxExxFxxxLWPxF

E/Trans-Prenyl Diphosphate Synthase

Region 1	Region 2	Region 3
GGxxKRxRPxV	ExIHxxSLVH**DD**xx**D**xxDxRRGxPT	GDxxxLxxxA

Region 4	Region 5
GQxxD	KTAxLxASxxxGA

Region 6	Region 7
GxxFQIx**DD**xx**D**xxxxxxxLGKPxGxD	LxxxxxxxxxR
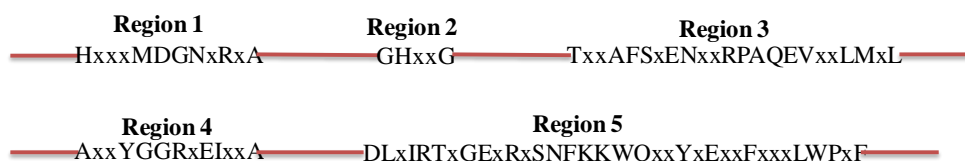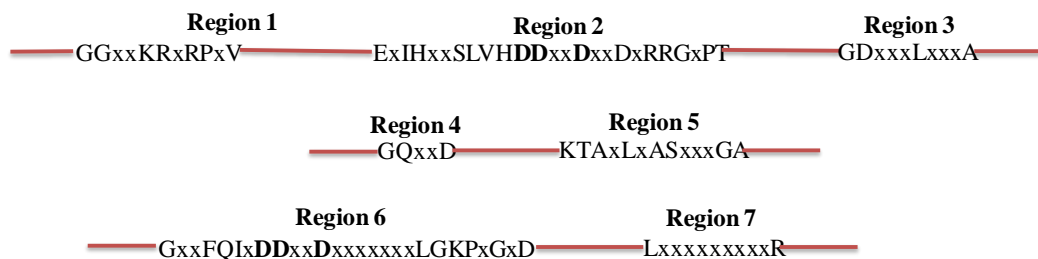
Figure 1. 11 Conserved Regions of (*E*)-Prenyl Diphosphate and (*Z*)-Prenyl Diphosphate Synthases.

An ionization–condensation–elimination mechanism has been proposed for (*E*)-prenyl diphosphate synthases. The catalytic reaction starts with cleavage of a carbon-oxygen bond in allylic prenyl diphosphate to form a cation. Addition of single IPP units occurs with the stereospecific removal of a proton from 2-position, allowing

a new C-C bond and new double bond in the product. The only difference is that in (*E*)-prenyl diphosphate synthases, a *pro-R* proton is eliminated from 2-position of IPP but in (*Z*)-prenyl diphosphate synthases *pro-S* proton is eliminated (Figure 1.12)[73,87,92].
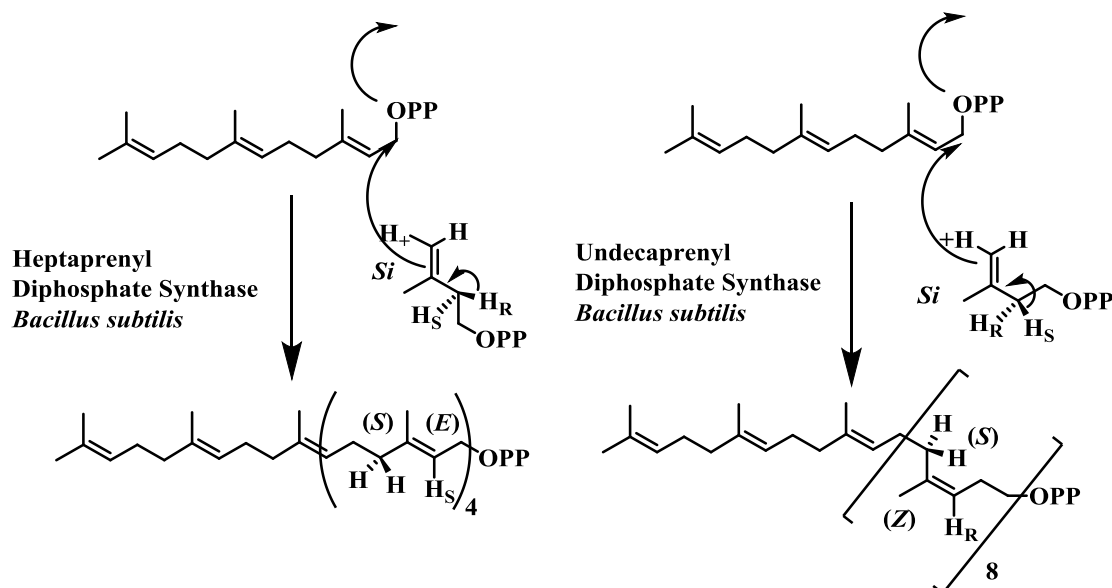
Figure 1. 12 Absolute Stereochemistry of (*E*)-(left) and (*Z*)-Prenyl Chain (right) Elongation Reactions by Preynl Diphosphate Synthases.

3-methyl-1-butyl is the basic unit of isoprenoid carbon skeletons, which are joined in nine different patterns. In four different ways, C1′ unit of 3-methyl-1-butyl joins with single carbon of another unit (1′-1, 1′-2, 1′-3, and 1′-4). In three different ways, C1′ unit joins in a cyclic structure manner with another unit (C1′-1-2, C1′-2-3, and C1′-2-3-2′) and in one case, C1′ unit inserted between other units (2-1′-3). The 4′-4 pattern is observed in archaeal lipids. The 1′-4 linkage is the most commonly seen in isoprenoid compounds, and this attachment between isoprene units are called as head-to-tail or regular linkage. Other patterns are designated as non-head-to-tail or irregular linkage. In regular isoprenoids, the chain elongation proceeds by a dissociative electrophilic alkylation of the double bond in IPP by the allylic cations generated from DMAPP/other diphosphates. In irregular isoprenoids, dissociative electrophilic alkylation of the double bond in DMAPP by the dimethylallyl cation results in a protonated cyclopropane intermediate, which was observed in first pathway-specific step in the biosynthesis of sterols and carotenoids. The

cyclopropane intermediate may undergo cyclization to form a cyclobutylcarbinyl cation, which has been observed in mealybug mating pheromones (Figure 1.13)[93].

A

1′-1

1′-2

1′- 3

1′- 4

2-1′-3

4′- 4

B

c1′-1-2

c1′-2-3

c1′-2-3-2′

Chain Elongation

IPP

DMAPP

GPP

Cyclopropanation, Branching, Cyclobutanation

DMAPP

DMAPP

(1R,3R)-CPP

(1R)-MPP

(1R,3R)-PPP
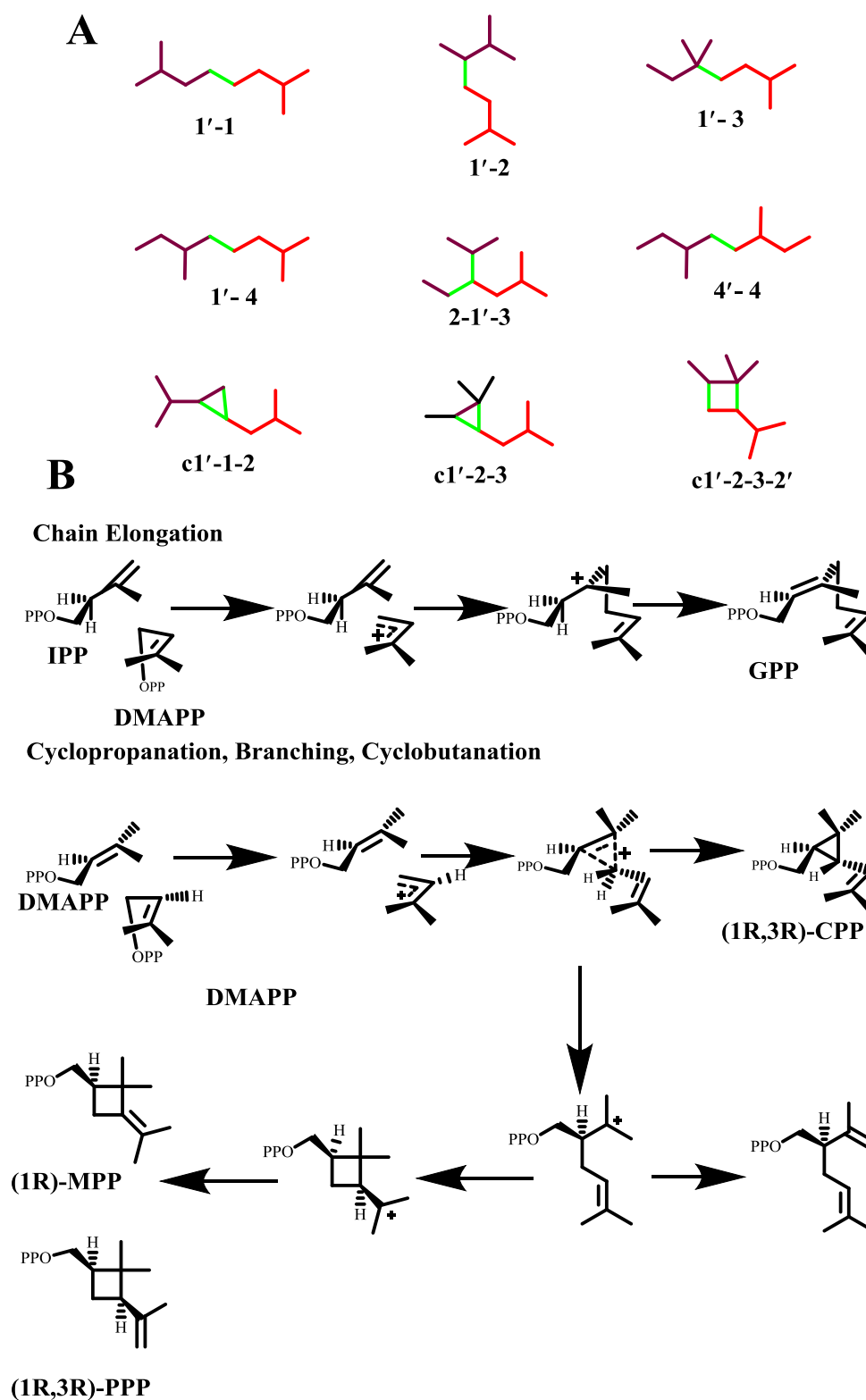
Figure 1. 13 Different Ways for Isoprene Units Joining During Chain Elongation.

A) Different ways of connecting isoprene units. Red and black colour indicates isoprene units and green indicated bonds joining isoprene units. B) Electrophilic alkylation mechanism for chain elongation, cyclopropanation, branching, and cyclobutanation.

1.4 Classification of Isoprenoids

Isoprenoids are divided into different classes based on carbon number and precursor allylic diphosphate.

1.4.1 Hemiterpenoids

Hemiterpenoids are the smallest known terpenoids that are synthesized from the C_5 isoprene units i.e. IPP or DMAPP and are of volatile in nature. Most of the plants emit the hemiterpenes such as isoprene and methylbutenol (MBO). These two compounds are synthesized from DMAPP in chloroplasts[94]. Methylbutenol synthase is a bifunctional enzyme that produces both MBO and isoprene in a ratio of ~90:1. Plants that emit isoprenes are able to tolerate sunlight-induced rapid heating of leaves (heat flecks)[95]. Isoprene moiety transfers to ATP or ADP and undergoes hydroxylation to synthesize zeatin, a precursor for cytokinins. These cytokinins act as hormones for plant development[96].
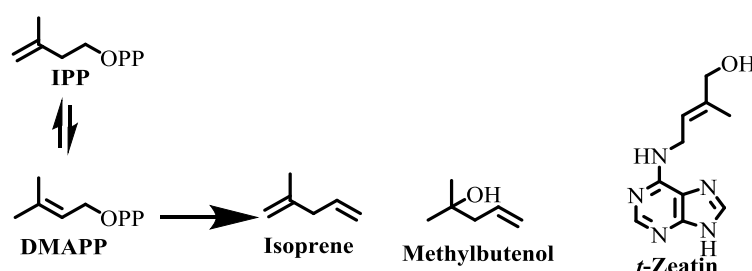
Figure 1. 14 Hemiterpenoids.

1.4.2 Monoterpenoids

Monoterpenoids are acyclic (linalool, myrcene and ocimene), or monocyclic (limonene, carveol and carvone) or bicyclic (camphene, sabinene and pinene) compounds synthesized from GPP catalyzed by monoterpene synthases. GPP binds to active site of monoterpene synthases as a complex with the divalent metal ion. GPP undergoes ionization to from linalyl cation which undergoes further isomerization and cyclization to form diverse monoterpenes[97]. In plants, monoterpenoids are synthesized through MEP pathway[98]. The storage sites are oil glands, glandular hairs

and trichomes. These compounds play a key role in herbivore defence. Most of the trees in forests release α-pinene into the troposphere, which plays a key role in ozone balance. Monoterpenoids also have commercial values like industrial raw materials, essential oil for perfumery and flavours[99].
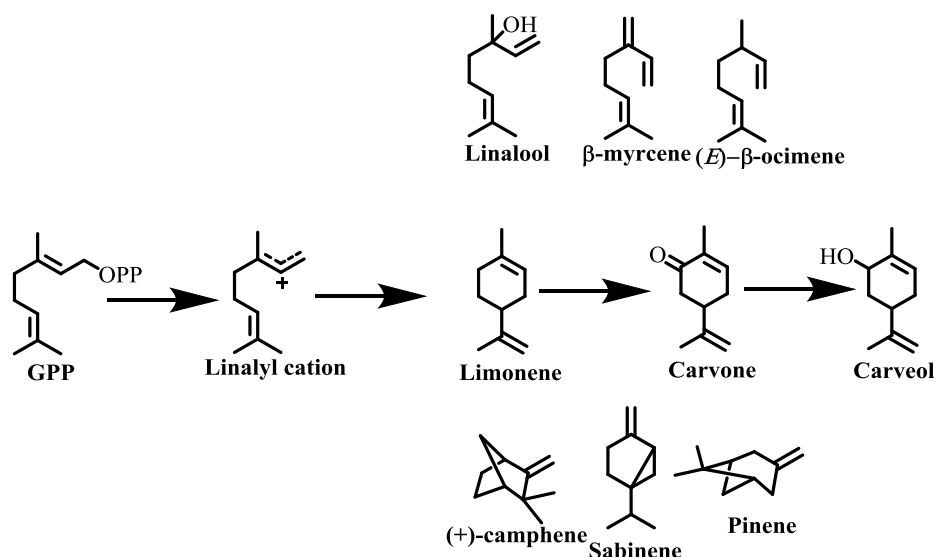
Figure 1. 15 Monoterpenoids.

1.4.2 Sesquiterpenoids

Sesquiterpenoids are one of the most diverse class of isoprenoids with more than 300 carbon skeletons and more than 7000 characterized compounds[100]. Sesquiterpenoids and their lactones have biological properties like anticancer (michelenolide and micheliolide[101]), cytotoxic, antiviral (triptofordin C-2[102]), antimalarial (artemisinin[103]) and antibiotic. These compounds are very well known for flavours and aromas. Sesquiterpenoids are synthesized by intramolecular cyclization of FPP. Sesquiterpenes can be divided into four groups: acyclic (farnesol, nerolidol), monocyclic (bisabolene, germacrene, curumene, zingiberene), bicyclic (selinene, β-santalene, bergamotene, aristolochene, valencene, caryophyllene) and tricyclic (α-cedrene, santalene, longifolene). In plants, sesquiterpenoid biosynthesis takes place in the cytosol through the MVA pathway[63]. Farnesene and bisabolene skeletons are used as advanced biofuels[34].
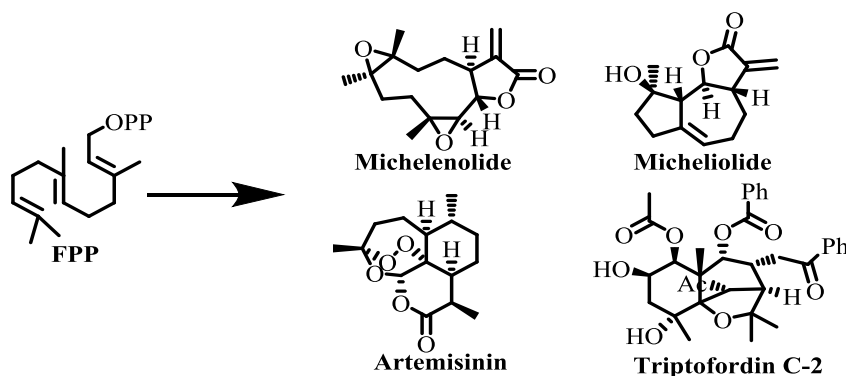
Figure 1. 16 Sesquiterpenoids.

1.4.3 Diterpenoids

Plants produce thousands of diterpenoids which are involved in primary and secondary metabolisms. Gibberellins (phytohormones) and phytol chains of chlorophyll are examples of diterpenes involved in primary metabolism. Most of the diterpenoids are secondary metabolites and limited to some specific plant taxonomy and acts as a signature for some plants. These metabolites are involved in defence against pest, pathogen (rhizathalene, phytoalexins, diterpene resin acids) and herbivores (17-hydroxygeranyllinalool glycosides). Major commercial diterpenoids include anti-cancer drug taxol (*Taxus brevifolia*), sclareol (*Salvia sclarea)* as a precursor for fragrance and fixatives in perfume manufacture, steviol (*Stevia rebaudiana*) as a natural sweetener, and diterpene resin acids (conifer trees) as feedstock for industrial coatings and inks[104]. MEP pathway provides the isoprene units for the synthesis of diterpenoids[35,97].
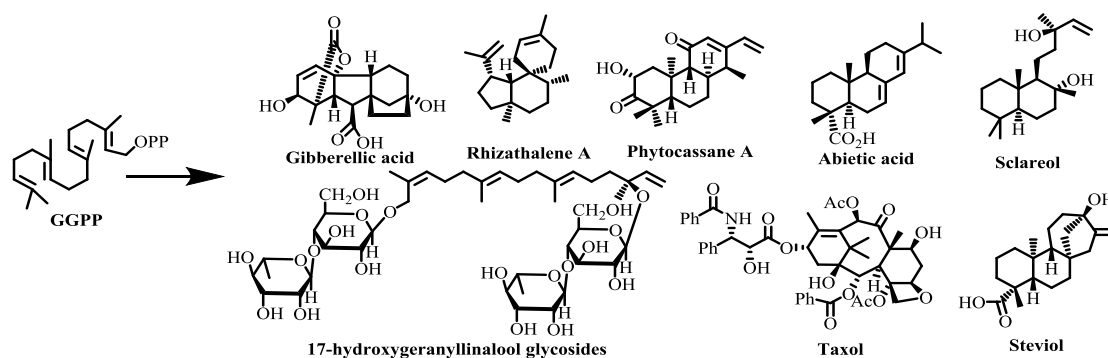
Figure 1. 17 Diterpenoids.

1.4.4 Triterpenoids

In plants, thousands of triterpenoids are identified, and they are synthesized from more than 100 different carbon skeletons. Triterpenoid biosynthesis takes place in the cytosol and endoplasmic reticulum. All triterpenoids are synthesized from cyclization of a common substrate, (*3S*)-oxidosqualene/2,3-oxidosqualene by triterpene synthases[105]. The cyclization of (*3S*)-oxidosqualene is activated by the cationic attack on the 2,3-peroxide bond, which is followed by the cascade of cation-olefin cyclizations resulting in cyclic carbocation. This cation undergoes 1,2 proton and methyl shifts and is finally the stable triterpene usually arises by deprotonation[106]. Triterpenoids are further classified into monocyclic (achilleol A, camelliol C), dicyclic (polypoda-7,13,17,21-tetraen-3-ol), tricyclic (Malabarica-14(27),17,21-trien-3-ol, malabarica-17,21-dien-3,14-diol), tetracyclic (lanosterol, cycloartenol, parkeol, cucurbitadienol, tirucallol, tirucalla-7,24-dien-3β-ol, euphol, butyrospermol, shionone) and pentacyclic (lupeol, α-amyrin, β-amyrin, taraxerol, isomultiflorenol, friedelin, Ψ-taraxasterol)[107]. Plants also accumulate triterpenoids in the form of glycosylated form, saponin. Lanosterol and cycloartenol are further modified into steroids and hormones which play a key role in the development and function of organisms. Other triterpenoids and saponins play a key role in plant defence against microbes, insects and herbivores. Triterpenoids have biological properties like anti-inflammatory, anti-cancer, anti-viral, and insecticidal. Furthermore, saponin, glycyrrhizin is used as natural sweetener[108].

The first committed step in triterpenoid biosynthesis is cyclization of 2,3-oxidosqualene which is catalyzed by triterpene synthases (TTS). In plants, microalgae and many protozoa have cycloartenol synthase (CAS), whereas animals and fungi have lanosterol synthase for the synthesis of steroids. In the active site of triterpene synthases, 2,3-oxidosqualene forced to take up a pre-organized conformation. The protonation of epoxide ring starts a cascade cyclization through intermediate carbocations. Aromatic amino acids present in the active site stabilize the intermediate carbocations through cation–π interactions. Skeletal rearrangement of a carbocation by 1,2 methyl and 1,2 proton shifts (from high to low π electron density) takes place and finally, deprotonation occurs to form the triterpene cyclic product[106-108].
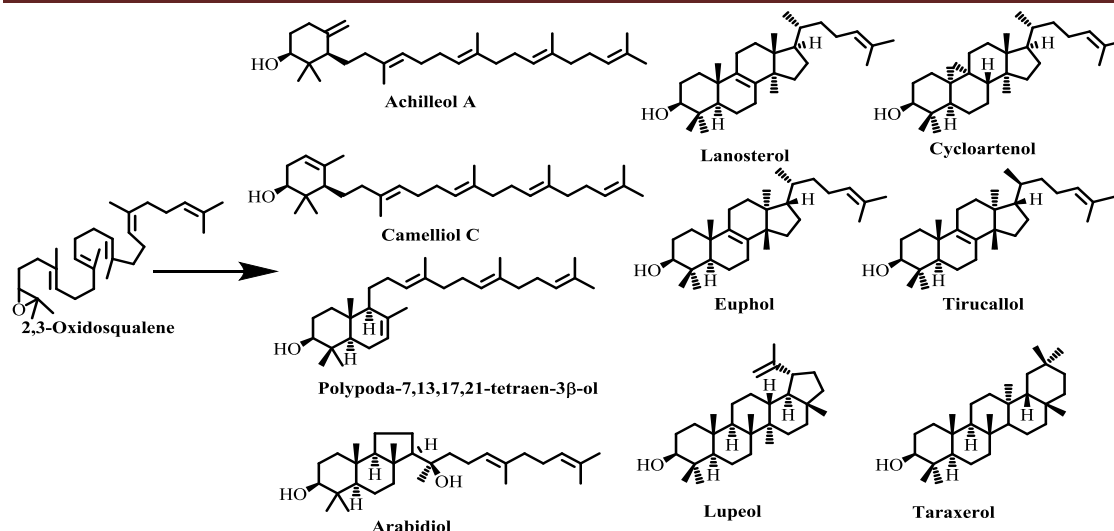
Figure 1. 18 Triterpenoids.

1.4.5 Tetraterpenoids

Tetraterpenoids are C_{40} compounds derived from phytoene formed by the head-to-head condensation of two molecules of GGPP[35]. Carotenoids are important metabolites in this group. In plants, carotenoids are involved in various biological processes like photoprotection, photomorphogenesis, photosynthesis, precursors for plant isoprenoid volatiles, phytohormones (abscisic acid and strigolactones) and colourants. MEP pathway provides a large part of carbon flux for carotenoid biosynthesis[109]. Carotenoids can be classified further as carotenes (β-carotene, α-carotene and lycopene) and xanthophylls (lutein, β-cryptoxanthin, zeaxanthin)[110]. Animals are incapable of producing carotenoids and provitamin A from plants acts as a precursor for the synthesis of retinol, retinoid in animals. These are critical antioxidants in the human diet. β-carotene, β-apo-8'-carotenal, and canthaxanthin are widely used as colouring agents for food[111].
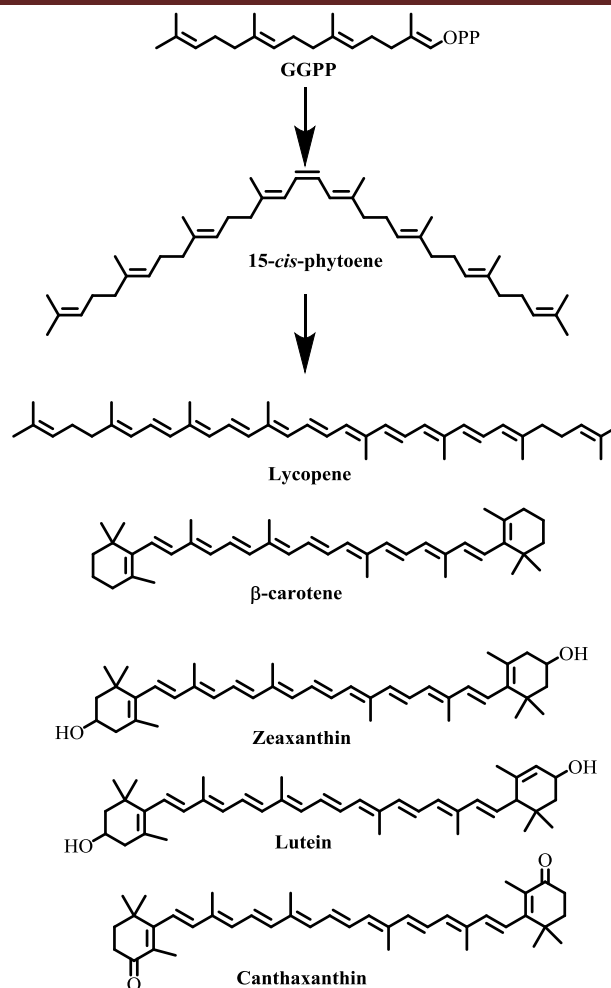
Figure 1. 19 Tetraterpenoids.

1.4.6 Polyprenoids

Polyisoprenoid alcohols are linear polymers of isoprene, consisting of 6-100 units with a terminal primary alcohol group. The reduction of a double bond at α-residue gives rise to dolichols. Polyprenoids can be classified into di-*trans*-poly-*cis* (dolichols, bacteria and plants origin), tri-*trans*-poly-*cis* and all-*trans* (solanesol, spadicol) [112]. The archaeal membrane contains polyprenyl through an ether bond, which enables the organisms to survive in extreme conditions. Polyprenoids are involved in cell wall synthesis of bacteria and yeast. Dolichols are present in the membrane and are involved in N-glycosylation of proteins[113,114]. Polyprenoids contain properties like antitumor, antithrombic and antiviral activities[115]. Quinones are one of the important groups in polyprenoids, containing polar head group and non-polar isoprenoid side chain. These quinones are involved in electron transport

chain. Menaquinones are most ancient quinones present in microbes and involved in respiration, photosynthesis. Phylloquinone (Vitamin K1) are present in cyanobacteria, algae and higher plants. Phylloquinone is required for blood clotting in mammals. Ubiquinones are present in eukaryotes and play a key role in electron transport chain. Rhodoquinone (RQ), the ubiquinone derivative with amino group substituting one of a methoxy group, present in purple bacteria *Rhodospirillaceae* family[116].
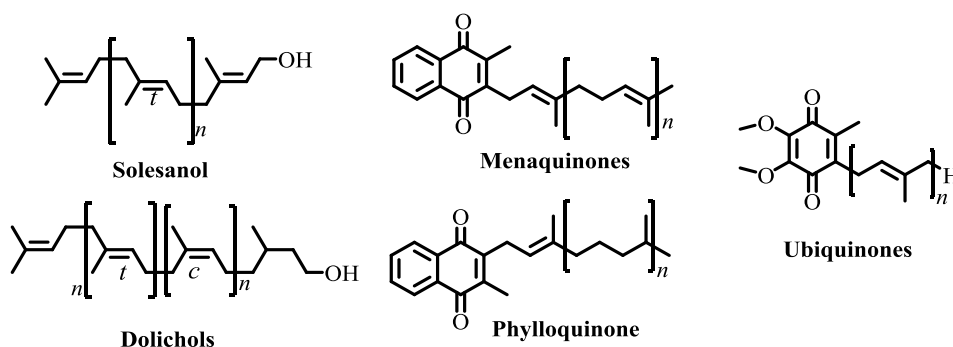
Figure 1. 20 Polypenoids.

1.5 *Azadirachta indica* (Neem)

Azadirachta indica is very well known in Indian and neighbouring countries for more than 2000 years as one of the most versatile medicinal plant[117,118]. In Sanskrit, the neem tree is called as 'Arishtha' meaning 'reliever of sickness' and hence is considered as 'Sarbaroganibarini'. The Persian name of neem is 'Azad-Darakth- E- Hind' which means 'free tree of India'. Apart from these names neem is famous with many other names like 'Indian lilac', 'Margosa' 'Divine Tree', 'Heal All', 'Nature's Drugstore', and 'Village Dispensary'. The medicinal properties of neem are listed in ancient documents 'Charak-Samhita' and 'Susruta-Samhita', which are foundations of Ayurveda. Along with Ayurveda, it was used in another medicinal system like Unani, Chinese, and European "Materia Medica"[119]. All the parts of neem are used in different medicinal systems and household remedy. It has great potential in the fields of pest management and environmental protection. From the 1970s, scientists in Europe and United States have been interested in neem because of its insecticidal properties and its low toxicity to mammals. In 1992, US National

Academy of Sciences published a report "Neem–A Tree for Solving Global Problems" about the importance of neem in medicine, agriculture and for environment protection[120].

Classification and distribution

Order: Rutales

Suborder: Rutinae

Family: Meliaceae

Subfamily: Melioideae

Tribe: Melieae

Genus: *Azadirachta*

Species: *indica*

Neem tree is indigenous to arid, semi-arid, wet-tropical, tropical and sub-tropical regions of South and Southeast Asia. It is widely distributed in Indonesia, Thailand, Myanmar, Bangladesh, Nepal, Sri Lanka, India and Pakistan. Neem has been introduced and established throughout the tropics and subtropics, especially in drier areas in Australia, North, South and Central America, the Caribbean, Africa and the Middle East[121]. Neem tree is evergreen, perennial, deep-rooted, glabrous, attains a height of 12-15 m, rarely up to 25 m and a girth of 1.8-3 m. Neem trees can grow well in dry, stony shallow soils and even on soils having hard calcareous or claypan and with rainfall ranges from 150 to 1200 mm. Neem thrives well in wide range of temperatures of 0 °C to 49 °C[119,122].

1.5.1 Neem Secondary Metabolites

Neem is widely used in medicine, agriculture and environment protection because of its secondary metabolites. Most of these belong to limonoids which are tetranortriterpenoids. The term limonoid was derived from limonin, the first tetranortriterpenoid obtained from citrus bitter principles. More than 1300 limonoids are identified till now with 35 different skeletons, out of which neem (*Azadirachta indica*) limonoids has 4,4,8-trimethyl-17-furanyl steroid skeletons and its derivatives[123,124]. The first limonoid from neem was nimbin isolated by Siddiqui in 1942[125]. The most potent insecticidal compound azadirachtin was isolated in 1968 by Butterworth and Morgan[126]. Then further isolation of limonoids was continued and

till date, around 150 compounds from neem have been identified[120]. Neem limonoids are highly complex, rearranged and oxygenated triterpenoids. Based on the skeleton, neem limonoids are classified into different groups[127].

A) **Protolimonoids:** Triterpenoids having hydroxylated or with cyclic furan side chain which is attached to C_{17} of oxidosqualene cyclic compounds might be euphol, tirucallol, butyrospermol, tirucalla-7,24-dien-3β-ol (Figure 1.21).

B) **Limonoids:** 4,4,8-trimethyl-17-furanyl steroid skeleton compounds belong to this group. Limonoids are formed from protolimonoids by the loss of four terminal carbon units and followed by furan ring formation (Figure 22-25). Limonoids are again divided into two groups, basic limonoids which contain all four rings intact (azadirone skeleton, gedunin skeleton, vilasinin skeleton, ring intact butenolide skeleton and other) and C-seco limonoids which have opened C-ring (nimbin skeleton, salannin skeleton, nimbolide skeleton, azadirachtin skeleton, nimbin skeleton and C-seco butenolide skeleton).
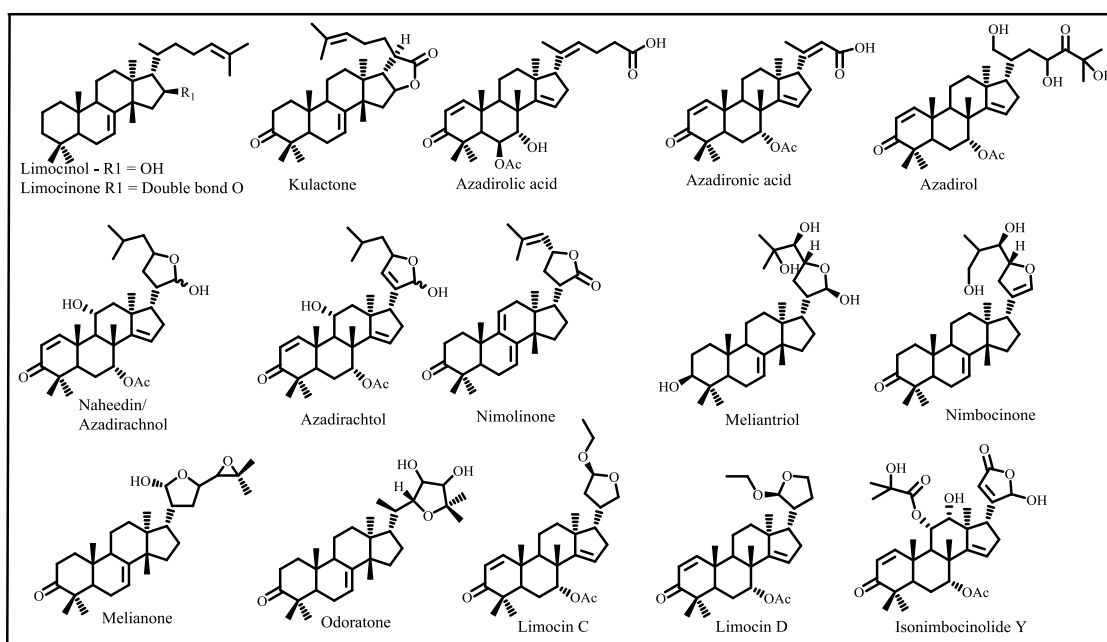
Figure 1. 21 Neem Protolimonoids.
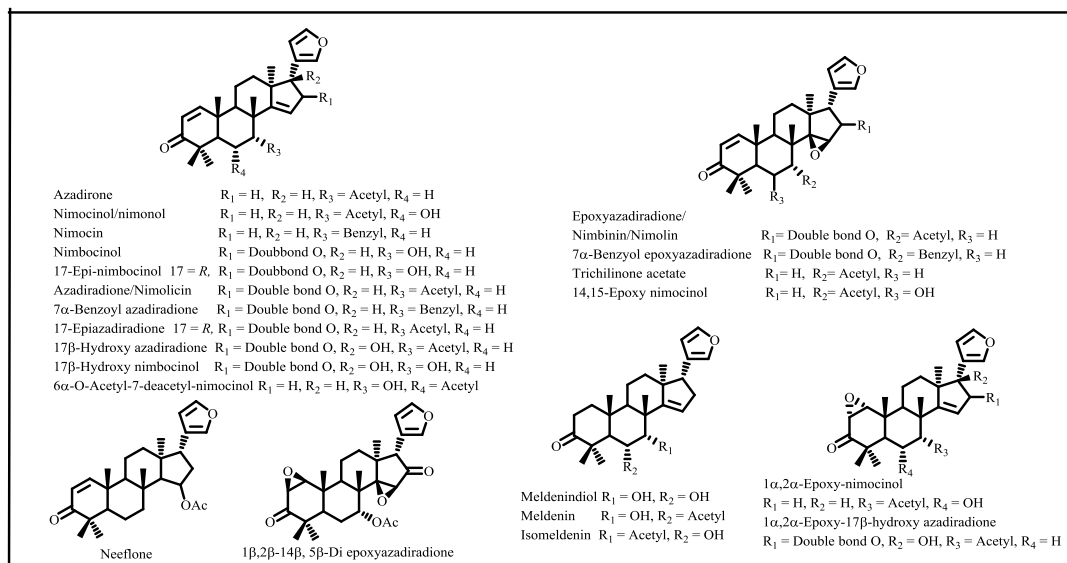
Figure 1. 22 Azadirone Skeleton Triterpenoids.
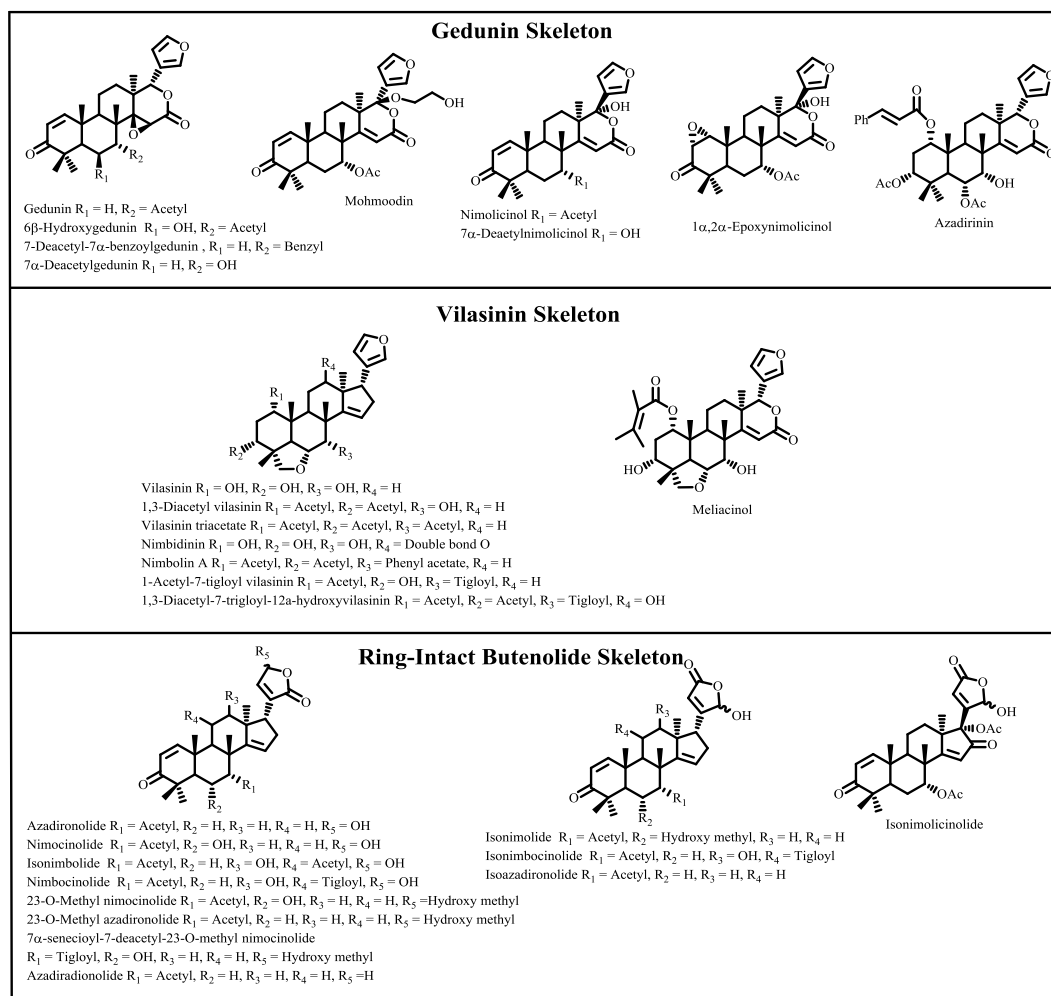
Figure 1. 23 Limonoids of Gedunin, Vilasinin and Ring intact Butenolide Skeleton.
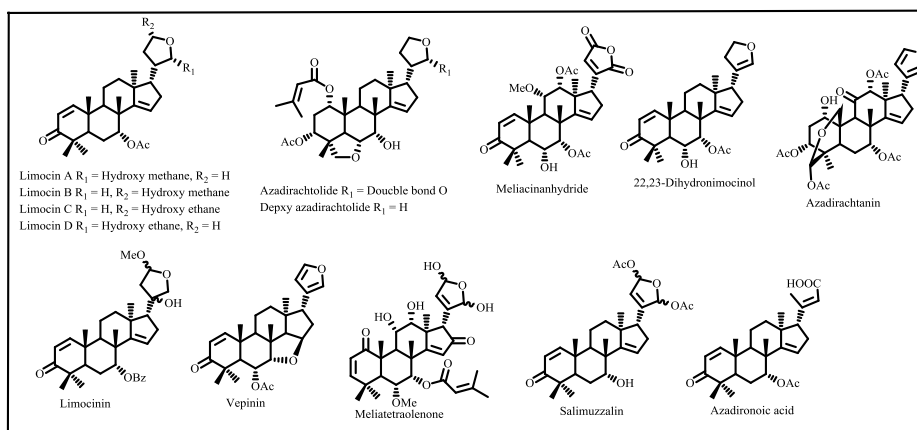
Figure 1. 24 Miscellaneous Ring intact Butenolide Skeleton.
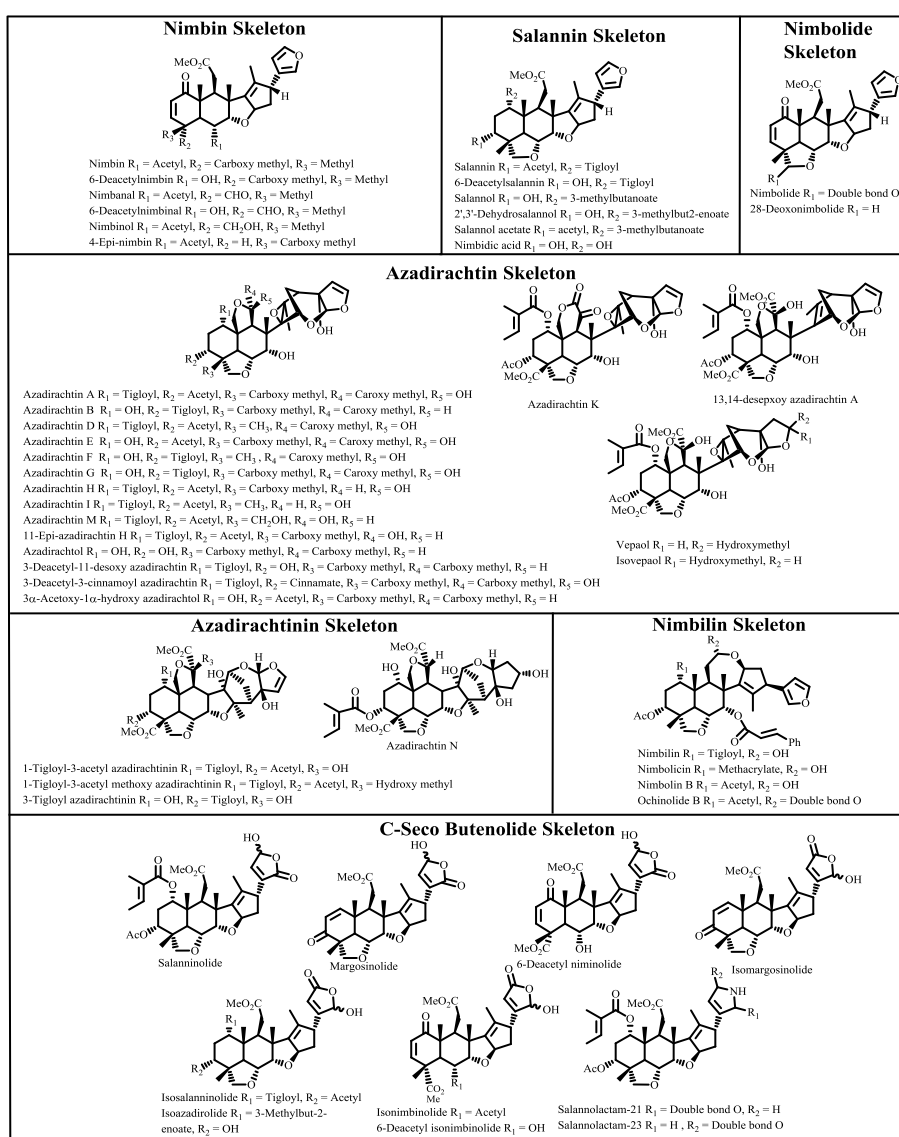
Figure 1. 25 C-Seco Limonoids.

1.5.2 Literature Studies on Limonoid Biosynthesis

The research in triterpenoid biosynthesis started in the 1950s with studies like the incorporation of $^{14}C/^3H$ labelled acetate and mevalonate[128-131]. $^{14}C/^3H$ labelled nimocinol, nimocinolide, nimbin, salannin, nimbolide and azadirachtin were observed when neem seeds were incubated with $^{14}C/^3H$ labelled acetate and mevalonate[132-134]. In general, isoprene units (IPP and DMAPP) are synthesized from MVA pathway and utilized in triterpenoid biosynthesis[63]. Head to tail condensation of DMAPP and IPP produced C_{10} GPP by the action of GDS/FDS. GPP combines with one more IPP to form C_{15} FPP by the action of FDS. Two molecules of FPP condense in 1-1' position to form C_{30} squalene by the action of squalene synthase (SQS) through presqualene intermediate[107]. Squalene is the first committed precursor for triterpenoid biosynthesis. Squalene epoxidase oxidizes squalene to (3S)-oxidosqualene. Basic triterpene cyclic product is formed by the action of triterpene synthase on (3S)-oxidosqualene[105]. In plant steroid biosynthesis cycloartenol synthase cyclizes (3S)-oxidosqualene into cycloartenol, which is further modified into phytosterol[106]. Based on oxygenated C_{30} compounds isolated from Meliaceae family, the precursor cyclic molecule (protolimonoid skeleton) for limonoids biosynthesis is speculated to be a euphol or tirucallol derivative. When tritium labelled euphol, tirucallol, Δ^7-tirucallol/(tirucalla-7,24-dien-3β-ol) and butyrospermol were fed to the neem leaves, all were found to be incorporated into the limonoids, nimbolide. However, euphol was more effectively incorporated into nimbolide as compared to others[134,135]. Involvement of $\Delta^{7,9(11)}$ system was confirmed by effective incorporation of 8α,9α-epoxyeuphol and diepoxyeuphol as compared to euphol[134]. The ^3H:^{14}C incorporation ratio is 2:5 in nimocinol and nimocinolide confirms the involvement of euphol/tirucallol in limonoid biosynthesis[133,136].

The biosynthesis of highly diverse, oxygenated complex limonoids was understood very poorly. Predicted limonoid biosynthesis starts from epoxidation of Δ^7-double bond and rearrangement of a C-14 methyl group by the opening of 7α,8α-epoxide ring to form protolimonoids[137]. Four terminal carbons from the side chain are removed by oxidative degradation, and then the side chain undergoes cyclization through cyclic hemiacetal to form basic limonoid skeletons[138]. Then, the third ring opens and oxidizes further to form C-seco limonoid skeletons. It is predicted that

azadirone undergoes third ring opening and further oxidation to form salannin which is then heavily oxidized and cyclized to from azadirachtin[139] (Figure 1.26).
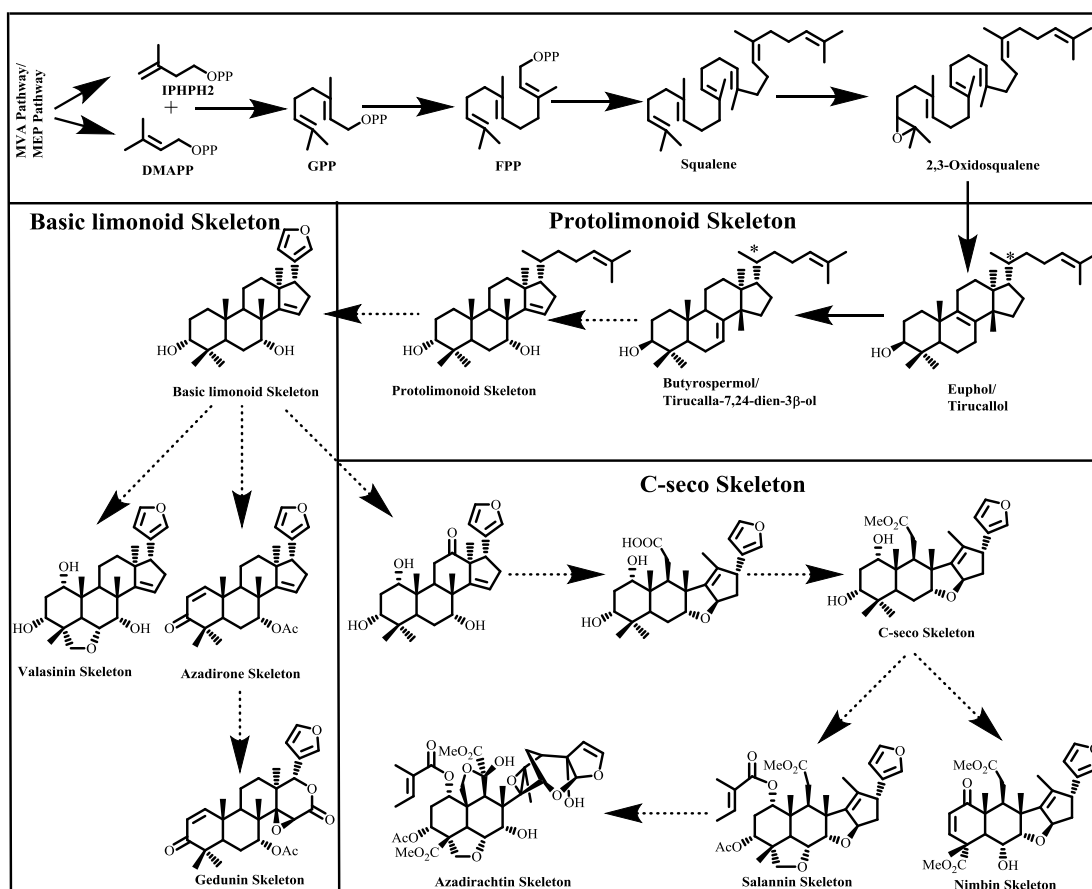
Figure 1. 26 Biosynthesis of Neem Limonoids.

Although few transcriptomic analysis of neem is done to predict the genes involved in limonoids biosynthesis, these predictions were upto oxidosqualene synthases only, and there is no further information about downstream enzymes. Neem is evolutionarily close to *Citrus* species, which was verified by both molecular phylogenetic analyses and sequence similarity. Comparative transcript expression analysis showed enhanced expression of initial genes involved in neem triterpenoid biosynthesis as compared to other sequenced angiosperms[140,141]. Assembly of Illumina and 454 sequencing genome reads resulted in 267 Mb, which accounts for 70 % of the estimated size of neem genome. Comparative analysis anchored 62 % (161 Mb) of assembled neem genomic contigs onto citrus chromomes[142]. From the transcriptome analysis, diphosphomevalonate decarboxylase and squalene epoxidase

were up-regulated in adventitious roots, at the same time geranylgeranyl diphosphate synthase type II and 1-deoxy-xylulose-5-phosphate synthase were down-regulated as compared to leaves[143]. The MEP pathway genes 1-deoxy-xylulose-5-phosphate synthase, 1-deoxy-xylulose-5-phosphate reductase showed low expression in all the stages of fruits as compared to leaves, while in the case of MVA pathway genes, 3-hydroxy-3-methylglutaryl-coenzyme A reductase was highly expressed in fruits as compared to leaves. In contrast, FDS and SQS were highly expressed in leaves as compared to fruits[144,145]. Still, further analysis is needed to identify the genes involved in limonoid biosynthesis in neem.

Very few investigations were present regarding metabolic fingerprinting of neem limonoids. The concentration of neem C-seco limonoids (azadirachtin, nimbin and salannin) were high in mature seed as compared to other tissues like shoot, root, cambium, pulp, flower, bark and leaf[142]. Azadirachtin was highly abundant in fruit stage three. Nimbin is several times higher in different stages of fruits as compared to leaves[145]. Neem kernel is the richest source of limonoids, including highly biologically active limonoid, azadirachtin. In the kernel, idioblast/secretory cells are the site for synthesis and storage of limonoids. After 40 days of flowering, idioblast cells are differentiated in the kernel and characterized by enlarged nucleus within vacuolated cells. Many terpenoid vesicles (50-100 nm diameter) are accumulated in the cytoplasm and originated as vesiculations of endoplasmic reticulum[146].

1.5.3 Biological Properties of Neem Limonoids

In 1959, Dr. Heinrich Schmutterer noticed that only neem tree remained green and healthy while all other vegetation was completely destroyed by the locust plague. In 1968, the azadirachtin was purified and showed to be the cause for insecticidal properties[126]. Neem limonoids got tremendous focus for identification of their biological properties. Neem limonoids are have properties like anti-carcinogenic, anti-inflammatory, anti-microbial, anti-viral, analgesic, immunomodulatory, anti-diabetic, insecticidal and anti-parasitic (Table 1.1).

1.5.3.1 Anti-Carcinogenic

Anticancer effects of neem extracts and neem limonoids such as azadirachtin, azadirone, 6-deacetyl nimbin, gedunin, nimbin, nimbolide and salannin are mediated

through preventing carcinogen activation; inhibiting cell proliferation, inflammation, invasion and angiogenesis, inducing apoptosis; enhancing host antioxidant and detoxification systems; modulating oncogenic transcription factors and signaling kinases, thereby influencing the epigenome[147,148].

Table 1. 1 Biological Properties of Neem Limonoids.

Biological Properties	Neem Limonoids
Anti-carcinogenic	Azadirachtin, azadiradione, epoxyazadiradione, nimbolide, 28-deoxonimbolide, gedunin, 7-deacetylgedunin, azadirone, nimbin, 6-deacetyl nimbin, salannin
Anti-inflammatory	Epoxyazadiradione, 17-epi-17-hydroxyazadiradione, 7-acetyl-16,17-dehydro-16-hydroxyneotrichilenone, 7-deacetylgedunin, nimolicinol and nimbin
Anti-microbial	Nimbolide, gedunin, mahmoodin, azadiradione
Anti-diabetic	Azadiradione, gedunin and extracts of neem leaves, stem, kernel and bark
Anti-malarial	Meldenin, isomeldnin, nimocinol, azadirachtin A, Nimbolide, 6-deacetylnimbin
Insecticidal	Azadirachtin A, azadirachtin B, salannin, azadirone, azadiradione, gedunin

1.5.3.2 Anti-Inflamatory

Neem limonoid extracts show anti-inflammatory effects by inhibiting the pro-inflammatory cytokine tumour necrosis factor (TNF-α), as well as NF-κB that plays a key role in inflammation[147]. In cells, epoxy azadiradione prevents the release of pro-inflammatory cytokines interleukins, TNF-α, inhibition of translocation of NF-κB and stimulates nitric oxide production. 17-epi-17-hydroxy azadiradione, 7-acetyl-16,17-dehydro-16-hydroxy neotrichilenone, 7-deacetyl gedunin, nimocinol and nimbin exhibits an inhibitory effect on 12-O-tetradecanoylphorbol-13-acetate (TPA)-induced inflammation[121].

1.5.3.3 Anti-Microbial

The oil from the neem leaves possesses antibacterial activity against *Mycobacterium tuberculosis,* streptomycin-resistant strains, *Vibrio cholerae*,

Klebsiella pneumonia and *Streptococcus pyogenes*. The antifungal activity of neem has been attributed to nimbolide and gedunin[149]. Mahmoodin and azadiradione isolated from neem showed significant antibacterial activity[150,151].

1.5.3.4 Anti-Diabetic activity

Neem leaves, stem, kernel and bark extract significantly decreases blood sugar level and prevents adrenaline as well as glucose-induced hyperglycaemia[152]. Neem leaves possess antihyperglycemic effect, which may be due to its antiserotonin activity[149]. The limonoids, azadiradione and gedunin could bind and inactivate α-amylase and may reduce/control post-prandial hyperglycemia [153].

1.5.3.5 Anti-Malarial activity

Neem seed fractions are not only active against the human parasite stages but also against the stages responsible for continued malaria transmission[120]. The limonoids (meldenin, isomeldenin and nimocinol) isolated from fresh neem leaves have been found to demonstrate anti-malarial activity against chloroquine-resistant *Plasmodium falciparum*. Azadirachtin A, nimbolide and 6-deacetylnimbin appeared to interfere with transmissible *Plasmodium* stages[154].

1.5.3.6 Insecticidal Activity

More emphasis is being given to the biopesticides because of hazardous side effects of chemically synthesized pesticides and insecticides. The World Health Organisation has banned the use of endosulfan, a synthetic pesticide that causes serious eye, kidney and liver problems. The need of eco-friendly pesticide for agriculture has led research experts to turn their attention to plants. Neem limonoids are extensively studied in the past 30 years and demonstrated to have insecticidal activity against 413 species in 16 different insect orders. The biological properties of limonoids against insects include repellence, feeding and oviposition deterrence, growth disruption, reduced fitness and sterility[120].

Neem seed oil shows insect repellent effect; for example, crude extracts of neem leaves repelled females of *Crocidolomia binotalis* from treated cabbage leaves at a distance of about 25 cm. Azadirachtin inhibits feeding of locust when it was offered with the concentration of 10-40 μg L^{-1}[126]. This antifeedant activity is due to the inhibition of firing of neurons that signal phago stimulants, whereas in the case of

larvae of *Pieris brassicae*, azadirachtin stimulates deterrent neuron of the sensilla[155]. Insect moulting process is affected by azadirachtin by lowering the dynamics of ecdysteroid hormone[156]. Apart from these effects, azadirachtin influence the fecundity of female insects by reduction of ovarian ecdysteroids[157,158].

1.6 Metabolic Engineering of Isoprenoids

Plants are the major source of isoprenoids which have a great impact on humans. However, the plant-based supply of isoprenoids is very low and isolation of them leads to consumption of a large number of natural sources. Many of these metabolites are very complex, and chemical synthesis often requires too many steps and difficult reactions, resulting in low yield or incorrect stereochemistry and high cost. For example, total chemical synthesis of highly complex triterpene azadirachtin took nearly two decades of research and yield of 0.00015 %[159,160]. These reasons result in an increase in the engineering of metabolic pathways for large-scale and cost-effective production of metabolites in heterologous hosts. Metabolic engineering of terpenoids includes following strategies:

a) Overexpression or increasing the copy number of MVA/MEP pathway genes to increase metabolic flux for isoprene unit production.

b) Up-regulating prenyl diphosphate synthases for increasing production of precursor prenyl diphosphates.

c) Knockdown of biosynthetic pathways which uses the competitive precursor prenyl diphosphates.

d) Manipulating transcription factors and promoters for rate-limiting enzymes and cofactors metabolism.

1.6.1 Metabolic Engineering in Bacteria

Artemisinin is a highly effective antimalarial drug and produced in *Artemisia annua*. But artemisinin occurs around 0.01-1% dry weight of leaves. An attractive route of production is metabolic engineering in microbes for production of artemisinin precursor such as amorpha-4,11-diene. The high level (24 mg L^{-1} in 14 h) of amorphadiene production occurred when MVA pathway genes from yeast were transplanted into *E.coli* to increase the flux of FPP biosynthesis and also by simultaneous optimization of fermentation conditions[161]. Additional improvements

(seven folds) was done by increasing the promoter strength and plasmid copy number of amorpha-4,11-diene synthase and mevalonate kinase. Further replacement[162] of HMGR and HMGS with more active homolog from *Staphylococcus aureus* eliminated the HMG-CoA accumulation resulting in 27 gm L^{-1} of amorphadiene production[45,163].

Taxadiene is a diterpene precursor for the production of anticancer drug taxol (paclitaxel). Earlier, 1.3 mg L^{-1} production of taxadiene in *E. coli* was achieved by expression of DXS, IPPI, GGDS and taxadiene synthase[164]. Optimization of copy number and promoter for MEP pathway genes and taxadiene synthase result in the production of 1 gm L^{-1} in fed-batch fermentation. Further introduction of chimeric fusion of taxadiene-5α-hydroxylase with CYP 450 reductase resulted in the production of 58 mg L^{-1} of taxadiene-5α-ol[165]. Apart from these medicinal natural products, other metabolites such as 8-epi-cedrol, δ-cadenin, epi-aristolochene, vetispiradiene and Δ-3-carene were successfully produced by metabolic engineering in bacteria[166] (Figure 1.27)

1.6.2 Metabolic Engineering in Yeast

Yeast (*Saccharomyces cerevisiae*) is a preferred host over *E.coli* as it contains native MVA pathway and is considered as a better system for cytochrome P450 enzymes[167]. Amorphadiene synthase expression in yeast produced 4.4 mg L^{-1} under GAL1 promoter[168]. Further production was increased (40 g L^{-1} on shaking flask) by switching host from S288C to CEN.PK2 strain, transcription factor upc2-1 to globally upregulate the MVA pathway expression, including integrating three copies of HMGR, down-regulation of squalene synthesis, deleting the *GAL1* gene to eliminate the utilization of galactose and optimizing fermentation conditions[169]. CYP71AV1 which catalyzes the oxidation of amorphadiene to artemisinic acid along with cytochrome P450 reductase expressed in yeast resulted in the production of 2.5 g L^{-1} of artemisinic acid. Further artemisinic acid production (7.7 g L^{-1}) was increased by expressing aldehyde dehydrogenase from *Artemisia annua*[166,170].

Triterpenoid production through metabolic engineering studies is in initial stages. β-amyrin production in yeast was improved to 6 mg L^{-1} (50%) by expression β-amyrin synthase (*Artemisia annua*) and HMGR under GAL promoter while down-regulating lanosterol synthase[171]. Overexpressing truncated 3-hydroxyl-3-

methylglutaryl-CoA reductase, farnesyl diphosphate synthase, squalene synthase and 2,3-oxidosqualene synthase genes, together with increasing protopanaxadiol synthase activity through codon optimization and two-phase extractive fermentation resulted in the production of protopanaxadiol (1.2 gm L^{-1}), dammarenediol-II (1.5 gm L^{-1})[172]. Yeast is being used for the production of wide variety of products like biofuels, protein drugs and non-ribosomal peptides[167].
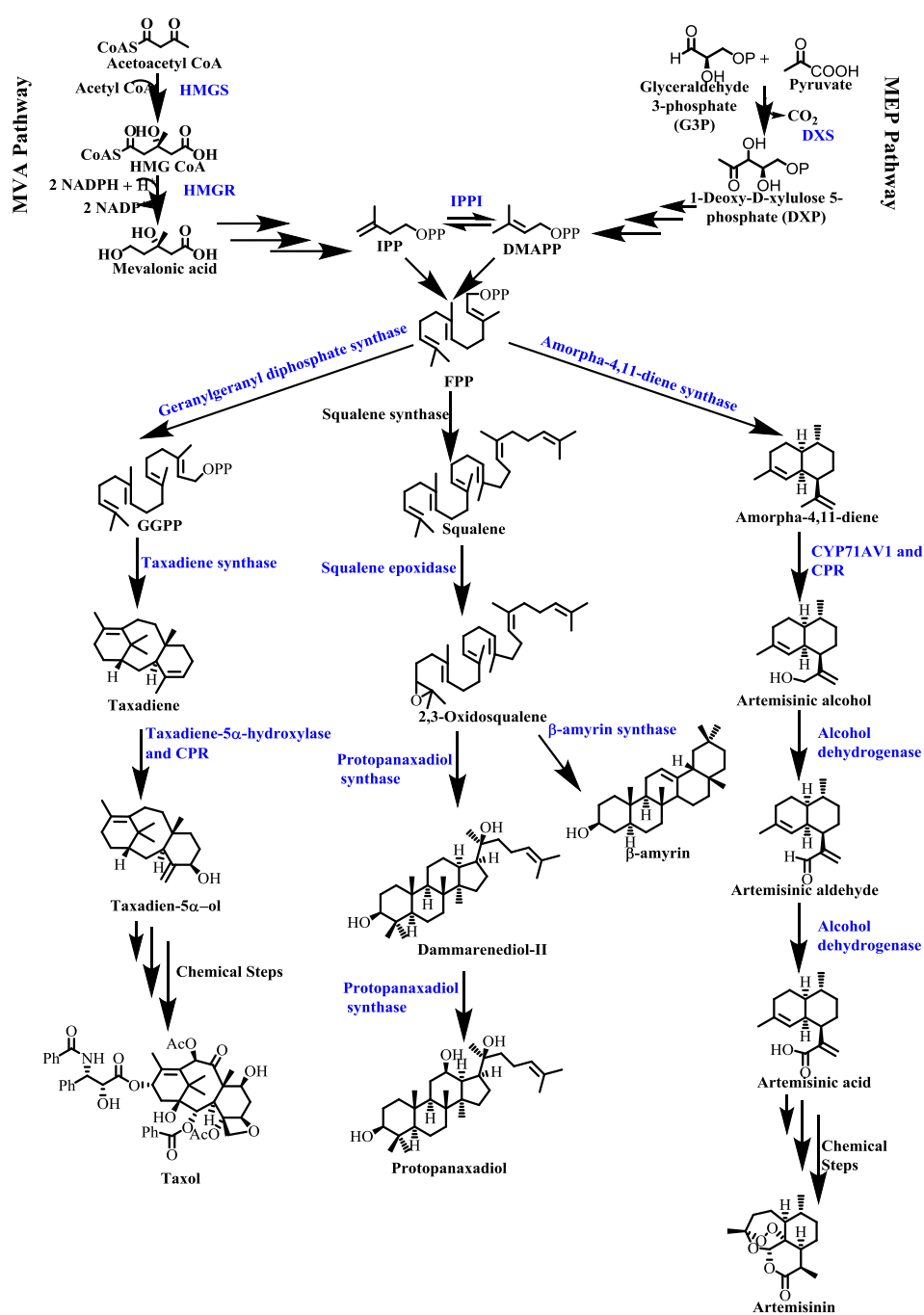
Figure 1. 27 Metabolic Engineering of Isoprenoids.

1.7 Current Status on Study of Secondary Metabolite Biosynthesis

Secondary metabolites are synthesized from the primary metabolites for enhancing functions and interaction with the environment in living systems. Plants synthesize very diverse, complex and most effective secondary metabolites as compared to other organisms[173]. Till now, over 200,000 secondary metabolites have been discovered from the plant kingdom. Plants synthesize these metabolites, in order to achieve fitness during evolution and to defend themselves from competing plants, herbivores and pathogens[174]. In addition to their biological functions, plant secondary metabolites are very useful for mankind due to their pharmacological and toxicological properties. One-half of all the licensed drugs are natural products or their synthetic derivatives [175].

In the past, identification and structural determination of natural products have been restricted by the limitation of available techniques comparatively. However, in the current era, high sensitive techniques like HRMS, NMR helps for the identification, structural determination and quantification of secondary metabolites across different tissues or conditions[176]. Biosynthesis of secondary metabolites is under tremendous progress from last two centuries. In the past, the prospects for identification of natural product biosynthesis have been limited by the availability of known enzymes to concatenate into pathways. In the current era, automation and decrease in sequencing costs help in identification and selection of enzymes that can be involved[177]. Bioinformatics tools like the basic local alignment search tool (BLAST) and protein domain recognition databases, including the protein families (Pfam) database, InterPro scan and the conserved domains database (CDD), give a rapid prediction of enzyme function. Tools like MetaCyc[178], KEGG[179], BRENDA[180] helps in the prediction of secondary metabolic pathways based on genomic sequencing data. Tools like differential gene expression[181] and coexpression analysis[182] help in fishing out the genes involved in specific metabolite biosynthesis. Correlating the expression profile data with metabolic profiling further helps in the prediction of genes involved in biosynthesis[183]. In addition, the number of secondary metabolite gene clusters present in the plant genome is now realized. Hence genome mining for clusters making an easy strategy for identification of genes related to metabolite biosynthesis[184,185].

In plants, the secondary metabolite synthesis is highly compartmentalized to a specific tissues/condition. Within the cells, their synthesis is again compartmentalized. For example, monoterpenoid and diterpenoid biosynthesis takes place in plastids, whereas sesqui- and triterpenoid biosynthesis takes place in the cytosol and endoplasmic reticulum membrane, respectivly[186]. Apart from compartmentalization, the enzymes involved in specific secondary metabolite forms a complex or metabolon[187]. The reasons for metabolon formation are to improve efficiency of synthesis by channeling an intermediate, to prevent kinetic constraints that result from the dilution of intermediates, to convert labile and/or toxic intermediates into more stable and less toxic metabolites, to control and co-ordinate metabolic cross-talk that is mediated either by enzymes that function in different pathways or by intermediates that are shared between different metabolic pathways. The advantages associated with the organization of a portion or an entire biosynthetic pathway in a metabolon are thus many-fold[188].

Plants synthesize secondary metabolites in low quantity and most of these metabolites are very complex, chemical synthesis often requires many steps and difficult reactions, resulting in low yield or incorrect stereochemistry and high cost[189]. These reasons result in the engineering of metabolic pathways for large-scale and cost-effective production of metabolites in heterologous hosts[190] or plants[191]. Advanced synthetic biology techniques like type IIS restriction endonuclease-based cloning, site-specific recombination-based cloning, long overlap based assembly, circular polymerase extension cloning and ligase cycling reactions[192] help for integrating the multigene cassette in heterologous host sytams[193]. CRISPR/Cas9[194], yeast homologous recombination and *Agrobacterium*-based transformation of multigene cassette help in metabolic engineering for the plant secondary metabolite production[195]. Further development related to proteomic-based analysis, transcription factor[196] and microRNAs[197] regulation of biosynthetic pathways, and high throughput metabolic engineering is required for industrial-scale production plant secondary metabolites.

1.8 Scope of Thesis

In plants, thousands of triterpenoids are identified and are synthesized from more than 100 different carbon skeletons. Neem has widely used in medicine,

agriculture and environment protection because of its secondary metabolites. Most of these belong to limonoids, a tetranortriterpenoid. More than 1300 limonoids are identified till now with 35 different skeletons out of which neem (*Azadirachta indica*) limonoids are of 4,4,8-trimethyl-17-furanyl steroid skeletons and its derivatives[123,124]. The most potent insecticidal compound azadirachtin was isolated in 1968 by Butterworth and Morgan[126]. Then further isolation of limonoids is continued, till now around 150 compounds were identified from neem[120]. Neem limonoids contain biological properties like anti-carcinogenic, anti-inflammatory, antimicrobial, antiviral, analgesic, immunomodulatory, anti-diabetic, insecticidal and antiparasitic.

The research in triterpenoid biosynthesis was started with studies like the incorporation of ^{14}C/^{3}H labelled acetate and mevalonate[128-131]. Based on tritium labelled studies euphol, tirucallol, Δ^7-tirucallol/(tirucalla-7,24-dien-3β-ol) and butyrospermol were predicted to be involved in limonoids biosynthesis[134,135]. Few genomic and transcriptomic analysis of neem are done to predict the genes involved in limonoids biosynthesis[140,141]. Still, there are no reports regarding the functional characterization of genes involved in neem limonoids biosynthesis. The main aim of this work is a systematic analysis of transcriptome to identify the putative genes involved in limonoid biosynthesis, functional characterization and to understand the mechanism of action. Targeted metabolic fingerprinting was carried out to identify tissue-specific production of limonoids. Based on metabolic profiling, tissues are selected for transcriptome analysis. The transcriptome is analysed by functional annotation tools like BLAST, KAAS and Pfam analysis to identify the triterpenoid biosynthetic genes. By applying differential gene expression analysis, the downstream enzymes are selected. Prenyl diphosphate synthases (AiGDS and AiFDS), squalene epoxidase (AiSQE1), triterpene synthesis (AiTTS1 and AiTTS2), NADPH-cytochrome P450 reductases (AiCPR1 and AiCPR2) and cytochrome P450 (AiCYP1 and AiCYP2) were cloned and functionally characterized. Further and ultimate scope of the work is the large-scale production of neem limonoids and their intermediates by heterologous expression of genes involved in biosynthesis.

1.9 References

1.   Vining, L.C. Functions of secondary metabolites. *Annual Reviews in Microbiology* **44**, 395-427 (1990).

2.   Oksman-Caldentey, K.M. & Inze, D. Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. *Trends in Plant Science* **9**, 433-440 (2004).

3.   Pichersky, E. & Gang, D.R. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in Plant Science* **5**, 439-445 (2000).

4.   Anarat Cappillino, G.l. & Sattely, E.S. The chemical logic of plant natural product biosynthesis. *Current Opinion in Plant Biology* **19**, 51-58 (2014).

5.   Poelman, E.H. & Kessler, A. Keystone herbivores and the evolution of plant defences. *Trends in Plant Science* **21**, 477-485 (2016).

6.   Zhao, N., Wang, G., Norris, A., Chen, X. & Chen, F. Studying plant secondary metabolism in the age of genomics. *Critical Reviews in Plant Sciences* **32**, 369-382 (2013).

7.   Wink, M. Evolution of secondary metabolites in legumes (*Fabaceae*). *South African Journal of Botany* **89**, 164-175 (2013).

8.   Stevenson, P.C., Nicolson, S.W. & Wright, G.A. Plant secondary metabolites in nectar: impacts on pollinators and ecological functions. *Functional Ecology* **31**, 65-75 (2017).

9.   Akula, R. & Ravishankar, G.A. Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signaling & Behavior* **6**, 1720-1731 (2011).

10.   Hartmann, T. Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomologia Experimentalis et Applicata* **80**, 177-188 (1996).

11.   Hartmann, T. From waste products to eco chemicals: fifty years research of plant secondary metabolism. *Phytochemistry* **68**, 2831-2846 (2007).

12.   Swain, T. Secondary compounds as protective agents. *Annual Review of Plant Physiology* **28**, 479-501 (1977).

13.   Verpoorte, R., Contin, A. & Memelink, J. Biotechnology for the production of plant secondary metabolites. *Phytochemistry Reviews* **1**, 13-25 (2002).

14.   Julsing, M.K., Koulman, A., Woerdenbag, H.J., Quax, W.J. & Kayser, O. Combinatorial biosynthesis of medicinal plant secondary metabolites. *Biomolecular Engineering* **23**, 265-279 (2006).

15.   Kennedy, D.O. & Wightman, E.L. Herbal extracts and phytochemicals: plant secondary metabolites and the enhancement of human brain function. *Advances in Nutrition: An International Review Journal* **2**, 32-50 (2011).

16.   Chae, L., Kim, T., Nilo Poyanco, R. & Rhee, S.Y. Genomic signatures of specialized metabolism in plants. *Science* **344**, 510-513 (2014).

17.   Cragg, G.M. & Newman, D.J. Natural products: a continuing source of novel drug leads. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1830**, 3670-3695 (2013).

18. Jain, P. Secondary metabolites for antiulcer activity. *Natural Product Research* **30**, 640-656 (2016).

19. Wink, M. Medicinal plants: a source of anti-parasitic secondary metabolites. *Molecules* **17**, 12771-12791 (2012).

20. Okwute, S.K. Plants as potential sources of pesticidal agents: a review. in *Pesticides-Advances in Chemical and Botanical Pesticides* (InTech, 2012).

21. Chaudhary, S. *et al*. Progress on *Azadirachta indica* based biopesticides in replacing synthetic toxic pesticides. *Frontiers in plant science* **8**, 610 (2017).

22. Kabera, J.N., Semana, E., Mussa, A.R. & He, X. Plant secondary metabolites: biosynthesis, classification, function and pharmacological properties. *Journal of Pharmacy and Pharmacology* **2**, 377-392 (2014).

23. Bernays, E.A., Driver, G.C. & Bilgener, M. Herbivores and plant tannins. *Advances in Ecological Research* **19**, 263-302 (1989).

24. Pereira, D., Valentao, P., Pereira, J. & Andrade, P. Phenolics: From Chemistry to Biology. *Molecules* **14**, 2202-2011 (2009).

25. Dai, J. & Mumper, R.J. Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. *Molecules* **15**, 7313-7352 (2010).

26. Kutchan, T.M. Alkaloid biosynthesis [mdash] the basis for metabolic engineering of medicinal plants. *The Plant Cell* **7**, 1059-1070 (1995).

27. Blakemore, P.R. & White, J.D. Morphine, the Proteus of organic molecules. *Chemical Communications*, 1159-1168 (2002).

28. Michael, J.P. Quinoline, quinazoline and acridone alkaloids. *Natural Product Reports* **16**, 697-709 (1999).

29. D O'Hagan, D. Pyrrole, pyrrolidine, pyridine, piperidine and tropane alkaloids. *Natural Product Reports* **17**, 435-446 (2000).

30. Facchini, P.J. Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annual Review of Plant Biology* **52**, 29-66 (2001).

31. Vranova, E., Coman, D. & Gruissem, W. Structure and dynamics of the isoprenoid pathway network. *Molecular Plant* **5**, 318-333 (2012).

32. Cornish, K. & Blakeslee, J.J. Rubber Biosynthesis in Plants. *American Oil Chemist Society, The Lipid Library* (2011).

33. Puskas, J.E., Gautriaud, E., Deffieux, A. & Kennedy, J.P. Natural rubber biosynthesis-A living carbocationic polymerization? *Progress in Polymer Science* **31**, 533-548 (2006).

34. Peralta Yahya, P.P., Zhang, F., del Cardayre, S.B. & Keasling, J.D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320-328 (2012).

35. Abdallah, I.I. & Quax, W.J. A Glimpse into the Biosynthesis of Terpenoids. *KnE Life Sciences* **3**, 81-98 (2017).

36. Roberts, S.C. Production and engineering of terpenoids in plant cell culture. *Nature Chemical Biology* **3**, 387-395 (2007).

37. Wouters, J. *et al*. Catalytic mechanism of *Escherichia coli* isopentenyl diphosphate isomerase involves Cys-67, Glu-116, and Tyr-104 as suggested by crystal structures of complexes with transition state analogues and irreversible inhibitors. *Journal of Biological Chemistry* **278**, 11903-11908 (2003).

38. Hemmi, H., Ikeda, Y., Yamashita, S., Nakayama, T. & Nishino, T. Catalytic mechanism of type 2 isopentenyl diphosphate: dimethylallyl diphosphate isomerase: verification of a redox role of the flavin cofactor in a reaction with no net redox change. *Biochemical and Biophysical Research Communications* **322**, 905-910 (2004).

39. Kuzuyama, T. & Seto, H. Two distinct pathways for essential metabolic precursors for isoprenoid biosynthesis. *Proceedings of the Japan Academy, Series B* **88**, 41-52 (2012).

40. Dellas, N., Thomas, S.T., Manning, G. & Noel, J.P. Discovery of a metabolic alternative to the classical mevalonate pathway. *Elife* **2**, e00672 (2013).

41. VanNice, J.C. *et al*. Identification in *Haloferax volcanii* of phosphomevalonate decarboxylase and isopentenyl phosphate kinase as catalysts of the terminal enzyme reactions in an archaeal alternate mevalonate pathway. *Journal of Bacteriology* **196**, 1055-1063 (2014).

42. Vinokur, J.M. *et al*. Structural analysis of mevalonate-3-kinase provides insight into the mechanisms of isoprenoid pathway decarboxylases. *Protein Science* **24**, 212-220 (2015).

43. Vinokur, J.M., Korman, T.P., Cao, Z. & Bowie, J.U. Evidence of a novel mevalonate pathway in archaea. *Biochemistry* **53**, 4161-4168 (2014).

44. Azami, Y. *et al*. (*R*)-mevalonate 3-phosphate is an intermediate of the mevalonate pathway in *Thermoplasma acidophilum*. *Journal of Biological Chemistry* **289**, 15957-15967 (2014).

45. Zhao, L., Chang, W. C., Xiao, Y., Liu, H. W. & Liu, P. Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annual Review of Biochemistry* **82**, 497-530 (2013).

46. Wolfertz, M., Sharkey, T.D., Boland, W. & Kuhnemann, F. Rapid regulation of the methylerythritol 4-phosphate pathway during isoprene synthesis. *Plant Physiology* **135**, 1939-1945 (2004).

47. Wolfertz, M. *et al*. Biochemical regulation of isoprene emission. *Plant, Cell & Environment* **26**, 1357-1364 (2003).

48. Eisenreich, W., Bacher, A., Arigoni, D. & Rohdich, F. Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cellular and Molecular Life Sciences* **61**, 1401-1426 (2004).

49. Banerjee, A. & Sharkey, T.D. Methylerythritol 4-phosphate (MEP) pathway metabolic regulation. *Natural Product Reports* **31**, 1043-1055 (2014).

50.   Chang, W.C., Song, H., Liu, H.W. & Liu, P. Current development in isoprenoid precursor biosynthesis and regulation. *Current Opinion in Chemical Biology* **17**, 571-579 (2013).

51.   Frank, A. & Groll, M. The methylerythritol phosphate pathway to isoprenoids. *Chemical Reviews* **117**, 5675-5703 (2016).

52.   Rohmer, M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* **16**, 565-574 (1999).

53.   Hemmerlin, A., Harwood, J.L. & Bach, T.J. A raison detre for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Progress in Lipid Research* **51**, 95-148 (2012).

54.   Hemmerlin, A.a. *et al*. Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *Journal of Biological Chemistry* **278**, 26666-26676 (2003).

55.   Kasahara, H. *et al*. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. *Journal of Biological Chemistry* **277**, 45188-45194 (2002).

56.   Bick, J.A. & Lange, B.M. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Archives of Biochemistry and Biophysics* **415**, 146-154 (2003).

57.   Burg, J.S. & Espenshade, P.J. Regulation of HMG-CoA reductase in mammals and yeast. *Progress in Lipid Research* **50**, 403-410 (2011).

58.   Goldstein, J.L., DeBose-Boyd, R.A. & Brown, M.S. Protein sensors for membrane sterols. *Cell* **124**, 35-46 (2006).

59.   Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. & Gruissem, W. GENEVESTIGATOR *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology* **136**, 2621-2632 (2004).

60.   Lynch, M. & Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155 (2000).

61.   Zhang, J. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18**, 292-298 (2003).

62.   James, A.B. *et al*. The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* **322**, 1832-1835 (2008).

63.   Vranova, E., Coman, D. & Gruissem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology* **64**, 665-700 (2013).

64.   Montiel, G.g. *et al*. Transcription factor Agamous-like 12 from *Arabidopsis* promotes tissue-like organization and alkaloid biosynthesis in *Catharanthus roseus* suspension cells. *Metabolic Engineering* **9**, 125-132 (2007).

65. Kim, C.Y. & Zhang, S. Activation of a mitogen-activated protein kinase cascade induces WRKY family of transcription factors and defence genes in tobacco. *The Plant Journal* **38**, 142-151 (2004).

66. Douglas, P., Pigaglio, E., Ferrer, A., Halford, N.G. & MacKintosh, C. Three spinach leaf nitrate reductase-3-hydroxy-3-methylglutaryl-CoA reductase kinases that are regulated by reversible phosphorylation and/or Ca^{2+} ions. *Biochemical Journal* **325**, 101-109 (1997).

67. Flores-Perez, U. *et al.* PLEIOTROPIC REGULATORY LOCUS 1 (PRL1) integrates the regulation of sugar responses with isoprenoid metabolism in *Arabidopsis*. *Molecular Plant* **3**, 101-112 (2010).

68. Tang, J., Kobayashi, K., Suzuki, M., Matsumoto, S. & Muranaka, T. The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. *The Plant Journal* **61**, 456-466 (2010).

69. Lynen, F., Eggerer, H., Henning, U. & Kessel, I. Farnesyl-pyrophosphate and 3-Methyl-delta3-butenyl-1-pyrophosphate, die biologischen Vorstufen des Squalens. Zur Biosynthese der Terpene, III. *Angewandte Chemie* **70**, 738-742 (1958).

70. Chaykin, S., Law, J., Phillips, A.H., Tchen, T.T. & Bloch, K. Phosphorylated intermediates in the synthesis of squalene. *Proceedings of the National Academy of Sciences* **44**, 998-1004 (1958).

71. Cornforth, J.W., Cornforth, R.H., Popjak, G. & Yengoyan, L. Studies on the biosynthesis of cholesterol XX. Steric course of decarboxylation of 5-pyrophosphomevalonate and of the carbon to carbon bond formation in the biosynthesis of farnesyl pyrophosphate. *Journal of Biological Chemistry* **241**, 3970-3987 (1966).

72. Popjak, G. & Cornforth, J.W. Substrate stereochemistry in squalene biosynthesis: The first Ciba medal lecture. *Biochemical Journal* **101**, 553. b4-568 (1966).

73. Poulter, C.D. & Rilling, H.C. The prenyl transfer reaction. Enzymic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway. *Accounts of Chemical Research* **11**, 307-313 (1978).

74. Clarke, C.F. *et al.* Molecular cloning and sequence of a cholesterol-repressible enzyme related to prenyltransferase in the isoprene biosynthetic pathway. *Molecular and Cellular Biology* **7**, 3138-3146 (1987).

75. Tarshis, L.C., Yan, M., Poulter, C.D. & Sacchettini, J.C. Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-.ANG. Resolution. *Biochemistry* **33**, 10871-10877 (1994).

76. Rai, A., Smita, S.S., Singh, A.K., Shanker, K. & Nagegowda, D.A. Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. *Molecular plant* **6**, 1531-1549 (2013).

77. Ogawa, T., Yoshimura, T. & Hemmi, H. Geranylfarnesyl diphosphate synthase from *Methanosarcina mazei*: different role, different evolution. *Biochemical and Biophysical Research Communications* **393**, 16-20 (2010).

78. Tachibana, A. *et al*. Novel prenyltransferase gene encoding farnesylgeranyl diphosphate synthase from a hyperthermophilic archaeon, *Aeropyrum pernix*. *The FEBS Journal* **267**, 321-328 (2000).

79. Zhang, Y.W. *et al*. Two subunits of heptaprenyl diphosphate synthase of *Bacillus subtilis* form a catalytically active complex. *Biochemistry* **37**, 13411-13420 (1998).

80. Gin, P. & Clarke, C.F. Genetic evidence for a multi-subunit complex in coenzyme Q biosynthesis in yeast and the role of the Coq1 hexaprenyl diphosphate synthase. *Journal of Biological Chemistry* **280**, 2676-2681 (2005).

81. Asai, K.I. *et al*. The identification of *Escherichia coli* ispB (cel) gene encoding the octaprenyl diphosphate synthase. *Biochemical and Biophysical Research Communications* **202**, 340-345 (1994).

82. Ohnuma, S.I., Koyama, T. & Ogura, K. Purification of solanesyl-diphosphate synthase from *Micrococcus luteus*. A new class of prenyltransferase. *Journal of Biological Chemistry* **266**, 23706-23713 (1991).

83. Schulbach, M.C., Brennan, P.J. & Crick, D.C. Identification of a short (C15) chainZ-Isoprenyl dphosphate synthase and a homologous long (C50) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **275**, 22876-22881 (2000).

84. Sato, T. Unique biosynthesis of sesquarterpenes (C35 terpenes). *Bioscience, Biotechnology, and Biochemistry* **77**, 1155-1159 (2013).

85. Apfel, C.M., Takacs, B., Fountoulakis, M., Stieger, M. & Keck, W. Use of genomics to identify bacterial undecaprenyl pyrophosphate synthetase: cloning, expression, and characterization of the essential uppS gene. *Journal of Bacteriology* **181**, 483-492 (1999).

86. Matsuoka, S., Sagami, H., Kurisaki, A. & Ogura, K. Variable product specificity of microsomal dehydrodolichyl diphosphate synthase from rat liver. *Journal of Biological Chemistry* **266**, 3464-3468 (1991).

87. Kharel, Y. & Koyama, T. Molecular analysis of *cis*-prenyl chain elongating enzymes. *Natural Product Reports* **20**, 111-118 (2003).

88. Thulasiram, H.V. & Poulter, C.D. Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity. *Journal of the American Chemical Society* **128**, 15819-15823 (2006).

89. Wang, K. & Ohnuma, S.i. Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends in Biochemical Sciences* **24**, 445-451 (1999).

90. Fujihashi, M. *et al*. Crystal structure of *cis*-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *Proceedings of The National Academy of Sciences* **98**, 4337-4342 (2001).

91. Guo, R.T. *et al*. Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate role of the metal ion and conserved residues in catalysis. *Journal of Biological Chemistry* **280**, 20762-20774 (2005).

92. Liang, P.H., Ko, T.P. & Wang, A.H.J. Structure, mechanism and function of prenyltransferases. *The FEBS Journal* **269**, 3339-3354 (2002).

93. Thulasiram, H.V., Erickson, H.K. & Poulter, C.D. Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science* **316**, 73-76 (2007).

94. Rosenstiel, T.N., Fisher, A.J., Fall, R. & Monson, R.K. Differential accumulation of dimethylallyl diphosphate in leaves and needles of isoprene- and methylbutenol-emitting and nonemitting species. *Plant Physiology* **129**, 1276-1284 (2002).

95. Sharkey, T.D., Wiberley, A.E. & Donohue, A.R. Isoprene emission from plants: why and how. *Annals of Botany* **101**, 5-18 (2007).

96. Kakimoto, T. Biosynthesis of cytokinins. *Journal of Plant Research* **116**, 233-239 (2003).

97. Dewick, P.M. The biosynthesis of C 5-C 25 terpenoid compounds. *Natural Product Reports* **19**, 181-222 (2002).

98. Banthorpe, D.V., Charlwood, B.V. & Francis, M.J.O. Biosynthesis of monoterpenes. *Chemical Reviews* **72**, 115-155 (1972).

99. Chern, L.Y. Monoterpenes in plants-a mini review. *Asian Journal of Plant Biology* **1**, 15-19 (2014).

100. Asadollahi, M.A. *et al*. Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnology and Bioengineering* **99**, 666-677 (2008).

101. Ogura, M., Cordell, G.A. & Farnsworth, N.R. Anticancer sesquiterpene lactones of *Michelia compressa* (*Magnoliaceae*). *Phytochemistry* **17**, 957-961 (1978).

102. Hayashi, K., Hayashi, T., Ujita, K. & Takaishi, Y. Characterization of antiviral activity of a sesquiterpene, triptofordin C-2. *Journal of Antimicrobial Chemotherapy* **37**, 759-768 (1996).

103. Teoh, K.H., Polichuk, D.R., Reed, D.W., Nowak, G. & Covello, P.S. *Artemisia annua* L.(*Asteraceae*) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS letters* **580**, 1411-1416 (2006).

104. Zerbe, P. & Bohlmann, J. Plant diterpene synthases: exploring modularity and metabolic diversity for bioengineering. *Trends in Biotechnology* **33**, 419-428 (2015).

105. Phillips, D.R., Rasbery, J.M., Bartel, B. & Matsuda, S.P.T. Biosynthetic diversity in plant triterpene cyclization. *Current Opinion in Plant Biology* **9**, 305-314 (2006).

106. Xu, R., Fazio, G.C. & Matsuda, S.P.T. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**, 261-291 (2004).

107. Hill, R.A. & Connolly, J.D. Triterpenoids. *Natural Product Reports* **30**, 1028-1065 (2013).

108. Sawai, S. & Saito, K. Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science* **2**, 25 (2011).

109. Giuliano, G. Plant carotenoids: genomics meets multi-gene engineering. *Current Opinion in Plant Biology* **19**, 111-117 (2014).

110. Noviendri, D., Hasrini, R.F. & Octavianti, F. Carotenoids: Sources, medicinal properties and their application in food and nutraceutical industry. *Journal of Medicinal Plants Research* **5**, 7119-7131 (2011).

111. Gordon, H.T., Bauernfeind, J.C. & Furia, T.E. Carotenoids as food colorants. *CRC Critical Reviews in Food Science and Nutrition* **18**, 59-97 (1983).

112. Swiezewska, E. & Danikiewicz, W. Polyisoprenoids: structure, biosynthesis and function. *Progress in Lipid Research* **44**, 235-258 (2005).

113. Ribeiro, N., Gotoh, M., Nakatani, Y., Deì•saubry, L. & Pramatarova, A.L.D. Biomimetic synthesis and properties of polyprenoid. in *On Biomimetics* Ch. 09 (InTech, Rijeka, 2011).

114. Rip, J.W., Rupar, C.A., Ravi, K. & Carroll, K.K. Distribution, metabolism and function of dolichol and polyprenols. *Progress in Lipid Research* **24**, 269-309 (1985).

115. Khidyrova, N.K. & Shakhidoyatov, K.M. Plant polyprenols and their biological activity. *Chemistry of Natural Compounds* **38**, 107-121 (2002).

116. Nowicka, B. & Kruk, J. Occurrence, biosynthesis and function of isoprenoid quinones. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1797**, 1587-1605 (2010).

117. Biswas, K., Chattopadhyay, I., Banerjee, R.K. & Bandyopadhyay, U. Biological activities and medicinal properties of neem (*Azadirachta indica*). *Current Science* **82**, 1336-1345 (2002).

118. Schmutterer, H. The Neem Tree. Edited by VCH Verlagsgesellschaft GmbH. (Weinheim, Germany, 1995).

119. Maithani, A., Parcha, V., Pant, G., Dhulia, I. & Kumar, D. *Azadirachta indica* (neem) leaf: A review. *Journal of Pharmacy Research* **4**, 1824-1827 (2011).

120. Brahmachari, G. Neem-an omnipotent plant: a retrospection. *Chembiochem* **5**, 408-421 (2004).

121. Lim, T.K. *Azadirachta indica*. in *Edible medicinal and non medicinal plants* 409-455 (Springer, 2014).

122. Girish, K. & Shankara Bhat, S. Neem-a green treasure. *Electronic Journal of Biology* **4**, 201-111 (2008).

123. Zhang, Y. & Xu, H. Recent progress in the chemistry and biology of limonoids. *RSC Advances* **7**, 35191-35220 (2017).

124. Yan, X.H. *et al*. Chemical constituents from fruits of *Harrisonia perforata*. *Phytochemistry* **72**, 508-513 (2011).

125. Siddiqui, S. A note on the isolation of three new bitter principles from the neem oil. *Current Science* **11**, 278-279 (1942).

126. Butterworth, J.H. & Morgan, E.D. Isolation of a substance that suppresses feeding in locusts. *Chemical Communications (London)*, 23-24 (1968).

127. Aktar, M.W. & Sengupta, D. Neem: A great boon to mankind. *Pesticide Residue Laboratory, Deptt. Of Agril. Chemicals Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia West Bengal, India* (2008).

128. Sharma, R.K. Biosynthesis of plant sterols and triterpenoids. The incorporation of (3RS)-[2-^{14}C,(4R)-4-^{3}H$_1$] mevalonate into α-spinasterol and β-amyrin in *Camellia sinensis*. *Phytochemistry* **9**, 565-568 (1970).

129. Dauben, W.G., Ban, Y. & Richards, J.H. The biosynthesis of the triterpene, eburicoic acid: the utilization of methyl-labeled acetate1 2. *Journal of The American Chemical Society* **79**, 968-970 (1957).

130. Rees, H.H., Mercer, E.I. & Goodwin, T.W. The stereospecific biosynthesis of plant sterols and α-and β-amyrin. *Biochemical Journal* **99**, 726 (1966).

131. Groeneveld, H.W. Biosynthesis of latex triterpenes in *Euphoriba*; Evidance for a dueal synthesis. *Acta Botanica Neerlandica* **25**, 459-473 (1976).

132. Akhila, A., Srivastava, M. & Rani, K. Production of radioactive azadirachtin in the seed kernels of *Azadirachta indica* (the Indian neem tree). *Natural Product Letters* **11**, 107-110 (1998).

133. Rani, K. & Akhila, A. Biosynthetic relationship between nemocinol and nimocinolide in *Azadirachta indica*. *Natural Product Letters* **4**, 179-182 (1994).

134. Ekundayo, O. Biosynthesis of nimbolide in *Azadirachta indica* A. Juss from (2-^{14}C) mevalonate and (2-^{14}C) acetate. *Zeitschrift Fur Pflanzenphysiologie* **112**, 139-146 (1983).

135. Ekong, D.E.U., Ibiyemi, S.A. & Olagbemi, E.O. The meliacins (limonoids). Biosynthesis of nimbolide in the leaves of *Azadirachta indica*. *Journal of the Chemical Society D: Chemical Communications*, 1117-1118 (1971).

136. Ekong, D.E.U. & Ibiyemi, S.A. Biosynthesis of nimbolide from [2-^{14}C,(4R) 4-^{3}H$_1$] mevalonic acid lactone in the leaves of *Azadirachta indica*. *Phytochemistry* **24**, 2259-2261 (1985).

137. Siddiqui, S., Siddiqui, B.S., Faizi, S. & Mahmood, T. Tetracyclic triterpenoids and their derivatives from *Azadirachta indica*. *Journal of Natural Products* **51**, 30-43 (1988).

138. Jones, P.S., Ley, S.V., Morgan, E.D. & Santafianos, D. The chemistry of the neem tree. *Focus on Phytochemical Pesticides* **1**, 19-45 (1989).

139. Champagne, D.E., Koul, O., Isman, M.B., Scudder, G.G.E. & Towers, G.H.N. Biological activity of limonoids from the Rutales. *Phytochemistry* **31**, 377-394 (1992).

140. Krishnan, N.M. *et al*. *De novo* sequencing and assembly of *Azadirachta indica* fruit transcriptome. *Current Science* **101**, 1553-1561 (2011).

141. Krishnan, N.M. *et al*. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *Bmc Genomics* **13**, 464 (2012).

142. Kuravadi, N.A. *et al*. Comprehensive analyses of genomes, transcriptomes and metabolites of neem tree. *PeerJ* **3**, e1066 (2015).

143. Wang, S., Zhang, H., Li, X. & Zhang, J. Gene expression profiling analysis reveals a crucial gene regulating metabolism in adventitious roots of neem (*Azadirachta indica*). *RSC Advances* **6**, 114889-114898 (2016).

144. Narnoliya, L.K., Rajakani, R., Sangwan, N.S., Gupta, V. & Sangwan, R.S. Comparative transcripts profiling of fruit mesocarp and endocarp relevant to secondary metabolism by suppression subtractive hybridization in *Azadirachta indica* (neem). *Molecular Biology Reports* **41**, 3147-3162 (2014).

145. Bhambhani, S. *et al*. Transcriptome and metabolite analyses in *Azadirachta indica*: identification of genes involved in biosynthesis of bioactive triterpenoids. *Scientific Reports* **7**, 5043 (2017).

146. Dayanandan, P. & Ponsamuel, J. Ultrastructure of Terpenoid Secretory Cells of Neem (*Azadirachta indica* A. Juss.). *Electron Microscopy in Medicine and Biology, PD Gupta and H. Yamamoto, eds.(New Delhi: Oxford & IBH Publishing Co. Pvt. Ltd.)*, 179-195 (2000).

147. Paul, R., Prasad, M. & Sah, N.K. Anticancer biology of *Azadirachta indica* L (neem): a mini review. *Cancer Biology & Therapy* **12**, 467-476 (2011).

148. Nagini, S. & Priyadarsini, R.V. *Azadirachta indica* (neem) and neem limonoids as anticancer agents: molecular mechanisms and targets. in *Perspectives in Cancer Prevention-Translational Cancer Research* 45-60 (Springer, 2014).

149. Subapriya, R. & Nagini, S. Medicinal properties of neem leaves: a review. *Current Medicinal Chemistry-Anti-Cancer Agents* **5**, 149-156 (2005).

150. Maneerat, W., Laphookhieo, S., Koysomboon, S. & Chantrapromma, K. Antimalarial, antimycobacterial and cytotoxic limonoids from *Chisocheton siamensis*. *Phytomedicine* **15**, 1130-1134 (2008).

151. Siddiqui, S., Faizi, S. & Siddiqui, B.S. Constituents of *Azadirachta indica*: isolation and structure elucidation of a new antibacterial tetranortriterpenoid, mahmoodin, and a new protolimonoid, naheedin. *Journal of Natural Products* **55**, 303-310 (1992).

152. Gupta, S.C., Prasad, S., Tyagi, A.K., Kunnumakkara, A.B. & Aggarwal, B.B. Neem (*Azadirachta indica*): An Indian traditional panacea with modern molecular basis. *Phytomedicine* **34**, 14-20 (2017).

153. Ponnusamy, S. *et al*. Gedunin and Azadiradione: Human Pancreatic α-Amylase Inhibiting Limonoids from Neem (*Azadirachta indica*) as Anti-Diabetic Agents. *PloS one* **10**, e0140113 (2015).

154. Tapanelli, S. *et al*. Transmission blocking effects of neem (*Azadirachta indica*) seed kernel limonoids on Plasmodium berghei early sporogonic development. *Fitoterapia* **114**, 122-126 (2016).

155. Schoonhoven, L.M. & Jermy, T. A behavioral and electrophysical analysis of insect feeding deter-rents. *Crop Protection Agents-their Biological Evaluation*, 133-146 (1977).

156. Sieber, K.P. & Rembold, H. The effects of azadirachtin on the endocrine control of moulting in *Locusta migratoria*. *Journal of Insect Physiology* **29**, 523-527 (1983).

157. Rembold, H. Azadirachtins: Their structure and mode of action. 150-163 (ACS Publications, 1989).

158. Schmutterer, H. Properties and potential of natural pesticides from the neem tree, *Azadirachta indica*. *Annual Review of Entomology* **35**, 271-297 (1990).

159. Veitch, G.E., Boyer, A. & Ley, S.V. The azadirachtin story. *Angewandte Chemie International Edition* **47**, 9402-9429 (2008).

160. Jauch, J. Total Synthesis of Azadirachtin-Finally Completed After 22 Years. *Angewandte Chemie International Edition* **47**, 34-37 (2008).

161. Newman, J.D. *et al*. High-level production of amorpha-4, 11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnology and Bioengineering* **95**, 684-691 (2006).

162. Kizer, L., Pitera, D.J., Pfleger, B.F. & Keasling, J.D. Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and Environmental Microbiology* **74**, 3229-3241 (2008).

163. Tsuruta, H. *et al*. High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PloS one* **4**, e4489 (2009).

164. Huang, Q., Roessner, C.A., Croteau, R. & Scott, A.I. Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorganic & Medicinal Chemistry* **9**, 2237-2242 (2001).

165. Ajikumar, P.K. *et al*. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70-74 (2010).

166. Withers, S.T. & Keasling, J.D. Biosynthesis and engineering of isoprenoid small molecules. *Applied Microbiology and Biotechnology* **73**, 980-990 (2007).

167. Hong, K.K. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences* **69**, 2671-2690 (2012).

168. Ro, D.K. *et al*. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).

169. Westfall, P.J. *et al*. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences* **109**, E111-E118 (2012).

170. Teoh, K.H., Polichuk, D.R., Reed, D.W. & Covello, P.S. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* **87**, 635-642 (2009).

171. Kirby, J., Romanini, D.W., Paradise, E.M. & Keasling, J.D. Engineering triterpene production in *Saccharomyces cerevisiae*-β-amyrin synthase from *Artemisia annua*. *The FEBS Journal* **275**, 1852-1859 (2008).

172. Dai, Z. *et al*. Metabolic engineering of *Saccharomyces cerevisiae* for production of ginsenosides. *Metabolic Engineering* **20**, 146-156 (2013).

173. Gandhi, S.G., Mahajan, V. & Bedi, Y.S. Changing trends in biotechnology of secondary metabolism in medicinal and aromatic plants. *Planta* **241**, 303-317 (2015).

174. Berdy, J. Bioactive microbial metabolites. *The Journal of antibiotics* **58**, 1 (2005).

175. Patridge, E., Gareiss, P., Kinch, M.S. & Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug discovery today* **21**, 204-207 (2016).

176. Dunn, W.B. & Ellis, D.I. Metabolomics: current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry* **24**, 285-294 (2005).

177. Egan, A.N., Schlueter, J. & Spooner, D.M. Applications of next-generation sequencing in plant biology. *American journal of Botany* **99**, 175-185 (2012).

178. Caspi, R. *et al*. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research* **36**, D623-D631 (2007).

179. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).

180. Schomburg, I. *et al*. BRENDA: a resource for enzyme data and metabolic information. *Trends in Biochemical Sciences.* **27**, 54-56 (2002).

181. Rapaport, F. *et al*. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology* **14**, 3158 (2013).

182. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**(2005).

183. Yonekura-Sakakibara, K. *et al*. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. *The plant cell* **20**, 2160-2176 (2008).

184. Qi, X. *et al*. A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8233-8238 (2004).

185. Field, B. & Osbourn, A.E. Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Science* **320**, 543-547 (2008).

186. Sweetlove, L.J. & Fernie, A.R. The spatial organization of metabolism within the plant cell. *Annual review of plant biology* **64**, 723-746 (2013).

187. Jorgensen, K. *et al*. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Current Opinion in Plant Biology* **8**, 280-291 (2005).

188. Ralston, L. & Yu, O. Metabolons involving plant cytochrome P450s. *Phytochemistry reviews* **5**, 459 (2006).

189. Farre, G., Twyman, R.M., Christou, P., Capell, T. & Zhu, C. Knowledge-driven approaches for engineering complex metabolic pathways in plants. *Current Opinion in Biotechnology* **32**, 54-60 (2015).

190. Luo, Y. *et al*. Engineered biosynthesis of natural products in heterologous hosts. *Chemical Society Reviews* **44**, 5265-5290 (2015).

191. Lau, W., Fischbach, M.A., Osbourn, A. & Sattely, E.S. Key applications of plant metabolic engineering. *PLoS Biology* **12**, e1001879 (2014).

192. Casini, A., Storch, M., Baldwin, G.S. & Ellis, T. Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology* **16**, 568 (2015).

193. Cobb, R.E., Ning, J.C. & Zhao, H. DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *Journal of Industrial Microbiology & Biotechnology* **41**, 469-477 (2014).

194. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).

195. Patron, N.J. *et al*. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist* **208**, 13-19 (2015).

196. Broun, P. Transcription factors as tools for metabolic engineering in plants. *Current Opinion in Plant Biology* **7**, 202-209 (2004).

197. Tang, G., Galili, G. & Zhuang, X. RNAi and microRNA: breakthrough technologies for the improvement of plant nutritional value and metabolic engineering. *Metabolomics* **3**, 357-369 (2007).

"When I started working, I thought of DNA as an inert substance. The notion of DNA containing all the information for making a complete organism would have been thought of as science fiction. But that is the way it is. The DNA is the genome and we now know how to read the information contained in it "

Sanger, quoted in Fletcher and Porter, 1997, p. 72.

Leonhard Euler.

Chapter 2

Neem Transcriptome

Analysis

Modified from Elsa Góngora-Castilloa and C. Robin Buell, 2013.

2.1 Introduction

2.1.1 Transcriptome

The order of nucleic acids in polynucleotide chains ultimately contains the information for the hereditary, biochemical and physiological properties of terrestrial life. Therefore sequencing of polynucleotides is very important in biological research[1]. Concepts of DNA replication and protein encoding by nucleic acids was understood very well followed by famous DNA structure solved by Watson and Crick in 1953[2] by working on the crystallographic data produced by Rosalind Franklin[3], Linus Pauling[4] and Maurice Wilkins[5]. Successful efforts in the sequencing of DNA came in 1977 by Sanger's chain-termination by dideoxynucleotides and Maxam, Gilbert's chemical cleavage techniques[6,7]. A number of improvements were made to Sanger sequencing in the following years, which primarily involved the replacement of cumbersome radiolabeling method with fluorometric based detection and improved detection through capillary-based electrophoresis. Both of these improvements contributed to the development of automated first-generation DNA sequencing machines[8] that has been used for Human Genome Project[9,10]. Till now, Sangers method gives the maximum sequencing length of 1000 bp with an accuracy of 99.999%. Despite these methods, Sangers method suffers from following disadvantages like gels or polymers used for separation of fluorescently labelled DNA fragments, in which, only a low number of samples could be analyzed parallelly and also the difficulty of total automation of sample preparation methods[11]. To overcome these disadvantages, the next generation sequencing (NGS) platforms were developed during the 20th century. 454 pyrosequencing was the 1st next-generation sequencing introduced in 2005 and with recent improvements, lead to read length of 400 - 500 bp[12]. In 2006, Illumina (Solexa) Genome Analyzer was introduced with the principle of sequencing by synthesis with reversible terminator nucleotides having fluorescent dye and containing maximum read length of 150 bp[13]. In October 2007, Applied Biosystems SOLiD sequencer was introduced with a principle of sequencing by oligo ligation and detection[14]. In 2010, ion torrent sequencing was introduced with the principle of detection of a change in conductivity due to hydrogen ions released during sequencing[15]. The limitations of next-generation sequences are low read length

and inaccurate calling of homopolymer length[16]. To overcome these problems, single molecular sequencers are under development. Heliscope single molecular sequencer was the first introduced in 2007, with the principle of true single molecule sequencing or one base at a time technology. Other technologies like single molecular real-time (SMRT[TM]) and nanopore DNA sequencers are under development[13,15]. Apart from genome sequencing, NGS has numerous applications in biological research[17] such as

a) Transcriptome (sequencing of cDNA from cell or group of cells) analysis helps in gene expression profiling, genome annotation, rearrangement detection.

b) Noncoding RNA discovery and profiling.

c) Chromatin immune precipitation followed by sequencing (CHiP –Seq) helps in identification of promoter regions, transcription factor or DNA binding sites and methylation pattern in the genome, posttranslational modification of histones and nucleosome positioning on a genome-wide scale.

d) Metagenomics.

Next-generation sequencing technologies are now in common use in biology. However, short reads produced are large and complex, interpretation is not a straightforward process. Bioinformatics tools used for analysis of NGS data are critical for analyzing short reads[18]. In general, due to short reads, NGS would restrict to resequencing applications such as single-nucleotide polymorphism, genetic variations and expression profile[19,20]. In the absence of reference sequences, the key step for analysis is *de novo* sequence assembly. This assembly is done by using overlap layout consensus (OLC), de Bruijn graph and string graph[21]. The first step in OLC is finding overlaps between sequence reads to generate a layout and then derive a consensus sequence. In de Bruijn graph, the short reads are cut into smaller pieces called *K*-mers. The graph is generated by taking *K*-1 mers as vertices or nodes and *K*-mers as edges and then finding the Eulerian path[22]. Mayer proposed the string graph, where the overlap graph is generated from all the sequence reads which are followed by merging or reducing redundant overlaps or edges and removing the false vertices and edges; finally assembly is done by identifying a Eulerian path in graph[23]. Paired reads which are a source of long-range information is applied to the above graphs and

scaffolds are generated. The functional annotation of genes or transcripts are then done by different software likes BLAST, KAAS, BLAST2GO, Pfam, InterPro scan, gene ontology and others.

2.1.2 Profiling of Neem Triterpenoids

Identification of gene function is the main focus in the post-genomic era. A major task in functional characterization of enzymes is the heterologous expression, purification and *in-vitro* or *ex-vivo* assay. Plant functional genomics couples the generation of transgenic, metabolic engineered and mutant plants to the multi-parallel analysis of gene function at different levels such as mRNA, protein and metabolite [24,25]. Metabolic profiling defines precisely the biochemical function of plant metabolism in a particular tissue or condition. These analyses improve the functional genomics methodologies while offering a direct link between a gene sequence and the function of the metabolic network in plants[26,27]. Targeted metabolomics is all about the identification and quantification of known metabolites and their time and space resolved distribution in a specific biological system[28]. Hyphenated mass spectrometry is a powerful and most utilized analytical technique in metabolomics due to its high sensitivity, accuracy, resolution, low sample requirement and ability to monitor a broad range of metabolites [29-31].

Triterpenoids in neem are diverse in skeletal architecture, huge in the count and their abundance is highly tissue-specific. From the different tissues amount of crude extract obtained (by solvent partition technique) was directly correlated with the triterpenoid content (Figure 2.1). The mature seed kernel and pericarp of initial stages were found to contain the highest amount of total triterpenoids. The quantitative level of individual fifteen triterpenoids across various tissues of neem has been represented in Figure 2.2. Among the fifteen triterpenoids, C-seco triterpenoids (nimbin, salannin, 6-deacetylnimbin, 3-deacetylsalannin, nimbanal, salannol acetate, nimbinene, 6-deacetylnimbinene, azadirachtin A and B) were observed in the kernel as compared to the other tissues. Fruit pericarp, flower and leaf contained majorly ring-intact triterpenoids (azadirone, azadiradione, gedunin, nimocinol, epoxyazadiradione). Analysis of neem triterpenoid profile indicates that limonoid biosynthesis takes place in different tissues of neem and is very high in fruits. The

genes which are involved in basic limonoids are highly expressed in fruit kernel and pericarp on the hypothesis of C-seco limonoids are synthesised from basic limonoids[32].
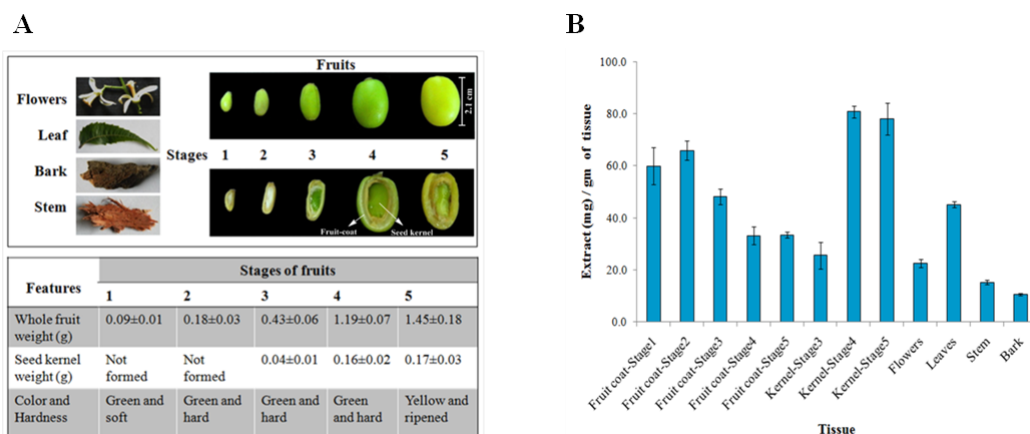
Figure 2. 1 Total Triterpenoid Content in Neem Tissues.
A) Different tissues of Neem and physical characteristics of Neem fruits at various stages. (B) Amount of triterpenoid extracts obtained from various tissues of Neem.
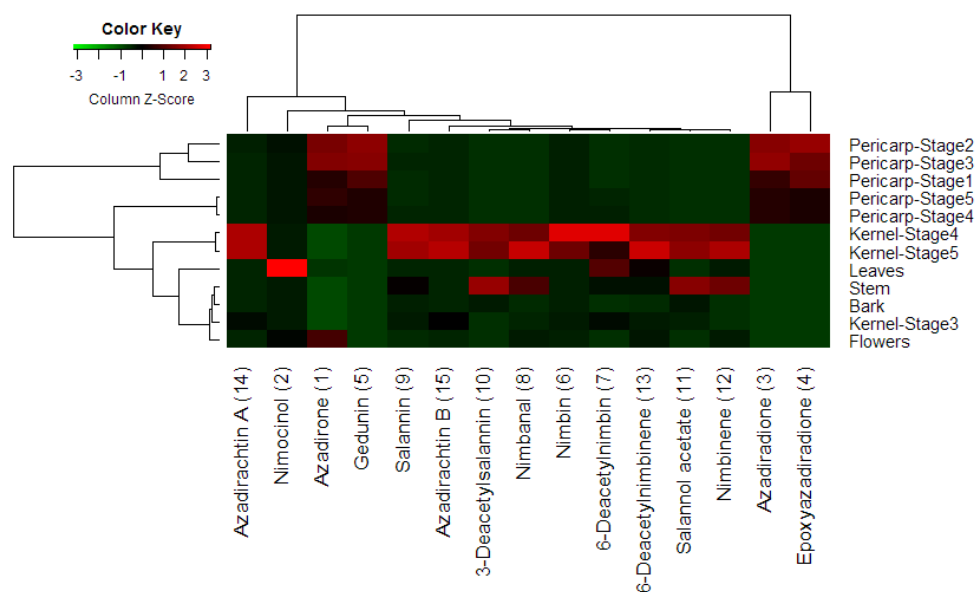
Figure 2. 2 Quantitative Abundance of Major Triterpenoids in Different Tissues of Neem.

The transcriptome is one of the best methods for functional genomics in the nonmodel organisms. Here, we present the isolation of RNA from tissues which showed different levels of triterpenoid content (fruit, flower and leaf). Initially, the

project was started with analyzing the transcriptome of pooled RNA from fruit, flower and leaves in equal quantity, which resulted in the identification of three triterpene synthases and CYP450s. However, which one of these involves in limonoids biosynthesis was not clear. In order to understand this, we proceeded for tissue-specific (pericarp, kernel, flower and leaves) transcriptome. The paired-end reads were generated by NGS and then assembled to identify transcripts. Transcriptome annotation was done by comparing with different databases such as NCBI, KEGG and Pfam leading to the identification of the putative genes involved in isoprenoid and triterpenoid biosynthesis.

2.2 Material and Methods

2.2.1 Plant Material

Neem tissues for transcriptome sequencing were collected from Pune region, Maharashtra, India in the period March to May. The tissues were flash frozen in liquid nitrogen and stored at -80 °C till further use.

2.2.2 Reagents and Kits

All the plastic and glassware used for RNA isolation were soaked in freshly prepared 0.1 % diethyl pyrocarbonate (DEPC) water, dried, and sterilized in an autoclave before use. All the solutions were prepared in 0.1 % DEPC and sterilized in an autoclave. Spectrum^{TM} Plant Total RNA Isolation Kit from Sigma-Aldrich, USA was used for RNA isolation. High Sensitivity Bioanalyzer Kit (Agilent) was used for analysis of RNA and transcriptome library quality. TruSeq RNA Sample preparation Kits (Illumina) was used for library preparation. SuperScript III Reverse Transcriptase (Invitrogen), HighPrep PCR (MAGBIO) and Agencourt AMPure XP SPRI beads (Beckman Coulter) were used for cDNA synthesis and PCR purification, respectively.

2.2.3 RNA Integrity

Total RNA was quantified using NanoDrop (Thermo Scientific) by measuring the optical density of isolated RNA in DEPC treated water. The integrity of total RNA was analyzed by running on 1 % agarose gel and High Sensitivity Bioanalyzer Chip (Agilent).

2.2.4 Transcriptome Sequencing

Library preparation was performed at Genotypic Technology's Genomics facility following Illumina TruSeq RNA library protocol outlined in "TruSeq RNA Sample Preparation Guide". 1μg of total RNA (kernel, pericarp, flowers and leaves and pooled RNA from seeds, flower and leaves) was subjected to Poly A purification for mRNA. Purified mRNA was fragmented for 2-8 minutes at 94 °C in the presence of divalent cations and reverse transcribed with Superscript III reverse transcriptase by priming with random hexamers. Second strand cDNA was synthesized in the presence of DNA polymerase I and RnaseH and cleaned up using HighPrep PCR. Illumina adapters were ligated to the cDNA molecules after end repair and addition of A base. SPRI cleanup was performed after ligation. The library was amplified using 8 cycles of PCR for the enrichment of adapter-ligated fragments. The prepared library was quantified using Nanodrop and validated for quality by running an aliquot on High Sensitivity Bioanalyzer Chip.

2.2.5 *De novo* Transcriptome Assembly

Paired-end reads were generated by using Illumina GAII analyser and Illumina Nextseq500 for pooled RNA and individual four tissues respectively. Raw read quality was checked by using low SeqQC-2.1 and RNAseq. Then reads were processed by the in-house script for adapters and low-quality bases trimming towards 3' end. The pooled RNA sample reads were assembled into contigs by using Velvet_1.1.05[33] with optimized hash length 41. Contigs were submitted to Oasis_0.2.01[34] to generate transcripts. Tissue-specific raw reads were assembled by Trinity with hash length 25[35]. By using CD-HIT transcripts from all the tissues, they were clustered. RPKM values were generated to the clustered transcripts based on raw reads.

2.2.6 Transcriptome Functional Annotation

Neem transcripts were submitted to Blastx against non-redundant database available at NCBI with an E-value cutoff of 10^{-5}. Pathway annotation was done by bidirectional best hit method of KAAS[36] (KEGG Automatic Annotation Server. http://www.genome.jp/kegg/kaas) by selecting plant reference database. Virtual ribosome[37], (http://www.cbs.dtu.dk/services/VirtualRibosome/) a web-based server,

was used for deducing the ORFs of these transcripts. The peptide sequences of transcripts with length more than 99 amino acids were submitted to Pfam batch search [38] (http://pfam.xfam.org/search#tabview=tab1). The differential gene expression analysis was done by using DESeq2[39] tools with p-adjust value <0.05.

2.3 Results and Discussion

2.3.1 RNA Isolation from Neem (*Azadirachta indica*)

Triterpenoid profiling in neem indicated that there is tissue-specific variation in their abundance. The mature seed kernel and initial stages of pericarp were found to contain the highest amount of limonoids. Furthermore, a wide diversity of triterpenoids, especially C-seco triterpenoids were observed in the kernel as compared to the other tissues. Pericarp, flower and leaf contained mainly ring-intact triterpenoids. Neem tissues such as seeds (kernel and pericarp), flower and leaves where the triterpenoids are abundant were selected further for transcriptome analysis. RNA was isolated from all the tissues using Spectrum™ Plant Total RNA Isolation Kit.
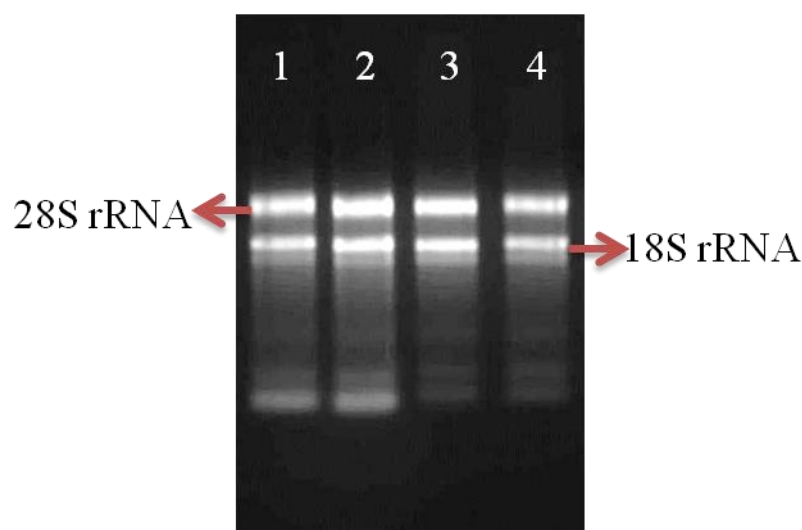
Figure 2. 3 Agarose Gel Electrophoresis of Total RNA Isolated from Different Tissues of Neem.
Lanes 1 – 4: Kernel, pericarp leaves and flowers.

Total RNA extracted from the seeds, kernel, pericarp, leaves and flower tissues of neem plant showed two distinct rRNA bands (28S rRNA and 18S rRNA) on a 1 % agarose gel electrophoresis in the ratio of 2:1, without degradation and also

showed no genomic DNA contamination (Figure 2.3). The A260/280 ratios were found to be 2.09, 2.07, 2.04, 2.04 and 2.01 and the A260/230 ratios were found to be 1.79, 2.43, 2.15, 2.38 and 2.15 respectively, for RNAs from pooled kernel, pericarp, leaves and flower tissues indicating no contamination of polysaccharides, protein, polyphenolics, salts and solvents. To check the integrity of isolated total RNA, an aliquot of the samples was run on an Agilent RNA Bioanalyzer chip (Figure 2.4).
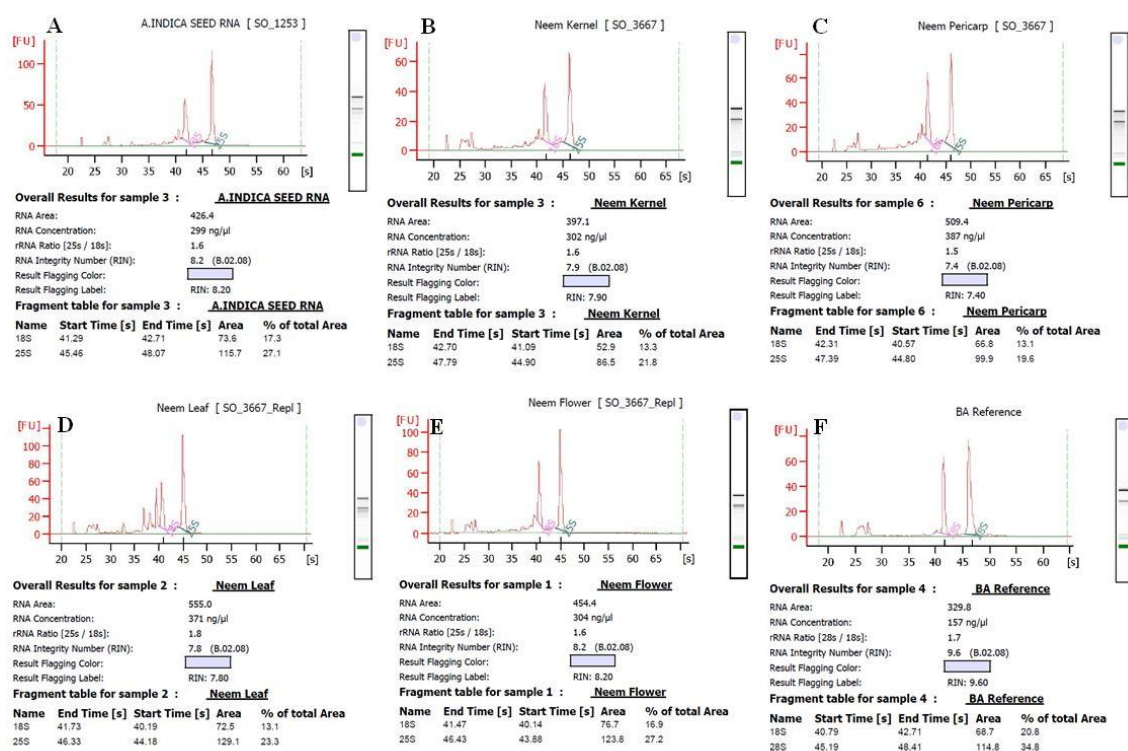
Figure 2. 4 Bioanalyzer Analyses of Total RNA from Different Neem Tissues.
A) Seeds, B) Kernel, C) Pericarp, D) Leaves, E) Flower and F) Reference.

2.3.2 Neem Transcriptome Sequencing and Assembly

2.3.2.1 Pooled RNA Transcriptome Sequencing and Assembly

Initially, the project was started by analyzing the transcriptome of pooled RNA from fruit, flower and leaves in equal quantity. TruSeq RNA Sample Preparation Guide was followed for production of sequencing library. The library was amplified using 8 cycles of PCR for the enrichment of adapter-ligated fragments. The library quality was checked by running a small aliquot on Bioanalyzer (Figure 2.5).
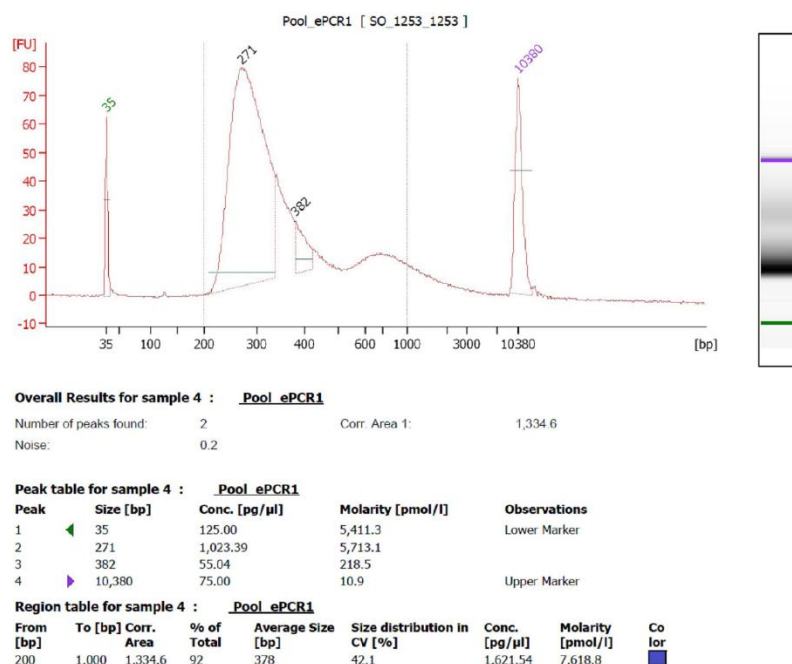
Figure 2. 5 Bioanalyzer Analyses of Pooled RNA Sequencing Library.

A total of 79,079,412 (79.08 million) paired-end reads each of 72 bp length were generated by Illumina GA II platform. 71,537,895 (90.46 %) high-quality reads were obtained with more than 20 phred score and reads of low quality were trimmed and used for further analysis. Total 27,390 contigs were generated using Velvet with a hash length of 41. These contigs were given as input for Oases to generate 41,140 transcripts. The average length of transcripts obtained was 1331 bp and the N50 length was 1953 bp (Table 2.1).

Table 2. 1 Statistics of Pooled RNA Transcriptome Sequencing and Assembly.

Total number of reads	79079412
Total Number of HQ Reads	71537895
Hash length	41
Contigs generated	27390
Average contig length	897.431
N50 length of contigs	1479
Transcripts generated	41140
N50 length of transcripts	1953
Number of assembled reads	68871778

2.3.2.2 Tissue-Specific Transcriptome Sequencing and Assembly

Tissue-specific (pericarp, kernel, flower and leaves) transcriptome was analyzed to understand the biosynthesis of secondary metabolites in neem focusing on triterpenoid biosynthesis. Tissue-specific sequencing library was constructed by Guidelines of TruSeq RNA Sample Preparation. The quality of prepared libraries was checked by running a small aliquot on Bioanalyzer (Figure 2.6). Illumina Nextseq500 was used to generate paired-end reads with read length 151 bp.
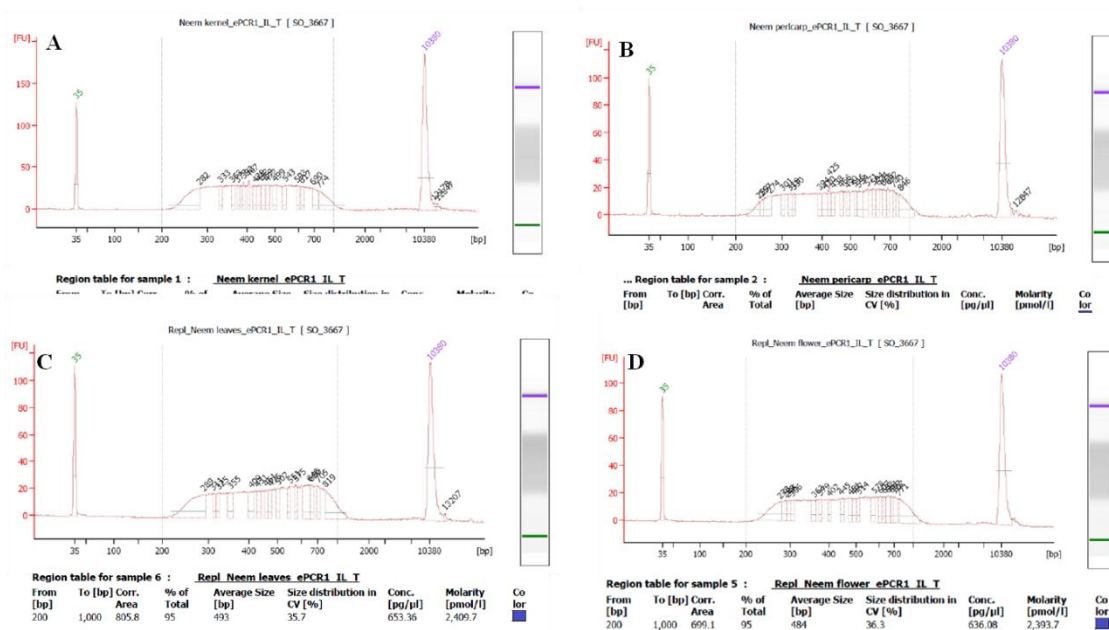
Figure 2. 6 Bioanalyzer Analyses of Sequencing Libraries.
A) kernel, B) pericarp, C) leaves and D) flowers.

A total of 209,882,798 (209.88 million) paired-end reads were generated out of which 192,730,444 (91.82 %) high-quality reads were obtained with more than 20 phred scores and read of low quality were trimmed and used for further analysis. Total 68111 (kernel), 66257 (pericarp), 70484 (leaves) and 73354 (flower) transcripts were generated by using Trinity with a hash length of 25. The average lengths of transcripts obtained were 1383 bp (kernel), 1370 bp (pericarp), 1386 bp (leaves) and 1288 bp (flower). Total 127,518 transcripts were generated by clustering all the transcripts from each tissue with CD-HIT[40] to remove the redundancy (Table 2.2).

Table 2. 2 Tissue-Specific Transcriptome Sequencing and Assembly Statistics of Neem.

Sample Name	Flower	Kernel	Leaf	Pericarp
Tool used		Trinity		
Hash length		25		
Transcripts Generated	73354	68111	70484	66257
Maximum Transcript Length	15952	15196	16031	15688
Minimum Transcript Length	300	300	300	300
Average Transcript Length	1335.5	1383.4	1386.7	1370.5
Median Transcript Length	1288	1227	424.5	3060
Total Transcripts Length	97960656	94225147	97737722	90806300
Transcripts >=300 b	73354	68111	70484	66257
Transcripts > 500 b	53148	51499	51851	48947
Transcripts > 1 Kb	35124	34432	35154	32687
Transcripts > 10 Kb	26	26	31	28
N50 value	2035	2039	2102	2057
Clustered Transcripts		127518		

2.3.3 Neem Transcriptome Functional Annotation

2.3.3.1 Pooled RNA Transcriptome Functional Annotation

All the transcripts were submitted to Blastx against non-redundant database available at NCBI with an E-value cutoff of 10^{-5}, where, a total of 32,856 (79.8 %) transcripts were annotated. Pathway annotation was carried out by KAAS (KEGG Automatic Annotation Server) with plant reference database. Out of the 41,140 transcripts, only 6281 transcripts were assigned 2749 unique KO numbers, which covered 223 pathways. The virtual ribosome, a web-based server, was used for finding the Open Reading Frame (ORF) of transcripts. 27,368 transcripts had an ORF with length more than 99 amino acids and 67 transcripts without any ORF. The peptide sequences of transcripts with length more than 99 amino acids were submitted to Pfam analysis. 18,807 transcripts were assigned different Pfam IDs. A total of 3467 different Pfam IDs were assigned to the transcripts (Figure 2.7).

Based on transcriptome annotation, all the genes involved in triterpenoid back-bone biosynthesis, right from isoprene units (MVA pathway and MEP pathway) to triterpene cyclases were found. A total of 134 transcripts predicted as cytochrome P450 monooxygenases and two transcripts as cytochrome P450 reductases were identified. Based on BLAST results, with reference to *Arabidopsis thaliana* cytochrome P450, Neem CYP450s were classified into 39 families and 78 subfamilies, out of which most of the CYP450 belonged to the CYP71 family. Seven transcripts were related to plant steroid biosynthesis and fifteen transcripts related to triterpenoid biosynthesis were predicted (Table 2.3).
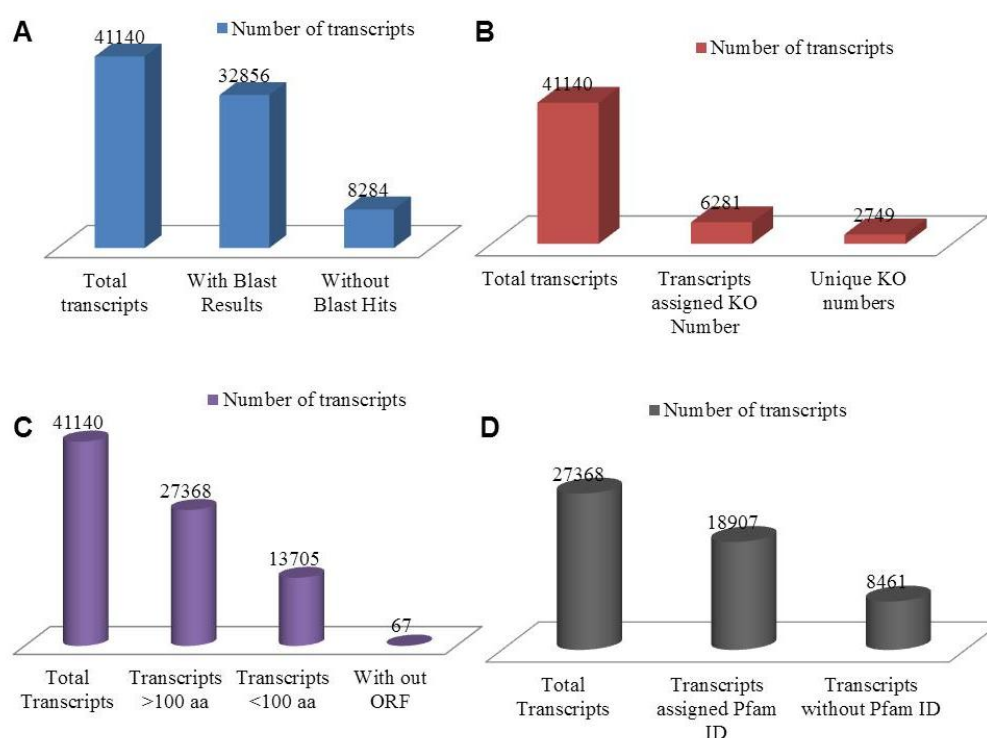
Figure 2. 7 Statistics in Functional Annotation of Neem Transcriptome.
A) Blastx analysis with NCBI nr database, B) KEGG functional annotation to assign KO (KEGG Orthology) numbers and pathways, C) Virtual ribosome analysis for identification of ORF and D) Transcript domain identification with Pfam database.

Table 2. 3 Predicted Genes Related to MVA, MEP and Triterpene Biosynthesis.

Predicted Genes for Triterpenoid Backbone Biosynthesis		
Mevalonate Pathway		Blastx Results
Acetyl-CoA C-acetyltransferase [EC: 2.3.1.9]	Neem_transcript_6172	96 % similarity with acetyl-CoA C-acetyltransferase *Hevea brasiliensis* [BAF98276]
	Neem_transcript_14672	90 % similarity with acetyl Co-A acetyltransferase *Hevea brasiliensis* [AAL18924]
	Master_Control_32747	75 % similarity with acetyl Co-A acetyltransferase, mitochondrial isoform X2 *Pieris rapae* [XP_022124008]
	Master_Control_32256	83 % similarity with acetyl Co-A acetyltransferase, mitochondrial isoform X1 *Dufourea novaeangliae* [XP_015433043]
Hydroxymethylglutaryl-CoA synthase [EC: 2.3.3.10]	Neem_transcript_13206	93 % similarity with hydroxy methylglutaryl-CoA synthase *Hevea brasiliensis* [BAF98279]
	Neem_transcript_36716	96 % similarity with hydroxy methylglutaryl-CoA synthase *Citrus clementina* [XP_024040489]
Hydroxymethylglutaryl-	Neem_transcript_11884	91 % similarity with 3-

CoA reductase (NADPH) [EC: 1.1.1.34]		hydroxy-3-methylglutaryl coenzyme A reductase 1 *Dimocarpus longan* [AET72044]
	Neem_transcript_21736	91 % similarity with 3-hydroxy-3-methylglutaryl coenzyme A reductase 2 *Dimocarpus longan* [AET72045]
	Master_Control_39226	85 % similarity with 3-hydroxy-3-methylglutaryl coenzyme A reductase *Salvia miltiorrhiza* [ACD37361]
	Master_Control_55031	60 % similarity with 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 *Nelumbo nucifera* [XP_010270571]
Mevalonate kinase [EC: 2.7.1.36]	Neem_transcript_9934	90 % similarity with mevalonate kinase *Hevea brasiliensis* [KM272630]
Phospho mevalonate Kinase [EC: 2.7.4.2]	Neem_transcript_27403	88 % similarity with PREDICTED: phospho mevalonate kinase *Vitis vinifera* [XP_002275808]
Diphosphomevalonate decarboxylase [EC: 4.1.1.33]	Neem_transcript_5109	93 % similarity with diphosphomevelonate decarboxylase *Hevea brasiliensis* [BAF98285]

Isopentenyl-diphosphate delta-isomerase [EC: 5.3.3.2]	Neem_transcript_31626	96 % similarity with isopentenyl diphosphate isomerase *Bupleurum chinense* [ACV74320]
Non-Mevalonate Pathway (MEP/DOXP pathway)		
1-deoxy-D-xylulose-5-phosphate synthase [EC: 2.2.1.7]	Neem_transcript_584	94 % similarity with 1-deoxyxylulose-5-phosphate synthase, putative *Ricinus communis* [XP_002516843]
	Neem_transcript_13351	96 % similarity with 1-deoxyxylulose-5-phosphate synthase, putative *Ricinus communis* [XP_002514364]
	Neem_transcript_23240	92 % similarity with 1-deoxyxylulose-5-phosphate synthase, putative *Ricinus communis* [XP_002532384]
1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC: 1.1.1.267]	Neem_transcript_31593	94 % similarity with 1-deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplast precursor, putative *Ricinus communis* [XP_002511399]
2-C-methyl-D-erythritol 4-phosphate cytidylyl Transferase [EC: 2.7.7.60]	Neem_transcript_19227	88 % similarity with 2-C-methyl-D-erythritol 4-phosphate cytidylyl

		transferase *Hevea brasiliensis* [BAF98291]
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [EC: 2.7.1.148]	Neem_transcript_4316	89 % similarity with 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, putative *Ricinus communis* [XP_002523216]
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [EC: 4.6.1.12]	Neem_transcript_24304	88 % similarity with 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase *Citrus jambhiri* [BAF73931]
(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase [EC: 1.17.7.1]	Neem_transcript_14312	95 % similarity with 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase *Hevea brasiliensis* [BAF98296]
4-hydroxy-3-methylbut-2-enyl diphosphatereductase [EC: 1.17.1.2]	Neem_transcript_350	95 % similarity with PREDICTED: 4-hydroxy-3-methylbut-2-enyl diphosphatereductase *Vitis vinifera* [XP_002284659]
Prenyl Pyrophosphate Synthase		
Geranyl diphosphate synthase [EC: 2.5.1.1]	Neem_transcript_10912 (AiGDS)	88 % similarity with geranyl diphosphat synthase *Quercus robur* [CAC20852]
	Neem_transcript_10001	95 % similarity with geranyl diphosphate synthase *Hevea brasiliensis* [BAF98300]

Farnesyl diphosphate synthase [EC: 2.5.1.10]	Neem_transcript_25722 (AiFDS)	90 % similarity with farnesyl pyrophosphate synthase *Cyclocarya paliurus* [ACY80695]
Geranylgeranyl diphosphate synthase [EC: 2.5.1.29]	Neem_transcript_1166	72 % similarity with geranylgeranyl pyrophosphate synthase 1 *Solanum lycopersicum* [NP_001234087]
	Neem_transcript_3894	86 % similarity with geranylgeranyl pyrophosphate synthase, putative *Ricinus communis* [XP_002529802]
	Neem_transcript_16200	93 % similarity with geranylgeranyl diphosphate synthase *Medicago sativa* [ADG01841]
	Neem_transcript_16736	69 % similarity with geranylgeranyl pyrophosphate synthase, *Jatropha curcas* [ADD82422]
	Neem_transcript_28215	72 % similarity with geranylgeranyl pyrophosphate synthase *Nicotiana tabacum* [ADD49735]

	Neem_transcript_30369	81 % similarity with geranylgeranyl pyrophosphate synthase 1 *Solanum pennellii* [ADZ24718]
	Neem_transcript_18547	79 % similarity with geranylgeranyl pyrophosphate synthase, putative *Ricinus communis* [XP_002531191]
Triterpenoid Biosynthesis		
Farnesyl-diphosphate farnesyl transferase [EC: 2.5.1.21]	Neem_transcript_33869	79 % Similarity with squalene synthase *Glycine max* [NP_001236365]
Squalene monooxygenase [EC: 1.14.13.132]	Neem_transcript_11067 (AiSQE1)	91 % Similarity with squalene monooxygenase putative *Ricinus communis* [XP_002530610]
	Neem_transcript_18229 (AiSQE2)	90 % Similarity with PREDICTED: squalene monooxygenase *Vitis vinifera* [XP_002271528]
	Neem_transcript_18980 (AiSQE3)	90 % Similarity with squalene monooxygenase, putative *Ricinus communis* [XP_002510043]
Triterpene cyclases	Neem_transcript_27436 (AiCAS)	92 % Similarity with cycloartenol synthase *Betula platyphylla* [Q8W3Z3]
	Neem_transcript_28920	86 % Similarity with β-amyrin synthase *Betula platyphylla*

	(AiTTS1)	[Q8W3Z1]
	Master_Control_74892 (AiTTS2)	77 % Similarity with β-amyrin synthase *Betula platyphylla* [Q8W3Z1]
	Master_Control_70149 (AiTTS3)	71 % Similarity with β-amyrin synthase *Kalopanax septemlobus* [ALO23119]
	Master_Control_70584 (AiTTS4)	70 % Similarity triterpene synthase *Eugenia uniflora* [AIK19225]
	Master_Control_101750 (AiTTS5)	77 % Similarity with β-amyrin synthase *Quercus suber* [POF03376]
Putative CYP related to triterpenoid biosynthesis	Neem_transcript_26034	44 % Similarity with β-amyrin 11-oxidase *Glycyrrhiza uralensis* [BAG68929]
	Neem_transcript_26318	65 % Similarity with Dammarenediol 12-hydroxylase *Panax ginseng* [AEY75213]
	Neem_transcript_34861 (AiCYP1)	62 % Similarity with Protopanaxadiol 6-hydroxylase *Panax ginseng* [AFO63031]
	Neem_transcript_10225	78 % Similarity with Cytochrome P450 CYP72A219 *Panax ginseng* [AEY75218]
	Neem_transcript_38933 (AiCYP2)	53 % Similarity with *Panax ginseng* [AFO63032]
	Neem_transcript_23030	59 % Similarity with Cytochrome P450 CYP736A12 *Panax ginseng* [AEY75215]

2.3.3.2 Tissue-Specific Transcriptome Functional Annotation

Tissue-specific transcriptomes were submitted to blastx against non-redundant database available at NCBI with an E-value cutoff of 10^{-5}. Total of 36546 (78.95 %), 33947 (77.04 %), 37163 (79.99 %) and 37477 (78.38 %) transcripts were annotated from kernel, pericarp, leaves and flower respectively (Figure 2.8 and 2.9). The clustered 127518 transcripts were submitted to KAAS analysis by taking plant reference database. 17823 (14 %) transcripts were assigned 4037 unique KO numbers, which covered 250 pathways. In virtual ribosome analysis, 126258 transcripts had an ORF length more than 20 amino acids and 4 transcripts without any ORF. The peptide sequences of transcripts with length more than 20 amino acids were submitted to Pfam analysis. 53168 transcripts were assigned unique Pfam IDs. A total of 4930 different Pfam IDs were assigned to the transcripts.
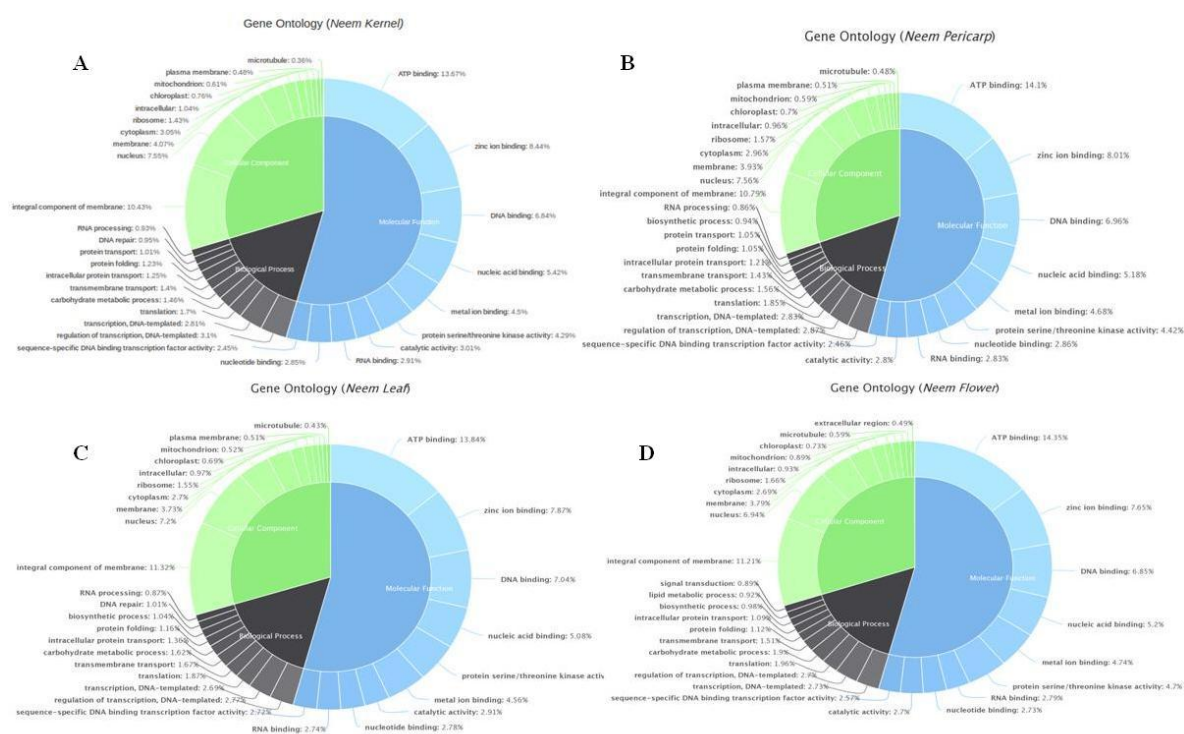
Figure 2. 8 Transcripts Distribution Based on Gene Ontology in Kernel, Pericarp, Leaves and Flowers Respectively.

KAAS analysis

Virtual Ribosome
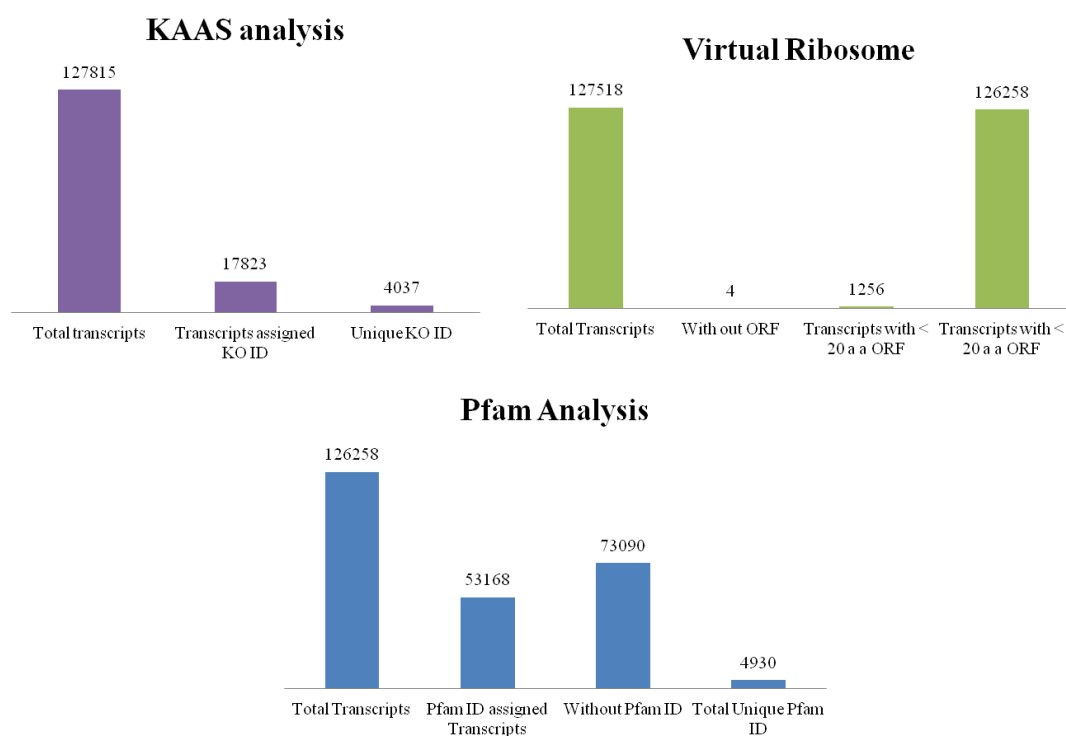
Pfam Analysis

Figure 2. 9 Functional Annotation of Tissue-Specific Transcriptome with KEGG and Pfam Database.

Tissue-specific transcriptome annotation helped in-depth analysis and identification of the genes related to MVA, MEP pathways and triterpenoid biosynthesis as compared to the pooled transcriptome. Two genes related to acetyl-CoA-acetyltransferase were known from pooled transcriptome and further two more transcripts were identified Master_Control_32747 and Master_Control_32256 from tissue-specific transcriptome. Master_Control_49393 was identified as hydroxymethyl glutaryl-CoA synthase in additional to two transcripts identified from pooled neem transcriptome (Table 2.3). Two more transcripts, Master_Control_39226 and Master_Control_55031 were identified as HMG reductases in addition to two identified hydroxymethyl glutaryl-CoA reductase from pooled transcriptome. Four triterpene synthases (Master_Control_74892, Master_Control_101750, Master_Control_70149 and Master_Control_70584) were identified in addition to two identified triterpene synthases from the pooled transcriptome.

2.3.4 Neem Terpenoid Metabolism

Secondary metabolites are synthesized from primary metabolites for enhancing particular biological function in living systems, mainly to interact with their environment. Analyzing the terpenoid metabolism in neem helps in understanding the efficient utilization of isoprene units for biosynthesis of triterpenoids from different tissues. The functional annotation of transcriptome by Blastx, KEGG and Pfam resulted in the identification of genes related to terpenoid biosynthetic pathways in neem such as MVA pathway, MEP pathway, hemiterpenoid, monoterpenoid, sesquiterpenoid, diterpenoid, triterpenoid, steroid biosynthesis, carotenoid biosynthesis and ubiquinone biosynthesis. RPKM (Reads Per Kilobase of transcript per Million mapped reads) values has helped to understand the differences in terpenoid biosynthetic genes expression in different tissues (Figure 2.10 - 2.12).

Isoprene units are synthesized through MVA and MEP pathways in cytosol and plastids, respectively. In MVA pathway, the rate-limiting enzymes HMG synthase and reductase were highly expressed in kernel and pericarp where limonoids are abundant. In MEP pathway, the rate-limiting enzyme 1-deoxy-D-xylulose-5-phosphate synthase was highly expressed in flowers. Cytokine synthase, and cytokinin dehydrogenase of hemiterpenoids/zeatin biosynthetic pathway genes are expressed highly in leaves and kernel followed by the flower. Only a few genes were identified related to monoterpenoid biosyntheses such as linalool synthase, terpineol synthase and neomenthol dehydrogenase and most of these genes were highly expressed in flowers and leaves. Germacrene D synthase, α-farnesene synthase and premnaspirodiene oxygenase were identified related to sesquiterpenoid biosynthesis. Germacrene D synthase was highly abundant in leaves. Gibberellin biosynthetic genes were identified related to diterpenoids. The Cytochrome P450 genes related to gibberellins biosynthesis had low abundance in kernel as compared to other tissues. All the genes related to steroid biosynthesis were observed and highly expressed in leaves, kernel and pericarp and same pattern of expression was observed in case of brassinosteroid biosynthesis. Most of the genes related to carotenoid biosynthesis were highly expressed in leaves (involves in photosynthesis) followed by flowers. Ubiquinone biosynthetic genes are almost expressed equally in all the tissues except in kernel, where higher expression was observed.

In triterpenoid biosynthesis, Farnesyl diphosphate (FDS) showed an almost same pattern of expression in all the tissues. Squalene synthase (SQS) showed the highest expression in kernel. Total three genes related to squalene epoxidase (SQE) were observed, out of which Master_Control_800013/ Neem_transcript_18980 was highly expressed in kernel and pericarp. Master_Control_31859/ Neem_transcript_11067 was highly expressed in flower and leaves. Total six triterpene synthases were observed in neem, out of these Master_Control_74065/ Neem_transcript_27436 showed homology with cycloartenol synthase whereas, others showed homology with β-amyrin synthase. Master_Control_24780/ Neem_transcript_28920 showed highest expression as compared to others, mainly in kernel and pericarp. Master_Control_74892 and Master_Control_101750 were highly expressed in flowers as compared to other tissues. Master_Control_70149 and Master_Control_70584 were showing higher expression in leaves as compared to other tissues. Cycloartenol synthase, which is involved in steroid biosynthesis, showed almost similar levels of expression in all the tissues. From these expression analyses, we were able to deduce that Master_Control_24780 might be involved in triterpenoid biosynthesis (Figure 2.11).
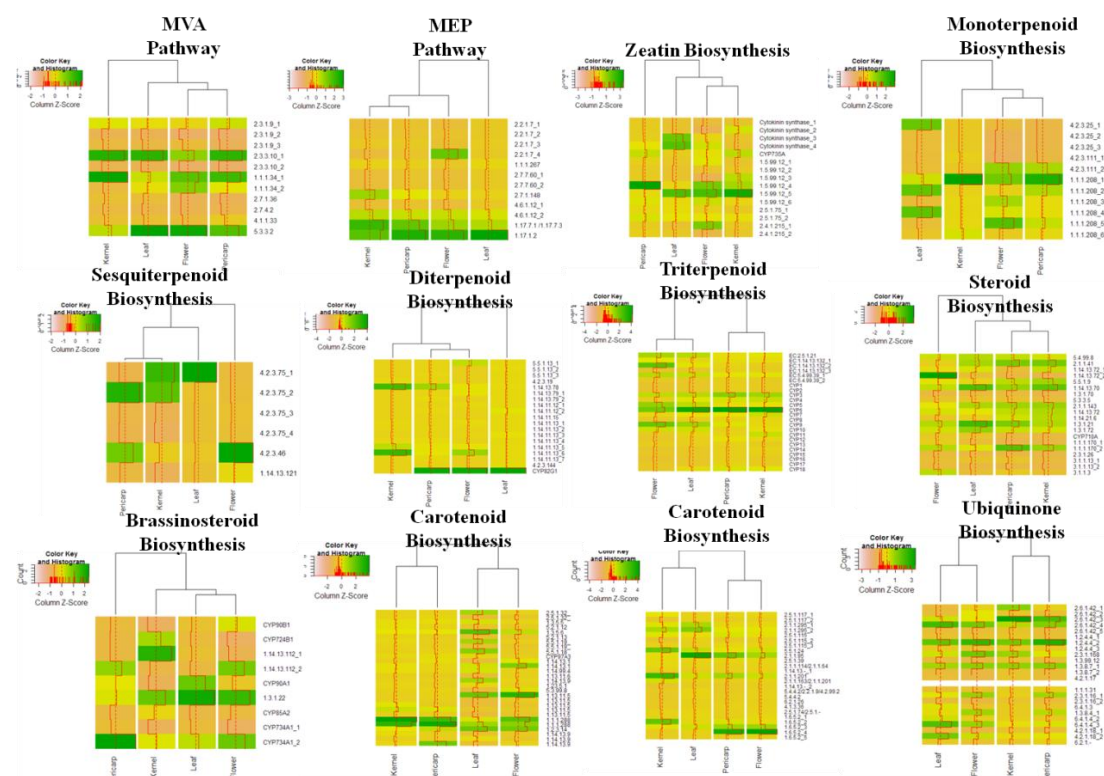
Figure 2. 10 RPKM Levels of Neem Terpenoid Metabolic Pathways.

Comparison of terpenoid metabolism in different tissues revealed that triterpenoid biosynthetic genes were highly expressed kernel and pericarp which is in line with metabolic profiling (Figure 2.11). MVA pathway genes are also highly expressed in pericarp and kernel. MEP pathway genes were expressed highly in flower and leaves. In leaves, the MEP pathway and diterpenoid biosynthetic genes showed high expression. In flowers MVA, MEP pathways, steroid, sesquiterpenoid and carotenoid biosynthetic genes were highly expressed as compared to other terpenoid biosynthetic pathways.
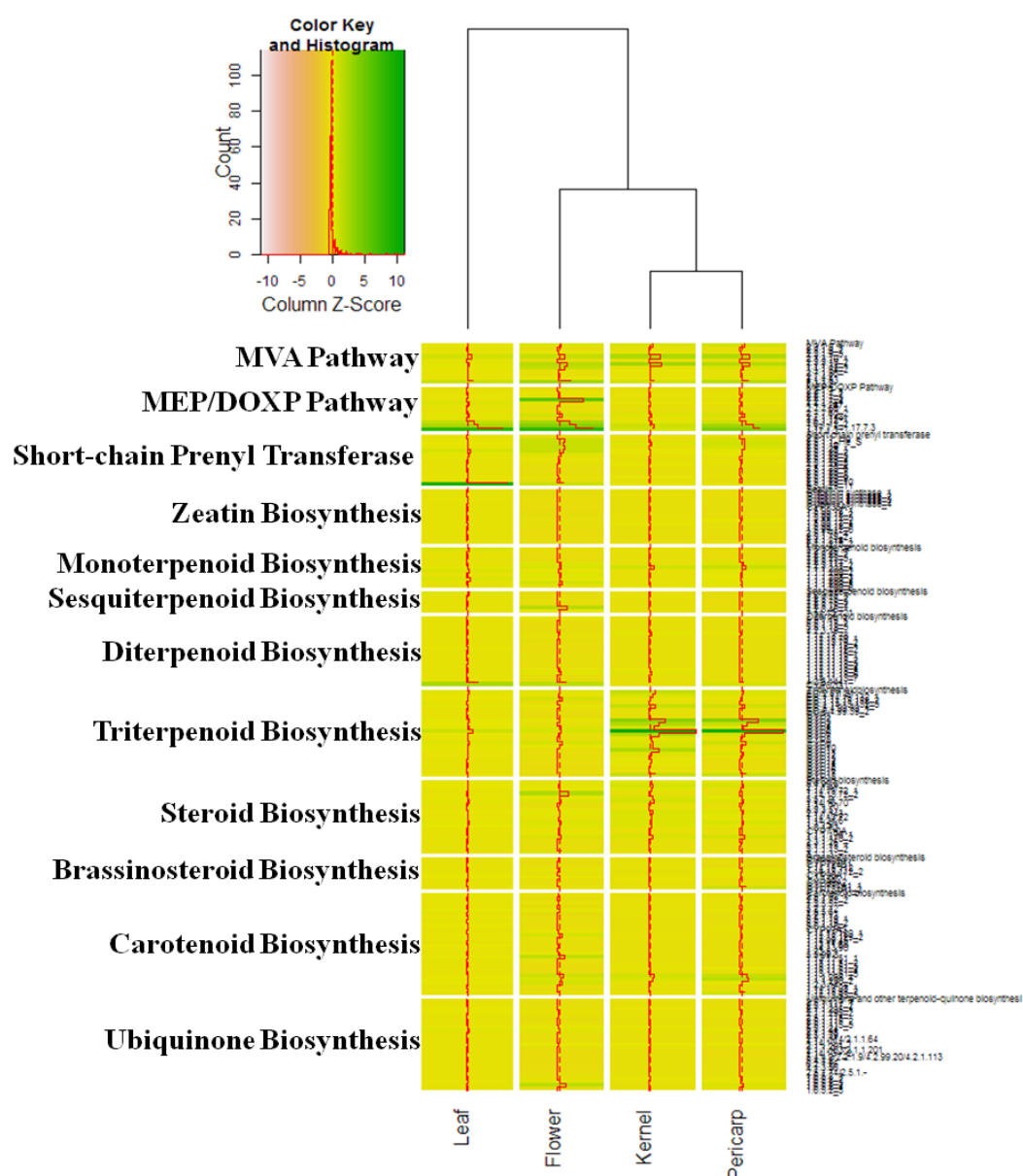
Figure 2. 11 Comparison of RPKM Levels of Neem Terpenoid Metabolic Pathways.

2.3.5 Differential Gene Expression Analysis between Flower and Kernel

RNA-Seq can be used for both discovery and quantification of transcripts in a single experiment. The expression level of each transcript is measured by the number of reads that map to it, which is expected to correlate directly with its abundance level. One of the applications of expression level of transcriptome is to identify differentially expressed genes in two or more conditions[41]. The selection of genes is based on combination of expression change and score cutoff then validated by *P* values generated by statistical modelling[42]. Since neem limonoids production is low in flower and highest in kernel, the transcript information of flower and kernel was used to analyze differentially expressed genes between these tissues. DESeq2 was used to predict the genes involved in limonoids biosynthesis[39].
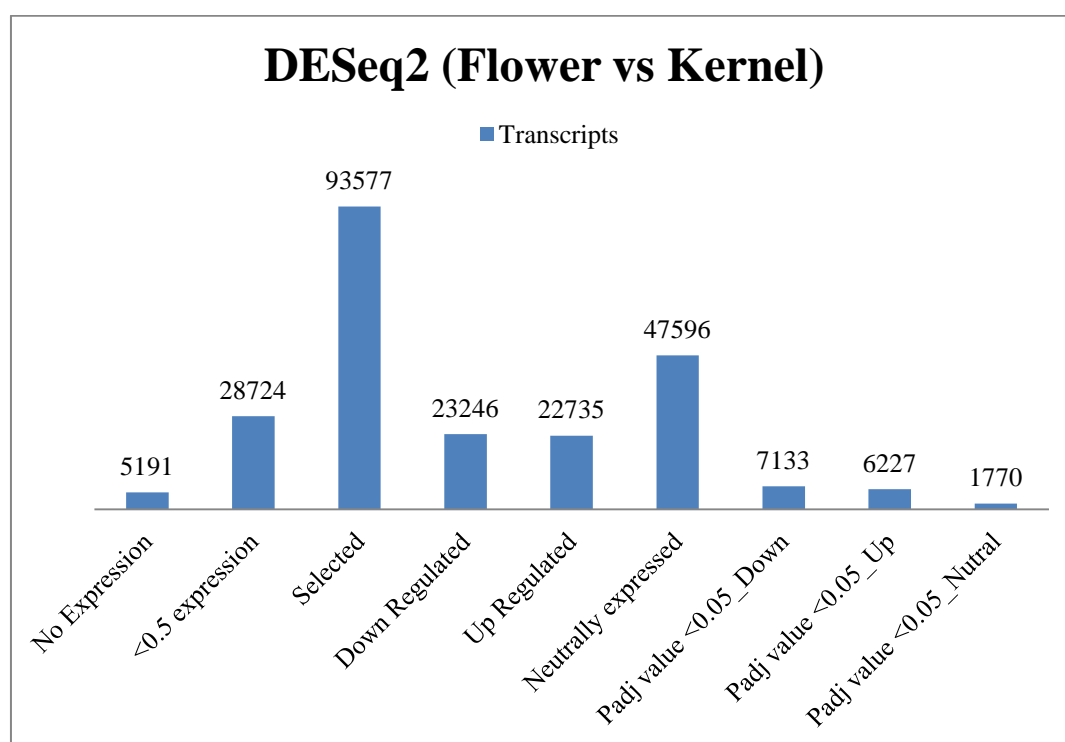
Figure 2. 12 Statistics of Differential Gene Expression Analysis Between Flower and Kernel.

Transcripts which have RPKM value more than 0.5 in both tissues were selected for differential gene expression analysis. Out of 93577 transcripts, 6227 transcripts were highly expressed in kernel as compared to flowers, 7133 transcripts were down-regulated and 1771 transcripts were neutral in expression with validation of adjusted *P* value <0.05 (Figure 2.12). In MVA pathway, HMG-CoA synthase,

HMG-CoA reductase and mevalonate kinase are up-regulated in kernel as compared to flowers, which is in line with triterpenoid content obtained through metabolic fingerprinting data. In triterpenoid biosynthesis, FDS, SQS, SQE and one of triterpene synthases were up-regulated in kernel which is in correlation with triterpenoid content obtained through metabolic fingerprinting (Figure 2.13). Differential gene expression analyzes helped in prediction of total 16 cytochrome P450 genes involved in limonoid biosynthesis (Table 2.4).
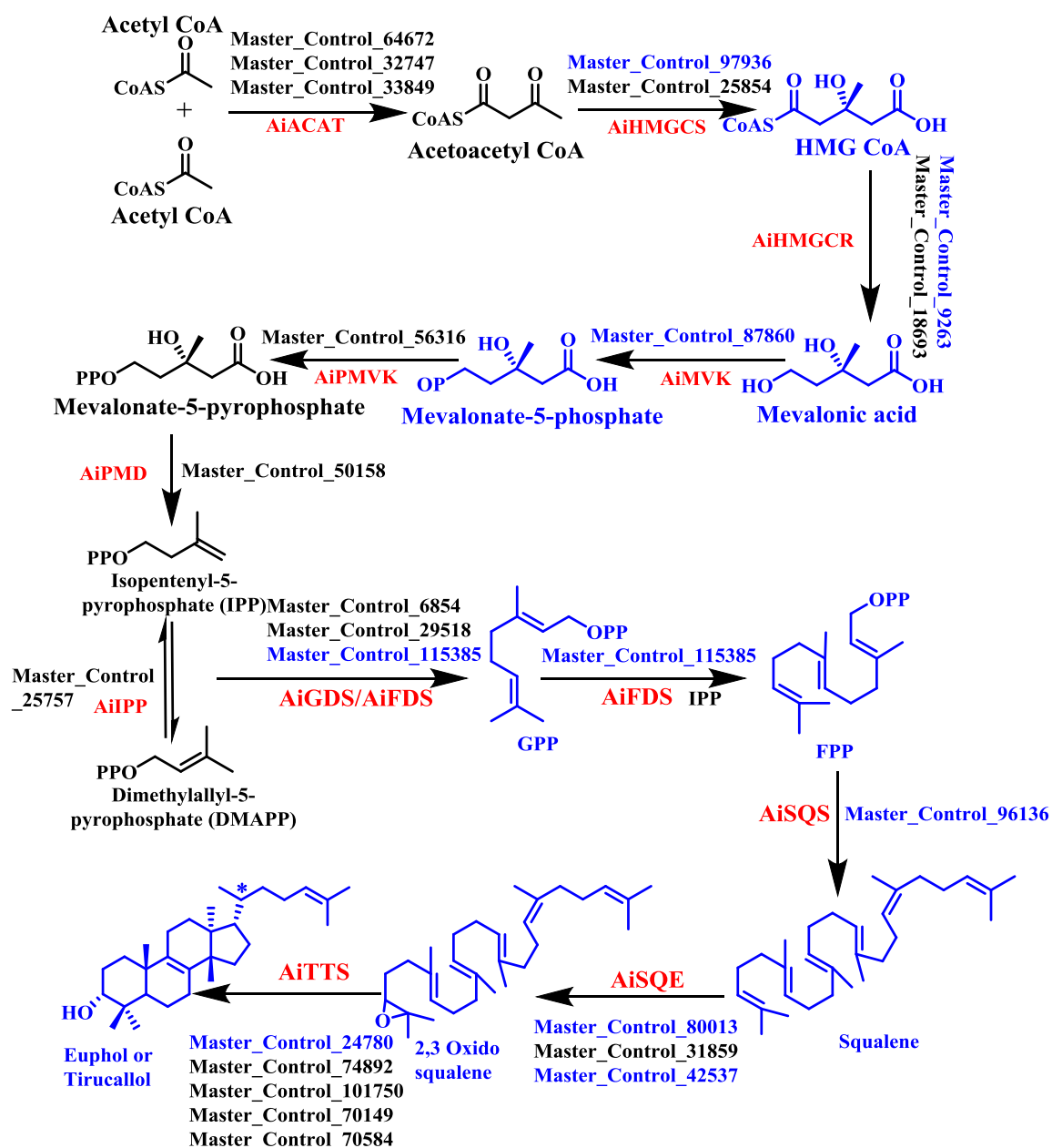
Figure 2. 13 Terpenoid Biosynthesis, Right from MVA Pathway to Triterpene Synthesis, Showing Transcripts Which are Up-regulated in Kernel as Compared to Flowers.

Table 2. 4 Predicted Cytochrome P450 Involved in Triterpene Biosynthesis.

Cytochrome P450	BLAST Results
Master_Control_119707	72 % Similarity with β-amyrin 24-hydroxylase *Glycine max* [Q9XHC6]
Master_Control_33323	89 % Similarity with β-amyrin 28-oxidase-like *Herrania umbratica* [XP_021297091]
Master_Control_120547	76 % Similarity with Cytochrome P450 CYP72A219 *Panax ginseng* [H2DH21]
Master_Control_57632 (AiCYP2)	71 % Similarity with Beta-amyrin 11-oxidase *Glycyrrhiza uralensis* [B5BSX1]
Master_Control_82633	90 % Similarity with Sterol 14-demethylase *Arabidopsis thaliana* [Q9SAA9]
Master_Control_84673 (AiCYP1)	63 % Similarity with Beta-amyrin 28-oxidase *Panax ginseng* [I7C6E8]
Master_Control_25760	69 % Similarity with Premnaspirodiene oxygenase *Hyoscyamus muticus* [A6YIH8]
Master_Control_8521	67 % Similarity with Premnaspirodiene oxygenase *Hyoscyamus muticus* [A6YIH8]
Master_Control_21505	87 % Similarity withFull=Cytochrome P450 86A22; AltName: Full=Long-chain acyl-CoA omega-monooxygenase *Petunia x hybrida* [B3RFJ6]
Master_Control_47318	64% Similarity with Full=Cytochrome P450 94A1; AltName: Full=P450-dependent fatty acid omega-hydroxylase *Vicia sativa* [O81117]
Master_Control_66055	69 % Similarity with Premnaspirodiene oxygenase *Hyoscyamus muticus* [A6YIH8]
Master_Control_106114	60 % Similarity with Isoflavone 2'-hydroxylase *Glycyrrhiza echinata* [P93147]
Master_Control_62316	89 % Similarity withEnt-kaurenoic acid oxidase 1 *Arabidopsis thaliana* [Ent-kaurenoic acid oxidase 1]
Master_Control_53970	90 % Similarity with Cytochrome P450 86A8

	Arabidopsis thaliana [O80823]
Master_Control_38507	82 % Similarity withCytochrome P450 78A7 *Arabidopsis thaliana* [Q9FIB0]
Master_Control_61421	72 % Similarity with Premnaspirodiene oxygenase *Hyoscyamus muticus* [A6YIH8]

2.4 Conclusion

Sequencing and transcriptome analysis are the primary tools for the discovery of novel genes, especially in nonmodel plants for which full genome sequencing is not economically feasible. Transcriptome sequencing represents relatively economical method for analyzing the transcribed region of the genome.

Triterpenoid profiling in neem indicated that there is tissue-specific variation in their abundance. The mature seed kernel and initial stages of pericarp were found to contain the highest amount of limonoids. Furthermore, a wide diversity of triterpenoids, especially C-seco triterpenoids were observed in kernel as compared to the other tissues. Pericarp, flower and leaf contained mainly ring-intact triterpenoids. Neem tissues such as seeds (kernel and pericarp), flower and leaves where the triterpenoids were deemed to be abundant were further selected for transcriptome analysis. A total of 79,079,412 (79.08 million) paired-end reads, each of 72 bp length was generated from pooled RNA and assembled to 41,140 transcripts. Tissue-specific (pericarp, kernel, flower and leaves) transcriptome helped in analyzing differential gene expression related to triterpenoid biosynthesis. A total of 209,882,798 (209.88 million) paired-end reads were generated and assembled to generate transcripts of 127,518.

Based on transcriptome annotation, all the genes involved in triterpenoid backbone biosynthesis, right from isoprene units (MVA pathway and MEP pathway) to triterpene cyclases were found. A total of 134 transcripts predicted as cytochrome P450 monooxygenases and two transcripts as cytochrome P450 reductases were identified. In triterpenoid biosynthesis, FDS showed almost same pattern of expression in all the tissues. Squalene synthase showed highest expression in kernel. Total three genes related to squalene epoxidase were observed, out of which

Master_Control_800013 was highly expressed in kernel and pericarp. Total six triterpene synthases were observed in neem, out of which Master_Control_24780 showed highest expression among all tissues, mainly in kernel and pericarp and might be involved in triterpenoid biosynthesis. Fifteen cytochrome P450 genes involved in limonoids biosynthesis were predicted based on differential gene expression analysis.

2.5 References

1. Garcia-Sancho, M. Sequencing as a way of work: a history of its emergence and mechanisation: from proteins to DNA, 1945-200.*University of London* (2007).
2. Watson, J.D. & Crick, F.H.C. Molecular structure of nucleic acids. *Nature* **171**, 737-738 (1953).
3. Franklin, R.E. & Gosling, R.G. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172**, 156-157 (1953).
4. Pauling, L. & Corey, R.B. A proposed structure for the nucleic acids. *Proceedings of the National Academy of Sciences* **39**, 84-97 (1953).
5. Wilkins, M.H.F. & Randall, J.T. Crystallinity in sperm heads: molecular structure of nucleoprotein *in vivo*. *Biochimica et Biophysica Acta* **10**, 192-193 (1953).
6. Maxam, A.M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**, 560-564 (1977).
7. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463-5467 (1977).
8. Hunkapiller, T., Kaiser, R.J., Koop, B.F. & Hood, L. Large-scale and automated DNA sequence determination. *Science* **254**, 59-67 (1991).
9. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
10. Venter, J.C. *et al*. The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
11. Ansorge, W.J. Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203 (2009).
12. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-1145 (2008).
13. Heather, J.M. & Chain, B. The sequence of sequencers: the history of sequencing DNA. *Genomics* **107**, 1-8 (2016).
14. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nature Methods* **5**, 16-18 (2008).
15. Pareek, C.S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**, 413-435 (2011).

16. Kircher, M. & Kelso, J. High-throughput DNA sequencing-concepts and limitations. *Bioessays* **32**, 524-536 (2010).

17. Buermans, H.P.J. & Den Dunnen, J.T. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular basis of Disease* **1842**, 1932-1941 (2014).

18. Garber, M., Grabherr, M.G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods* **8**, 469-477 (2011).

19. Horner, D.S. *et al*. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in Bioinformatics* **11**, 181-197 (2009).

20. Alkan, C., Sajjadian, S. & Eichler, E.E. Limitations of next-generation genome sequence assembly. *Nature Methods* **8**, 61-65 (2011).

21. Paszkiewicz, K. & Studholme, D.J. *De novo* assembly of short sequence reads. *Briefings in Bioinformatics* **11**, 457-472 (2010).

22. Compeau, P.E.C., Pevzner, P.A. & Tesler, G. How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology* **29**, 987-991 (2011).

23. Myers, E.W. Toward simplifying and accurately formulating fragment assembly. *Journal of Computational Biology* **2**, 275-290 (1995).

24. Fernie, A.R., Trethewey, R.N., Krotzky, A.J. & Willmitzer, L. Metabolite profiling: from diagnostics to systems biology. *Nature Reviews Molecular Cell Biology* **5**, 763-769 (2004).

25. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics* **2**, 155-168 (2001).

26. Fiehn, O. *et al*. Metabolite profiling for plant functional genomics. *Nature Biotechnology* **18**, 1157-1161 (2000).

27. Martzen, M.R. *et al*. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153-1155 (1999).

28. Rochfort, S. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *Journal of Natural Products* **68**, 1813-1820 (2005).

29. Allwood, J.W. & Goodacre, R. An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis* **21**, 33-47 (2010).

30. Kueger, S., Steinhauser, D., Willmitzer, L. & Giavalisco, P. High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. *The Plant Journal* **70**, 39-50 (2012).

31. Gika, H.G., Theodoridis, G.A., Plumb, R.S. & Wilson, I.D. Current practice of liquid chromatography-mass spectrometry in metabolomics and

metabonomics. *Journal of Pharmaceutical and Biomedical Analysis* **87**, 12-25 (2014).

32. Pandreka, A. *et al.* Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*). *BMC Plant Biology* **15**, 214 (2015).

33. Zerbino, D.R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829 (2008).

34. Schulz, M.H., Zerbino, D.R., Vingron, M. & Birney, E. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092 (2012).

35. Grabherr, M.G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).

36. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**, W182-W185 (2007).

37. Wernersson, R. Virtual Ribosome-a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Research* **34**, W385-W388 (2006).

38. Finn, R.D. *et al.* Pfam: the protein families database. *Nucleic Acids Research* **42**, D222-D230 (2013).

39. Love, M.I., Anders, S. & Huber, W. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

40. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

41. Vijay, N., Poelstra, J.W., KÃ¼nstner, A. & Wolf, J.B.W. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Molecular Ecology* **22**, 620-634 (2013).

42. Rapaport, F. *et al.* Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology* **14**, 3158 (2013).

Chapter 3
Cloning and Functional Characterization of Prenyltransferases
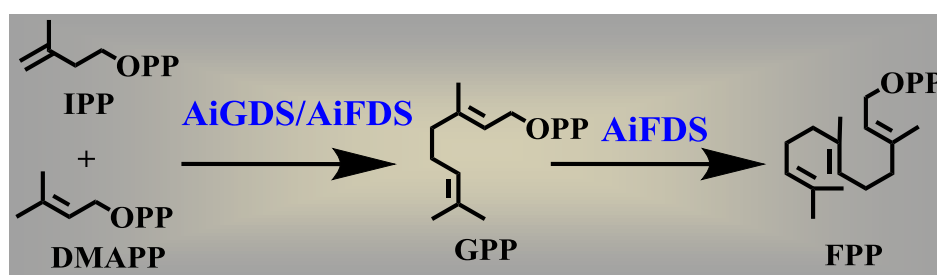
3.1 Introduction

Metabolites present in all the living systems and crucial for life are classified as primary metabolites. Whereas, secondary metabolites are synthesized by a single or a group of organism which helps them in enhancing their interaction with the environment towards survival. Terpenoids are one of the main classes of secondary metabolites. Till now more than 75,000 compounds have been identified (http://dnp.chemnetbase.com). Terpenoids are involved in membrane fluidity, respiration, photosynthesis, growth regulation, plant-pathogen interaction, protection of plants against herbivores and pathogens, the attraction of pollinators and seed dispersal animals[1]. Secondary metabolites from neem are widely used in medicine, agriculture and environment protection. Most of these belong to limonoids which are tetranortriterpenoids[2]. Neem (*Azadirachta indica*) contains more than 150 limonoids which are 4,4,8-trimethyl-17-furanyl steroid skeletons and its derivatives[3].

Terpenoid biosynthesis begins with isoprene units such as IPP and DMAPP. These are allylic diphosphates synthesized through the MVA and MEP pathway[4]. Allylic diphosphate, DMAPP condenses with one or more IPP molecules in head-to-tail fashion to produce diphosphates such as geranyl diphosphate, farnesyl diphosphate, geranylgeranyl diphosphate and long-chain prenyl diphosphates. Allylic diphosphates condensation reaction is catalyzed by prenyltransferases such as geranyl diphosphate synthase (GDS), farnesyl diphosphate synthase (FDS) and geranylgeranyl diphosphate synthase (GGDS) and long chain diphosphate synthases[5,6]. Among different prenyl diphosphates, FPP undergoes head to head 1-1′ condensation to form presqualene diphosphate which undergoes NADPH dependent reduction and rearrangement to form squalene[7]. In most of the organisms, squalene acts as a precursor for the synthesis of primary metabolites such as steroids which controls growth and division[8]. This molecule is the first committed precursor for the biosynthesis of secondary metabolites such as triterpenoids. This indicates that squalene acts as an important intermediate governing the balance between steroids and triterpenoid biosynthesis. Squalene undergoes oxidation to 2,3-oxidosqualene catalyzed by squalene epoxidase[9]. 2,3-oxidosqualene cyclizes to basic triterpene skeletons catalyzed by triterpene cyclases. Furthermore, addition of functional groups

and modification on the parental backbone of basic triterpene skeletons results in diverse triterpenoids[10].

Prenyltransferases, FDS is shown to play a key regulatory role in triterpenoid and phytosterol biosynthesis. To show some instances, FDS was up-regulated when hairy root culture of *Panax ginseng* was treated with methyl jasmonate (MJ) to enhance the production of triterpenoids[11]. In *Panax ginseng* hairy root culture, overexpression of FDS and mevalonate-5-diphosphate decarboxylase resulted in increased accumulation of phytosterols and triterepenes[12]. In *Centella asiatica*, overexpression of *Panax ginseng* FDS resulted in enhanced production of triterpenes[13]. Therefore, identification and functional characterization of prenyltransferases help in the understanding of triterpenoid biosynthesis.

3.2 Neem Prenyltransferases

Prenyltransferases catalyse head to tail condensation and chain elongation reaction of isoprene units such as DMAPP and IPP. These prenyltransferases act at the branching point of terpenoid metabolism and play a regulatory role in the distribution of isoprene units into various terpenoid pathways. A total of ten prenyltransferases was obtained from neem transcriptome (Chapter 2, Table 2.3). Based on functional annotation studies, two geranyl diphosphate synthase (GDS), one farnesyl diphosphate synthase (FDS) and seven putative geranylgeranyl diphosphate synthases (GGDS) were identified. BLAST and KAAS studies indicated that Neem_transcript_10912 (NCBI accession no - KM108315) was a homomeric GDS and Neem_transcript_10001 may be the smaller subunit of heteromeric GDS.

According to *in silico* (targetP analysis) cellular localization study Neem_transcript_10912 was highly localized to mitochondria, which is not involved in triterpenoid biosynthesis. Analysis of characterized homomeric GDS from other plants (*Solanum lycopersicum*[14], *Arabidopsis thaliana*[15], *Catharanthus roseus*[16], *Mangifera indica*[17] and *Picea abies* (PaGDS3)[18]) showed mitochondrial localization. Heteromeric GDS form other plants (smaller subunits of *C.roseus*[16], *Madia sativa*, *Salvia miltiorrhiza*[19] and larger subunits of *Humulus lupulus*[20], *Antirrhinum majus*[21], *Mentha piperita* and *C. roseus*[16]) showed plastidial localization. Neem_transcript_10001 (smaller subunit of heteromeric GDS) was localized to mitochondria but RC

score (5) was very poor (Table 3.1). Lager subunit of heteromeric GDS might be one of seven predicted GGDS but only one showed plastidial localization (Table 3.1). Subcellular localization of heteromeric GDPS will explain whether it is involved in triterpenoid biosynthesis. In this study, Neem_transcript_10912 (AiGDS) and Neem_transcript_25722 (NCBI accession no - KM108316, AiFDS) were selected for cloning and functional characterization.

Table 3. 1 TargetP Analysis Prenyl Diphosphate Synthases.

Name	Len	cTP	mTP	SP	otherLoc	RC	
Neem_transcript_10912	420	0.068	0.882	0.007	0.064	M	1
Neem_transcript_10001	306	1.80	0.290	0.106	0.204	M	5
Neem_transcript_25722	342	0.056	0.093	0.106	0.891	-	2
Neem_transcript_1166	351	0.035	0.521	0.029	0.605	-	5
Neem_transcript_3894	333	0.238	0.194	0.260	0.155	S	5
Neem_transcript_16200	232	0.052	0.021	0.343	0.861	-	3
Neem_transcript_16736	286	0.124	0.071	0.094	0.863	-	2
Neem_transcript_18547	134	0.111	0.116	0.122	0.872	-	2
Neem_transcript_28215	355	0.301	0.352	0.022	0.116	M	5
Neem_transcript_30369	366	0.798	0.027	0.123	0.214	C	3
cutoff		0.000	0.000	0.000	0.000		

3.3 Materials and Methods

3.3.1 Materials Used in this Study

3.3.1.1 Bacterial Strains and Plasmids Used in the Study

Escherichia coli TOP10 (Invitrogen/ Life Technologies, USA) cloning cells were used for transformation of plasmids or ligation mixtures. pET32a expression vector (Novagen, Addgene, USA) was used for cloning of AiFDS and AiGDS. Expression of prenyl diphosphate synthases was done in *E. coli* BL21DE3 (Novagen, Addgene, USA) and *E. coli* Lemo21DE3 (New England Biolabs, USA).

3.3.1.2 Kits and Reagents Used in the Study

SuperScript® III Reverse Transcriptase (ThermoFisher Scientific, USA) was used for cDNA synthesis. JumpStart™ Taq DNA Polymerase (Sigma-Aldrich, USA) and AccuPrime™ *Pfx* DNA Polymerase (ThermoFisher Scientific, USA) were used for amplification of prenyl diphosphate synthases. PCR products were gel eluted by using GenElute™ PCR Clean-Up Kit (Sigma-Aldrich, USA). Plasmids were isolated by using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, USA). GelRed™ (Biotium Inc., USA) was used for nucleic acid staining. Restriction enzymes (New England Biolabs, USA) and T_4 DNA ligase (Invitrogen/ Life Technologies, USA) used for cloning. SuperScript® III Platinum® SYBR® Green One-Step qRT-PCR Kit (ThermoFisher Scientific, USA) was used for real-time PCR.

3.3.1.3 Primers

3.3.1.3.1 Primers for AiGDS

Neem_Transcript_10912 was used as a template to design primers for real-time PCR and cloning primers with restriction sites *BamHI* and *SacI*. Primers were analyzed by using software Oligoanalyzer.

Table 3. 2 Primers Used for AiGDS.

Name	Primer Sequence
AiGDS_FP	ATGACC**GGATCC**ATGTTATTTTCTCGTG
AiGDS_RP	CATGTC**GAGCTC**CTATTTATTTCTTGTGATG
AiGDS_RT_FP	AGTTCCCTGAGTTGCGTAAAG
AiGDS_RT_RP	TCATCGTTGCTTTCTGGTAGAG

3.3.1.3.2 Primers for AiFDS

Neem_Transcript_25722 was used as a template to design primers for real-time PCR and cloning primers with restriction sites *BamHI* and *XhoI*. Primers were analyzed by using software Oligoanalyzer.

Table 3. 3 Primers Used for AiFDS.

Name	Primer Sequence
AiFDS_FP	ATGAGC**GGATCC**ATGAGTGATCTGCATTCC

AiFDS_RP	ACAGAT**CTCGAG**TTACTTCTGCCTCTTG
AiFDS_RT_FP	GGTGCATCGAATGGCTTCAA
AiFDS_RT_RP	GTGCACATGGTTGCGTAGAA

3.3.1.3.3 Primers for GAPDH

Glyceraldehyde 3-phosphate dehydrogenase was used as intrinsic gene for real-time PCR analysis. Primers were analyzed by using software Oligoanalyzer.

Table 3. 4 Primers Used for GAPDH.

Name	Primer Sequence
AiFDS_RT_FP	TCGGAATCAACGGTTTTGGAA
AiFDS_RT_RP	CACTTGACCGTGAACACTGT

3.3.1.4 Buffer Compositions

3.3.1.4.1 Buffers Used for Characterization of AiGDS
AiGDS_Lysis buffer

100 mM MOPSO buffer containing NaCl (400 mM) with detergent (0.5 % CHAPS), protease inhibitor (0.5 mM PMSF) and 1 mg/mL lysozyme in 10 % glycerol and the pH was adjusted to 7.4 using 0.1 M NaOH.

AiGDS_Wash buffer

100 mM MOPSO buffer containing NaCl (400 mM) and imidazole (100 mM) in 10 % glycerol and the pH was adjusted to 7.4 using 0.1 M NaOH.

AiGDS_Elution buffer

50 mM MOPSO buffer containing NaCl (300 mM), imidazole (250 mM), and 0.2 % CHAPS in 10 % glycerol and the pH was adjusted to 7.4 using 0.1 M NaOH.

AiGDS_Desalting buffer

50 mM MOPS buffer containing KCl (100 mM) in 10 % glycerol and the pH was adjusted to 7.4 using 0.1 M NaOH.

AiGDS_Enzyme assay buffer

25 mM HEPES buffer containing KCl (100 mM), a reducing agent (2 mM DTT) and 10 mM $MgCl_2$ in 10% glycerol and the pH was adjusted to pH 7.4 using 0.1 M NaOH.

3.3.1.4.2 Buffers Used for Characterization of AiFDS

AiFDS_Lysis buffer

50 mM phosphate buffer containing NaCl (300 mM), 10 mM $MgCl_2$ with detergent (0.2 % CHAPS), protease inhibitor (0.5 mM PMSF) and 1 mg/mL lysozyme and the pH was adjusted to 7.4 with 0.1 M NaOH.

AiFDS_Wash buffer

50 mM phosphate buffer containing NaCl (300 mM) and imidazole (50 mM) in 10 % glycerol and the pH was adjusted to 7. 4 with 0.1 M NaOH.

AiFDS_Elution buffer

50 mM phosphate buffer containing NaCl (300 mM) and imidazole (250 mM) in 10 % glycerol and the pH was adjusted to 7. 4 with 0.1 M NaOH.

AiFDS_Desalting buffer

25 mM HEPES buffer containing KCl (100 mM) in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.

AiFDS_Enzyme assay buffer

25 mM HEPES buffer containing KCl (100 mM), $MgCl_2$ (10 mM) and reducing agent (2 mM DTT) in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.

3.3.1.5 GC and GC-MS analysis

1 µL of AiGDS and AiFDS enzyme assay extracts were injected in GC/GC-MS equipped with 30 m × 0.25 mm × 0.25 µm capillary columns (HP-5 and HP-5 MS, J & W Scientific). The column was equilibrated at 70 °C followed by a temperature gradient from 5 °C/min rise till 150 °C, 10 °C/min rise till 270 °C and final hold for 5 min. Helium was used as the carrier gas with flow rate of 1 mL/min.

Product formation were characterized by retenction time and by co-injection with standard samples.

3.3.2 Cloning and Characterization of AiGDS and AiFDS

3.3.2.1 Cloning and Characterization of AiGDS

The Neem seed RNA was used for the synthesis of cDNA using SuperScript$^\circledR$ III First-Strand Synthesis System (Invitrogen). Full-length primers (Table 3.2) for predicted AiGDS ORF were designed using their transcripts (Neem_Transcript_10912) as a template. cDNA was used as a template for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen) to amplify AiGDS. The program for PCR was the following, 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 56 °C for 30 sec, 68 °C for 1 min followed by final extension at 68 °C for 5 min. Full-length AiGDS PCR product was cloned into *BamHI* and *SacI* sites of pET32a expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter and T7 reverse primers) to identify the positive colonies. AiGDS cloning was confirmed by analyzing the Sanger sequencing of plasmids obtained from positive clones using T7 promoter forward and T7 reverse primers.

The expression of the recombinant plasmids containing AiGDS was carried out in BL21 (DE3) cells under 1 mM IPTG overnight induction at 16 °C. Bacterial cells were collected by centrifugation at 5000 × *g* for 10 min. The cell pellet was resuspended in different lysis buffer (Table 3.5) at a ratio of 5 ml per gram of cell pellet. Sonication was done for 10 cycles (30 sec pulse ON/ 30 sec pulse OFF, Amplitude- 75%) then centrifuged at 10,000 × *g* for 10 min at 4 °C. The solubility of recombinant AiGDS was analyzed by running on 12 % SDS-PAGE gel.

Table 3. 5 Buffers Used for AiGDS Solubility.

Name	Buffer Composition
Buffer 1	50 mM phosphate buffer containing NaCl (300 mM) and detergent 0.2% CHAPS in 10% glycerol and the pH was adjusted to 7. 4 with 0.1 M NaOH.
Buffer 2	50 mM phosphate buffer containing NaCl (300 mM) and detergent 0.2% Triton X 100 in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH
Buffer 3	50 mM Tris buffer containing NaCl (300 mM) and detergent 0.2% CHAPS in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M HCl.
Buffer 4	50 mM Tris buffer containing NaCl (300 mM) and detergent 0.2% Triton X 100 in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M HCl.
Buffer 5	50 mM MOPSO buffer containing NaCl (300 mM) and detergent 0.2% CHAPS in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.
Buffer 6	50 mM MOPSO buffer containing NaCl (300 mM) and detergent 0.2% Triton X 100 in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.
Buffer 7	50 mM HEPES buffer containing NaCl (300 mM) and detergent 0.2% CHAPS in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.
Buffer 8	50 mM HEPES buffer containing NaCl (300 mM) and detergent 0.2% Triton X 100 in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.
Buffer 9	100 mM MOPSO buffer containing NaCl (400 mM) and detergent 0.5% CHAPS in 10% glycerol and the pH was adjusted to 7.4 with 0.1 M NaOH.

In most of the buffers, recombinant AiGDS was insoluble. Expression of AiGDS was changed to Lemo 21 (DE3) cells which did not show any improvement in solubility. To obtain the soluble AiGDS protein, the bacterial cell pellet was resuspended in lysis buffer which contains higher concentration of buffering agent and salt (AiGDS_lysis Buffer) and proceed for sonication as mentioned above. The pellet obtained after centrifugation at $10,000 \times g$ was resuspended in lysis buffer. The pH of resuspended pellet was increased to 11.0 with 0.1 M NaOH and then reduced to 7.0 with 0.1 M HCl (pH adjustment was done on ice with continuous stirring). The resulting solution was centrifuged at $10,000 \times g$ for 10 min at 4 °C. The supernatant containing AiGDS protein was purified over Ni-NTA (1 mL resin / g cell pellet) affinity chromatography. The column was washed with wash buffer (AiGDS_Wash Buffer) till the OD reaches to zero at 280 nm. The protein was eluted out with elution buffer (AiGDS_Elution Buffer). Recombinant AiGDS was desalted on Hi-Prep^{TM} 26/10 Desalting Columns with desalting buffer (AiGDS_Desalting Buffer) using AKTA Avant (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and checked on 12 % SDS gel.

Enzyme assays for AiGDS were performed in assay buffer (AiGDS_Assay Buffer) with DMAPP (100 µM) and IPP (100 µM) as substrates. The reaction mixtures were incubated at 30 °C for 2 h. Further alkaline phosphatase (6 U) was added and incubated at 37 °C for 1 h. Reaction mixtures were extracted thrice using n-hexane. Samples were concentrated in a stream of dry nitrogen and analyzed by GC-MS (GC_MS Program-3.3.1.5) by injecting 1µL of concentrated assay extract. Geraniol formation was confirmed by co-injection with authentic standards and comparing the mass fragmentation pattern and retention time.

3.3.2.2 Cloning and Characterization of AiFDS

Full-length primers (Table 3.3) for AiFDS ORF (Chapter 2, Table 2.1) were designed using their transcript (Neem_Transcripts_25722) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 56 °C for 30 sec, 68 °C for 1.2 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *BamHI* and *XhoI* cloning sites of pET32a expression vector using T$_4$

DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter and T7 reverse primers) to identify the positive colonies. Cloning was confirmed by analyzing the Sanger sequencing of plasmids obtained from positive clones using T7 promoter forward and T7 reverse primers.

The expression of the recombinant plasmids containing AiFDS was carried out in BL21 (DE3) cells under 1 mM IPTG at 16 °C for overnight. Bacterial cells were collected by centrifugation at $5000 \times g$ for 10 min. Cell pellet was resuspended in lysis buffer (AiFDS_Lysis Buffer) at a ratio of 5 ml per gram of cell pellet. Sonication was done for 10 cycles (30 sec pulse on/ 30 sec pulse off, Amplitude- 75 %) then centrifuged at $10,000 \times g$ for 10 min. The supernatant containing AiFDS protein was purified over Ni-NTA (1 mL resin / g cell pellet) affinity chromatography. The column was washed with wash buffer (AiFDS_Wash Buffer) till the O.D. reaches to zero at 280 nm. The protein was eluted out with elution buffer (AiFDS_Elution Buffer). Protein was desalted on Hi-Prep^{TM} 26/10 Desalting Columns with desalting buffer (AiFDS_Elution Buffer) using AKTA Avant (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and checked on 12 % SDS gel.

Enzyme assays for AiFDS were performed in assay buffer (AiFDS_Assay Buffer) with DMAPP (100 µM)/GPP (100 µM) and IPP (100-200 µM) as substrates. The reaction mixtures were incubated at 30 °C for 2 h. Further alkaline phosphatase (6 U) was added and incubated at 37 °C for 1 h. Reaction mixtures were extracted thrice using n-hexane. Samples were concentrated in a stream of dry nitrogen and analyzed by GC-MS (GC_MS Program-3.3.1.5) by injecting 1µL of concentrated assay extract. Farnesol formation was confirmed by co-injection with authentic standards and comparing the mass fragmentation pattern and retention time.

3.3.3 RT-PCR Analysis of AiGDS and AiFDS

Real-time PCR was carried out using SuperScript^{TM} III platinum^{TM} SYBR^{TM} Green One-step qRT-PCR Kit (Invitrogen, USA). In brief, for AiGDS and AiFDS real-time analysis, 100 ng of DNase treated total RNA was added with AiGDS (Table

3.2) and AiFDS (Table 3.3) primers and for intrinsic control, GAPDH primers (Table 3.4) were used. Quantification was performed as follows: Initial cDNA synthesis at 50 °C for 20 min, followed by 95 °C for 5 min, 40 cycles of 95 °C for 10 sec and 60 °C for 30 sec. GAPDH primers were used as an endogenous control to normalize the expression levels between different tissues. Threshold (Ct) values were obtained and ΔCt was calculated as Ct target gene – Ct endogenous reference gene. Relative fold difference was calculated using $2^{\Delta Ct}$. Experiments were carried out using three biological replicates with five technical replicates each.

3.4 Results and Discussion

3.4.1 Cloning and Characterization of AiGDS and AiFDS

3.4.1.1 Cloning and Characterization of AiGDS

The ORF of AiGDS [GenBank: KM108315] was 1,263 bp, which coded for a protein of 420 amino acids with theoretical molecular weight and calculated pI as 46.1 kDa and 6.33, respectively. AiGDS had a maximum identity with several plants characterized homomeric GDSs such as 90 % identity to homomeric GDS from *Citrus sinensis* [GenBank: CAC16851][15], 86 % identity to GDS from *Mangifera indica* [GenBank: AFJ52721][17] and 76 % identity to GDS from *Catharanthus roseus* [GenBank: AGL91647][16]. The percentage identity matrix of AiGDS with another plant homomeric GDS and heteromeric GDS larger subunits indicated that AiGDS possesses 71 % to 89 % identity with homomeric GDS (Table 3.6).

Table 3. 6 Identity Matrix of AiGDS with Homomeric GDS and Heteromeric GDS Larger Subunits.

		1	2	3	4	5	6	7	8	9	10
1	AiGDS	100									
2	CsGDS	89.72	100								
3	MiGDS	83.57	88.79	100							
4	CrGDS	76.74	83.18	75.78	100						
5	AtGDS	71.15	80.69	70.74	68.02	100					
6	TcGDS1	71.25	80.69	70.59	70.52	65.53	100				
7	CrGDS.LSU	26.22	27.86	25	26.59	23.7	26.04	100			
8	HlGD.SLU	27.81	29.39	26.25	27.76	26.57	28.27	72.48	100		
9	MpGDS.LSU	25.87	28.63	24.93	26.1	24.63	27.03	71.39	67.57	100	
10	AmGDS.LSU	24.41	27.86	25.22	25.37	24.48	25.23	75	69.97	70.65	100

Amino acid sequences of AiGDS(*A. indica*, KM108315), CsGDS (*C. sinensis*, CAC16851), MiGDS (*M. indica*, AFJ52721.), CrGDS (*C. roseus*, AGL91647), AtGDS (*Arabidopsis thaliana*, CAC16849), TcGDS1 (*Theobroma cacao*, EOY33650), CrGDS.LSU (*C. roseus*, AGL91645), HlGD.LSU (*Humulus lupulus*, ACQ90682), MpGDS.LSU (*Mentha x piperita*, AAF08793) and AmGDS.LSU (*Antirrhinum majus*, AAS82860) are used for identity matrix generation.

```
AtGDS      1  MLFTRSVARISSKFLRNRSFYGSSQSLASHRFA-IIPDQG-HSCSDS----PHKG--YVC
TcGDS1     1  ------------MARA--ALLHLL-RHRSVATATA-PLSAYKCLSSNSKTPSGIRWTSIC
MiGDS      1  MLFSYGLSRISINPRA--SLLTCR-WLLSHLTGSLSPSTSSHTISDS----VHKVW--GC
AiGDS      1  MLFSRGLSRISRIPRN--SLIGCR-WLVSYRPDTI-LSGSSHSVGDS----TQKVL--GC
CsGDS      1  ------------------------------------------------------------

AtGDS     53  RTTYS-LKSPVFGGFSHQLYHQS--SSLVEEELDPFSLVADELSLLSNKLREMVLAEVPK
TcGDS1    45  RAFSSKAAVNDLIGIDMANTDSGAVMEEKERLDPFSLVADELSLIANRLRSMVVTEVPN
MiGDS     52  REAY-TWSVPALHGFRHQIHHQS--SSLIEDQLDPFSLVADELSLVANRLRSMVVTEVPK
AiGDS     51  REAY-LWSLPALHGIRHQIHQQS--SSLIEEELDPFSLVADELSLVANRLRSMVVTEVPK
CsGDS      1  ------------------------------------------MVIAEVPK

AtGDS    110  LASAAEYFFKRGVQGKQFRSTILLLMATALDVRVPEALIGESTDIVTSELRVRQRGIAEI
TcGDS1   105  LASAAEYFFKIGAEGKRFRPTVLLLMATALSVRIPELPPAGVGDTLPTDLRTSQQRIAEI
MiGDS    109  LASAAEYFFKMGVEGKRFRPAVLLLMATALNVHVLEPLPEGAGDALMTELRTRQQCIAEI
AiGDS    108  LASAAEYFFKMGVEGKRFRPTVLLLMASALNVQVPQPLSDGVGDALTTELRTRQQCIAEI
CsGDS      9  LASAAEYFFKMGVEGKRFRPTVLLLMATALNVRVPEPLHDGVEDASATELRTRQQCIAEI

AtGDS    170  TEMIHVASLLHDDVLDDADTRRGVGSLNVVMGNKMSVLAGDFLLSRACGALAALKNTEVV
TcGDS1   165  TEMIHVASLLHDDVLDDADKRRGICSLNAVMGNKLAVLAGDFLLSRACVSLASLKNTEVV
MiGDS    169  TEMIHVASLLHDDVLDDADTRRGIGSLNLVMGNKLAVLAGDFLLSRACVALASLKNTEVV
AiGDS    168  TEMIHVASLLHDDVLDDADTRRGIGSLNFVMGNKLAVLAGDFLLSRACVALASLKNTEVV
CsGDS     69  TEMIHVASLLHDDVLDDADTRRGIGSLNFVMGNKLAVLAGDFLLSRACVALASLKNTEVV

AtGDS    230  ALLATAVEHLVTGETMEITSSTEQRYSMDYYMQKTYYKTASLISNSCKAVAVLTGQTAEV
TcGDS1   225  TLIATVVENLVTGETMQLTTASKQRFSMDYYMQKTYNKTASLISNSCKSIAILAGHTAEV
MiGDS    229  SLLATVVEHLVTGETMQMTTSSDQRCSMEYYMQKTYYKTASLISNSCKAIALLAGQSAEV
AiGDS    228  SLLATVVEHLVTGETMQMTTTAEQRRSMDYYMQKTYYKTASLISNSCKAIALLAGTTEV
CsGDS    129  TLLATVVEHLVTGETMQMTTSSDQRCSMDYYMQKTYYKTASLISNSCKAIALLAGQTAEV

AtGDS    290  AVLAFEYGRNLGLAFQLIDDILDFTGTSASLGKGSLSDIRHGVITAPILFAMEEFPQLRE
TcGDS1   285  AMLAFEYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFAMEEFPELHA
MiGDS    289  AMLAFEFGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILFAMEEFPQLRA
AiGDS    288  AMLAFDYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILFAMEEFPELRK
CsGDS    189  AILAFDYGKNLGLAYQLIDDVLDFTGTSASLGKGSLSDIRHGIITAPILFAMEEFPQLRT

AtGDS    350  VVDQVEKDPRNVDIALEYLGKSKGIQRARELAMEHANLAAAAIGSLPETDNEDVKRSRRA
TcGDS1   345  VVDQGFKNPANVDIALGFLGKSGIQRTKELAMKHANLAAQVIESLPESDDANVIRSRQA
MiGDS    349  VIDQGFENPSNVDVALEYLGKSRGIQRTRELATNHANLAAAAIDALPKTDNEEVRKSRRA
AiGDS    348  VVDKGFDDPSNVDIALEYLGKSRGIQRTRELAQKHANLATVALDSLPESNDDDVKKSRRA
CsGDS    249  VVEQGFEDSSNVDIALEYLGKSRGIQKTRELAVKHANLAAAAIDSLPENNDEDVTKSRRA

AtGDS    410  LIDLTHRVITRNK
TcGDS1   405  LIDLTQRVLTRNK
MiGDS    409  LLDLTQRVITRNK
AiGDS    408  LLDLAQRVITRNK
CsGDS    309  LLDLTHRVITRNK
```

Figure 3. 1 Multiple Sequence Alignment of *A. indica* Geranyl Diphosphate Synthases (AiGDS).

Amino acid sequences of TcGDS1 (*T. cacao*, XP_007016031), AtGDS (*A. thaliana*, CAC16849), MiGDS (*M. indica*, AFJ52721.), AiGDS (*A. indica*, KM108315) and CsGDS (*C. sinensisi*, CAC16851) are used for multiple sequence alignment. The highly conserved Asp-rich motifs of prenyltransferases are indicated by solid line.

The multiple sequence alignment of AiGDS consisted of two aspartate-rich motifs DDX$_{(2-4)}$D and DDXXD which are highly conserved motifs in prenyltransferases and involved in substrate and metal ion binding (Figure 3.1). CxxxC motifs were not observed in AiGDS, which play a key role in the interaction of heteromeric GDS[20]. The ORF of AiGDS was cloned into a pET32a expression vector having an N-terminal thioredoxin domain and subsequently expressed in BL21

(DE3) cells. However, recombinant AiGDS protein was found in inclusion bodies even after using different buffers (Table 3.1, Figure 3.3).
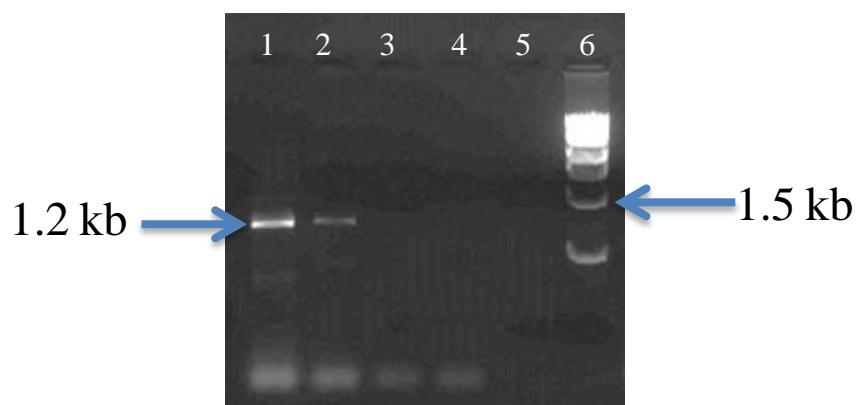
Figure 3. 2 AiGDS ORF Amplification.
Lane 1: AiGDS PCR product at 52 °C, **Lane 2:** AiGDS PCR product at 54 °C, **Lane 3:** AiGDS PCR product at 56 °C, **Lane 4:** AiGDS PCR product at 58 °C, **Lane 5:** Negative control and **Lane 6:** 1 kb DNA ladder Sigma (Addendum Figure A1.A).
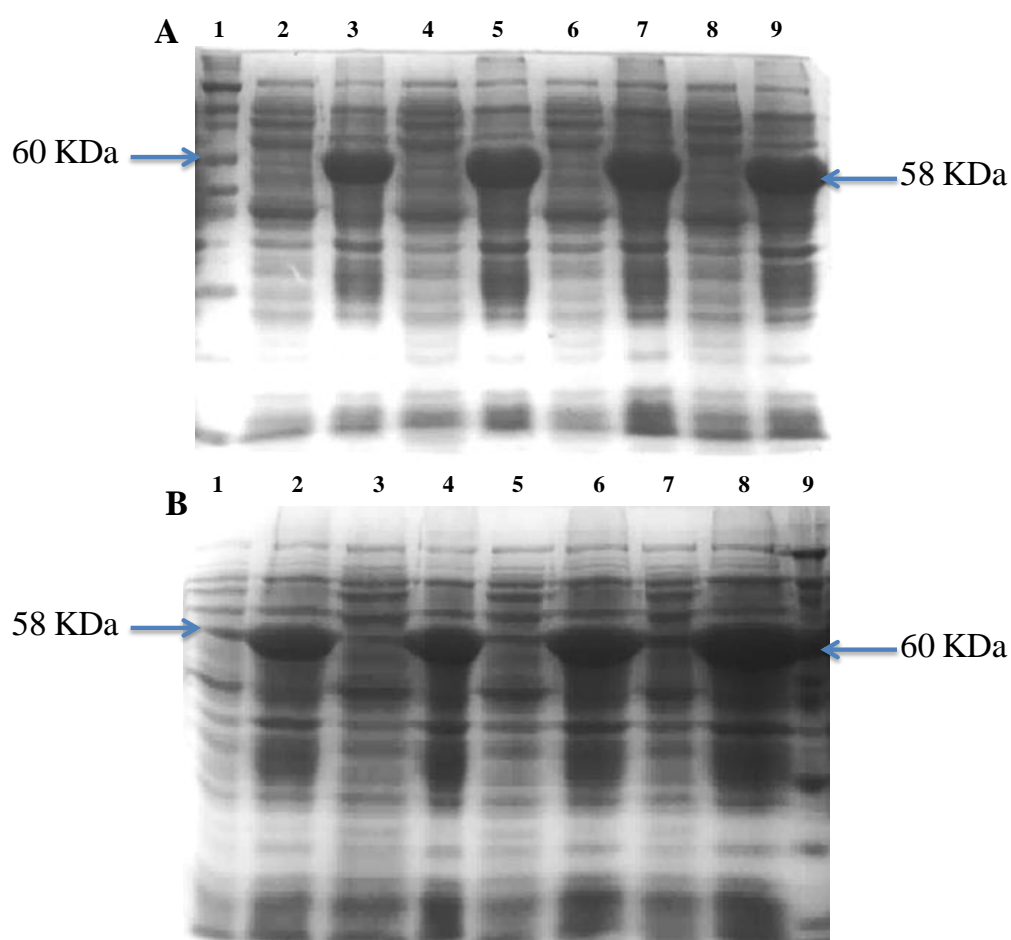
Figure 3. 3 AiGDS Solubility in Different Buffers.

A) Lane 1: Novex® sharp pre-stained protein standard (Addendum Figure A4.A), **Lane 2:** Buffer 1 supernatant fraction, **Lane 3:** Buffer 1 pellet fraction, **Lane 4:** Buffer 2 supernatant fraction, **Lane 5:** Buffer 2 pellet fraction, **Lane 6:** Buffer 3 supernatant fraction, **Lane 7:** Buffer 3 pellet fraction, **Lane 8:** Buffer 4 supernatant fraction, Lane 9: Buffer 4 pellet fraction. **B) Lane 1:** Buffer 5 supernatant fraction, **Lane 2:** Buffer 5 pellet fraction, **Lane 3:** Buffer 6 supernatant fraction, **Lane 4:** Buffer 6 pellet fraction, **Lane 5:** Buffer 7 supernatant fraction, **Lane 6:** Buffer 7 pellet fraction, **Lane 7:** Buffer 8 supernatant fraction, **Lane 8:** Buffer 8 pellet fraction, **Lane 9:** Novex® sharp pre-stained protein standard (Addendum Figure A2.A).

To enhance solubility, AiGDS cloned construct was transformed into Lemo 21 (DE3) cells[22] and expression was carried out. Recombinant AiGDS protein remained solely in the insoluble portion of the pellet. Eventually, we were able to obtain soluble active AiGDS by re-suspending the pellets in lysis buffer and adjusting the pH to 11.0 by drop-wise addition of 0.1 M NaOH with constant swirling on ice till the solution became clear. The pH was then reduced to 7.0 using 0.1 M HCl under similar conditions[23]. The resulting solution was centrifuged at 10,000 × g and subjected to 12 % SDS-PAGE analyses. After this processing, AiGDS was found to be in the soluble form in the supernatant, which was subjected to purification by Ni-NTA affinity chromatography. The recombinant protein was approximately 94 % pure as analyzed by SDS-PAGE (Figure 3.4).
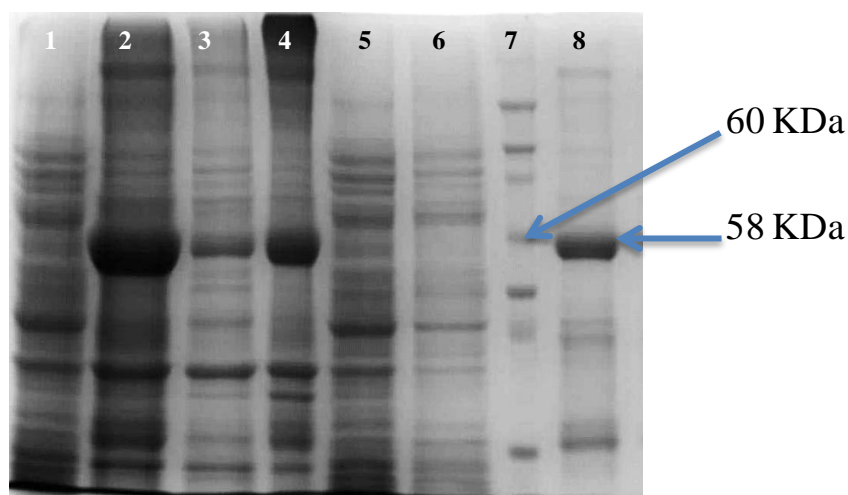
Figure 3. 4 SDS-PAGE for AiGDS Protein Purification in pET32a.
Lane 1: Supernatant fraction, **Lane 2:** Pellet fraction, **Lane 3:** Supernatant fraction after pH adjustment, **Lane 4:** Pellet fraction after pH adjustment, **Lane 5:** Unbound fraction, **Lane 6:** Wash fraction, **Lane 7:** Novex® Sharp Pre-stained Protein Standard (Addendum Figure A4.A) and **Lane 8:** Elution fraction.

Purified recombinant AiGDS was incubated with equimolar concentrations of IPP and DMAPP (Figure 3.5) followed by treatment with alkaline phosphatase to hydrolyze the diphosphate esters to their corresponding alcohols. The extracted assay mixture was analyzed by GC-MS and the products formed were confirmed by comparing the retention time and co-injection studies with standard geraniol (Figure 3.6). GC-MS analyses of the extracts of alkaline phosphatase treated assay mixture of AiGDS with GPP/FPP and IPP indicated that AiGDS failed to synthesize chain elongation products FPP (C_{15}) or GGPP (C_{20}) suggesting that AiGDS can catalyze the chain elongation reaction to produce GPP (C_{10}) as the sole enzymatic product.
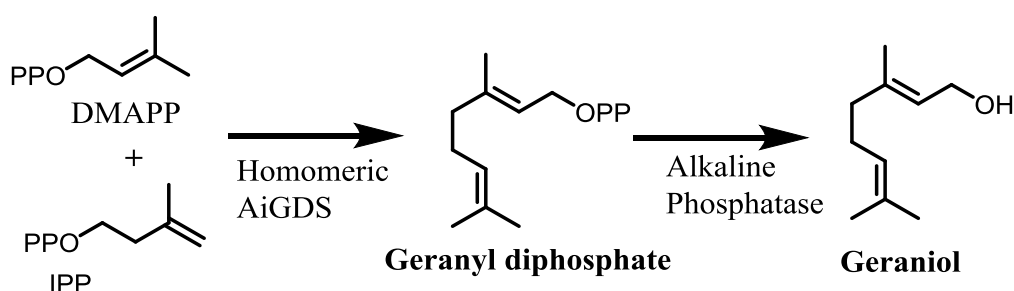
Figure 3. 5 Schematic Representation of AiGDS Assay.
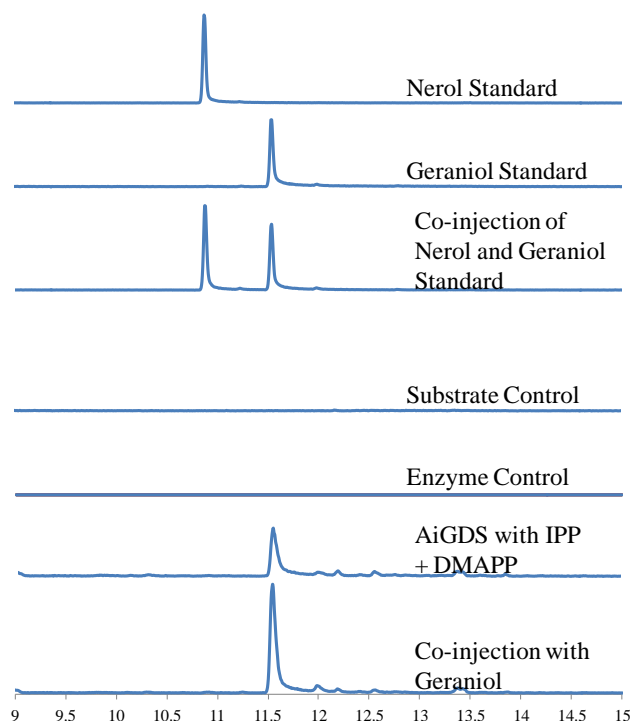
Figure 3. 6 Total Ion Chromatograms of AiGDS Assays With IPP and DMAPP as Substrates.

3.4.1.2 Cloning and Characterization of AiFDS

AiFDS [GenBank: KM10831] ORF of 1,029 bp length was found to be encoded for a protein of 342 amino acids. The theoretical molecular weight and pI for this polypeptide were 39.5 kDa and 5.59, respectively. The sequence comparison of AiFDS exhibited 83 % identity with FDS from *Mangifera indica* [GenBank: AFJ52720][17], 82 % identity with that from *Santalum album* [GenBank: AGV01244.1] and 81 % identity with FDS from *C. roseus* [GenBank: ADO95193.1][24]. The multiple sequence alignment of AiFDS consisted of two aspartate-rich motifs DDX$_{(2-4)}$D and DDXXD (Figure 3.7) which were highly conserved motifs in prenyltransferases.

Figure 3. 7 Multiple Sequence Alignment of *A. indica* Farnesyl Diphosphate Synthases (AiFDS).

Amino acid sequences of AtFDS1 (*A. thaliana*, NP_199588), AtFDS2 (*A. thaliana*, AAB07248), SaFDS (*Santalum album*, AEY80378), CrFDS (*C. roseus*, ADO95193), AiFDS

(*A. indica*, KM108316) and MiFDS (*M. indica,* AFJ52720) are used for multiple sequence alignment. The highly conserved Asp-rich motifs of prenyltransferases are indicated by solid line.
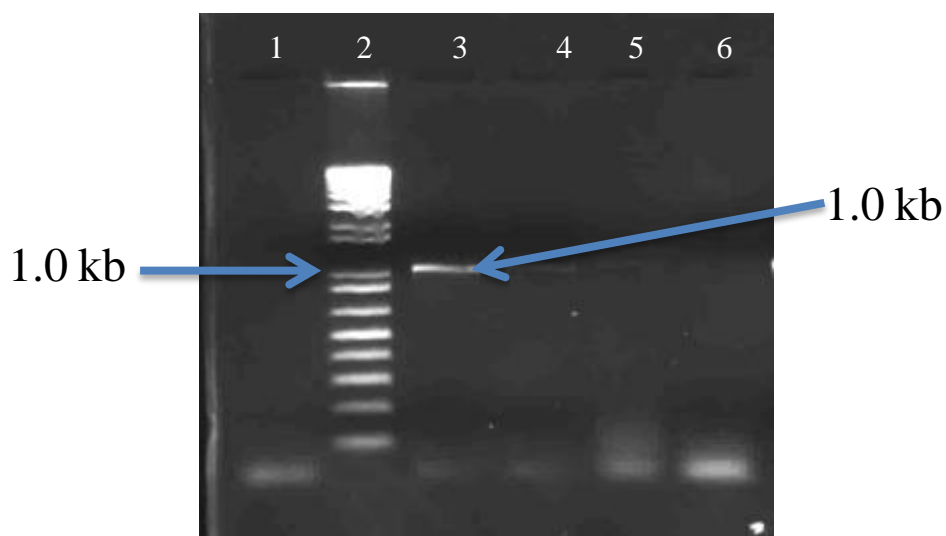
Figure 3. 8 AiFDS ORF Amplification.

**Lane 1:** Negative control, **Lane 2:** 1 kb DNA ladder Invitrogen (Addendum Figure A2.C), **Lane 3:** AiFDS PCR product at 52 °C, **Lane 4:** AiFDS PCR product at 54 °C, **Lane 5:** AiFDS PCR product at 56 °C and **Lane 6:** AiFDS PCR product at 58 °C.
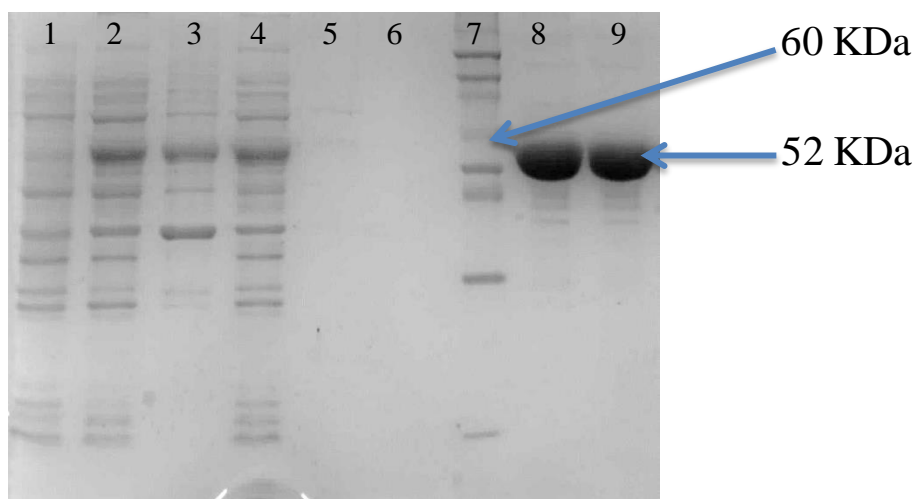
Figure 3. 9 SDS-PAGE for AiFDS Protein Purification in pET32a.
Lane1: Un-induced fraction, **Lane 2:** Supernatant fraction, **Lane 3:** Pellet fraction, **Lane 4:** Unbound fraction, **Lane 5:** Wash fraction 1, **Lane 6:** Wash fraction 2, **Lane 7:** Novex® Sharp Pre-stained Protein Standard (Addendum Figure A4.A), **Lane 8 and 9:** Elution fractions.

AiFDS was cloned into a pET32a expression vector. The cloned construct was transformed into BL21 (DE3) cells and expressed. AiFDS was obtained as soluble

form and purified by Ni-NTA affinity column chromatography. The recombinant protein was approximate 98 % pure as analyzed by SDS-PAGE (Figure 3.9). The purified prenyltransferase was incubated with DMAPP/GPP and IPP (Figure 3.10) followed by treatment with alkaline phosphatase. GC-MS analyses of the assay extracts indicated the formation of FPP which was further confirmed by comparing the retention time, mass fragmentation pattern and co-injection studies with standard (*E,E*)-farnesol (Figure 3.11). Further GC-MS analysis of alkaline phosphatase treated assay mixture of AiFDS with FPP and IPP did not show the formation of geranylgeraniol indicating that AiFDS catalyzes the chain elongation reaction to produce FPP as the sole enzymatic product.
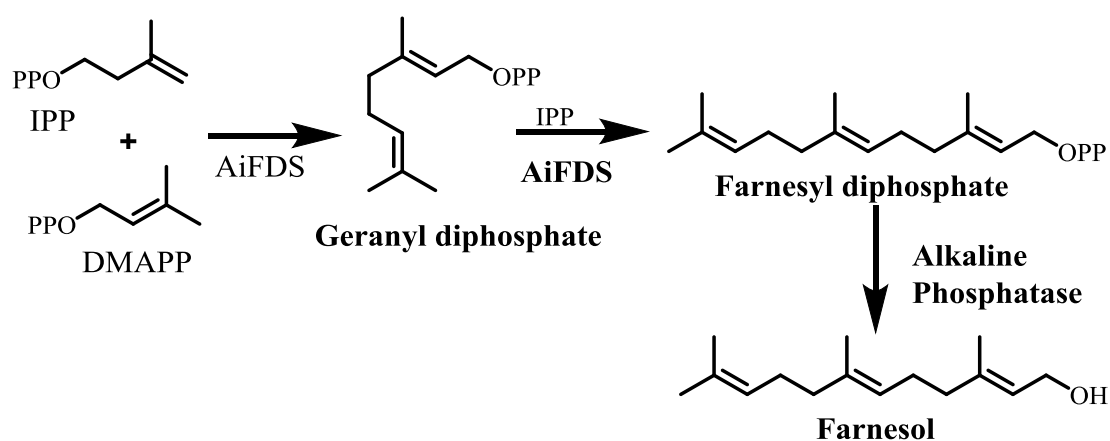
Figure 3. 10 Schematic Representation of AiFDS Assay.
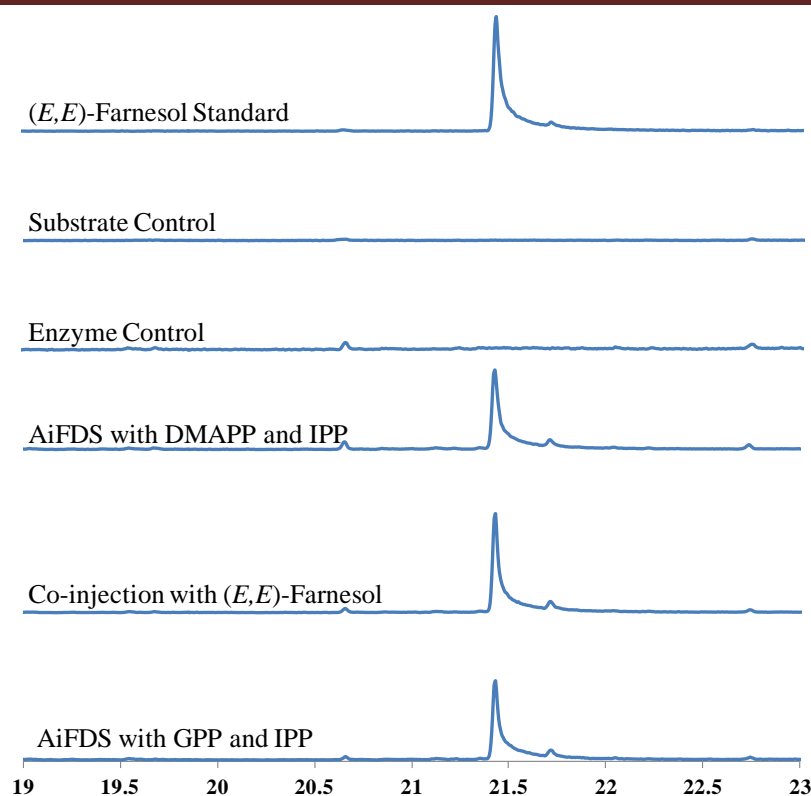
Figure 3. 11 Total Ion Chromatograms of AiFDS Assays with DMAPP/GPP and IPP as Substrates.

3.4.2 Real-Time PCR Analysis

To determine the role of prenyl diphosphate synthases in triterpenoid biosynthesis, real-time PCR analysis of AiGDS and AiFDS was carried out by taking GAPDH as an intrinsic control.

AiFDS (Figure 3.12) showed very high expression level in seeds as compared to other tissues. Expression patterns of AiFDS were found to be similar to metabolic profiling suggesting that these genes is involved in triterpenoid biosynthesis. Overexpression of FDS in *Panax ginseng* and *Centella asiatica* resulted in increased accumulation of phytosterols and triterepenes[12], which further confirms the involvement of AiFDFS in triterpenoid biosynthesis. AiGDS (Figure 3.12) showed very high expression in leaf and flower, compared to other tissues. These results indicate that AiGDS might be involved in other class of terpene biosynthesis in neem.
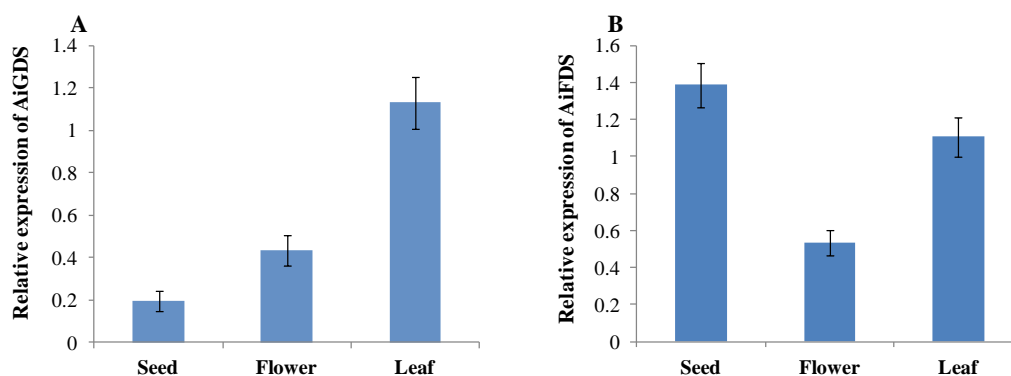
Figure 3. 12 Real-Time PCR Analysis of AiGDS and AiFDS.
A: AiGDS was highly expressed in leaf. **B:** AiGFS was highly expressed in seed.

3.5 Conclusion

Prenyltransferases catalyse head-to-tail condensation and chain elongation reactions of DMAPP and IPP to produce prenyl diphosphates like geranyl diphosphate, farnesyl diphosphate, geranylgeranyl diphosphate and other diphosphates. Prenyltransferases are involved in branching point of terpenoid metabolism and play a regulatory role in the distribution of isoprene units into various terpenoid biosynthesis. A total of ten prenyltransferases were obtained from neem transcriptome (Chapter 2, Table 2.3). Two geranyl diphosphate synthase (GDS), one farnesyl diphosphate synthase (FDS) and seven putative geranylgeranyl diphosphate synthases (GGDS) were identified. BLAST studies indicated that Neem_transcript_10912 was a homomeric GDS and Neem_transcript_10001 may be the smaller subunit of heteromeric GDS. Neem_transcript_10912 (AiGDS) and Neem_transcript_25722 (AiFDS) were cloned into pET32a vectors and protein was purified from expression system. DMAPP/GPP and IPP were used as substrates for assay and enzyme activity was confirmed by GC-MS and co-injection with authentic standards. AiFDS showed higher expression in seeds and its expression pattern matches with limonoids profile, which indicates its involvement in the triterpenoid biosynthesis.

3.6 Appendix: Agarose Gel Electrophoresis for Colony PCR Screening

3.6.1 Cloning of AiGDS in pET32a Vector

Colony PCR with T7 promoter forward and T7 reverse primer for the screening of geranyl diphosphate synthase cloned into pET32a vector.
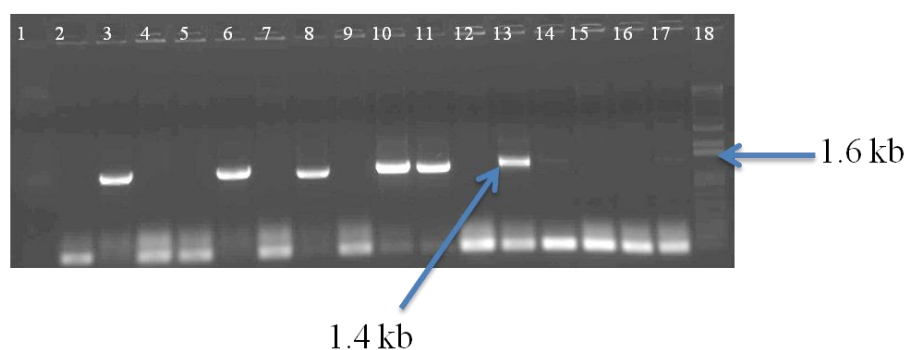
Figure 3. 13 Colony PCR Screening for AiGDS Cloned in pET32a on an Agarose Gel. **Lane 1**: Negative control, **Lanes 2-17:** PCR with T7 promoter and T7 reverse primer and **Lane 18**: 1 Kb DNA ladder Invitrogen (Addendum Figure A2.A).

3.6.2 Cloning of AiFDS in pET32a Vector

Colony PCR with T7 promoter and T7 reverse primer for the screening of farnesyl diphosphate synthase cloned into pET32a vector.
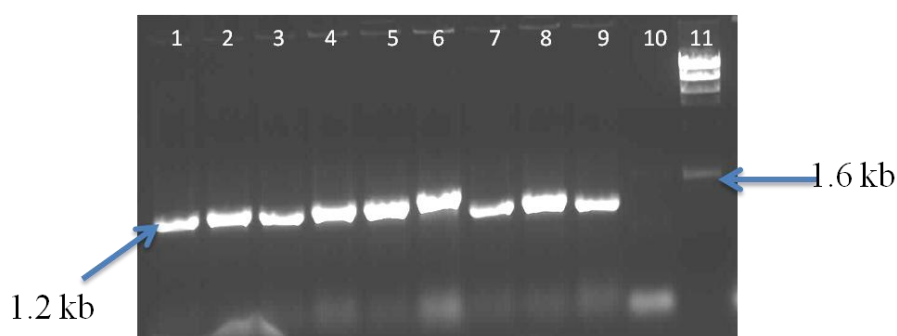
Figure 3. 14 Colony PCR Screening for AiFDS Cloned in pET32a on an Agarose Gel. **Lane 1-9**: PCR with T7 promoter and T7 reverse primer negative control, **Lanes 10:** Negative control and **Lane 11**: 1 Kb DNA ladder Sigma (Addendum Figure A1.A).

3.7 References

1. Bourgaud, F., Gravot, A., Milesi, S. & Gontier, E. Production of plant secondary metabolites: a historical perspective. *Plant Science* **161**, 839-851 (2001).

2. Nicoletti, M., Maccioni, O., Coccioletti, T., Mariani, S. & Vitali, F. Neem tree (*Azadirachta indica* A. Juss) as source of bioinsectides. in *Insecticides-Advances in Integrated Pest Management* (InTech, 2012).

3. Dai, J., Yaylayan, V.A., Raghavan, G.S.V. & Pare, J.R. Extraction and colorimetric determination of azadirachtin-related limonoids in neem seed kernel. *Journal of Agricultural and Food Chemistry* **47**, 3738-3742 (1999).

4. Lange, B.M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences* **97**, 13172-13177 (2000).

5. Thulasiram, H.V., Erickson, H.K. & Poulter, C.D. A common mechanism for branching, cyclopropanation, and cyclobutanation reactions in the isoprenoid biosynthetic pathway. *Journal of American Chemical Society* **130**, 1966-1971 (2008).

6. Thulasiram, H.V. & Poulter, C.D. Farnesyl diphosphate synthase: the art of compromise between substrate selectivity and stereoselectivity. *Journal of American Chemical Society* **128**, 15819-15823 (2006).

7. Mookhtiar, K.A., Kalinowski, S.S., Zhang, D. & Poulter, C.D. Yeast squalene synthase. A mechanism for addition of substrates and activation by NADPH. *Journal of Biological Chemistry* **269**, 11201-11207 (1994).

8. Schaller, H. New aspects of sterol biosynthesis in growth and development of higher plants. *Plant Physiology and Biochemistry* **42**, 465-476 (2004).

9. Kim, T.D., Han, J.Y., Huh, G.H. & Choi, Y.E. Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. *Plant and Cell Physiology* **52**, 125-137 (2010).

10. Suzuki, H., Achnine, L., Xu, R., Matsuda, S. & Dixon, R.A. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Journal* **32**, 1033-1048 (2002).

11. Kim, O.T. *et al*. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. *Biologia Plantarum* **54**, 47-53 (2010).

12. Kim, Y.K. *et al*. Enhanced Triterpene Accumulation in *Panax ginseng* Hairy Roots Overexpressing Mevalonate-5-pyrophosphate Decarboxylase and Farnesyl Pyrophosphate Synthase. *ACS Synthetic Biology* **3**, 773-779 (2014).

13. Kim, O.T. *et al.* Upregulation of phytosterol and triterpene biosynthesis in *Centella asiatica* hairy roots overexpressed ginseng farnesyl diphosphate synthase. *Plant Cell Rep* **29**, 403-11 (2010).

14. Ament, K., Van Schie, C.C., Bouwmeester, H.J., Haring, M.A. & Schuurink, R.C. Induction of a leaf specific geranylgeranyl pyrophosphate synthase and emission of (*E*, *E*)-4, 8, 12-trimethyltrideca-1, 3, 7, 11-tetraene in tomato are dependent on both jasmonic acid and salicylic acid signaling pathways. *Planta* **224**, 1197-1208 (2006).

15. Bouvier, F., Suire, C., d'Harlingue, A., Backhaus, R.A. & Camara, B. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. *Plant Journal* **24**, 241-252 (2000).

16. Rai, A., Smita, S.S., Singh, A.K., Shanker, K. & Nagegowda, D.A. Heteromeric and homomeric geranyl diphosphate synthases from *Catharanthus roseus* and their role in monoterpene indole alkaloid biosynthesis. *Molecular Plant* **6**, 1531-1549 (2013).

17. Kulkarni, R. *et al*. Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit. *Plant Physiology Biochemistry* **71**, 121-31 (2013).

18. Schmidt, A. & Gershenzon, J. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). *Phytochemistry* **69**, 49-57 (2008).

19. Ma, Y. *et al.* Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *Journal of Experimental Botany* **63**, 2809-2823 (2012).

20. Wang, G. & Dixon, R.A. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *Proceedings of the National Academy of Sciences* **106**, 9914-9919 (2009).

21. Tholl, D. *et al*. Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. *The Plant Cell* **16**, 977-992 (2004).

22. Wagner, S. *et al.* Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences* **105**, 14371-6 (2008).

23. Gennadios, H.A. *et al.* Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry.* **48**, 6175-83 (2009).

24. Thabet, I. *et al*. The subcellular localization of periwinkle farnesyl diphosphate synthase provides insight into the role of peroxisome in isoprenoid biosynthesis. J*ournal Plant Physiol.* **168**, 2110-6 (2011).

Leopold Ruzicka   **Konrad E. Bloch**   **Feodor Lynen**   **Robert B. Woodward**

Chapter 4

Cloning and Functional Characterization of Squalene Epoxidase and Triterpene Synthases

4.1 Introduction

Limonoids are a class of triterpenoids which are highly modified and oxygenated with basic skeletons of 4,4,8-trimethyl-17-furanylsteroid and its derivatives. These are present abundantly in plant Meliaceae family and are known as meliacins[1]. *Azadirachta indica* (neem) belongs to Meliaceae family and contains more than 150 triterpenoids or meliacins of which some are more complex and highly oxygenated[2]. The abundance of total limonoids is high in neem kernel and pericarp and lower in flower, stem and bark. Broadly neem limonoids were divided into ring intact or basic type and C-seco type with opened C ring. These limonoids distributions differ across neem tissues. For example, the basic limonoids are high in pericarp and C-seco limonoids are high in the kernel[3]. The tissue-specific distribution of limonoids in neem shows that their biosynthesis is very well regulated.

Farnesyl diphosphate, which is synthesized from allelic diphosphates (IPP and DMAPP) act as the precursor diphosphate metabolite for the synthesis of steroids and triterpenoids. Head-to-head condensation of two molecules of FPP results in the formation of squalene, which is catalyzed by squalene synthase (SQS) in the presence of cofactor NADPH. In 1953, Woodward R.B. and Bloch K proposed the involvement of squalene in cholesterol biosynthesis[4]. Squalene epoxidase (SQE) oxidizes squalene into 2,3-oxidosqualene, which act as the substrate for synthesis of steroids and triterpenoids. Independent research from Corey and Van Tamelen stated that 2,3-oxidosqualene acts as a substrate for lanosterol biosynthesis[5,6] and further in 1975 Barton confirmed the same[7,8]. In *Panax ginseng*, two SQE genes have been identified. When PgSQE1 transcription was suppressed through RNA interference a reduction of ginsenoside production was found. Whereas, in roots, upregulation of PgSQE2 and cycloartenol synthase resulted in enhanced phytosterol, which indicates that PgSQE1 is involved in triterpenoid biosynthesis[9]. Similarly, in *Arabidopsis thaliana*, three SQE genes were identified and out of the three, AtSQE1 play a key role in seed and root development. AtSQE1-3 mutant, accumulates squalene indicating the blockage of triterpenoid biosynthesis[10,11]. Thus, we can interpret that multiple SQE play a key regulatory role in steroid and triterpenoid biosynthesis in plants.

The first committed step in triterpenoid biosynthesis is cyclization of 2,3-oxidosqualene which is catalyzed by triterpene synthases (TTS). Plants, microalgae and many protozoa possess cycloartenol synthase (CAS), whereas animals and fungi have lanosterol synthase for the synthesis of the respective steroids[12]. In animals, lanosterol further modified into steroids, which play key role in membrane fluidity and act as hormones. In plants, cycloartenol further modified into brassinosteroids, play a key role in cell elongation and seed germination. In most of the plants, more than one TTS are found to be present in the genome. For example, two TTS are observed in *Ocimum basilicum*[13] whereas thirteen TTS were observed in *A. thaliana*. These TTS can cyclize the 2,3-oxidosqualene into single or multiple triterpene skeleton products. *A. thaliana* triterpene synthase PEN3 (at5g36150) makes tirucalla-7,24-dien-3β-ol (~85%) and seven minor products[14]. *Euphorbia tirucalli* (AB206469) β-amyrin synthase (EtBAS) can produce β-amyrin, tirucalla-7,24-dien-3β-ol and butyrospermol[15].

(*S*)2,3-oxidosqualene undergoes a series of cyclization, ring expansion, rearrangement (1,2 methyl/hydride shifts) and deprotonation to form cyclic triterpenes[16]. Protonation of epoxide ring triggers cyclization in 2,3-oxidosqualene. Neighbouring group participation (NGP) from double bond (C-6/C-7) facilitates the initial protonation as well as A ring formation, which results in monocyclic C-10 cation. Deprotonation of C-19 and C-1 from monocyclic C-10 cation results in camelliol C and achilleol A, respectively. Bicyclic C-8 cation B ring form when monocyclic C-10 cation undergoes NGP from double bond (C-8/C-9). Bicyclic triterpene, polypoda-7,13,17,21-tetraen-3β-ol, is generated by deprotonation of C-7 from bicyclic C-8 cation. C ring as well as 6,6,5-tricyclic C-14 cation forms when NGP from double bond (C-13/C-14) on bicyclic C-8 cation. Protosteryl C-20 cation and dammarenyl C-20 cation are formed from tricyclic cation having either C-B-C (chair-boat-chair) or C-C-C (chair-chair-chair)[17] conformation (Figure 4.1).
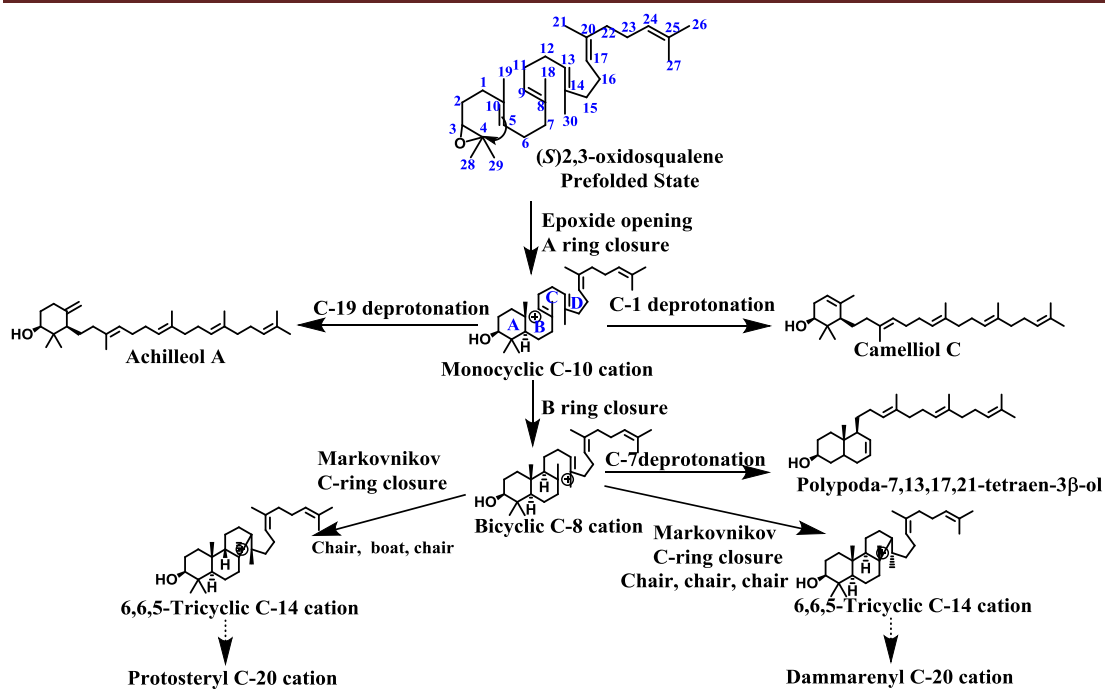
Figure 4. 1 Scheme for Cyclization of (*S*)2,3-Oxidosqualene into Tricyclic Cation and Derivatives.

6,6,5-tricyclic C-14 cation (C-B-C conformation) undergoes C-ring expansion (anti-markovnikov reaction) to from 6,6,6-tricyclic C-13 cation. Further, tetracyclic protosteryl C-20 cation is formed by NGP from double bond (C-17/C-21) on tricyclic cation. The formation of lanosteryl C-8 cation is a result of 1,2-shift of hydride ion (H17-Cβ20, H13α-C17) and two methyl group (Me30β-C13, Me18α-C14) from tetracyclic cation. Parkeol and lanosterol are formed by deprotonation at C-11 and C-9 from lanosteryl C-8 cation, respectively[17]. Furthermore, lanosteryl cation undergoes 1,2 shifts of hydride (H9β-C8) resulting in lanosteryl C-9 cation. The rearrangement (Me19β-C9, H5α-C10) and deprotonation C-6 of lanosteryl cation results in formation of cucurbitadienol. The deprotonation of C-19 from lanosteryl C-9 cation results in cycloartenol[18] (Figure 4.2).
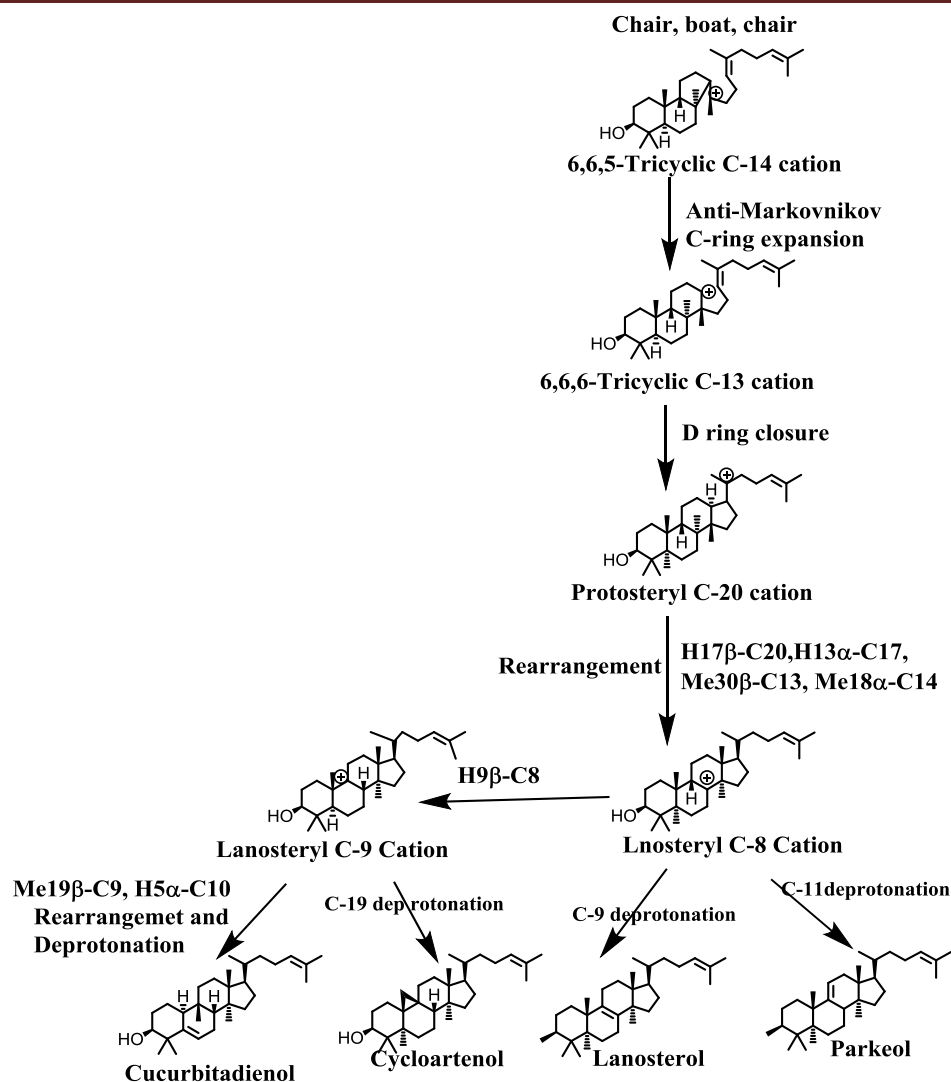
Figure 4. 2 Protosteryl C-20 Cation Synthesis and its Tetracyclic Triterpenes.

Triterpene malabarica-14(27),17,21-trien-3-ol synthesized is from 6,6,5-tricyclic C-14 cation (C-C-C conformation) by deprotonation at C-30[19]. C-ring expansion through anti-markovnikov reaction forms 6,6,6-tricyclic C-14 cation. The NGP from double bond (C-17/C-21) on tricyclic cation results in formation of tetracyclic dammarenyl C-20 cation (C-C-C-B) and 17-*epi*-dammarenyl C-20 cation (C-C-C-C). Tetracyclic triterpenes such as dammara-20(21),24-dien-3β-ol, (20*E*)-dammara-20(22),24-dien-3β-ol and (17*E*)-dammara-17(20),24-dien-3β-ol are formed by C-21, C-22 and C-17 deprotonation of dammarenyl C-20 cation, respectively. Further, the 1,2-shift of hydride and methyl groups (H17α –C20, H13β-C17, Me30α-C13, Me18β-C14) from (17*E*)- or dammarenyl C-20 cation results in tirucallane C-8

cation and euphane C-8 cation. The deprotonation at C-9 and C-7 from tirucallane C-8 cation produces tirucallol and tirucalla-7,24-dien-3β-ol, respectively. Similarly, deprotonation of C-9 and C-7 from euphane C-8 cation produces euphol and butyrospermol, respectively[20] (Figure 4.3).
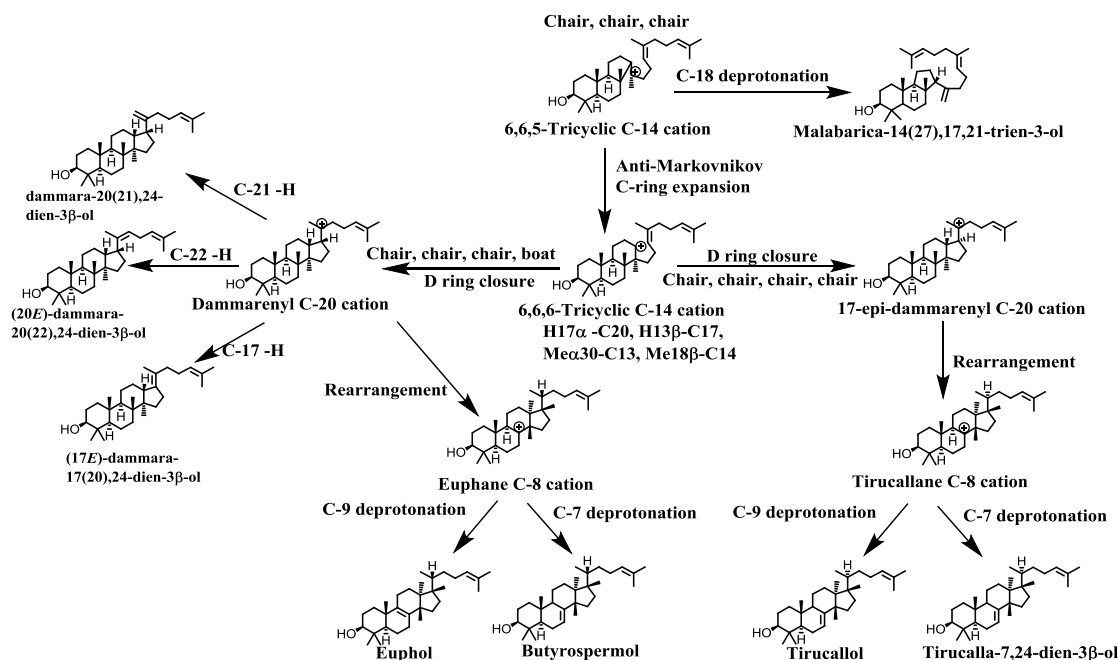
Figure 4. 3 Dammarenyl C-20 Cation Synthesis and its Tetracyclic Triterpenes.

Dammarenyl C-20 cation undergoes D ring expansion leading to the 6,6,6,6-fused baccharenyl C-18 cation. Further, E ring closure forms 6,6,6,6,5-fused lupanyl C-20 cation. Deprotonation from Met29 gives lupeol whereas, its rearrangement and deprotonation leads to formation of lup-12-en-3β-ol. Germanicyl C-19 cation is produced by E ring expansion from lupanyl cation and the deprotonation from C-18 results in germanicol. 1,2-shift of hydrid ions (H18α-C19, H13β-C18) produces oleanyl C-13 cation, C-12 deprotonation results in β-amyrin. When Oleanyl C-13 cation undergoes 1,2-shift of methyl group (Me27α-C13) it results in taraxareyl C-14 cation and the further deprotenation from C-15 leads to taraxerol formation. Rearrangement of taraxareyl C-14 cation (Me26β-C14, H9α-C8, Me25β-C9, H5α-C10) results in glutinyl C-5 cation, deprotonation from C-6 leads to glutinol. Further, glutinyl cation undergoes rearrangement (Me24β-C5) resulting in fridelyl C-4 cation. This rearrangement in turn leads to friedelin synthesis[21]. Taraxasteryl C-20 cation is

produced when germanicyl C-19 cation undergoes rearrangement (Me29β-C19). C-23 and C-30 deprotonation of taraxasteryl C-20 cation results in Ψ-taraxasterol and taraxasterol, respectively. Hydrid shift (H19-C20, H18α-C19, H13α-C18) in tarasasteryl cation leads to the formation of ursanyl C-13 cation. Deprotonation from C-12 and rearrangement leads to α-amyrin and isoursenol, respectively[20] (Figure 4.4).
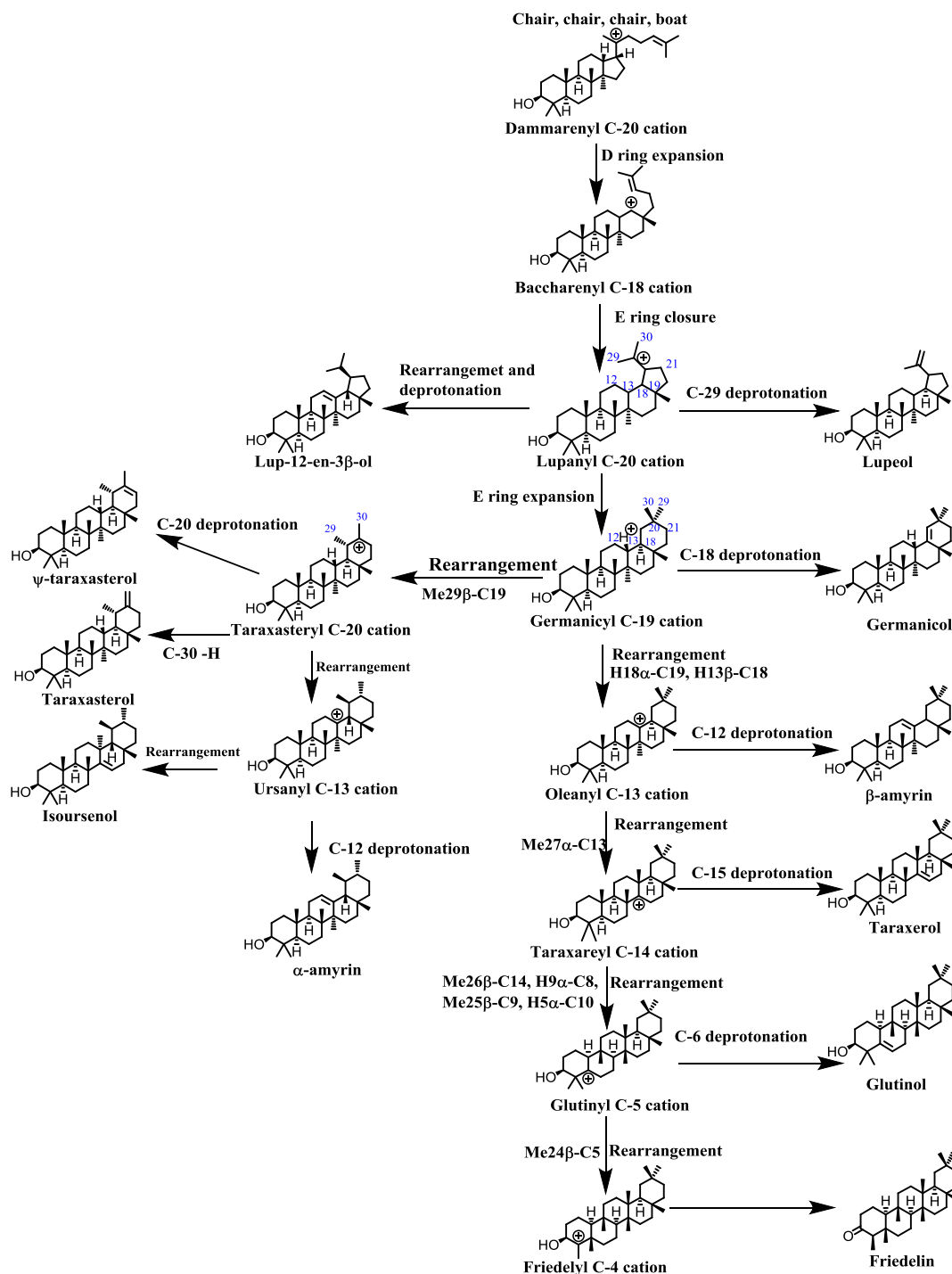
Figure 4. 4 Mechanism for Pentacyclic Triterpene Synthesis

Euphol and tirucallol derivatives are predicted to be involved in limonoid biosynthesis based on oxygenated C_{30} compounds (protolimonoid skeleton) isolated from Meliaceae plants[22]. The difference between euphol (C-20 *R*) and tirucallol (C-20 *S*) is the orientation of C-22 with respect to C-13. These two are epimers at C-20 and are formed by C17 (20)-bond rotation. Butyrospermol and tirucalla-7,24-dien-3β-ol are Δ^7-euphol and Δ^7-tirucallol, respectively and are synthesized from 6/6/6/5-fused tetracyclic dammarenyl cation[23]. The confirmation of dammarenyl cation is in euphol/Δ^7-euphol and tirucallol/Δ^7-tirucallol is chair-chair-chair-boat and chair-chair-chair-chair, respectively. The dammarenyl cation undergoes 1,2-migration of hydride and methyl groups (H17α,β→C20, H13β→C17, Me30α→C13, Me18β-C14) to form C-8 carbocation[24]. C-7 deprotonation and C-17 (20)-bond rotation results in the formation of butyrospermol and tirucalla-7,24-dien-3β-ol. In case of euphol and tirucallol, deprotonation occurs at C-9 position. Tritium labelled nimbolide was observed when labelled euphol, tirucallol, Δ^7-tirucallol and butyrospermol were fed into leaves of neem and euphol was found to be more effectively incorporated. From these experiments, either of Δ^7-isomer of euphane (butyrospermol) or tirucallane (tirucalla-7,24-dien-3β-ol) is predicted to be the protolimonoid skeleton for limonoid biosynthesis in neem[25,26]. Characterization of triterpene synthases from neem helps in understanding the type of protolimonoid skeleton involved in neem limonoid biosynthesis.

4.2 Neem Squalene Epoxidase and Triterpene Synthases

Squalene epoxidase catalyzes the oxidation of squalene into 2,3-oxidosqualene and plays a key role in the synthesis of steroids and triterpenoids. In neem, a total of three SQE were observed. AiSQE1 (Master_Control_31859/ Neem_transcript_11067; NCBI JX997152) is highly expressed in flower and leaves. AiSQE2 (Master_Control_42537/ Neem_transcript_18229) is expressed low as compared to other AiSQEs. AiSQE3 (Master_Control_80013/ Neem_transcript_18980) shows considerably high expression in kernel and pericarp. Based on the comparative expression levels, AiSQE3 is predicted to be involved in limonoid biosynthesis. Triterpene synthase catalyzes the first committed step in triterpenoid and steroid biosynthesis by cyclizing 2,3-oxidosqualene. In neem, six

triterpene synthases were observed. AiCAS (Master_Control_74065/ Neem_transcript_27436) is cycloartenol synthase which is expressed almost equally in all the tissues. A correlation can be drawn between expression profile of AiTTS1 and neem limonoids profile (Master_Control_24780/ Neem_transcript_28920) which are highly expressed in seed. AiTTS1 is differently expressed in the kernel as compared to flowers which give us an indication of its involvement in limonoid biosynthesis. AiTTS2 (Master_Control_74892) is highly expressed in flower and leaves. AiTTS3 (Master_Control_70149), AiTTS4 (Master_Control_70584) and AiTTS5 (Master_Control_101750) show a least expressions as compared to other AiTTSs. One squalene epoxidase (AiSQE1) and two triterpene synthases (AiTTS1 and AiTTS2) were selected for functional characterization.
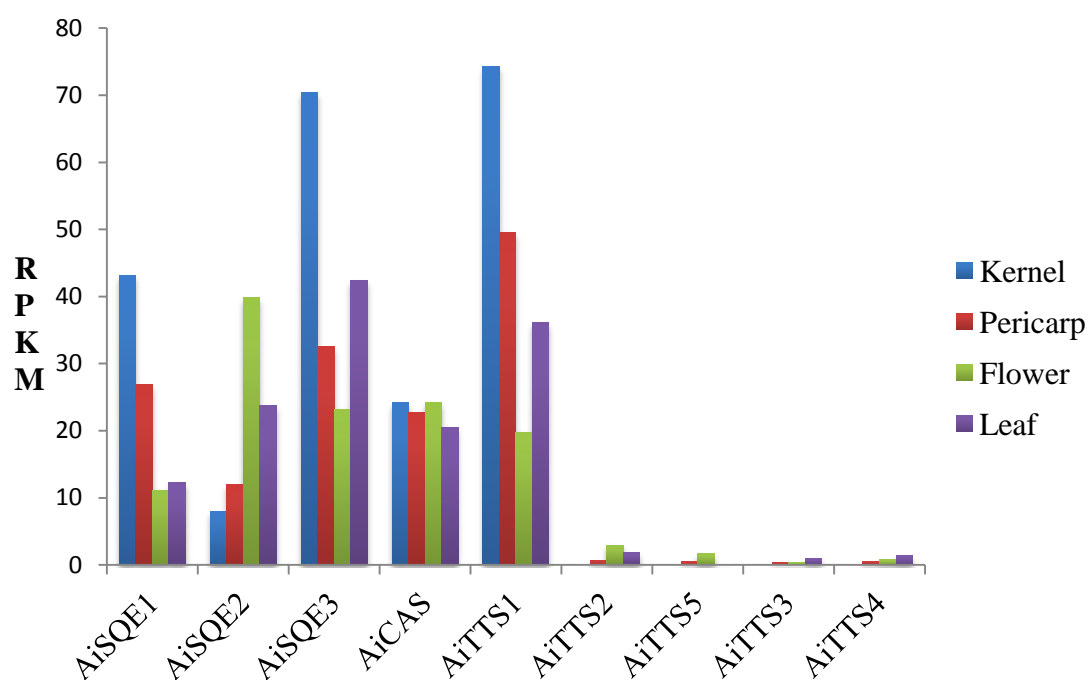
Figure 4. 5: **RPKM of Squalene Epoxidases (AiSQE) and Triterpene Synthase (AiTTS) Across Different Tissues in Neem.**

4.3 Materials and Methods

4.3.1 Materials Used in this Study

4.3.1.1 Bacterial Strains and Plasmids Used in the Study

Escherichia coli Mach1™ T1^R (ThermoFisher Scientific, USA) cloning cells were used for transformation of plasmids or ligation mixtures. Zero Blunt™ PCR Cloning Kit (ThermoFisher Scientific, USA) was used for cloning of RACE products. pYES2/CT Yeast Expression Vector (ThermoFisher Scientific, USA), pESC-LEU Yeast Expression Vector (Agilent Technologies, USA) and pRS315-TEF Yeast Expression Vector (obtained from Dr. Anand Bacchwat, IISER, Mohali) were used of cloning of AiSQE1 and AiTTSs. Expression of AiSQE and AiTTSs were done in INVSc1, *Saccharomyces cerevisiae* yeast Strain (ThermoFisher Scientific, USA).

4.3.1.2 Kits and Reagents Used in the Study

SuperScript® III Reverse Transcriptase (ThermoFisher Scientific, USA) was used for cDNA synthesis. JumpStart™ Taq DNA Polymerase (Sigma-Aldrich, USA) and AccuPrime™ *Pfx* DNA Polymerase (ThermoFisher Scientific, USA) were used for amplification of AiSQE1 and AiTTSs. PCR products were gel eluted by using PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (ThermoFisher Scientific, USA). Plasmids were isolated by using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, USA). GelRed^{TM} (Biotium Inc., USA) was used for nucleic acid staining. Restriction enzymes (New England Biolabs, USA) and T$_4$ DNA ligase (Invitrogen/ Life Technologies, USA) used for cloning. Postive clones were transformed into yeast by using *S.c.* EasyComp^{TM} Transformation Kit (ThermoFisher Scientific, USA). Mutagenesis was performed by using QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies, USA). DNA contamination from RNA was removed by using DNase I Amplification Grade Kit (Sigma-Aldrich, USA). FastStart Universal SYBR Green Master (Rox) (Roche, Switzerland) was used for real-time PCR.

4.3.1.3 Primers

4.3.1.3.1 Primers for AiSQE1

Table 4. 1 Primers Used for Cloning of AiSQE1.

Primer Name	Primer Sequence
AiSQE1_pESC-LEU_FP	ACACA**GGATCC**GATGGCGGCTGTGATTGAT
AiSQE1_pESC-LEU_RP	ATATTC**AAGCTT**TCAA**TGGTGATGGTGATGAT G**ATCGTCAACAGGAGGAGC
AiSQE1_ pRS315-TEF _FP	ACGATC**GGATCC**CACACAATGTCCACCGGAAA CTGCAATATCCAAAATG
AiSQE1_ pRS315-TEF _RP	GATTGC**AAGCTT**TTA**ATGGTGATGATGATGAT GAGAAGAACC**ATCGTCAACAGGAGGAGCTCTG
GAL1_FP	TCTGGGGTAATTAATCAGCGAAGCGATG
GAL1_RP	CTAGACTTCAGGTTGTCTAACTCCTTC
GAL10_FP	GATAATGGGGCTCTTTACATTTCCACAAC
GAL10_RP	CAACGATTTGACCCTTTTCCATCTTTTCG
CYC Reverse Primer	GCGTGAATGTAAGCGTGAC

4.3.1.3.2 Primers for AiTTS1

Table 4. 2 Primers Used for Cloning of AiTTS1.

Primer Name	Primer Sequence
AiTTS1 _FP	ATGTGGAAGCTGAAGATTGCAGAG
AiTTS1 _RP	TTAATTAGGCAATGGAACTTTTCTTCTATATT C
AiTTS1_ pYES2/CT _FP	GATCAC**GAATTC**AACACAATGTCT ATGTGG AAGCTGAAGATTG

AiTTS1_ pYES2/CT _RP	CATGCA**GCGGCCGC**ATTAGGCAATGGAACTTTTC
AiTTS1_ RT-PCR _FP	GACCTTATTGTTGAGCATACTTA
AiTTS1_ RT-PCR _RP	CCATCAGCAGTTTGTTCAT

4.3.1.3.3 Mutation Primers for AiTTS1

Table 4. 3 Primers Used for Mutagenesis of AiTTS1.

Mutation	Primer Sequence
AITTS1_L553F_FP	CCTGTAGAGTTT**TTC**GAGGACCTTATT
AITTS1_L553F_RP	AATAAGGTCCTC**GAA**AAACTCTACAGG
AITTS1_Y125F_FP	TCTGGCCCTATG**TTT**TTCCTTCCTCCA
AITTS1_Y125F_RP	TGGAGGAAGG**AAA**AACATAGGGCCAGA
AITTS1_ F260Y_FP	CAAATGTGGTGC**TAC**TGCCGGCTGGTT
AITTS1_ F260Y_RP	AACCAGCCGGCA**GTA**GCACCACATTTG
AITTS1_V550T_FP	TGGCTCAATCCT**ACA**GAGTTTCTGGAG
AITTS1_V550T_RP	CTCCAGAAACTC**TGT**AGGATTGAGCCA
AITTS1_V534A_FP	GGTGGAATAGCA**GCA**TGGGAGAAAGCTG
AITTS1_V534A_RP	CAGCTTTCTCCCA**TGC**TGCTATTCCACC
AITTS1_T413S_FP	GAATGAAAGTTCAG**TCA**TTTGGCAGTCAAAC
AITTS1_T413S_RP	GTTTGACTGCCAAA**TGA**CTGAACTTTCATTC
AITTS1_L556I_FP	GTTTCTGGAGGAC**ATA**ATTGTTGAGCATAC
AITTS1_L556I_RP	GTATGCTCAACAAT**TAT**GTCCTCCAGAAAC
AITTS1_V484L_FP	GATCATGGTTGGCAA**CTT**TCAGATTGTACTGCAG
AITTS1_V484L_RP	CTGCAGTACAATCTGA**AAG**TTGCCAACCATGATC

4.3.1.3.4 Primers for AiTTS2

Table 4. 4 Primers Used for Cloning of AiTTS2.

Primer Name	Primer Sequence

AiTTS2_3′RACE_FP	TATGGGTTGGCGAAGATGGAATGAAG
AiTTS2_3′RACE_NFP	GAATGGGATGCTGGTTTTGCTATTC
Universal 3′RACE RP	GCTGTCAACGATACGCTACGTAACG
Universal 3′RACE Nested RP	CGCTACGTAACGGCATGACAGTG
AiTTS2_ pYES2/CT _FP	GATCAC**GCGGCCGC**AACACAATGTCTATGTGGAAGCTTAAGGTTGCAGATG
AiTTS2_ pYES2/CT _RP	CAGTAC**TCTAGA**GCTACTTAGCTGCACGGTGAAAAGG
AiTTS2_ RT-PCR _FP	AGGTCAAAGCTGGGGAGGAAATCAC
AiTTS2_ RT-PCR _RP	GTCTTGGCAACCATCAGGTCCTTCT

4.3.1.3.5 Primers for Actin

Table 4. 5 Primers Used for Actin

Primer Name	Primer Sequence
Actin_FP	AGGCATCCACGAGACCACTT
Actin_RP	TGGCGCTAGAGCAGAAATTTC

4.3.1.6 GC-FID/GC-MS Program

Functional characterization of AiSQE1 and AiTTS1 were carried out on GC-FID or GC-MS equipped with 30 m × 0.25 mm × 0.25 µm capillary columns (HP-5 MS, J & W Scientific and Restek Rtx-5) using the program: 80 °C for 1 min, 5 °C/min rise till 290 °C and hold for 20 min was used.

4.3.2 Cloning and Characterization of AiSQE1, AiTTS1 and AiTTS2

4.3.2.1 Homology Model of AiTTS1 and AiTTS2

Easymodler 4.0 was used to generate homology model of AiTTS1 and AiTTS2. Human lanosterol synthase coordinates were used as a template. The

homology models were validated by using PROCHECK and images were generated by PYMOL (v. 1.8).

4.3.2.2 Cloning and Characterization of AiTTS1

Full-length primers (Table 4.2, AiTTS1_ pYES2/CT _FP and AiTTS1_ pYES2/CT _RP) for AiTTS1 ORF were designed from its transcripts (Master_Control_24780/ Neem_transcript_28920) as a template. cDNA was used as a template for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen) to amplify full-length AiTTS1using the PCR program: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 2 min 20 sec followed by final extension at 68 °C for 5 min. Full-length AiTTS1 PCR product was cloned into *EcoRI* and *NotI* cloning sites of pYES2/CT expression vector using T_4 DNA ligase. The ligation mixture was transformed into Mach1™ T1^R competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were identified by carrying out colony PCR using with T7 forward and CYC reverse primers. Full-length AiTTS1 cloning was confirmed by analyzing the sequences obtained by Sanger sequencing of positive clones by using T7 forward and CYC reverse primers.

The expression of AiTTS1 was carried out in INVSc1, *S. cerevisiae* yeast strain. The cloned AiTTS1 plasmid was transformed using *S.c.* EasyComp^{TM} Transformation Kit and plated on CSM-URA plates. A single colony was inoculated into 2 % glucose contain CSM-URA (50 ml) and incubated at 30 °C. The overnight grown culture was induced by transferring it into 2 % galactose containing CSM-URA (500 mL) and incubated at 30 °C for 24 h. Cells were collected by centrifugation at 1500 × *g* for 10 min at 4 °C. Saponification was done with 10 % KOH in 80 % ethanol at 70 °C for 2 h and then extracted thrice with equal volumes of n-hexane. Metabolite extracts were passed through anhydrous sodium sulphate and concentrated to 50 µL. The extracted samples were analyzed by GC-MS (4.3.1.6-GC-FID/GC-MS Program).

To characterize the AiTTS1 metabolite, 25 L of yeast culture was grown in 2 % galactose containing CSM-URA medium. Crude metabolite extract was obtained

from n-hexane extract of saponified yeast pellet as described above. 120 mg of crude metabolite was loaded onto the column (1×50 cm) containing 5 % silver nitrate impregnated silica gel (230 – 400 mesh size) and dichloromethane (DCM) was used as a solvent. The column was passed through 100 % DCM in which lanosterol and AiTTS1 metabolic product were eluted. Squalene and ergosterol were eluted when the polarity was increased to 5 % by methanol. All the eluted fractions were analyzed by GC-MS. The AiTTS1 metabolite mass, structure and stereochemistry were analyzed using different techniques such as by GC-MS, ^{1}H-NMR, ^{13}C-NMR, DEPT, HMBC, HSQC, COSY and NOSY.

4.3.2.3 Cloning and Characterization of AiTTS2

In order to get a full-length gene of AiTT2, 3' RACE PCR was performed by using AiTTS2 3' RACE forward primer (Table 4.4, AiTTS2_3' RACE_FP and AiTTS2_3'RACE_NFP) and universal 3' RACE reverse primer. RACE cDNA was used as a template for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen) to amplify 3' RACE product. A gradient PCR (95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 58 - 66 °C for 30 sec, 68 °C for 1 min 30 sec and a final extension at 68 °C for 5 min) was performed in order to amplify 3' RACE fragment. The amplified 3'RACE product was used as a template with AiTTS2 3' nested RACE forward primer and universal nested 3' RACE reverse primer to get desired 3' RACE product. The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 1 min 30 sec and the final extension at 68 °C for 5 min. The 3' RACE product was cloned into pCR™-blunt vector. The ligation mixture was transformed into Mach1™ T1[R] competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were identified by carrying out colony PCR using M13 forward and M13 reverse primers. The sequence of AiTTS 3' RACE was obtained by analyzing the sequences obtained by Sanger sequencing of positive clones by M13 forward and M13 reverse primers.

Full-length primers for AiTTS2 ORF were designed using their transcript (Master_Control_74892) and AiTTS2 3' RACE product as a template. cDNA was used as a template for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen) to amplify full-length AiTTS2. The program for PCR was 95 °C for 5 min, followed by

35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 2 min 20 sec followed by final extension at 68 °C for 5 min. Full-length AiTTS2 PCR product was cloned into *NotI* and *XbaI* cloning sites of pYES2/CT expression vector using T_4 DNA ligase. The ligation mixture was transformed into Mach1™ T1^R competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (T7 forward and CYC reverse primers) to identify the positive colonies. Full-length AiTTS1 cloning was confirmed by analysing the sequences of positive clones by Sanger sequencing using T7 forward and CYC reverse primers.

The expression of AiTTS2 was carried out in INVSc1, *S. cerevisiae* yeast strain. The cloned AiTTS2 plasmid was transformed by using *S.c.* EasyComp™ Transformation Kit and plated on CSM-URA plates. A single colony was inoculated into 2 % glucose contain CSM-URA and incubated at 30 °C. The overnight grown culture was induced by transfer to 2 % galactose containing CSM-URA and incubated at 30 °C for 24 h. Cells were collected by centrifugation at 1500 × *g* for 10 min at 4 °C. Saponification was done with 10 % KOH in 80 % ethanol at 70 °C for 2 h and then extracted thrice with equal volumes of n-hexane. Metabolite extracts were passed through anhydrous sodium sulphate and concentrated. The extracted samples were analyzed by GC-MS (4.3.1.6-GC-FID/GC-MS Program).

4.3.2.4 Cloning and Characterization of AiSQE1

Full-length primers (Table 4.2: AiSQE1_pESC-LEU_FP and AiSQE1_pESC-LEU_RP) for AiSEQ1 ORF were designed from their transcripts (Master_Control_31859/Neem_transcript_11067) as template. Apart from this, primers (Table 4.2, AiSQE1_pRS315-TEF_FP and AiSQE1_pRS315-TEF_RP) were designed by removing the putative membrane-binding domain of AiSQE1 to obtain truncated form of the gene. cDNA was used as a template for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen) to amplify full-length AiSQE1. The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 62 °C for 30 sec, 68 °C for 1 min 30 sec followed by final extension at 68 °C for 5 min. Full-length AiSQE1 PCR product was cloned into *BamHI* and *SacI* cloning sites of pESC-LEU expression vector using T_4 DNA ligase. The ligation mixture was transformed

into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Further colony PCR was carried out (with GAL1 forward and GAL1 reverse primers) to identify the positive clones. Full-length AiSQE1 cloning was confirmed by analyzing the sequences obtained by Sanger sequencing of positive clones by using GAL1 forward and GAL1 reverse primers.

The truncated AiSQE1 was amplified using AccuPrime *Pfx* Supermix (Invitrogen) using PCR program as follows: 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 56 °C for 30 sec, 68 °C for 1 min 30 sec min followed by final extension at 68 °C for 5 min. Truncated AiSQE1 PCR product was cloned into *BamHI* and *SacI* cloning sites of pRS315-TEF expression vector using T_4 DNA ligase. The ligation mixture was transformed into Mach1™ T1^R competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Positive clones were identified by colony PCR with AiSQE1 forward primer and CYC reverse primers. AiSQE1 cloning was confirmed by analyzing the sequences obtained by Sanger sequencing of positive clones by using CYC reverse primer.

The expression of AiSQE1 was carried out in INVSc1, *Saccharomyces cerevisiae* yeast strain. AiSQE1 positive clone of pESC-LEU and pRS315-TEF plasmids were transformed using *S.c.* EasyComp^{TM} Transformation Kit and plated on complete supplement mixture without URA (CSM-URA) and complete supplement mixture without LEU (CSM-LEU) plates, respectively. Along with this, AiSQE1 positive clone was transformed into INVSc1 cells containing AiTTS1-pYES2/CT. A single colony was inoculated into 2 % glucose containing complete supplement mixture (CSM) without respective amino acid selection marker and incubated at 30 °C. An overnight grown culture was induced by transferring it into 2 % galactose containing (CSM, 500 ml) without respective amino acid selection marker (O.D at 600 nm was maintained as 0.4) and incubated at 30 °C for 24 h. Cells were collected by centrifugation at 1500 × *g* for 10 min at 4 °C. To quantify the changes in the level of sterol, 200 µg of cholesterol (Sigma-Aldrich) was added as an internal control to each cell pellet. Saponification was done with 10 % KOH in 80 % ethanol at 70 °C for 2 h and then extracted thrice with equal volumes of n-hexane. Metabolite extracts were passed through anhydrous sodium sulphate and concentrated to 500 µL. The

extracted samples were analyzed by GC-FID (4.3.1.6-GC-FID/GC-MS Program). Quantification of metabolites was done by comparing with a standard graph of cholesterol ranging from 0.1 to 0.75 mg/mL.

4.3.3 Mutational Analysis of AITTS1

4.3.3.1 Phylogenetic Analysis

AiTTS1, AiTTS2 and triterpene synthases characterized from other organisms were considered for phylogenetic analysis. Sequences were aligned by ClustalW and neighbour joining tree was constructed with MEGA version 7.0.21 software. Confidence levels for branches were obtained by bootstrap analyses with 1000 replicates.

4.3.3.2 Site-Directed Mutagenesis

The mutation primers (Table 4.3) were designed based on phylogenetic analysis. AiTTS1 in the pYES2/CT plasmid was used as a template for site-directed mutagenesis. The mutations were carried out by using QuikChange Lightning site-directed mutagenesis kit from Agilent. The mutation reactions were transformed into XL10 gold ultra-competent cells and plated on LA containing 100 μg/mL of ampicillin and incubated overnight at 37 °C. Plasmids were isolated from the individual colonies and mutation was confirmed by analysing sequences obtained through Sanger sequencing.

The expression of mutated AiTTS1 was carried out in INVSc1, *S. cerevisiae* yeast strain. The mutated AiTTS1 plasmid was transformed using *S.c.* EasyComp^{TM} Transformation Kit and plated on complete supplement mixture without URA (CSM-URA). Expression and metabolite extraction was done as mentioned above for AiTTS1.

4.3.4 RT-PCR Analysis of AiTTS1 and AiTTS2

qPCR primers were designed for AiTTS1, AiTTS2 (Table 4.2 and 4.4) and actin (Table 4.5) as endogenous gene. RNA was isolated from neem kernel, pericarp, leaves and flowers using the Spectrum^{TM} plant total RNA isolation kit (Sigma). DNase treatment was done for all RNA samples using DNase I Amplification Grade

Kit (Sigma) followed by cDNA synthesis using SuperScript® III First-Stand Synthesis System (Invitrogen). Real-time PCR was carried out with FastStart Universal SYBR Green Master (Rox) (ROCHE, Switzerland) in AriaMx Real-time PCR System (Agilent Technologies). Quantification was performed as follows: 95 °C for 5 min, 40 cycles of 95 °C for 30 sec and 57 °C for 1 min. Actin primers were used as an endogenous control to normalize the expression levels between different tissues. Threshold (Ct) values were obtained and ΔCt was calculated as Ct target gene − Ct endogenous reference gene. Relative fold difference was calculated using $\Delta\Delta C_T$ method. Experiments were carried out using three biological replicates with five technical replicates each.

4.4 Results and Discussion

4.4.1 Homology Model of AiTTS1 and AiTTS2

AiTTS1 and AiTTS2 crystal structure are not available; hence the model was generated by using crystal structure co-ordinates from homologous proteins. Human lanosterol synthase is the only available crystal structure among the 2,3-oxidosqualene cyclases (triterpene synthases). AiTTS1 and AiTTS2 showed 36.7 % and 33.4 % identity with human lanosterol synthase (PDB ID:1W6K), respectively. Homology model of AiTTS1 and AiTTS2 were validated by PROCHEK. Ramachandran plot suggested that AiTTS1 (Figure 4.6) has 561 residues (84.5 %) were present in the most favoured region, 80 residues (12 %) were in additional allowed regions and 17 residues (2.6 %) were in generously allowed regions. Only 6 residues (0.9 %) were in disallowed regions. In homology modeled AiTTS2 (Figure 4.7), 525 residues (86.3 %) were present in most favoured region, 66 residues (10.9 %) were in additional allowed regions and 11 residues (1.8 %) were in generously allowed regions. Only 6 (1 %) residues of AiTTS2 were in disallowed regions. These results clearly indicate that AiTTS1 (Figure 4.8) and AiTTS2 (Figure 4.9) models are valid and can be used for further studies.
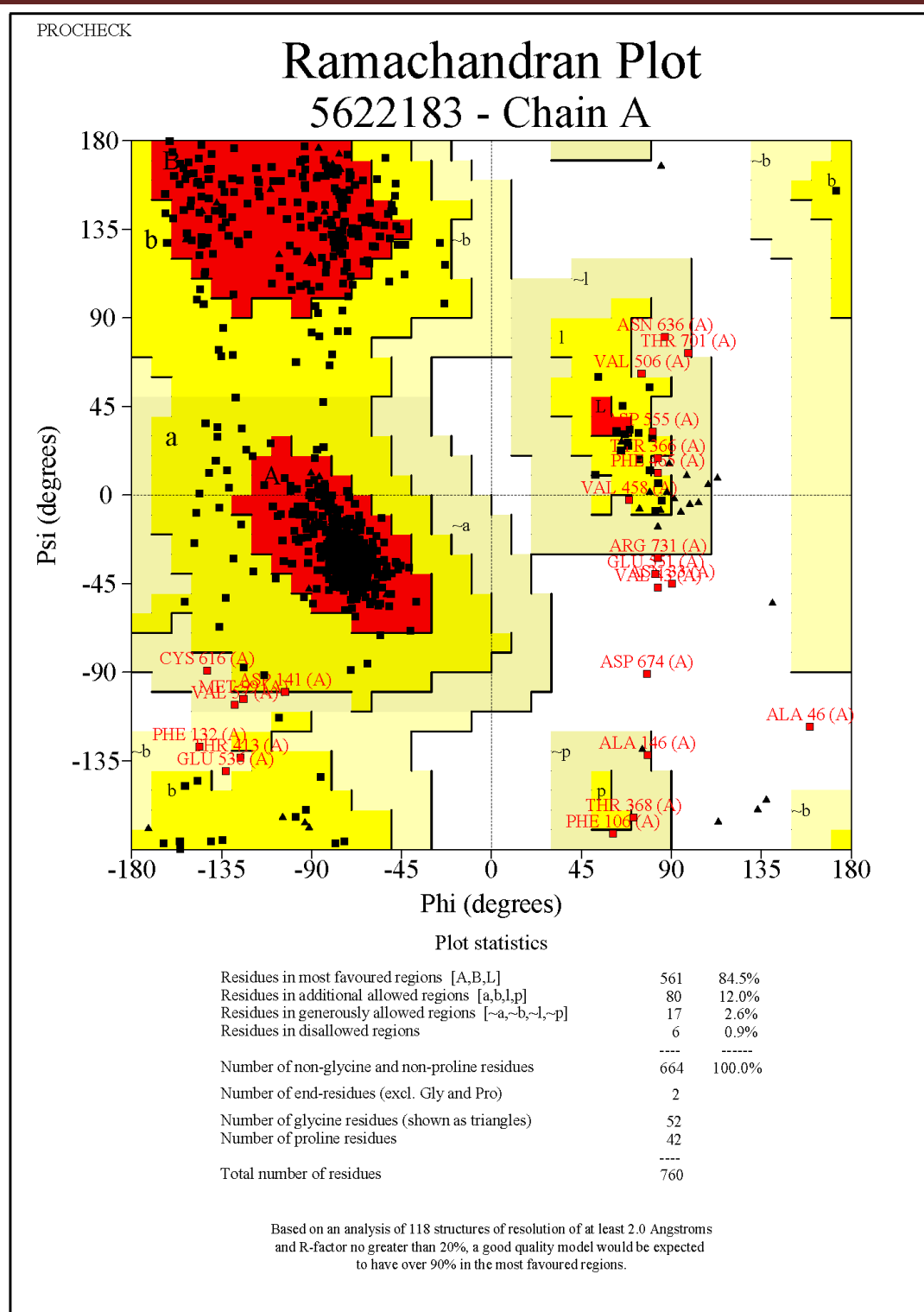
Figure 4. 6 Ramachandran Plot for AiTTS1.

PROCHECK

Ramachandran Plot
7765049 - Chain A

Plot statistics

Residues in most favoured regions [A,B,L]	525	86.3%
Residues in additional allowed regions [a,b,l,p]	66	10.9%
Residues in generously allowed regions [~a,~b,~l,~p]	11	1.8%
Residues in disallowed regions	6	1.0%
	----	------
Number of non-glycine and non-proline residues	608	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	53	
Number of proline residues	41	

Total number of residues	704	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 4. 7 Ramachandran Plot for AiTTS2.
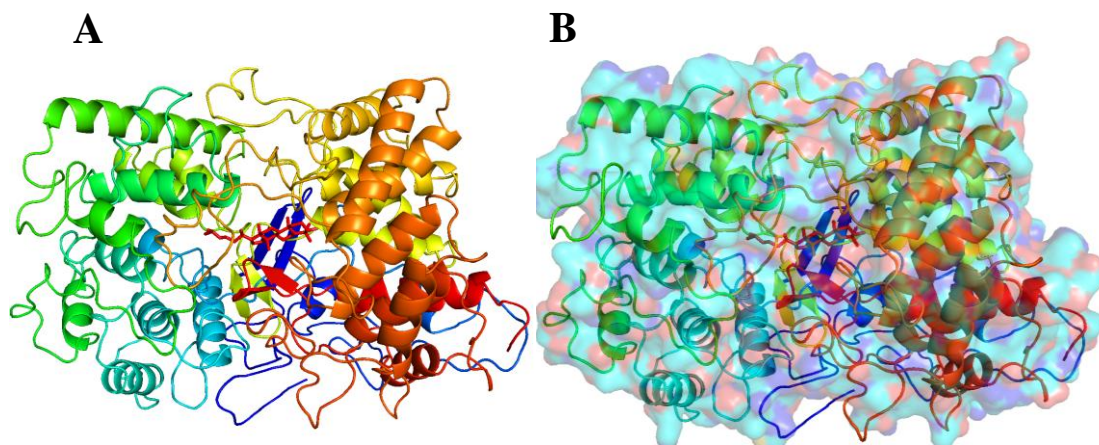
Figure 4. 8 Homology Model of AiTTS1.

A) AiTTS1 model B) Superimposed image of human lanosterol synthase (1W6K; Surface view and AiTTS1; helices and sheet view).
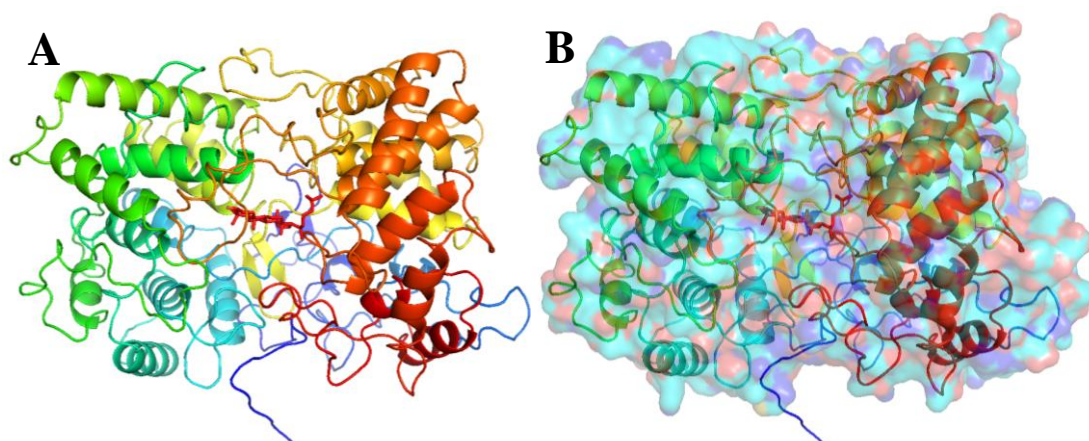
Figure 4. 9 Homology Model of AiTTS2

A) AiTTS2 model B) Superimposed image of human lanosterol synthase (1W6K; Surface view and AiTTS2; helices and sheet view).

4.4.2 Cloning and Characterization of AiTTS1

The ORF of AiTTS1 was 2,283 bp, which coded for a protein of 760 amino acids with theoretical molecular weight and calculated pI as 87.2 kDa and 6.7, respectively. AiTTS1 showed a maximum identity with several characterized TTS such as 71 % identity with β-amyrin synthase from *Betula platyphylla* [UniProt: Q8W3Z1][32], *Glycyrrhiza glabra* [UniProt: Q9MB42][33] and 70 % identity with taraxerol synthase from *Kalanchoe daigremontiana* [UniProt: E2IUA6][34]. The

multiple sequence alignment of AiTTS1 consisted of DCTAE motif and six copies of QW [(K/R)(G/A)XX(F/Y/W)(L/I/V)XXXQXXXGXW] motifs. The aspartate residue in DCTAE motif involves in the protonation of epoxide ring of 2,3-oxidosqualene in order to start a cascade cyclization. QW motifs are the structural elements present in all triterpene synthase[15] (Figure 4.11).

AiTTS1 was cloned into a pYES2/CT vector (Figure 4.10) and expressed in INVSc1 yeast strain (Figure 4.12). Crude n-hexane metabolite extracts of saponified AiTTS1 yeast cells were analyzed on GC-MS along with standard triterpenes such as tirucallol, euphol, lanosterol, ergosterol, lupeol, α-amyrin and β-amyrin. The metabolite formed by AiTTS1 showed same retention time with respect to α-amyrin and lupeol (Figure 4.13). To characterize the metabolite produced from AiTTS1, large-scale yeast expression was carried for metabolite purification.
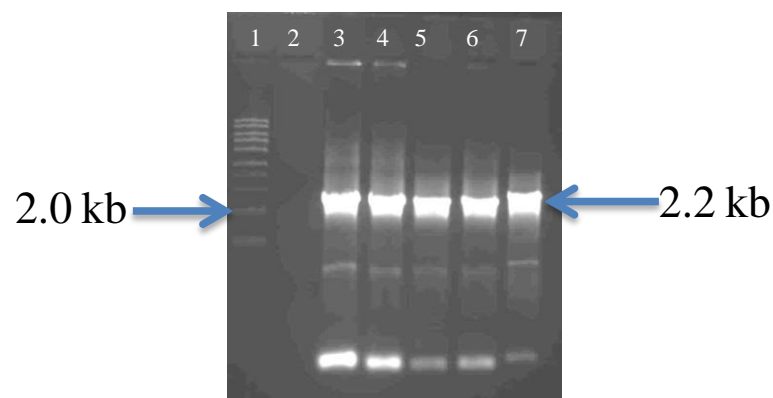
Figure 4. 10 AiTTS1 ORF PCR Amplification,

Lane 1: 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C), **Lane 2:** Negative control, **Lane 3:** AiTTS1 PCR product at 54 °C, **Lane 4:** AiTTS1 PCR product at 56 °C and **Lane 5:** AiTTS1 PCR product at 58 °C, **Lane 6:** AiTTS1 PCR product at 60 °C, **Lane 7:** AiTTS1 PCR product at 62 °C.

Figure 4. 11 Multiple Sequence Alignment of *A. indica* Triterpene synthase 1 (AiTTS1)

Amino acid sequences of AiTTS1 (*A. indica*), AiTTS2 (*A. indica*), AiCAS (*A. indica*, AGC82085), EtTTS (*Euphorbia tirucalli*, AB206469), RsTTS (*Rhizophora stylosa*,

BAF80442), KcTTS (*Kandelia candel,* BAF35580) and AaTTS (*Ailanthus altissima,* DD135972) are used for multiple sequence alignment. The highly conserved DCTAE motif is indicated in orange, QW motifs are indicated in blue and active site residues in red colour letters.
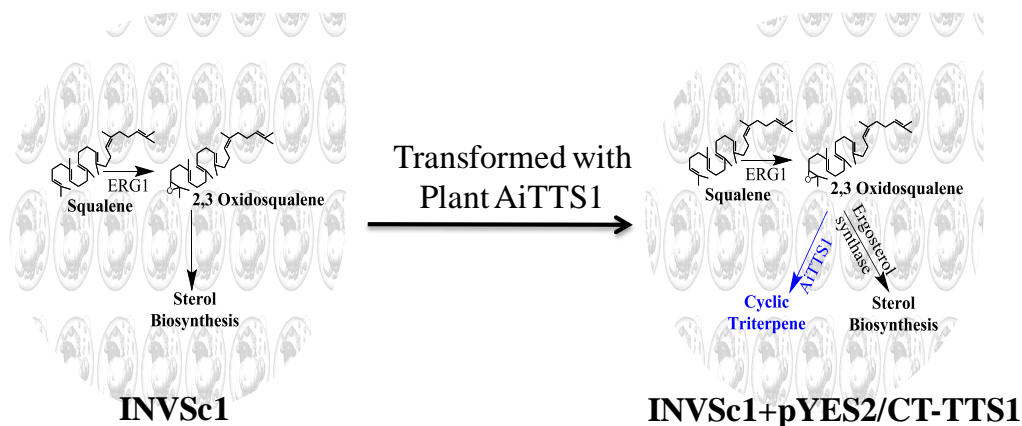
Figure 4. 12 Schematic Representation of the AiTTS1 Expression in INVSc1 Yeast Strain.
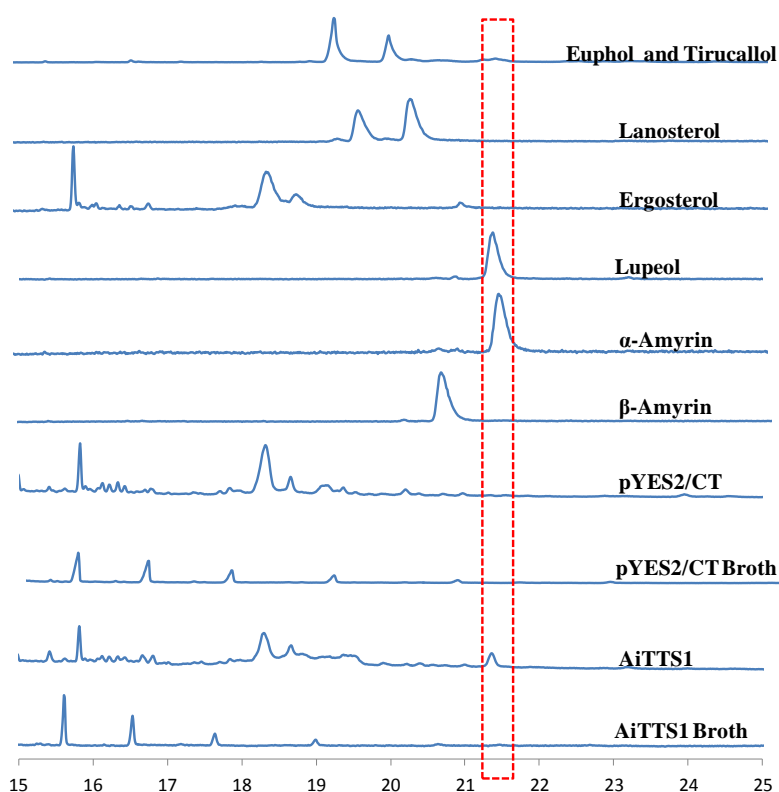
Figure 4. 13 Total Ion Chromatograms of AiTTS1-INVSc1 Metabolite Extract with Different Triterpene Standards.

To purify AiTTS1 metabolite, 25 L of yeast culture was grown in synthetic medium without uracil containing 2 % galactose. 120 mg of crude extract was

obtained from saponification of 130 g of yeast cell pellet followed by n-hexane. The crude extract was loaded onto the column (1×50 cm) containing 5% silver nitrate impregnated silica gel (230 – 400 mesh size) and dichloromethane (DCM) was used as a solvent. Pure DCM running on the column resulted in elution of lanosterol from fraction 44 to 52. Further running the column with 100 % DCM, AiTTS1 metabolite was eluted with fraction numbers 64 to 77. Based on the GC-MS analysis fractions 64 to 74 were combined and the pure AiTTS1 metabolite was around 6 mg. When the polarity was increased to 5 % by methanol, squalene (fractions 110 to 113) was eluted and finally, ergosterol (fractions 119 to 123) was eluted. All the eluted fractions were analyzed by GC-MS (Figure 4.14 and 4.15). The AiTTS1 metabolite was analyzed by GC-MS, ^{1}H-NMR, ^{13}C-NMR, DEPT, HMBC, HSQC, COSY and NOESY. The AiTTS1 product was identified as tirucalla-7,24-dien-3β-ol based on spectral data. The spectarl data matched well with that of reported for tirucalla-7,24-dien-3β-ol (Table 4.7)[14]. Two enzymes which produce tirucalla-7,24-dien-3β-ol synthases from *A. thaliana* (AtPEN3)[14] and *Ailanthus altissima* (AaTDS) were reported till now. AiTTS1shows 66 % identity with AtPEN3 which produces 85% tirucalla-7,24-dien-3β-ol and several minor products. AiTTS1 shows 85 % identity with AaTDS which produces tirucalla-7,24-dien-3β-ol.

Table 4. 6 Tirucalla-7,24-dien-3β-ol Synthases from Other Plants

Gene Name	Identity with AiTTS1	Products fromed
A. thaliana (AtPEN3)	66%	85% Tirucalla-7,24-dien-3β-ol and several minor products
Ailanthus altissima (AaTDS)	85%	Tirucalla-7,24-dien-3β-ol
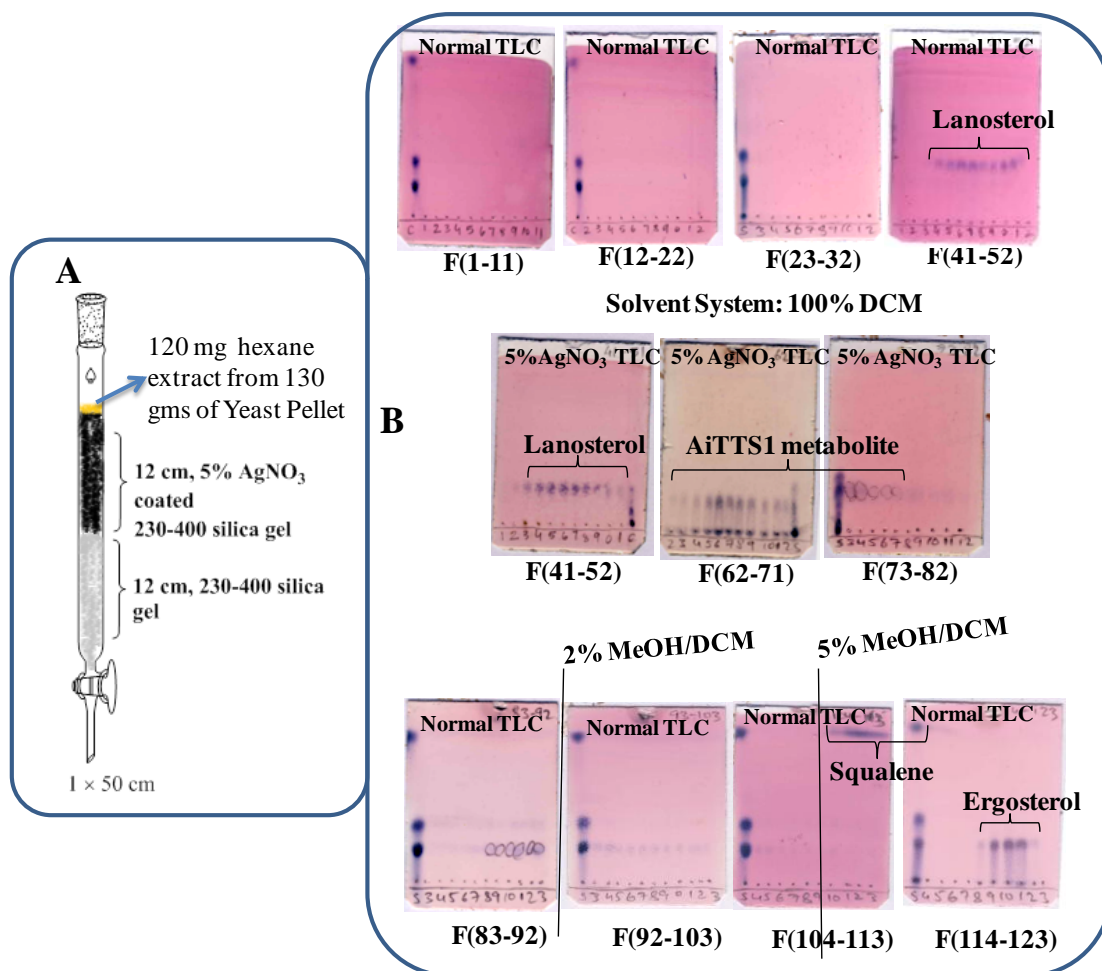
Figure 4. 14 5 % AgNO$_3$ Silica Column of AiTTS1 Metabolite Purification.

A) Schematic representation of AiTTS1 metabolite column, **B)** TLC plates of different AiTTS1 5 % AgNO$_3$ silica column fractions.
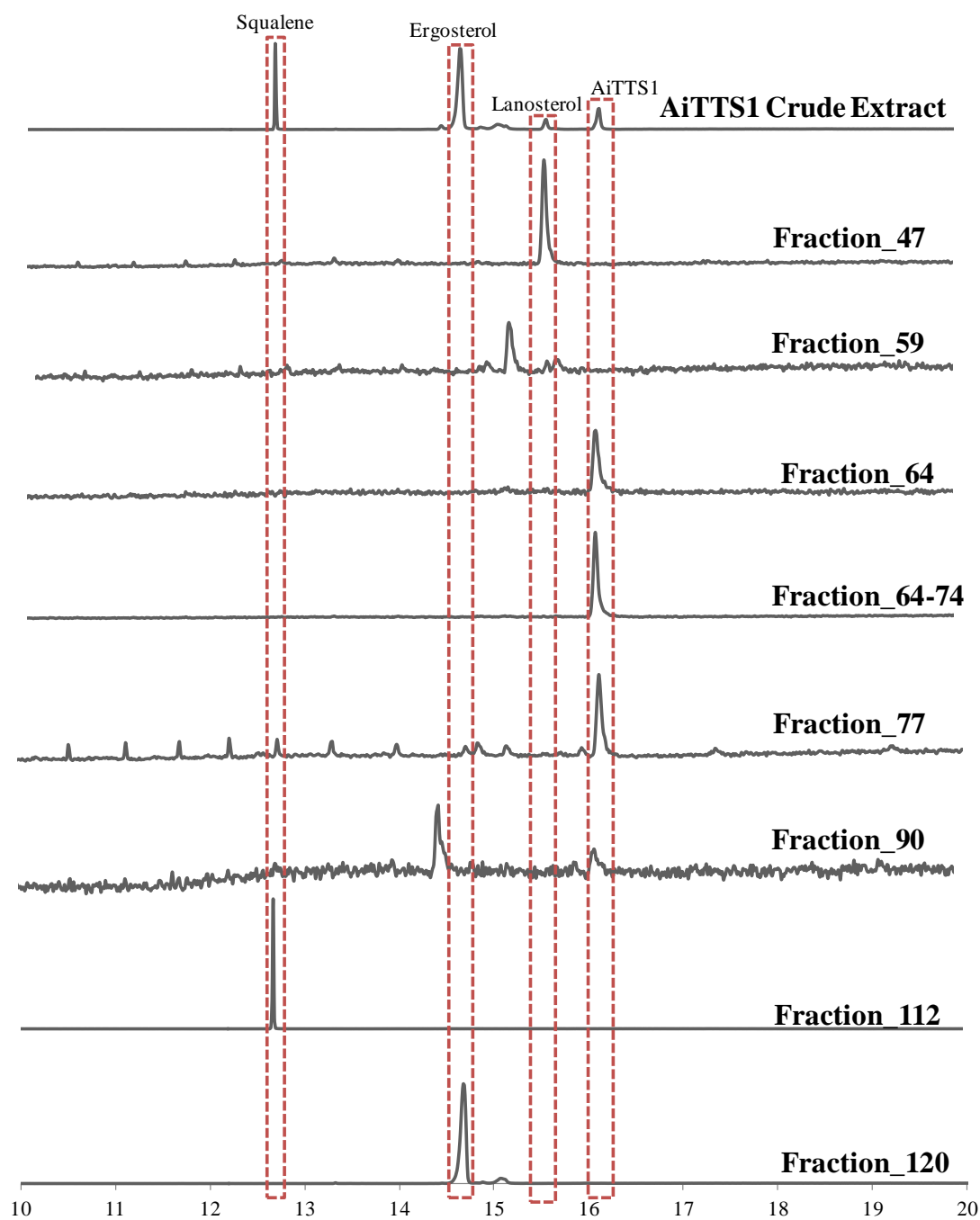
Figure 4. 15 Total Ion Chromatograms of Different 5% AgNO₃ Silica Column Fractions from AiTTS1 Yeast Extraction.

4.4.2.1 Characterization of AiTTS1 Metabolite

Table 4. 7 Comarison of NMR Chemical Shifts and Coupling Constants for Tirucalla-7,24-dien-3β-ol from AiTTS1 and AtPEN3.

	^{13}C Chemical Shifts			^{1}H Chemical Shifts			
	Tirucalla-7,24-dien-3β-ol					Tirucalla-7,24-dien-3β-ol	
	AiTTS1	AtPEN3[14]		AiTTS1		AtPEN3[14]	
Carbon	δ_C	δ_C	Proton	δ_H	Scalar ^{1}H-^{1}H couplings	δ_H	Scalar ^{1}H-^{1}H couplings
C-1	37.19	37.2	H-1α	1.12 - 1.18	m, 1 H	1.143	td, 13.3, 4.0
			H-1β	1.68	br. s., 1 H	1.681	dt, ~13, 3.5
C-2	27.67	27.69	H-2α	1.66	d, J=3.75 Hz, 1 H	1.657	dq, 13.0, 3.8
			H-2β	1.59	d, J=3.75 Hz, 1 H	1.598	dddd, 13.9, 13.0, 11.8, 3.7
C-3	79.25	79.26	H-3α	3.25	dd, J=11.41, 3.92 Hz, 1 H	3.245	ddd, 11.6, 5.7, 4.0
C-4	38.94	38.96	−				
C-5	50.6	50.62	H-5α	1.32	dd, J=12.09, 5.62 Hz, 1 H	1.317	dd, 12.1, 5.6
C-6	23.92	23.93	H-6α	2.12 - 2.17	m, 1 H	2.139	dddd, 17.8, 5.6, 4.3, 2.6
			H-6β	1.91 - 2.00	m, 2 H	1.964	dddd, 17.8, 12.1, 2.9, 1.1
C-7	117.78	117.79	H-7	5.26	dt, J=3.41 Hz, 2.95 Hz 1 H	5.256	dt, 4.3, 2.9

C-8	145.88	145.89	–				
C-9	48.93	48.96	H-9α	2.18 - 2.24	m, 1 H	2.203	dtq, ~13.4, 3.7, 2.8
C-10	34.92	34.93	–				
C-11	18.11	18.12	H-11α	1.45	dt, J=11.84, 2.26 Hz, 2 H	1.526	m
			H-11β	1.53	m, 2 H	1.489	m
C-12	33.77	33.78	H-12α	1.62 - 1.64	m, 1 H	1.619	ddd, 14.2, 10.3, 8.8
			H-12β	1.79	dd, J=13.62, 9.88 Hz, 1H	1.782	br dd, 14.2, 9.9
C-13	43.5	43.51	–				
C-14	51.13	51.14	–				
C-15	34.01	34.02	H-15α	1.53	m, 2 H	1.49	m
			H-15β	1.47	m, 1 H	1.434	ddd, 12.4, 9.3, 2.3
C-16	28.2	28.21	H-16α	1.26	br. s., 1 H	1.265	dddd, 13.7, 11.0, 8.3, 2.5
			H-16β	1.91 - 2.00	m, 2 H	1.939	dtd, 13.6, 9.4, 7.3
C-17	52.93	52.94	H-17α	1.49	d, 1 H	1.472	br q, ~9
C-18	21.89	21.9	H-18	0.82	s, 3 H	0.809	d, 0.7
C-19	13.1	13.11	H-19	0.75	s, 3 H	0.747	d, 0.8
C-20	35.95	35.96	H-20	1.39	m., 1 H	1.374	tqd, 9.1, 6.5, 2.7
C-21	18.31	18.32	H-21	0.89	d, J=6.47 Hz, 3 H	0.882	d, 6.4
C-22	36.17	36.18	H-22R	1.01 - 1.08	m, 1 H	1.033	dddd, 13.7, 10.0, 8.9, 5.0
			H-22S	1.45	dt, J=11.84, 2.26 Hz, 2 H	1.433	dddd, ~13.7, 10.4, 6.4, 2.7
C-23	25	25.01	H-23R	2.01 - 2.08	m, 1 H	2.039	m

			H-23S	1.83 - 1.90	m, 1 H	1.859	dq, ~14, 8
C-24	125.2	125.21	**H-24**	5.11	t, J=6.47 Hz, 1 H	5.099	t of septet, 7.1, 1.4
C-25	130.9 1	130.93	–				
C-26	25.71	25.72	**H-26**	1.69	s, 3 H	1.684	qd, 1.3, 0.5
C-27	17.62	17.63	**H-27**	1.61	s, 3 H	1.604	m
C-28	27.59	27.6	**H-28**	0.98	s, 6 H	0.971	d, 0.4
C-29	14.71	14.72	**H-29**	0.87	s, 3 H	0.861	s
C-30	27.25	27.26	**H-30**	0.98	s, 6 H	0.968	d, 1.2

^{1}H NMR (CDCl$_3$, 700 MHz, ppm): δ 5.26 (1H, dt, J=3.41 Hz,2.95 Hz, H-7), 5.11 (1H, t, J=6.47 Hz, H-24), 3.25 (1H, dd, J=11.41, 3.92 Hz, H-3α), 2.18-2.24 (1H, m, H-9), 2.12-2.17 (1H, m, H-6α), 2.01-2.08 (1H, m, H-23R), 1.91-2.00 (2H, m, H-6β and H-16β), 1.83-1.90 (1H, m, H-23S), 1.79 (1H, dd, J=13.62, 9.88 Hz, H-12β), 1.69 (3H, s, H-26), 1.68 (1H, br. s, H-1β), 1.66 (1H, d, J=3.75 Hz, H-2α), 1.62-1.64 (1H, m, H-12α), 1.61 (3H, s, H-27), 1.59 (1H, d, J=3.75 Hz, H-2β), 1.53 (2H, m, H-11β and H-15α), 1.49 (1H, d, H-17β), 1.47 (1H, m, H-15β), 1.45 (2H, m, H-11α and H-22S), 1.39 (1H, m, H-20α), 1.32 (1H, dd, J=12.09, 5.62 Hz, H-5α), 1.26 (1H, br. s, H-16α), 1.12-1.18 (1H, m, H-1α), 1.01-1.08 (1H, m, H-22R), 0.98 (6H, m, H-28 and H-30), 0.89 (3H, d, J=6.47 Hz, H-21), 0.87 (3H, s, H-29), 0.82 (3H, s, H-18), 0.75 (3H, s, H-19).

^{13}C NMR (CDCl$_3$, 700 MHz, ppm): δ 145.88 (C-8), 130.91 (C-25), 125.2 (C-24), 117.78 (C-7), 79.25 (C-3), 52.93 (C-17), 51.13 (C-14), 50.6 (C-5), 48.93 (C-9), 43.5 (C-13), 38.94 (C-4), 37.17 (C-1), 36.17 (C-22), 35.95 (C-20), 34.92 (C-10), 34.01 (C-15), 33.77 (C-12), 28.2 (C-16), 27.67 (C-2), 27.59 (C-28), 27.25 (C-30), 25.71 (C-26), 25 (C-23), 23.92 (C-6), 21.89 (C-18), 18.31 (C-21), 18.11 (C-11), 17.62 (C-27), 14.71 (C-29), 13.1 (C-19).
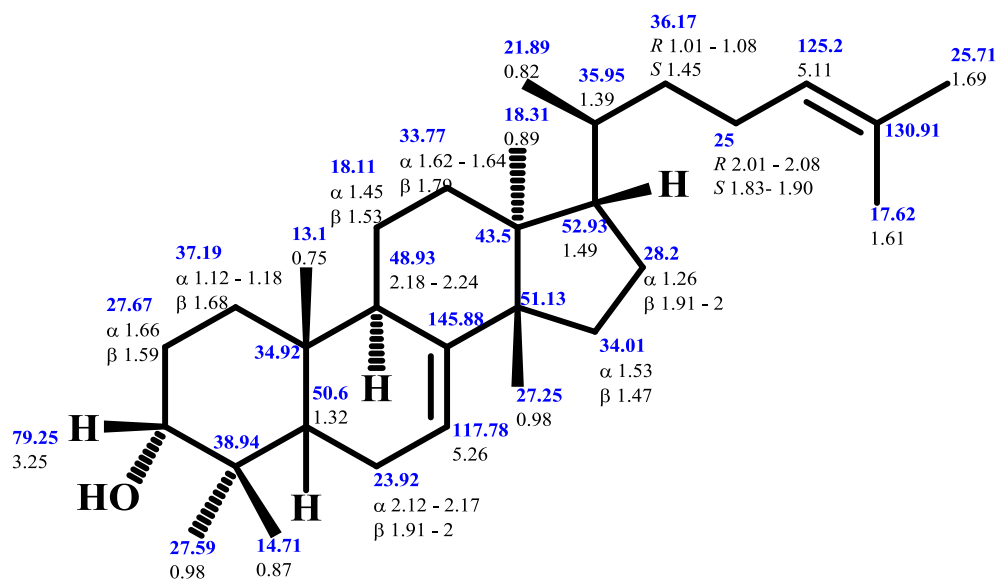
Figure 4. 16 **^{1}H and ^{13}C NMR Chemical Shifts of Tirucalla-7,24-dien-3β-ol. ^{13}Cδ are Shown in Blue colour and ^{1}Hδ are Shown in Black Colour.**
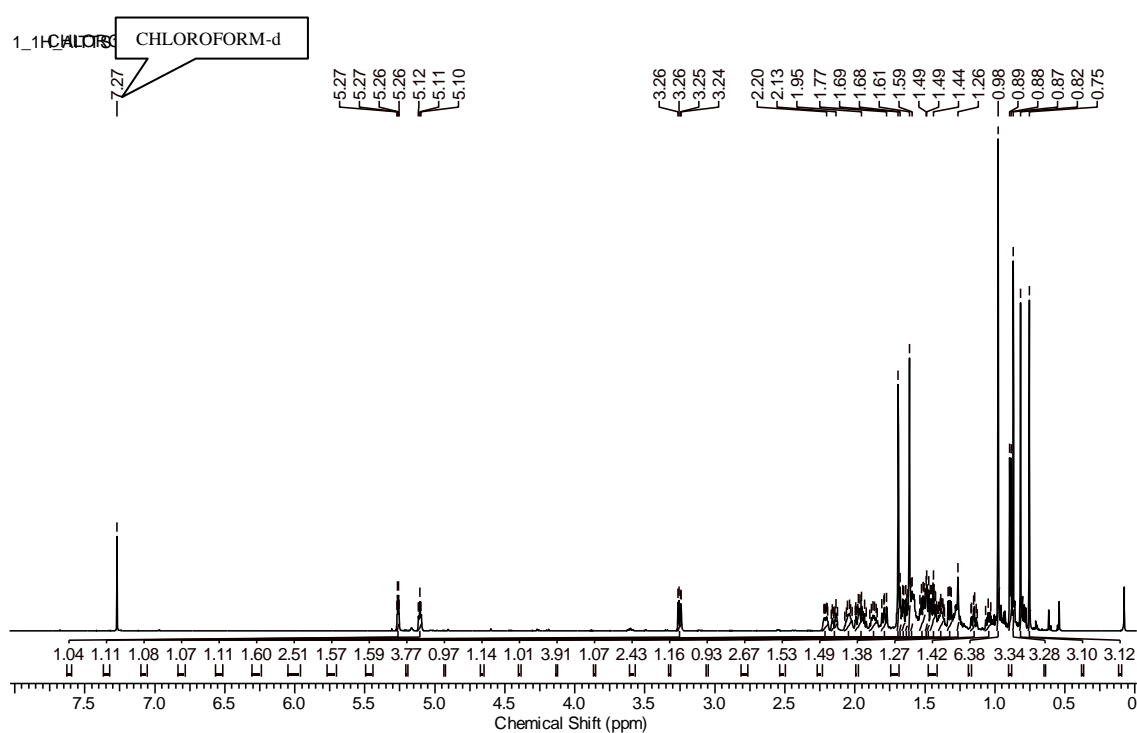
Figure 4. 17 **^{1}H NMR Spectrum of Tirucalla-7,24-dien-3β-ol.**

2_13C_AITTS1.ESP

145.88
130.91
125.20
117.78
79.25
77.18
77.00
76.82
52.93
51.13
50.60
48.93
38.94
37.19
36.17
35.95
34.01
33.77
28.20
27.67
27.59
25.71
25.00
23.92
18.31
18.11
14.71

CHLOROFORM-d

Chemical Shift (ppm)
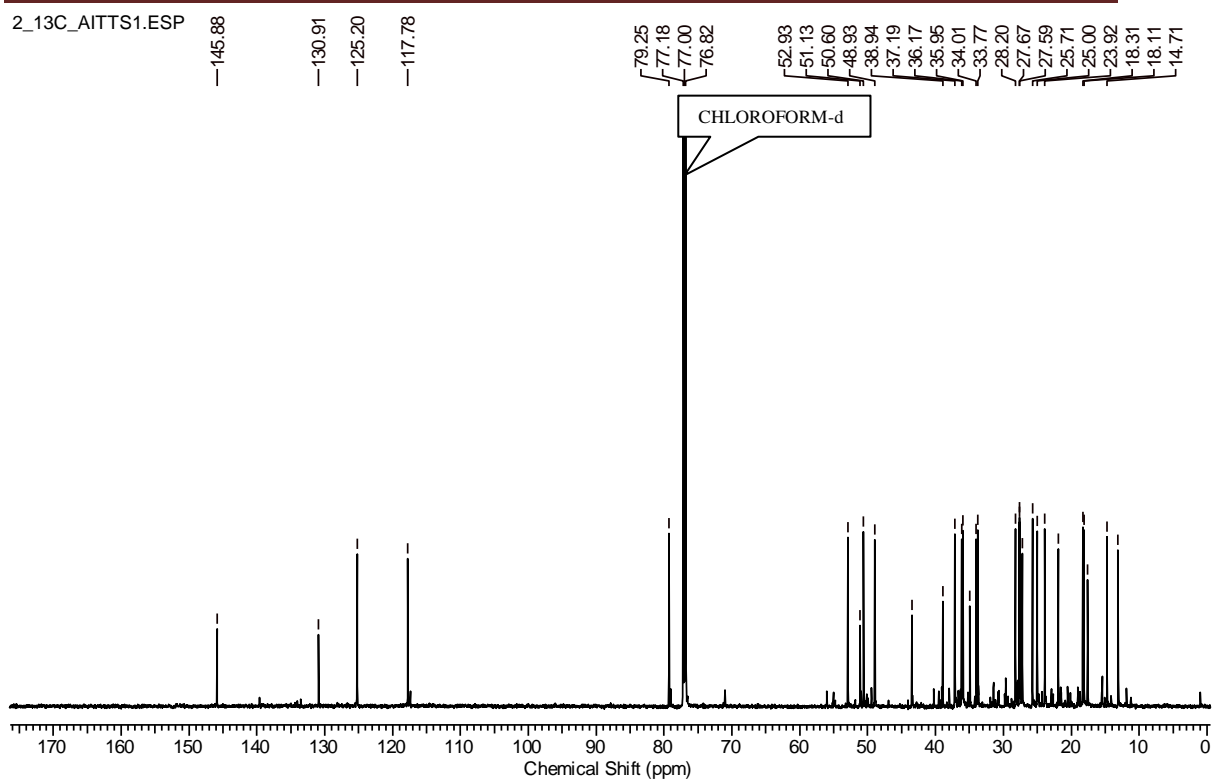
Figure 4. 18 ¹³C NMR Spectrum of Tirucalla-7,24-dien-3β-ol.

3_DEPT135_AITTS1.ESP

125.22
117.79
79.27
52.95
50.62
48.95
35.67
27.61
27.27
21.91
18.33
14.74
13.12
17.65
36.19
37.21
33.78
34.03
28.22
27.69
23.95
25.02
18.13

Chemical Shift (ppm)
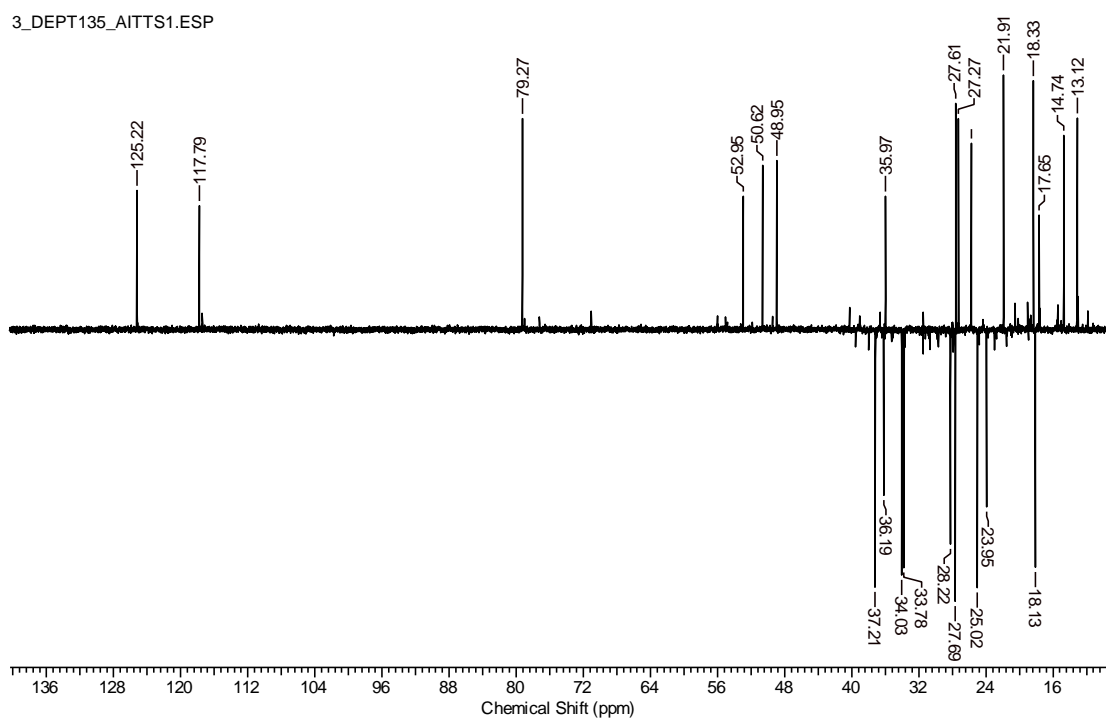
Figure 4. 19 DEPT-135 NMR Spectrum of Tirucalla-7,24-dien-3β-ol.
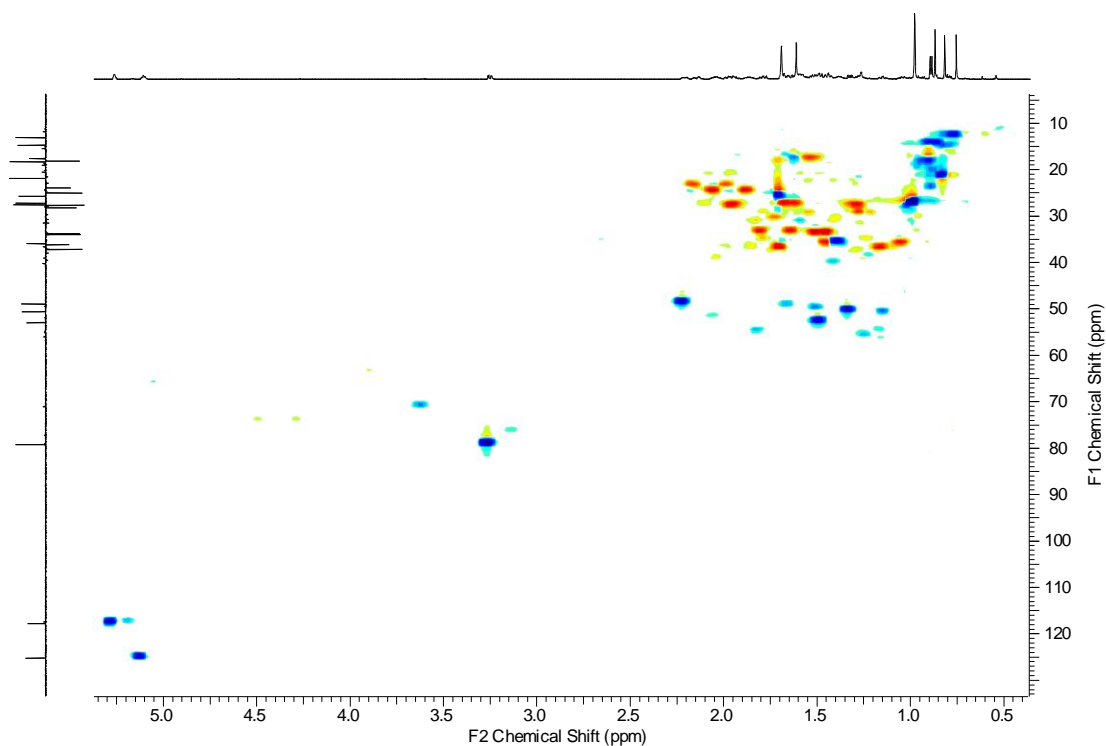
Figure 4. 20 HSQC DEPT-135 NMR Spectrum of Tirucalla-7,24-dien-3β-ol.
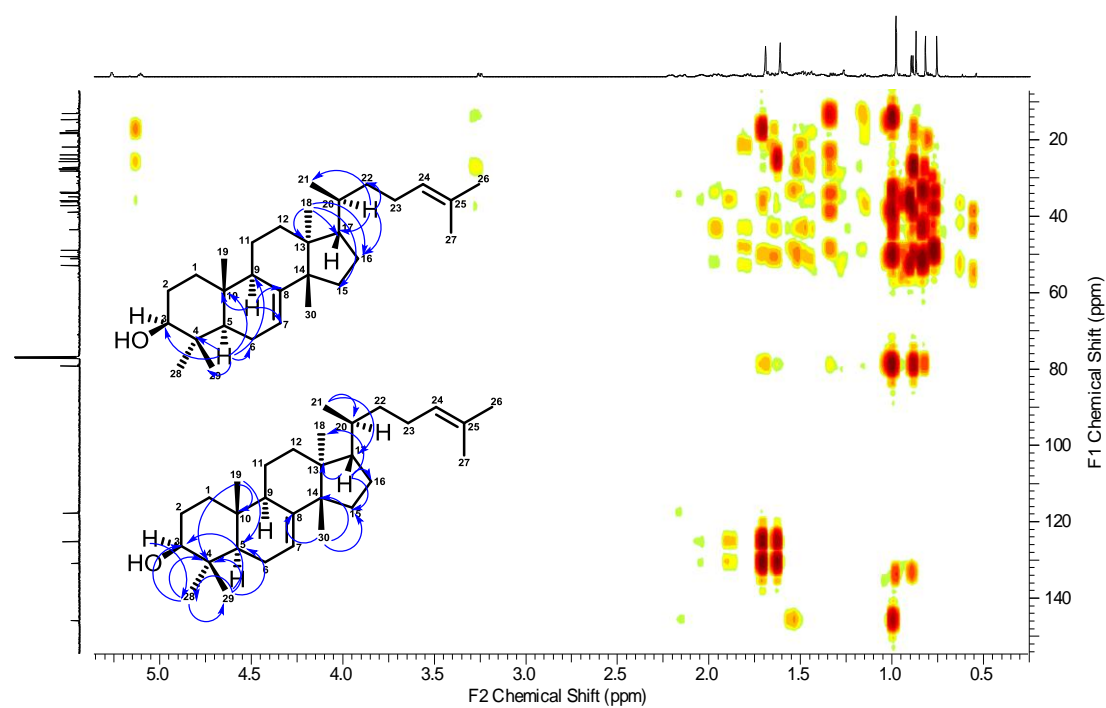
Figure 4. 21 HMBC NMR Spectrum of Tirucalla-7,24-dien-3β-ol.
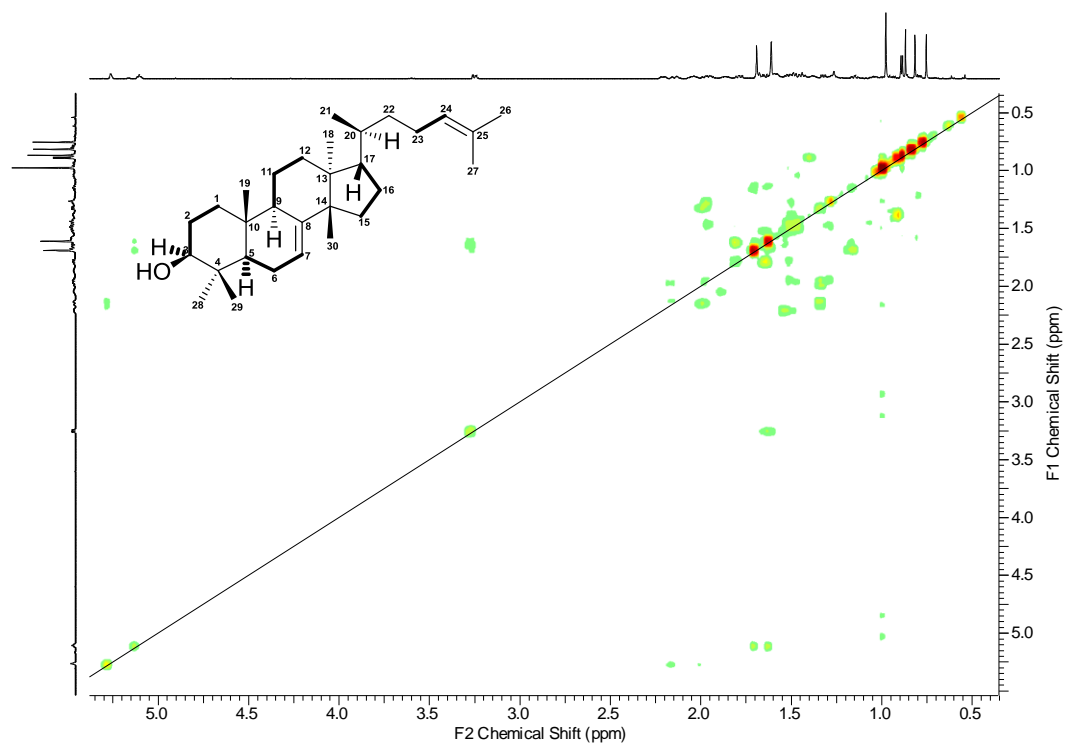
Figure 4. 22 COSY NMR Spectrum of Tirucalla-7,24-dien-3β-ol.
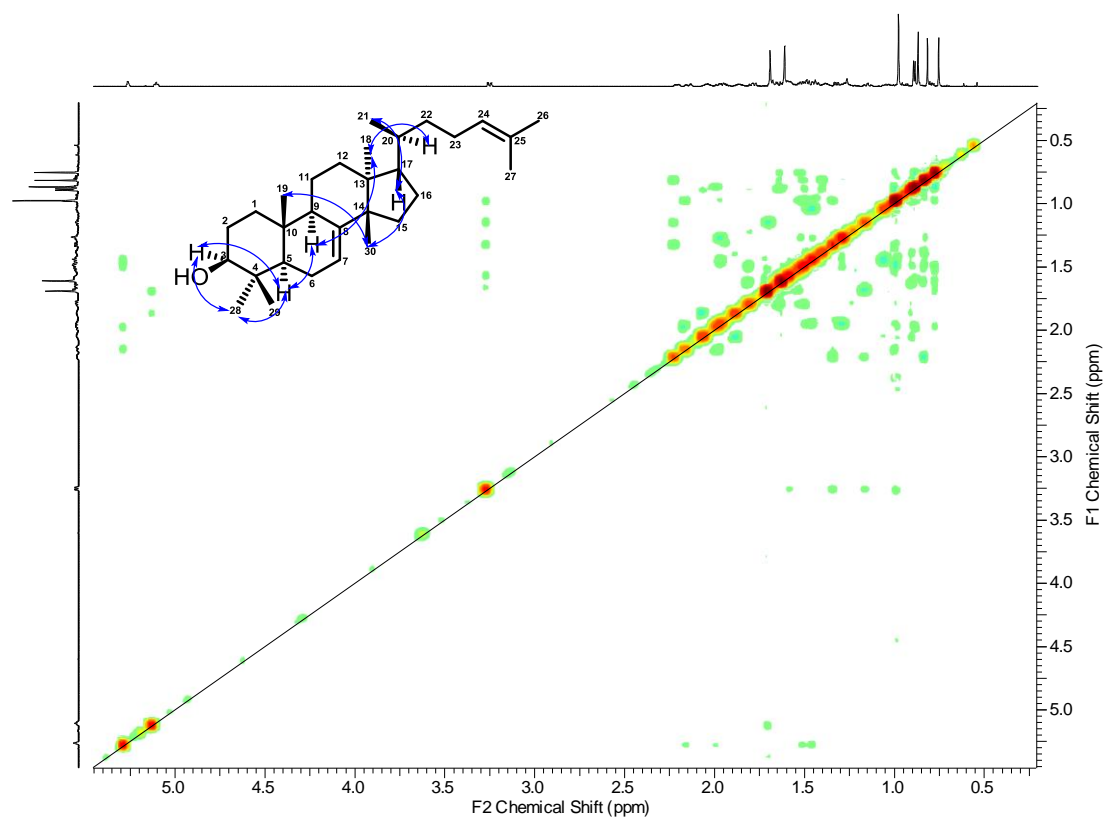
Figure 4. 23 NOESY Spectrum of Tirucalla-7,24-dien-3β-ol.

4.4.3 Cloning and Characterization of AiTTS2

AiTTS2 (Master_Control_74892) was showing 87 % match with as β-amyrin synthase from *Betula platyphylla* [UniProt: Q8W3Z1][32] and lacked 3' sequence. To obtain full-length AiTTS2, 3' RACE was performed and obtained the expected 1.2 kb fragment. AiTTS2 3' RACE fragment was cloned into pCR™-Blunt vector and sequenced with M13 forward and reverse primers. Sequence analysis of 3' RACE fragment showed 72 % match with Germanicol synthase from *Rhizophora stylosa* [UniProt: A8C980].
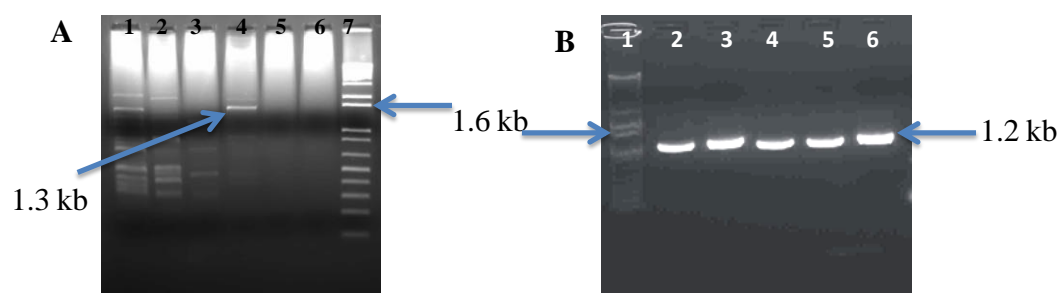
Figure 4. 24 AiTTS2 3'RACE PCR Product Amplification

A) AiTTS2_3'RACE product amplification with 3'RACE primers. **Lane 1:** AiTTS2 3'RACE PCR product at 58 °C,**Lane 2:** AiTTS2 3'RACE PCR product at 60 °C, **Lane 3:** AiTTS2 3'RACE PCR product at 62 °C, **Lane 4:** AiTTS2 3'RACE PCR product at 64 °C, **Lane 5:** AiTTS2 3'RACE PCR product at 66 °C, **Lane 6:** Negative control and **Lane 7:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C). **B)** AiTTS2_3'RACE product amplification with 3' Nested(N) RACE primers. **Lane 1:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C), **Lane 2:** AiTTS2 3'NRACE PCR product at 58 °C, **Lane 3:** AiTTS2 3'NRACE PCR product at 60 °C, **Lane 4:** AiTTS2 3'NRACE PCR product at 62 °C, **Lane 5:** AiTTS2 3'NRACE PCR product at 64 °C and **Lane 6:** AiTTS2 3'NRACE PCR product at 66 °C.

The ORF of AiTTS2 is 2,115 bp, which coded for a protein of 704 amino acids with theoretical molecular weight and calculated pI as 81.0 kDa and 6.28, respectively. AiTTS2 had a maximum identity with several characterized TTS such as 77 % identity with β-amyrin synthase from *Betula platyphylla* [UniProt: Q8W3Z1][32], germanicol synthase from *Rhizophora stylosa* [UniProt: A8C980] and 78% identity with β-amyrin synthase from *Bruguiera gymnorhiza* [UniProt: A8CDT2]. The multiple sequence alignment of AiTTS2 consists of DCTAE motif and five copies of QW [(K/R)(G/A)XX(F/Y/W)(L/I/V)XXXQXXXGXW] motifs. The aspartate residue in DCTAE motif involves in protonation of epoxide ring of 2,3-oxidosqualene in

order to start a cascade cyclization. QW motifs are the structural elements present in all triterpene synthases (Figure 4.11)[15]. Apart from one missing QW motif, the C terminal β-sheet which provide key active site residues for cyclization of 2,3-oxidosqualene were missing from AiTTS2. This clearly states that AiTTS2 is an inactive enzyme (Figure 26). To further confirm, AiTTS2 was cloned into a pYES2/CT vector (Figure 4.25) and expressed in INVSc1 yeast strain. Crude n-hexane metabolite extracts of saponified AiTTS1 yeast cells were analyzed on GC-MS. No triterpene cyclic product was observed from GC-MS data analysis (Figure 4.27). This confirms that AiTTS2 is inactive in nature.
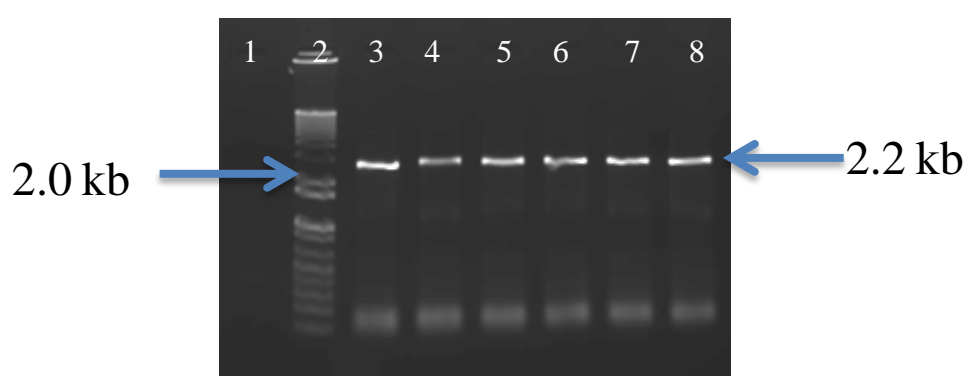
Figure 4. 25 AiTTS2 ORF PCR Amplification.

Lane 1: Negative control **Lane 2:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C), **Lane 3:** AiTTS2 PCR product at 54 °C, **Lane 4:** AiTTS2 PCR product at 56 °C, **Lane 5:** AiTTS2 PCR product at 58 °C, **Lane 6:** AiTTS2 PCR product at 60 °C,   **Lane 7:** AiTTS2 PCR product at 62 °C and **Lane 8:** AiTTS2 PCR product at 64 °C.
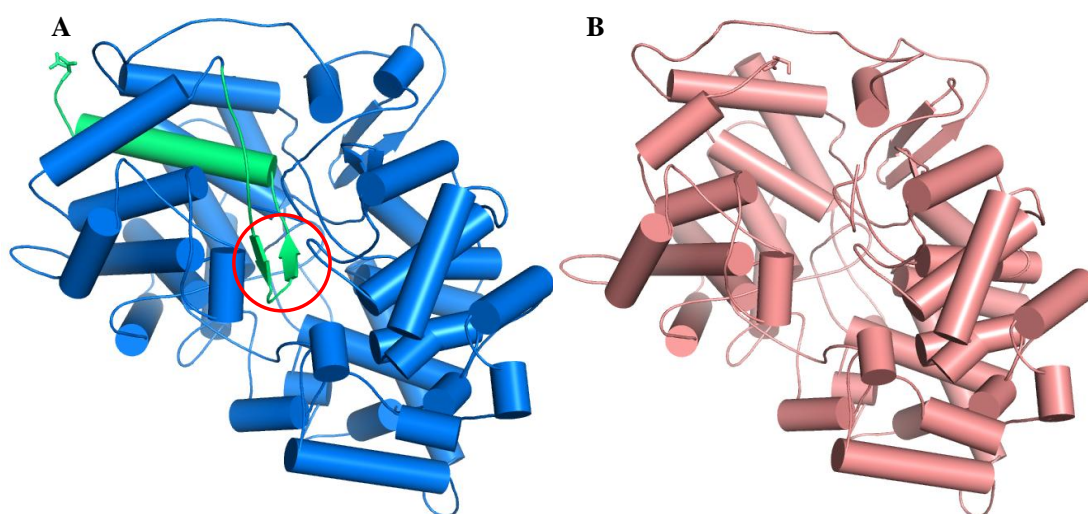
Figure 4. 26 Predicted Structures of AiTTS1 and AiTTS2.

A) Structure of AiTTS1 with C-terminal β-sheet (highlighted in the red circle), **B)** Structure of AiTTS2 which does not have C-terminal β-sheet.
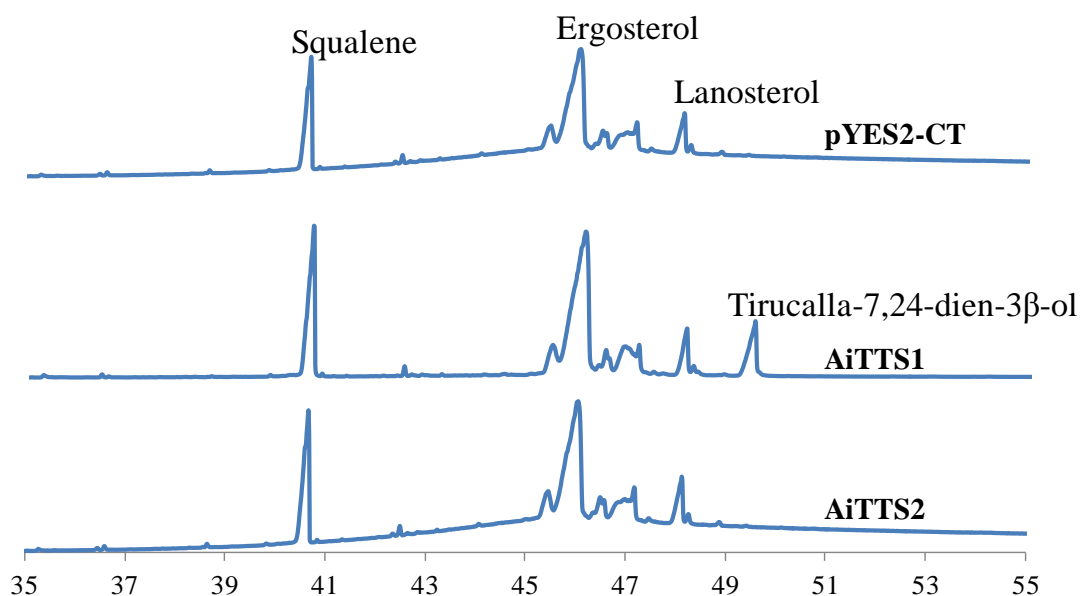
Figure 4. 27 Comparison of Total Ion Chromatograms of AiTTS1-INVSc1, AiTTS2-INVSc1 Metabolite Extract with that of the Control Vector.

4.4.4 Cloning and Characterization of AiSQE1

The ORF of AiSQE1 [GenBank: JX997152] was 1,593 bp, which codes for a protein of 530 amino acids with theoretical molecular weight and calculated pI as 57.8 kDa and 6.28, respectively. AiSQE showed a maximum identity with several characterized SQEs such as 78 % identity to SQE from *Arabidopsis thaliana* [UniProt: O81000][11], 46 % identity to SQE from *Homo sapiens* [UniProt: Q14534][27] and 37 % identity to SQE from *Candida albicans* SC5314 [UniProt: Q92206][28]. The multiple sequence alignment of AiSQE consisted of Rossman fold GXGXXG motif, DG and GD motifs, as present in fungus and vertebrates. The GXGXXG motif binds with dinucleotide of a FAD, DG motif interacts with diphosphate group of FAD and GD motif may interact with ribityl moiety of a FAD. Photoaffinity labeling[29,30] and site-directed mutagenesis[31] experiments rat and human SQE helped in identification of active site residues and these were observed in AiSQE (Figure 4.28).
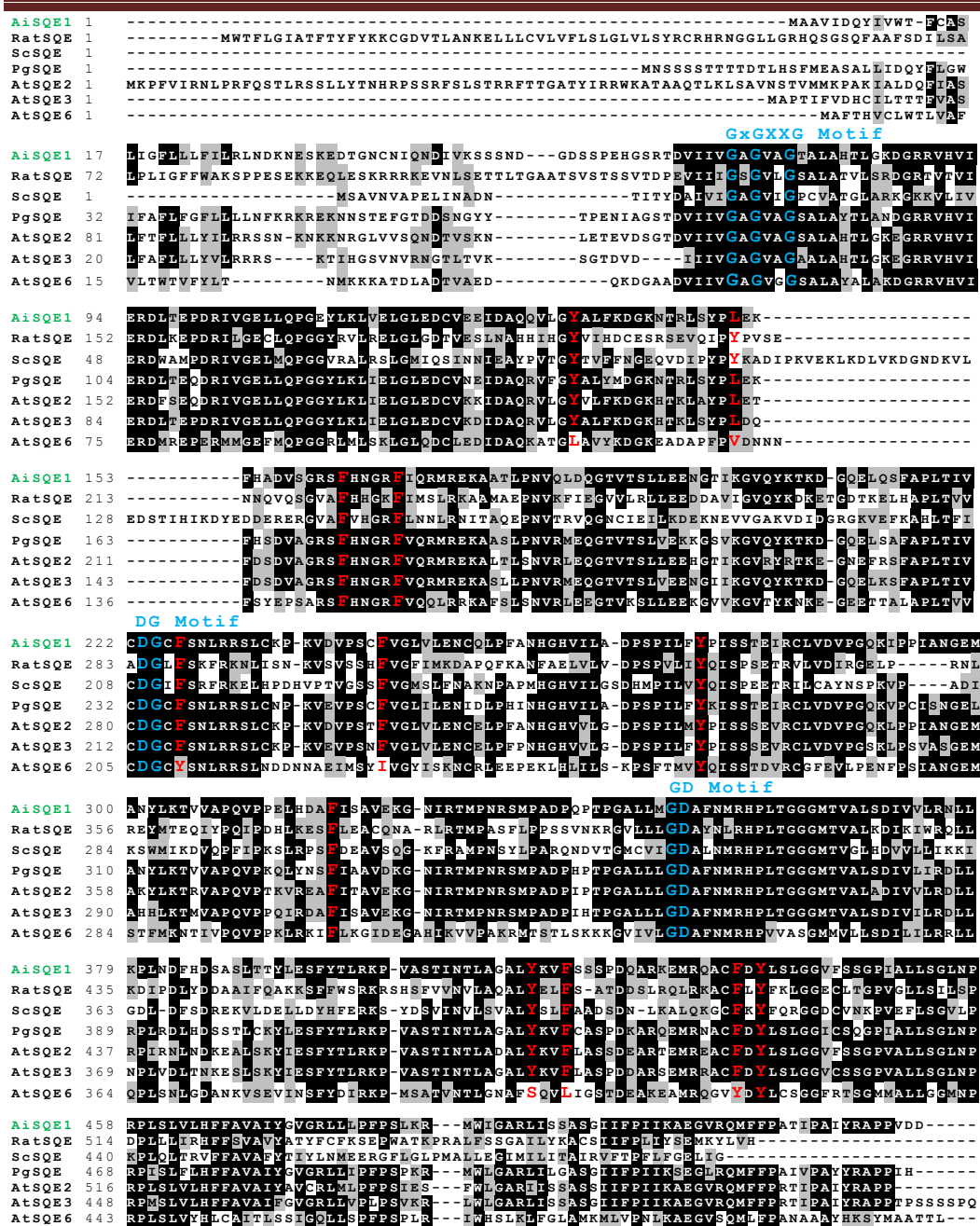
Figure 4. 28 Multiple Sequence Alignment of *A. indica* Squalene Epoxidase 1 (AiSQE1)

Amino acid sequences of AiSQE1 (*A. indica*, AGC82087), RatSQE (*Rattus norvegicus*, NP_058832), ScSQE (*S. cerevisiae*, NP_011691), PgSQE(*Panax ginseng*, O48651), AtSQE2 (*A. thaliana,* NP_179868), AtSQE3 (*A. thaliana,* NP_568033) and AtSQE6 (*A. thaliana,* NP_197804) are used for multiple sequence alignment. The highly conserved GXGXXG motif, DG and GD motifs are indicated in blue colour and active site residues are in red colour letters**.**

The full-length ORF of AiSQE1 was cloned into a pESC-LEU expression vector. Further, a truncated[37] ΔAiSQE1 without N-terminal putative membrane-

binding domain was cloned into a pRS315-TEF vector. Both the constructs were expressed in INVSc1 yeast strain. Saponification of expressed INVSc1 cells was done to extract metabolites such as sterol and other triterpenes.
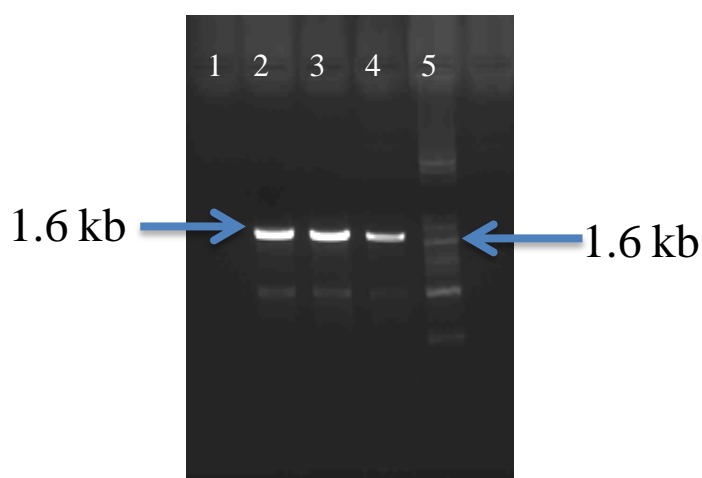
Figure 4. 29 AiSQE1ORF PCR Amplification

Lane 1: Negative control, **Lane 2:** AiSQE1 PCR product at 54 °C, **Lane 3:** AiSQE1 PCR product at 56 °C, **Lane 4:** AiSQE1 PCR product at 58 °C and **Lane 5:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C).
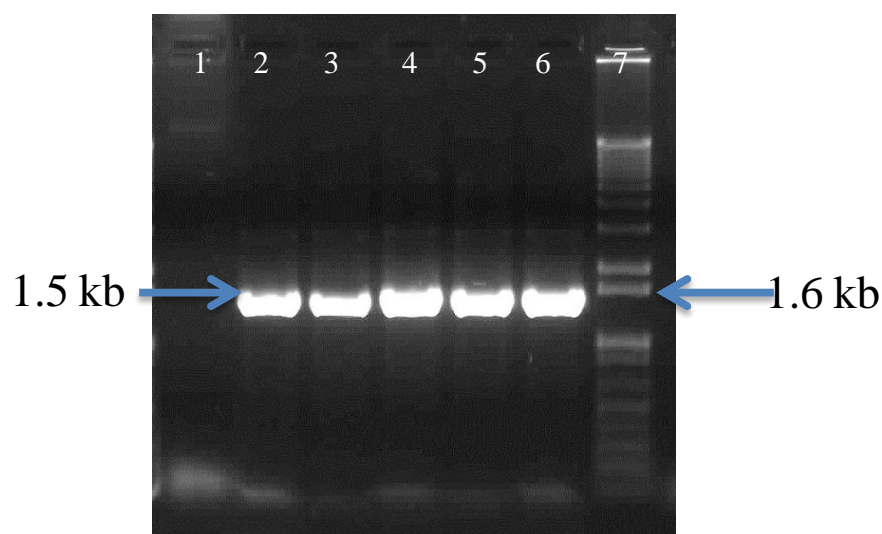
Figure 4. 30 AiSQE1 Truncated PCR Amplification

Lane 1: Negative control, **Lane 2:** AiSQE1 PCR product at 50 °C, **Lane 3:** AiSQE1 PCR product at 52 °C, **Lane 4:** AiSQE1 PCR product at 54 °C, **Lane 5:** AiSQE1 PCR product at 56 °C, Lane **6:** AiSQE1 PCR product at 58 °C and **Lane 7:** 1 kb Plus DNA ladder Invitrogen (Addendum Figure A1.C).

The relative abundance of sterols including the enzymatic product was estimated by comparing with a standard graph drawn for the internal standard, cholesterol (Figure 4.31). Expression of AiSQE1 (pESC-LEU-AiSQE1) under Gal promoter showed a 45 % increase in lanosterol production as compared to vector control. While co-expressing with pYES2/CT-AiTTS1, there was a two-fold increase in tirucalla-7,24-dien-3β-ol production and 23 % reduction in ergosterol production (Figure 4.32 and 4.33). These results indicate that AiSQE1 converts squalene into 2,3-oxidosqualene. Furthermore, it is clear that co expression of AiTTS1 and AiSQE1 in yeast resulted in production of tirucalla-7,24-dien-3β-ol around 1 mg/g of the biomass.
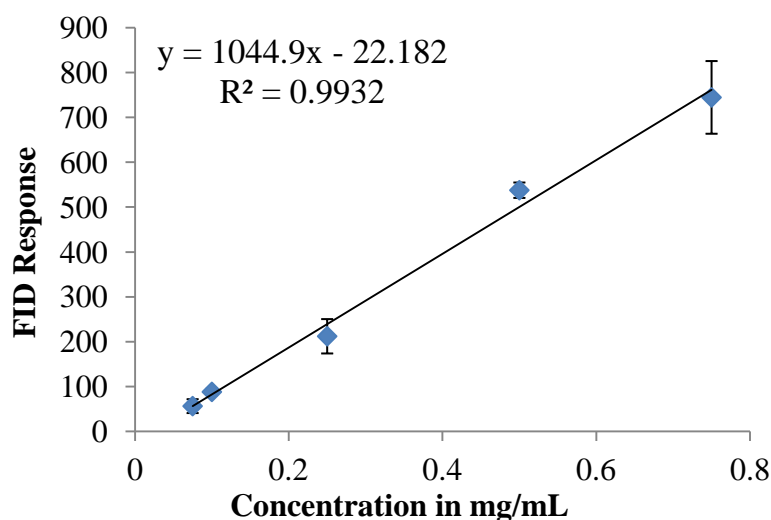
Figure 4. 31 Standard Graph of Cholesterol Obtained from GC-FID at Injection Volume of 1 μL.
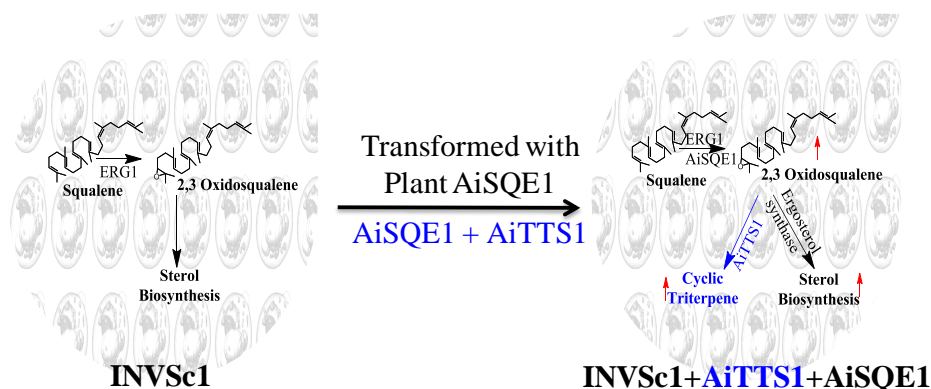
Figure 4. 32 Schematic Representation of the AiSQE1 Expression in INVSc1 Yeast Strain.
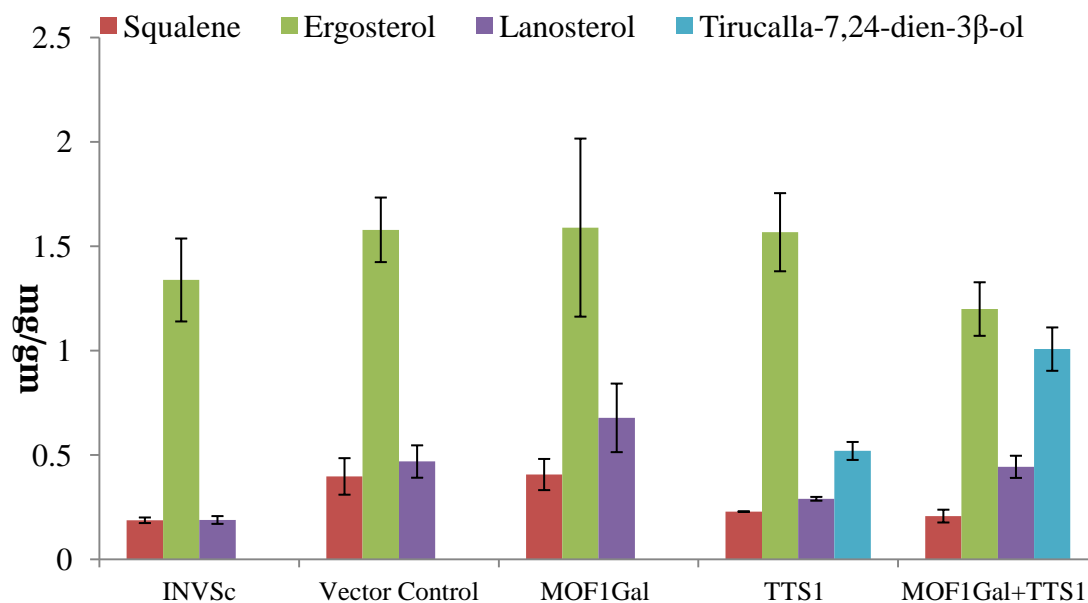
Figure 4. 33 Relative Fold Change of Squalene, Ergosterol, Lanosterol and Tirucalla-7,24-dien-3β-ol in AiSQE1 Expression in INVSc1.

4.4.5 Mutational Analysis of AiTTS1

4.4.5.1 Phylogenetic Analysis

Triterpenes are one of the largest and structurally diverse groups derived from squalene or 2,3-oxidosqualene. In general, squalene or 2,3-oxidosqualene undergoes cationic attack, followed by a cascade of cation-olefine cyclization to generate cyclic carbocations which can further modify into triterpene cyclic product. The enzymes which catalyze these reactions are known as triterpene synthases. Squalene or 2,3-oxidosqualene first gain a predefined structure, then undergoes cation attack, to initiate cascade cyclization in order to generate carbocation. These carbocations are stabilized by π interactions from active site residues. Based on the type of carbocation stabilized in the active site, respective type of triterpene cyclic products are synthesized. For example, the protosteryl C-20 cation is stabilized in the active site, which leads to the synthesis of lanosterol and cycloartenol. Dammarenyl C-20 cation stabilization leads to the synthesis of parkeol, euphol, tirucallol, butyrospermol and tirucalla-7,24-dien-3β-ol. Baccharenyl C-18 cation stabilization leads to shionone and baccharis oxide. Lupanyl or lupyl C-20 cation stabilization leads to lupeol.

Germanicyl C-19 cation stabilization leads to the synthesis of amyrins, taraxerol and friedelin (Figure 4.1 - 4.4)[16,35].

More than 50 triterpene synthases have been isolated and characterized. To understand the evolution of triterpene synthases towards stabilizing different cations, a phylogenetic analysis was carried out (Figure 4.34). This analysis revealed that tetracyclic carbocations stabilizing triterpene synthases evolved to stabilize the pentacyclic carbocations. In brief, the protosteryl C-20 cation stabilizing enzymes (lanosterol synthase and cycloartenol synthase) evolved to stabilize dammarenyl C-20 cation as performed by tirucalla-7,24-dien-3β-ol synthase and parkeol synthase. These dammarenyl C-20 cation stabilizing enzymes are evolved into multi-product forming enzymes, which can stabilize dammarenyl C-20, lupanyl C-20 and germanicyl C-19 cation. Further multiproduct forming triterpene synthases have evolved into enzymes which stabilize germanicyl C-19 cation such as amyrins, taraxerol and friedelin synthases. The aromatic amino acids play an important role in carbocation stabilization through π interactions. During the course of an evolution, increase of aromatic amino acids is seen in the active site such that pentacyclic carbocation will be stabilized (Figure 4.35). Eight active site amino acids of AiTTS1 (Y155, F260, T 413, V484, V534, V550, L553 and L556) showed a high variation in comparison with β-amyrin synthases.
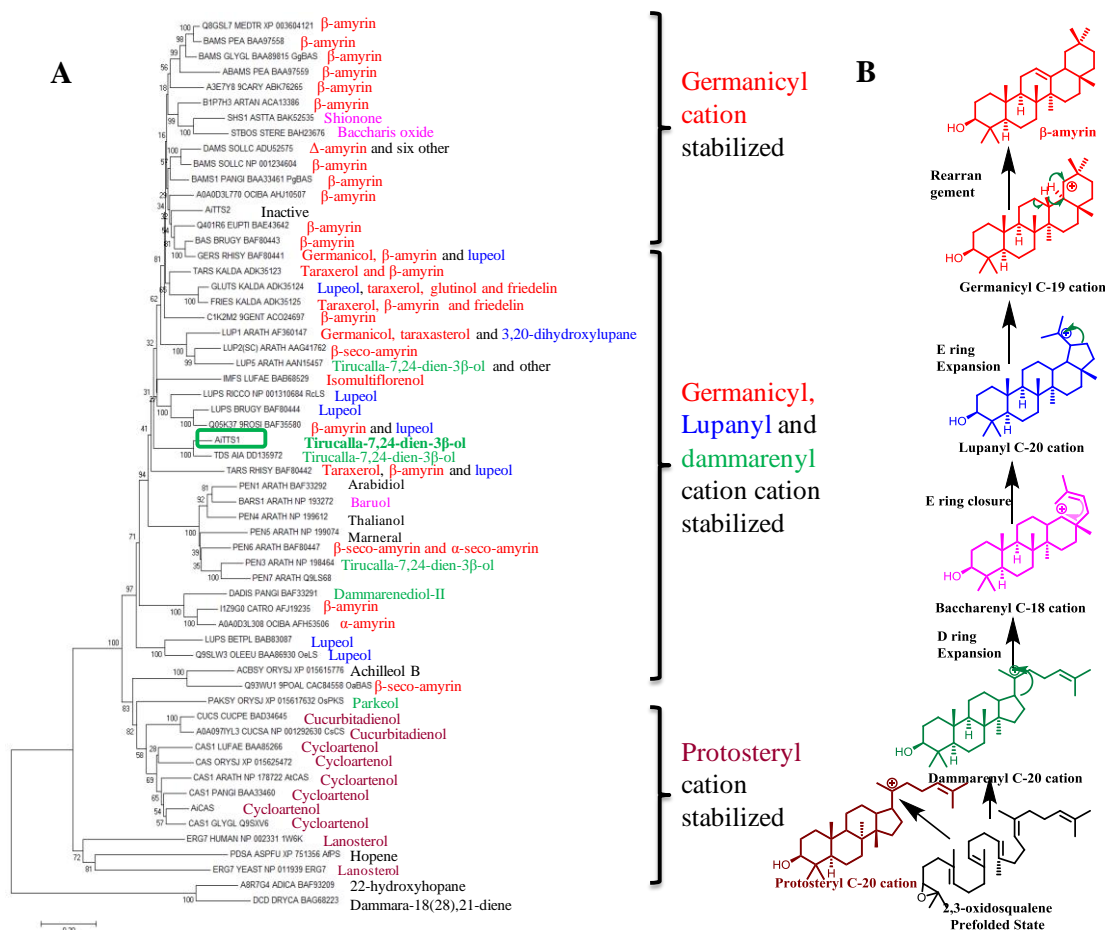
Figure 4. 34 Phylogenetic Analysis of Characterized Triterpene Synthases with Neem Triterpene Synthases.

A) Phylogenetic analysis showing the evolution of triterpene synthases to stabilize different carbocations, **B)** Schematic representation of different carbocations involved in cyclization of 2,3-oxidosqualene

| Gene Name & ID	Active Site Amino Acid								Product
Q8GSL7_MEDTR_XP_003604121	F	Y	S	V	A	T	F	I	β-amyrin
BAMS_PEA_BAA97558	F	Y	S	V	A	T	F	I	β-amyrin
BAMS_GLYGL_BAA89815_GgBAS	F	Y	S	V	A	T	F	I	β-amyrin
ABAMS_PEA_BAA97559	F	Y	S	V	A	I	L	I	β-amyrin
A3E7Y8_9CARY_ABK76265	F	Y	S	V	A	T	F	I	β-amyrin
B1P7H3_ARTAN_ACA13386	Y	Y	S	V	A	T	F	I	β-amyrin
SHS1_ASTTA_BAK52535	F	Y	T	L	A	I	I	I	Shionone
STBOS_STERE_BAH23676	F	Y	S	T	G	S	F	I	Baccharis oxide
DAMS_SOLLC_ADU52575	F	Y	G	V	A	T	F	I	Δ-amyrin and six other
BAMS_SOLLC_NP_001234604	F	Y	S	V	A	T	F	I	β-amyrin
BAMS1_PANGI_BAA33461_PgBAS	F	Y	S	V	A	T	F	I	β-amyrin
A0A0D3L770_OCIBA_AHJ10507	F	Y	S	V	A	T	F	I	β-amyrin
AiTTS2	F	Y	S	V	A	T	F	I	
Q401R6_EUPTI_BAE43642	F	Y	S	V	A	T	F	I	β-amyrin
BAS_BRUGY_BAF80443	F	Y	S	V	A	T	F	I	β-amyrin
GERS_RHISY_BAF80441	F	Y	S	V	A	T	F	I	Germanicol, β-amyrin and lupeol
TARS_KALDA_ADK35123	F	Y	S	V	A	T	F	I	Taraxerol and β-amyrin
GLUTS_KALDA_ADK35124	F	Y	S	L	G	T	F	I	Lupeol, taraxerol, glutinol and friedelin
FRIES_KALDA_ADK35125	F	Y	S	L	G	T	F	I	Taraxerol, β-amyrin and friedelin
C1K2M2_9GENT_ACO24697	F	Y	S	V	G	T	F	L	β-amyrin
LUP1_ARATH_AF360147	F	Y	S	V	A	T	M	T	Germanicol, taraxasterol and 3,20-dihydroxylupane
LUP2(SC)_ARATH_AAG41762	F	Y	S	V	A	T	F	V	β-seco-amyrin
LUP5_ARATH_AAN15457	C	Y	C	V	A	T	F	A	Tirucalla-7,24-dien-3β-ol and other
IMFS_LUFAE_BAB68529	F	Y	S	V	A	V	L	L	Isomultiflorenol
LUPS_RICCO_NP_001310684_RcLS	F	Y	S	V	A	V	M	L	Lupeol
LUPS_BRUGY_BAF80444	F	Y	S	V	A	V	L	L	Lupeol
Q05K37_9ROSI_BAF35580	F	Y	S	V	A	V	L	L	β-amyrin and lupeol
AiTTS1	Y 125	F 260	T 413	V 484	V 534	V 550	L 553	L 556	Tirucalla-7,24-dien-3β-ol
TDS_AIA_DD135972	Y	Y	T	V	V	V	L	L	Tirucalla-7,24-dien-3β-ol
TARS_RHISY_BAF80442	F	Y	G	A	C	L	F	I	Taraxerol, β-amyrin and lupeol
PEN1_ARATH_BAF33292	F	Y	S	V	A	V	V	T	Arabidiol
BARS1_ARATH_NP_193272	F	Y	S	V	A	V	I	A	Baruol
PEN4_ARATH_NP_199612	F	Y	S	V	A	V	L	T	Thalianol
PEN5_ARATH_NP_199074	F	Y	L	I	V	V	V	T	Marneral
PEN6_ARATH_BAF80447	F	Y	S	V	V	V	M	T	β-seco-amyrin and α-seco-amyrin
PEN3_ARATH_NP_198464	F	Y	T	V	I	I	I	T	Tirucalla-7,24-dien-3β-ol
PEN7_ARATH_Q9LS68	F	Y	S	A	V	I	M	T	
DADIS_PANGI_BAF33291	L	Y	S	V	V	S	F	I	Dammarenediol-II
I1Z9G0_CATRO_AFJ19235	F	Y	S	V	I	S	F	I	β-amyrin
A0A0D3L308_OCIBA_AFH53506	F	Y	S	V	I	S	F	I	α-amyrin
LUPS_BETPL_BAB83087	F	Y	S	V	A	T	F	T	Lupeol
Q9SLW3_OLEEU_BAA86930_OeLS	F	Y	S	V	A	T	F	V	Lupeol
ACBSY_ORYSJ_XP_015615776	F	F	V	V	S	S	F	I	Achilleol B
Q93WU1_9POAL_CAC84558_OaBAS	F	F	V	V	T	S	F	I	β-seco-amyrin
PAKSY_ORYSJ_XP_015617632_OsPKS	F	Y	I	V	A	T	F	V	Parkeol
CUCS_CUCPE_BAD34645	F	H	G	I	S	A	F	I	Cucurbitadienol
A0A097IYL3_CUCSA_NP_001292630_CsCS	F	H	G	I	S	A	F	I	Cucurbitadienol
CAS1_LUFAE_BAA85266	F	H	G	I	T	A	F	I	Cycloartenol
CAS_ORYSJ_XP_015625472	F	H	G	I	T	A	F	I	Cycloartenol
CAS1_ARATH_NP_178722_AtCAS	F	H	G	I	T	A	F	I	Cycloartenol
CAS1_PANGI_BAA33460	F	H	G	I	T	A	F	I	Cycloartenol
AiCAS	F	H	G	I	T	A	F	I	Cycloartenol
CAS1_GLYGL_Q9SXV6	F	H	G	I	T	A	F	I	Cycloartenol
ERG7_HUMAN_NP_002331_1W6K	F	H	G	V	T	S	F	I	Lanosterol
PDSA_ASPFU_XP_751356_AfPS	F	H	G	V	A	T	Y	V	Hopene
ERG7_YEAST_NP_011939_ERG7	F	H	G	V	T	A	F	I	Lanosterol
A8R7G4_ADICA_BAF93209	T	W	S	D	A	G	K	I	22-hydroxyhopane
DCD_DRYCA_BAG68223	T	W	S	D	A	S	K	V	Dammara-18(28),21-diene

Figure 4. 35 Amino Acids Which Showed High Variation in AiTTS1 as Compared to β-amyrin Synthase.

4.4.5.2 Mechanism of Action of AiTTS1

Herewith, to study its mechanism of action of the characterized AiTTS1, it was compared with human lanosterol synthase (HLS)[35]: Change in the confirmation of the residues Tyr 237, Cys 233 and Ile 524 or rearrangement of strained loops 516-524 and 697-699 in HLS allows 2,3-oxidosqualene to enter the active site pocket. Similar to HLS, AiTTS1 contains Tyr 265, Cys 261 and Leu 556 whose conformation changes when 2,3-oxidosqualene enters the active site (Figure 4.36). In the cyclization process, 2,3-oxidosqualene is forced to take up a pre-organized conformation, following which protonation of epoxide ring occurs in order to start a

cascade cyclization and form the dammarenyl cation. Skeletal rearrangement of a cation by 1,2 methyl and 1,2 proton shifts from high to low π electron density takes place and finally, deprotonation occurs to form the cyclic product. The cyclization reaction was initiated by protonation of epoxide ring of pre-folded 2,3-oxidosqualene by Asp 486 (Asp 455 in HLS), which is a part of the highly conserved DCTAE motif of triterpene synthases. The Cys 565 and Cys 487 act as hydrogen-bonding partners with Asp 486 for the protonation of peroxide ring. In HLS, during A ring formation, the C-6 cation is stabilized by Trp 581 located below the plane of the molecule by π interaction, whereas in the case of AiTTS1, this C-6 cation is stabilized by Trp 613. In HLS, during B ring formation, the positive charge shifts from C-6 to C-10. This cation is stabilized by Phe 444 and Trp 581, and both these residues were conserved in AiTTS1 as Phe 475 and Trp 613. The boat conformation of B ring of lanosterol in HLS is stabilized by the Tyr 98. In AiTTS1, the corresponding residue at the 120th position is asparagine, due to which the chair conformation of B ring is established (Figure 4.38). Furthermore, to stabilize A-ring and B-ring of lanosterol, additional negative charges are provided by Tyr 707, Tyr 704 and Tyr 503 in HLS. In case of AiTTS1, the corresponding residues Tyr 740 and Tyr 737 provide the additional negative charge but Tyr 503 (HLS) was replaced by Trp. Trp 535 was conserved in lupeol and β-amyrin synthases (BAS) except for BAS from oats. In HLS, during C and D ring formation, the cation is formed at C-14 which then shifts to C-20. These are stabilized by π interactions of Phe 696 and His 232. In AiTTS1, Phe 260 and Phe 729 stabilizes the Markovnikov secondary cation created at C-14 then C-20. HLS cyclization is terminated at C-20 cation because of instability provided by His 232. But Phe 260 of AiTTS1 stabilizes the cation at C14 and C20 more strongly, which provides for further cyclization instead of termination like that in HLS. However, dammarenyl cation in the AiTTS1 is not cyclized further.
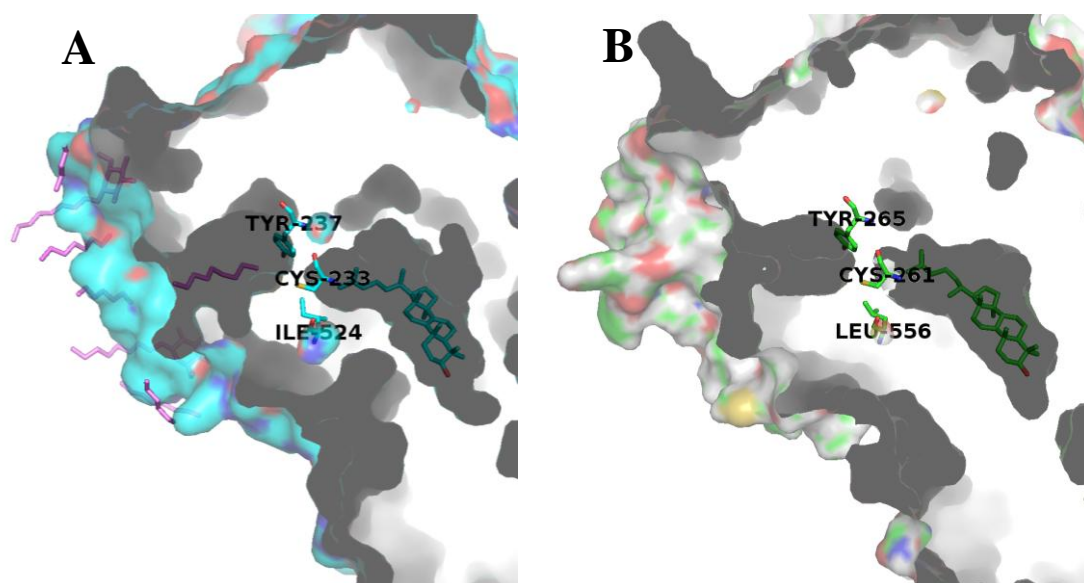
Figure 4. 36 Residues Which Allow 2,3-Oxidosqualene Entrance to Active Site.

A) Amino acids which involved in HLS, Violet colour highlighted molecules are detergents which stabilizing membrane binding region and channel for 2,3-oxidosqualene. B) Amino acids which involved in AiTTS1 for 2,3-oxidosqualene entrance.
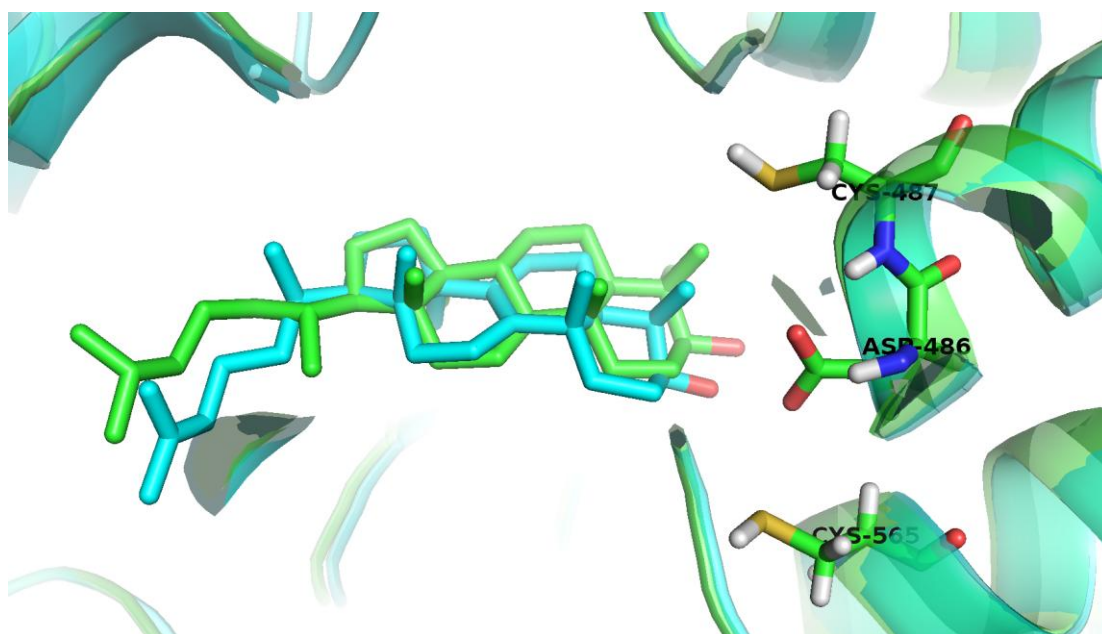
Figure 4. 37 AiTTS1 (Green) Superimposed Model with HLS (Blue) Showing Amino Acids Residues Involved in Initiation of Cyclization in 2,3-Oxiosqualene.
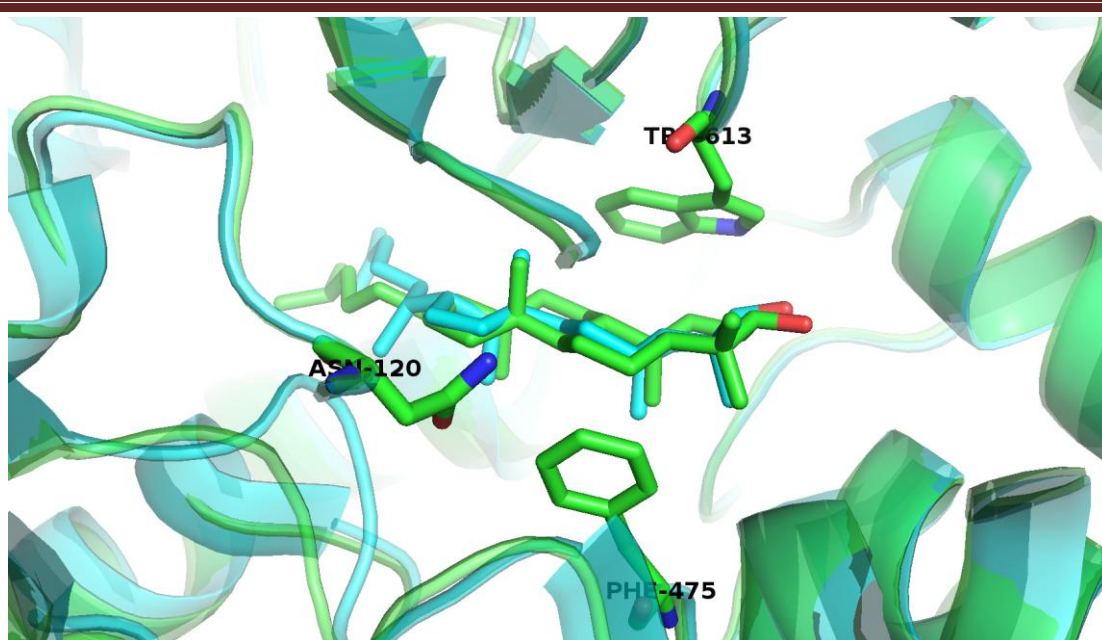
Figure 4. 38 AiTTS1 (Green) Superimposed Model with HLS (Blue) Showing Active Site Amino Acids which Involved in A and B Ring Stabilization.
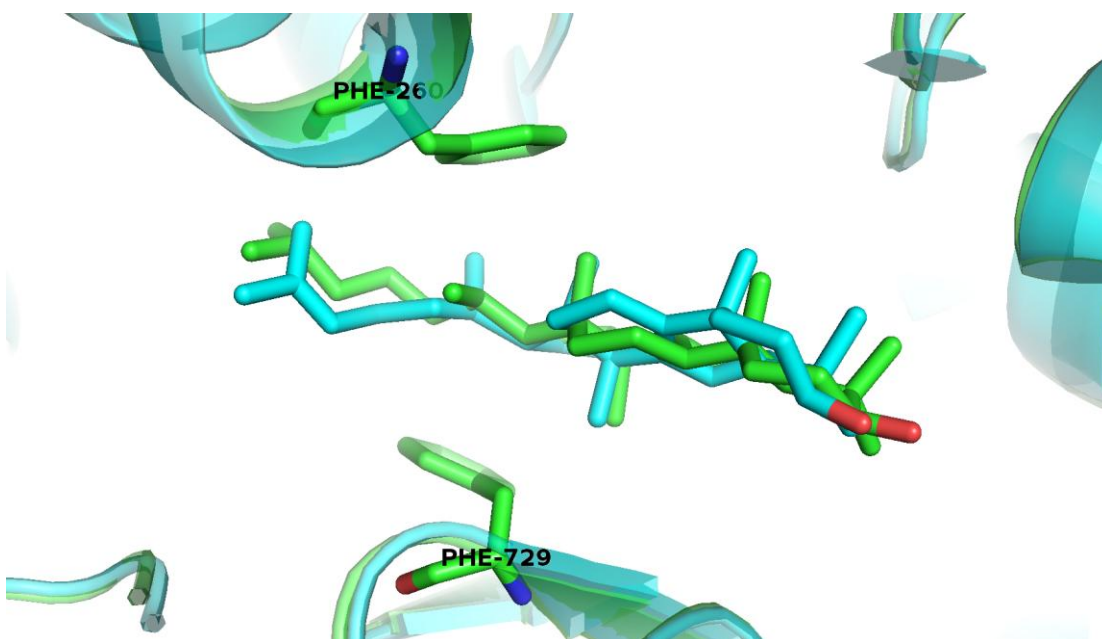
Figure 4. 39 AiTTS1 (Green) Model Superimposed with HLS (Blue) Showing Active Site Amino Acids which Involved in C and D Ring Stabilization.

In β-amyrin synthase of *E. tirucalli* (EtBAS), oleanyl cation C-25 (as per dammarenyl cation) positive charge is stabilized by the Trp 257 and Tyr 259[36]. Dammarenyl cation derived triterpenes were produced when Tyr 259 was replaced with His or any aliphatic amino acids, because of the instability of C-20 cation. In

EtBAS, when Trp 257 is replaced with another amino acid, in addition to β-amyrin, lupeol formation was also observed, which explains that Trp 257 stabilizes the C-25 cation. When Leu 256 of lupeol synthase of *Olea europaea* (OeLS) was mutated to Trp[37], lupanyl cation was further cyclized to oleanyl cation to form 75 % β-amyrin. When OeLS Tyr 258 was replaced with His, lupeol production was decreased by 50 % and remaining 50 % were dammarenyl cation derivatives. Based on these results EtBAS Trp 257 and Tyr 259 are found to be important for dammarenyl cation stabilization and its further cyclization to the oleanyl cation. In AiTTS1, the corresponding residues were found as Trp 258 and Phe 260. This confirms that AiTTS1 contains the residues which can stabilize the dammarenyl cation formation and further cyclization to the oleanyl cation, but metabolic profiling shows the presence of only dammarenyl cation derivatives.
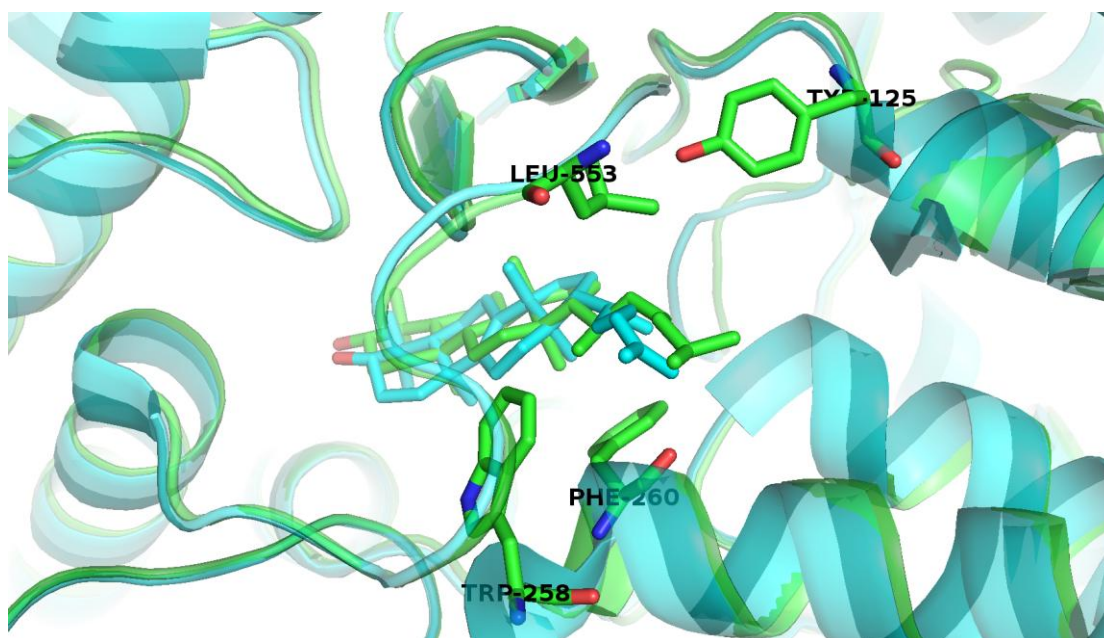
Figure 4. 40 AiTTS1 (Green) model superimposed with HLS (Blue) showing active site amino acids which involves in amyrin synthesis.

On comparison of active site residues, it was observed that Phe 552 and Phe 124 of EtBAS varied as Leu 553 and Tyr 125 in AiTTS1. EtBAS Phe 124 and Phe 552 are highly conserved residues in all triterpene synthases, but when aliphatic residues were observed in 552nd position, cyclization was terminated at lupeol cation. In OeLS, which contains Phe at 550th position, when Lue 256 was replaced by the aromatic amino acid tryptophan (L256W), it resulted in 75% β-amyrin production, suggesting that both these aromatic residues help in stabilizing the oleanyl cation. On

the other hand, lupeol synthase1 of *Arabidopsis thaliana* (AtLS1) containing Met at 550th position (corresponding to Phe 550 of OeLS) produced equal amounts of β-amyrin and lupeol when Leu 255 was replaced with Tryptophan[37], providing the proof once again, the role of this aromatic residue in the stabilization of oleanyl cation. These results, in comparison with EtBAS, aid to establish the importance of Phe 552 for oleanly cation stabilization. But, in AiTTS1, instead of aromatic amino acid at the corresponding position, Leu is present (Leu 553), due to which we can predict that the oleanyl cation stabilization is affected. Tirucalla-7,24-dien-3β-ol synthases from *A. thaliana* (AtPEN3)[14] and *Ailanthus altissima* (AaTDS) contain Leu 553 and Ile 555, respectively, that are conserved with respect to Leu at 553th in AiTTS1 hence the oleanyl cation may not stabilize. In addition to this, AiTTS1 may not stabilize the lupanyl cation because of the presence of Tyr at 125th position, as compared to the conserved Phe in all other triterpene synthases.

4.4.5.3 AiTTS1 Mutations

In AiTTS1, seven active site residues were considered for mutational analysis based on phylogenetic. When (*Maytenus ilicifolia*) friedelin synthase was mutated at Leu 482 to Val, the production of β-amyrin was increased[21]. Similarly, in AiTTS1, Val 484 was mutated to Leu resulting in the loss of enzyme activity. As mentioned above, Phe550 of OeLS may be involved in germanicyl C-19 cation stabilization[37]. Similarly in AiTTS1, when Leu 553 was mutated to Phe, it resulted in drastic reduction of enzyme activity. Phylogenetic analysis revealed that Val 550 (AiTTS1) was changed to Thr in β-amyrin synthases. V550T mutation of AiTTS1 resulted in drastic reduction of enzyme activity. From these results, V484, F550 and V550 are predicted to play a key role in stabilizing the dammarenyl C-20 cation. Change in these amino acids results in a drastic change in AiTTS1 activity. In EtBAS, Tyr 259 involved in π interaction with dammarenyl C-20 cation and allows its further cyclization to the oleanyl cation. Similarly in AiTTS1, Phe 260 was mutated to Tyr resulting in increased production of tirucalla-7,24-dien-3β-ol. Furthermore, Tyr 125 was mutated to Phe which resulted in increased production of tirucalla-7,24-dien-3β-ol. In AiTTS1, Y125F and F260Y resulted in an increased stability of dammarenyl C-20 cation but this mutation is not stabilizing for further cyclization to oleanyl cation

(Figure 4.41). Further analysis of mutations in the active site of AiTTS1 is required to get the stabilize at oleanyl cation and germanicyl C-19 cation.
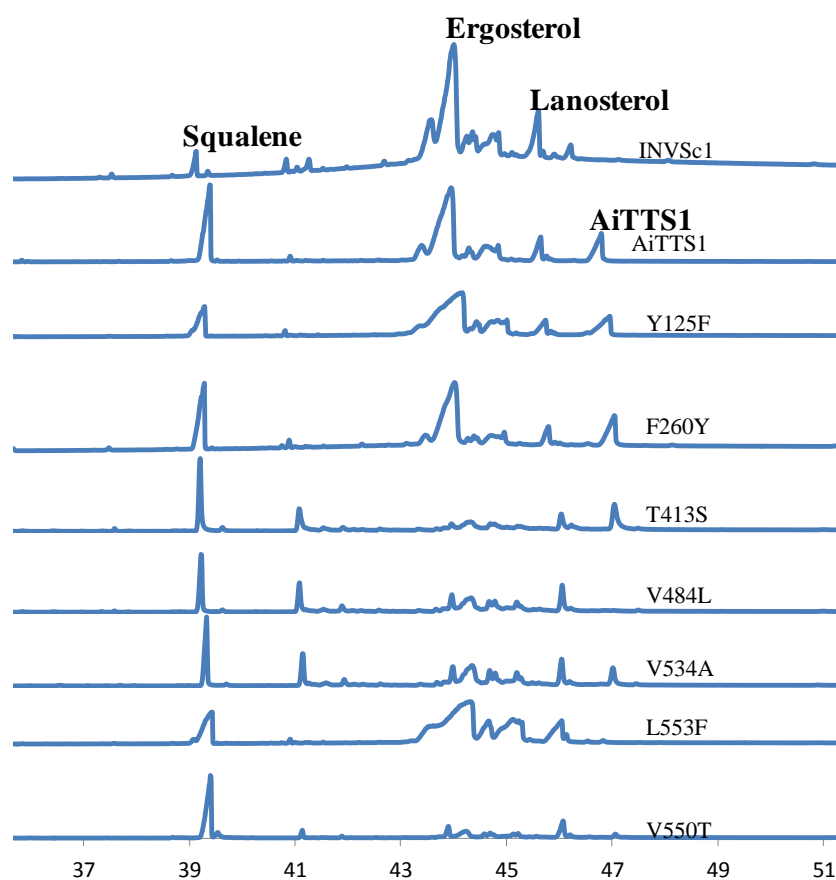
Figure 4. 41 Total Ion Chromatograms of Mutated AiTTS1 Yeast Extracts.

Table 4. 8 Tirucalla-7,24-dien-3β-ol Production in AiTTS1 Mutations as Compared to Lanosterol

AiTTS1 Mutation	% Lanosterol	% Tirucalla-7,24-dien-3β-ol
AiTTS1	36.6	63
Y125F	31.5	68.4
F260Y	21.32	78.68
T413S	29.89	70
V484L	99.02	1
V534A	57.04	42.6
L553F	95.6	4.3
V550T	84	15

4.4.6 RT-PCR Analysis of AiTTS1 and AiTTS2

To determine the expression level of triterpene synthases, so as to correlate with its involvement in limonoid biosynthesis, real-time PCR analysis of AiTTS1 and AiTTS2 was carried out by taking actin as an intrinsic control.

AiTTS1 (Figure 4.42) showed very high expression level in seeds (kernel followed by pericarp) as compared to other tissues as observed. Expression patterns of AiTTS1 were found to be similar to metabolic profiling suggesting its involvement in limonoid biosynthesis. AiTTS2 (Figure 3.42) showed very high expression in pericarp and flower, compared to other tissues. These results indicate that AiGDS might not be involved in triterpene biosynthesis in Neem.
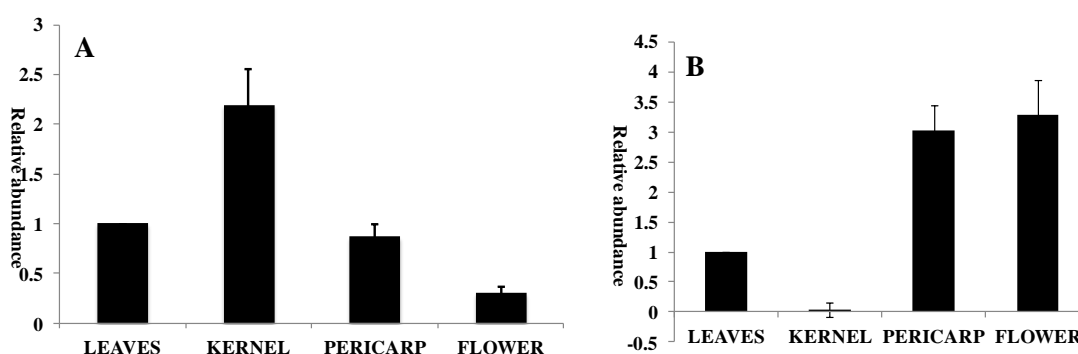
Figure 4. 42 Real-Time PCR Analysis of AiTTS1 and AiTTS2.

A) AiTTS1 was highly expressed in Kernel. **B)** AiTTS2 was highly expressed in pericarp and flowers.

4.5 Conclusion

Neem is one of the richest sources for limonoids. The first committed step in limonoid biosynthesis is cyclization 2,3-oxidosqualene which is catalyzed by triterpene synthases (TTS). In steroid biosynthesis, cycloartenol synthase (CAS) is involved in plants, microalgae and many protozoa whereas lanosterol synthase is mostly involved in animals and fungi[12]. In most of the plants more than one TTS are present in the genome.

Total six triterpene cyclases were identified from transcriptome analysis of neem. However, AiTTS1 showed very high expression as compared to other and its

expression is high in the seed. The expression profile of AiTTS1 was in line with the limonoids metabolic profiling. The AiTTS1 was cloned into a pYES2/CT vector and expression yeast. Based on metabolite GC-MS and NMR data analysis, AiTTS1 is tirucalla-7,24-dien-3β-ol synthase and this confirms its involvement in limonoid biosynthesis. The next enzyme AiTTS2 showed high expression in leaves and flowers. When we analysed its conversed domains, AiTTS2 exhibited lack of C-terminal β-sheets, which predicts its inactive nature. AiTTS2 was cloned and expressed in yeast but there is no formation of triterpene cyclic product, which confirms its inactive in nature. AiTTS1 showed the lowest activity when mutated at V484L, V550T and L553F, this explains their key role in stabilizing dammarenyl C-20 cation. F260Y and Y125F increased the stability of dammarenyl C-20 cation. Further analysis are required for understanding the mechanism and stabilization of dammarenyl C-20 cation in AiTTS1.

Euphol and tirucallol derivatives are predicted to be involved in limonoid biosynthesis based on oxygenated C_{30} compounds (protolimonoid skeleton) isolated from Meliaceae plants[22]. From the present work, product of AiTTS1, Tirucalla-7,24-dien-3β-ol can be proposed to be an intermediate in limonoid biosynthesis.

Squalene epoxidase is the enzyme which involved in the synthesis of 2,3-oxidosqualene from squalene. In neem, total three squalene epoxidases were observed. Out of this, AiSQE1 was cloned into pESC-LEU and pRS315-TEF vector. When we expressed in yeast, there is an increase in production of lanosterol and ergosterol. When coexpressed with AiTTS1, the production of tirucalla-7,24-dien-3β-ol increased by two folds. This coexpression states that, apart from HMGR overexpression, SQE overexpression enhances the production of triterpenoids engineering in yeast.

4.6 Appendix: Agarose Gel Electrophoresis for Colony PCR Screening

4.6.1 Cloning of AiSQE1 in pESC-LEU vector

Colony PCR with Gal1 forward and reverse primers for the screening of AiSQE1 (squalene epoxidase 1) cloned in the pESC-LEU vector.
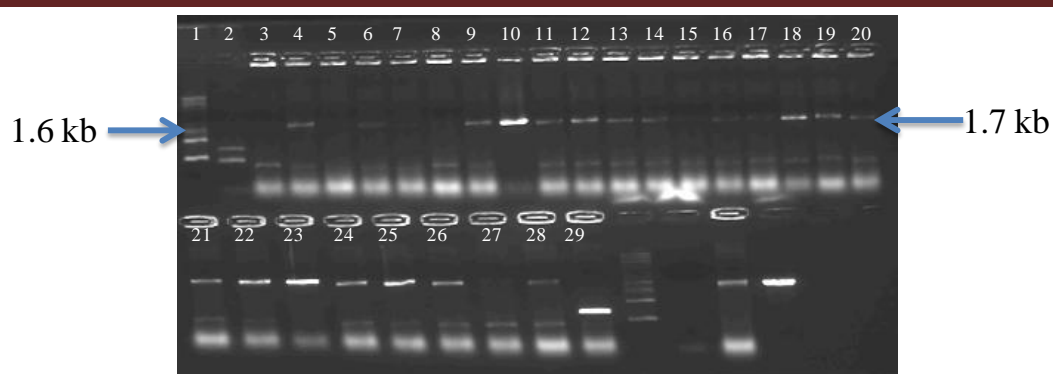
Figure 4. 43 Colony PCR Screening for AiSQE1 Cloned into pESC-LEU in an Agarose Gel.
Lane 1: 1 kb DNA ladder Sigma (Addendum Figure A1.B), **Lane 2**: Negative control and **Lanes 3-29:** PCR with Gal1 forward and reverse primers.

4.6.2 Cloning of AiSQE1 in pRS315-TEF vector

Colony PCR with AiSQE1_pRS315-TEF_FP and CYC reverse primer for the screening of squalene epoxidase 1 cloned in the pRS315-TEF vector.
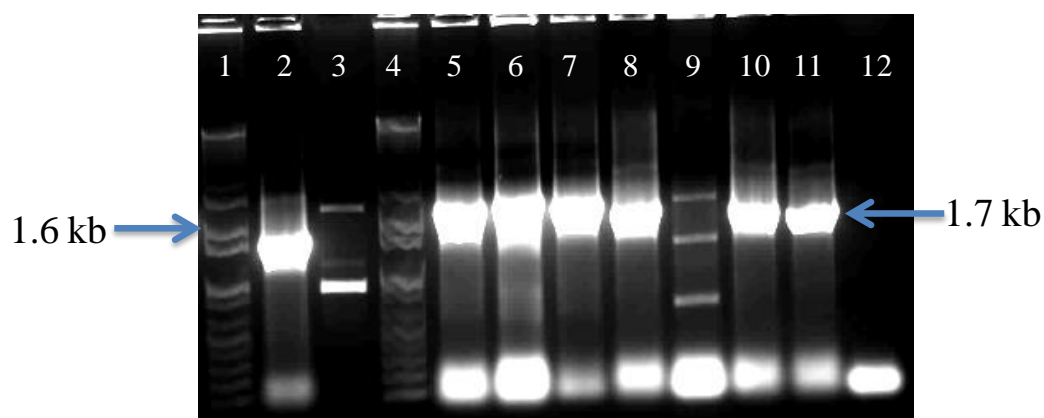
Figure 4. 44 Colony PCR Screening for AiSQE1 Cloned into pRS315-TEF in an Agarose Gel.

 Lane 1: 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C), **Lanes 2-11:** PCR with AiSQE1_pRS315-TEF_FP and CYC reverse primer and **Lane 12**: Negative control.

4.6.3 Cloning of AiTTS1 in pYES2/CT vector

Colony PCR with T7 promoter and CYC reverse primer for the screening of AiTTS1 (triterpene synthase 1) cloned into the pYES2/CT vector.
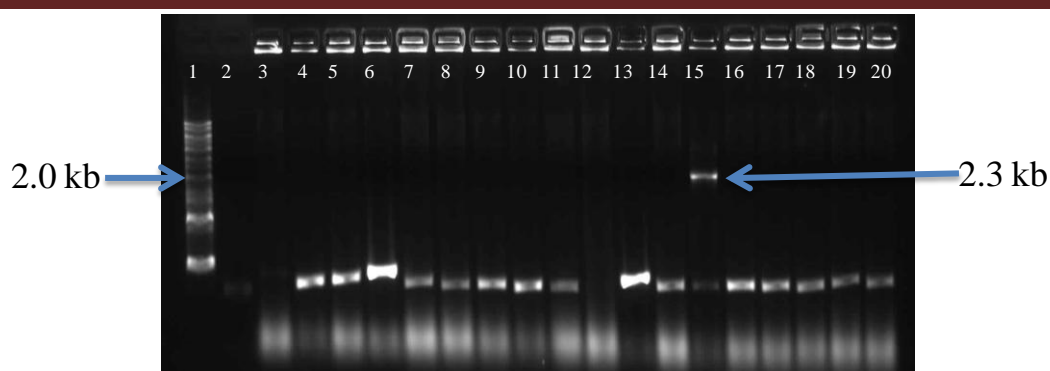
Figure 4. 45 Colony PCR Screening for AiTTS1 Cloned into pYES2/CT in an Agarose Gel.

Lane 1: 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C), **Lane 2**: Negative control and **Lanes 3-20:** PCR with T7 promoter and CYC reverse primer.

4.6.4 Cloning of AiTTS2 3' RACE product in pCR™-Blunt vector

Colony PCR with M13 forward and reverse primer for the screening of AiTTS2 3' RACE product cloned in pCR™-Blunt vector.
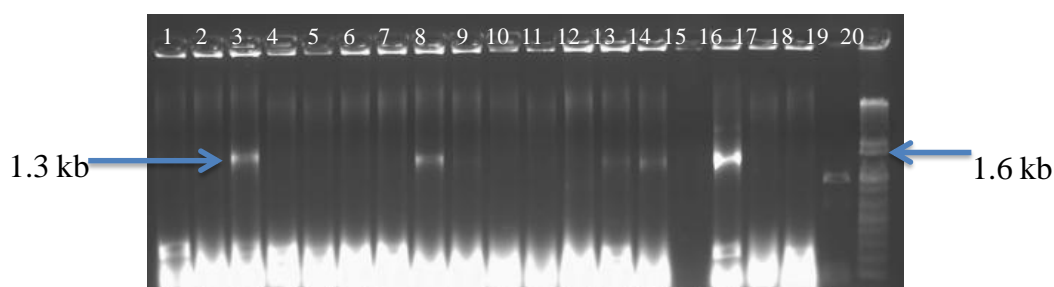
Figure 4. 46 Colony PCR Screening for AiTTS2 3' RACE Product Cloned into pCR™-Blunt Vector in an Agarose Gel.

Lane 1-18: PCR with M13 forward and reverse primer, **Lane 19**: Negative control and **Lanes 20:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C)

4.6.5 Cloning of AiTTS2 Product in pYES2/CT Vector

Colony PCR with T7 promoter and CYC reverse primer for the screening of AiTTS2 (triterpene synthase 2) cloned into pCR™-Blunt vector.
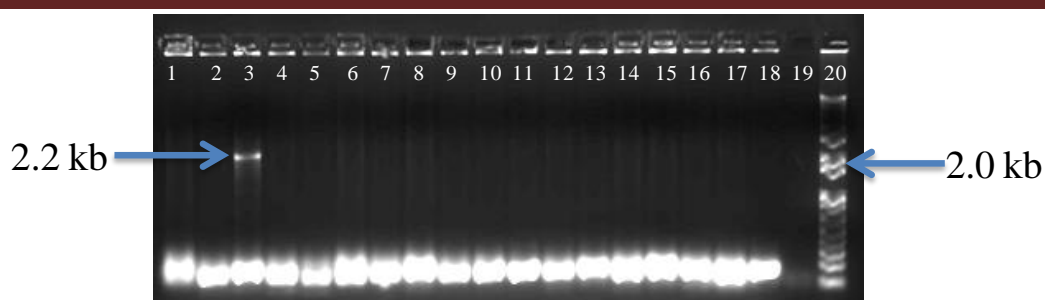
Figure 4. 47 Colony PCR Screening for AiTTS2 Product Cloned into pYES2/CT Vector in an Agarose Gel.

Lane 1-18: PCR with T7 promoter and CYC reverse primer, **Lane 19**: Negative control and **Lanes 20:** 1 kb plus DNA ladder Invitrogen (Addendum Figure A1.C)

4.7 References

1. Da Silva, M.F.D.G.F., Gottlieb, O.R. & Dreyer, D.L. Evolution of limonoids in the Meliaceae. *Biochemical Systematics and Ecology* **12**, 299-310 (1984).

2. Siddiqui, S., Siddiqui, B.S., Faizi, S. & Mahmood, T. Tetracyclic triterpenoids and their derivatives from *Azadirachta indica*. *Journal of Natural Products* **51**, 30-43 (1988).

3. Pandreka, A. *et al*. Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*). *BMC Plant Biology* **15**, 214 (2015).

4. Woodward, R.B. & Bloch, K. The cyclization of squalene in cholesterol synthesis. *Journal of the American Chemical Society* **75**, 2023-2024 (1953).

5. Corey, E.J., Russey, W.E. & de Montellano, P.R.O. 2,3-Oxidosqualene, an Intermediate in the Biological Synthesis of Sterols from Squalene. *Journal of the American Chemical Society* **88**, 4750-4751 (1966).

6. Van Tamelen, E.E., Willett, J.D., Clayton, R.B. & Lord, K.E. Enzymic conversion of squalene 2,3-oxide to lanosterol and cholesterol. *Journal of the American Chemical Society* **88**, 4752-4754 (1966).

7. Abe, I., Rohmer, M. & Prestwich, G.D. Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chemical Reviews* **93**, 2189-2206 (1993).

8. Barton, D.H.R. *et al*. Investigations on the biosynthesis of steroids and terpenoids. Part XII. Biosynthesis of 3β-hydroxy-triterpenoids and -steroids from (3*S*)-2,3-epoxy-2,3-dihydrosqualene. *Journal of the Chemical Society, Perkin Transactions 1*, 1134-1138 (1975).

9. Han, J.Y., In, J.G., Kwon, Y.S. & Choi, Y.E. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. *Phytochemistry* **71**, 36-46 (2010).

10. Laranjeira, S. *et al*. *Arabidopsis* Squalene Epoxidase 3 (SQE3) complements sqe1 and is important for embryo development and bulk squalene epoxidase activity. *Molecular plant* **8**, 1090-1102 (2015).

11. Rasbery, J.M. *et al*. *Arabidopsis thaliana* squalene epoxidase 1 is essential for root and seed development. *Journal of Biological Chemistry* **282**, 17002-17013 (2007).

12. Volkman, J.K. Sterols and other triterpenoids: source specificity and evolution of biosynthetic pathways. *Organic Geochemistry* **36**, 139-159 (2005).

13. Misra, R.C., Maiti, P., Chanotiya, C.S., Shanker, K. & Ghosh, S. Methyl jasmonate-elicited transcriptional responses and pentacyclic triterpene biosynthesis in sweet basil. *Plant Physiology* **164**, 1028-1044 (2014).

14. Morlacchi, P. *et al*. Product profile of PEN3: the last unexamined oxidosqualene cyclase in *Arabidopsis thaliana*. *Organic letters* **11**, 2627-2630 (2009).

15. Kajikawa, M. *et al*. Cloning and characterization of a cDNA encoding β-amyrin synthase from petroleum plant *Euphorbia tirucalli* L. *Phytochemistry* **66**, 1759-1766 (2005).

16. Xu, R., Fazio, G.C. & Matsuda, S.P.T. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**, 261-291 (2004).

17. Wu, T.K., Chang, C.H., Liu, Y.-T. & Wang, T.T. *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase: A chemistry-biology interdisciplinary study of the protein's structure-function-reaction mechanism relationships. *The Chemical Record* **8**, 302-325 (2008).

18. Takase, S. *et al*. Control of the 1, 2-rearrangement process by oxidosqualene cyclases during triterpene biosynthesis. *Organic & Biomolecular Chemistry* **13**, 7331-7336 (2015).

19. Wu, T.K. *et al*. Alteration of the substrate's prefolded conformation and cyclization stereochemistry of oxidosqualene-lanosterol cyclase of *Saccharomyces cerevisiae* by substitution at phenylalanine 699. *Organic Letters* **12**, 500-503 (2010).

20. Ito, R., Hashimoto, I., Masukawa, Y. & Hoshino, T. Effect of Cation-π Interactions and Steric Bulk on the Catalytic Action of Oxidosqualene Cyclase: A Case Study of Phe728 of β-Amyrin Synthase from *Euphorbia tirucalli* L. *Chemistry-A European Journal* **19**, 17150-17158 (2013).

21. Souza-Moreira, T.M. *et al*. Friedelin Synthase from *Maytenus ilicifolia*: Leucine 482 Plays an Essential Role in the Production of the Most Rearranged Pentacyclic Triterpene. *Scientific reports* **6**, 36858 (2016).

22. Tan, Q.G. & Luo, X.D. Meliaceous limonoids: chemistry and biological activities. *Chemical Reviews* **111**, 7437-7522 (2011).

23. Jones, P.S., Ley, S.V., Morgan, E.D. & Santafianos, D. The chemistry of the neem tree. *Focus on phytochemical pesticides* **1**, 19-45 (1989).

24. Augustin, J.r.M., Kuzina, V., Andersen, S.B. & Bak, S.r. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**, 435-457 (2011).

25. E.U. Ekong, D. & A. Ibiyemi, S. Biosynthesis of nimbolide from [2-^{14}C,(4R)4-^{3}H$_1$] mevalonic acid lactone in the leaves of *Azadirachta indica*. *Phytochemistry* **24**, 2259-2261 (1985).

26. Ekong, D.E.U., Ibiyemi, S.A. & Olagbemi, E.O. The meliacins (limonoids). Biosynthesis of nimbolide in the leaves of *Azadirachta indica*. *J Chem Soc Chem Commun* **18**(1971).

27. Nakamura, Y., Sakakibara, J., Izumi, T., Shibata, A. & Ono, T. Transcriptional regulation of squalene epoxidase by sterols and inhibitors in HeLa cells. *Journal of Biological Chemistry* **271**, 8053-8056 (1996).

28. Favre, B. & Ryder, N.S. Cloning and expression of squalene epoxidase from the pathogenic yeast *Candida albicans*. *Gene* **189**, 119-126 (1997).

29. Lee, H.K., Denner Ancona, P., Sakakibara, J., Ono, T. & Prestwich, G.D. Photoaffinity labeling and site-directed mutagenesis of rat squalene epoxidase. *Archives of Biochemistry and Biophysics* **381**, 43-52 (2000).

30. Lee, H.K. *et al*. Photoaffinity labeling identifies the substrate-binding site of mammalian squalene epoxidase. *Biochemical and Biophysical Research Communications* **315**, 1-9 (2004).

31. Abe, I., Abe, T., Lou, W., Masuoka, T. & Noguchi, H. Site-directed mutagenesis of conserved aromatic residues in rat squalene epoxidase. *Biochemical and biophysical research communications* **352**, 259-263 (2007).

32. Zhang, H., Shibuya, M., Yokota, S. & Ebizuka, Y. Oxidosqualene cyclases from cell suspension cultures of *Betula platyphylla* var. japonica: molecular evolution of oxidosqualene cyclases in higher plants. *Biological and Pharmaceutical Bulletin* **26**, 642-650 (2003).

33. Hayashi, H. *et al*. Cloning and characterization of a cDNA encoding β-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. *Biological and Pharmaceutical Bulletin* **24**, 912-916 (2001).

34. Wang, Z., Yeats, T., Han, H. & Jetter, R. Cloning and characterization of oxidosqualene cyclases from *Kalanchoe daigremontiana* enzymes catalyzing up to 10 rearrangement steps yielding friedelin and other triterpenoids. *Journal of Biological Chemistry* **285**, 29703-29712 (2010).

35. Thoma, R. *et al*. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **432**, 118 (2004).

36. Ito, R., Nakada, C. & Hoshino, T. β-Amyrin synthase from *Euphorbia tirucalli* L. functional analyses of the highly conserved aromatic residues Phe413, Tyr259 and Trp257 disclose the importance of the appropriate steric bulk, and cation-π and CH-π interactions for the efficient catalytic action of the polyolefin cyclization cascade. *Organic & biomolecular chemistry* **15**, 177-188 (2017).

37. Kushiro, T., Shibuya, M., Masuda, K. & Ebizuka, Y. Mutational studies on triterpene synthases: engineering lupeol synthase into β-amyrin synthase. *Journal of the American Chemical Society* **122**, 6816-6824 (2000).
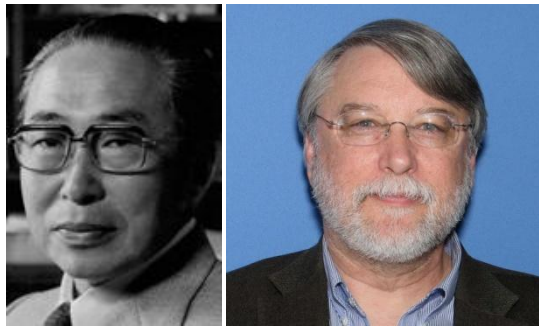
Osamu Hayaishi David R. Nelson

Chapter 5

Cloning and Functional

Characterization of

Cytochrome P450 Systems

5.1 Introduction

Neem (*Azadirachta indica*) is very well known for its medicinal and agricultural use. Neem possesses limonoids which have 4,4,8-trimethyl-17-furanyl steroid skeleton. Based on the skeleton, neem limonoids are further divided into protolimonoids, basic limonoids and C-seco limonoids. As discussed in chapter 4, neem limonoids are synthesized from tirucalla-7,24-dien-3β-ol. Predicted limonoid biosynthesis starts from the rearrangement of C-14 methyl group by opening of 7α, 8α-epoxide ring to form protolimonoids. Four terminal carbons from the side chain are removed by oxidative degradation, and then the side chain undergoes cyclization through cyclic hemiacetal to form basic limonoid skeletons. Then the third ring opens and oxidizes further to form C-seco limonoid skeletons. Hence, cytochrome P4540 enzymes (CYP450) are mainly involved in the tailoring of tirucalla-7,24-dien-3β-ol to synthesize limonoids.

CYP450 belongs to enzyme class monooxygenase (E.C: 1.1.14.-). The characteristic feature of these enzymes is a shift in absorbance from 413 nm to 450 nm in the carbon monoxide bound form[1]. Omura and Santo identified these as heme proteins and have given the name cytochrome P450[2]. The enzymes which have more than 40 % similarity are classified into the same family. The plant enzyme families start from CYP71 to CYP99 and continue again from CYP701-999. The nomenclature starts with root symbol 'CYP' (CYtochrome P450) followed by a number denoting family, a letter designating the subfamily and a number to represent individual gene[3]. These enzymes are found in all kingdoms of life including viruses. CYP450 gene content in plant (mainly in angiosperm) genome can reach up to 1%, whereas in vertebrates CYP450 content is less than 100 genes[4].
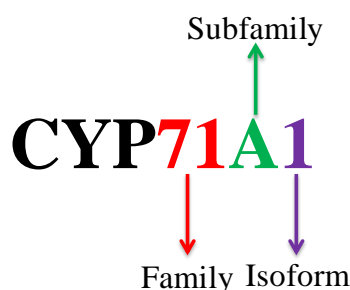
Figure 5. 1 Nomenclature of CYP450s.

Monooxygenases incorporate a single atom of molecular oxygen into a substrate with the concomitant reduction of the other atom to water. The monooxygenases are divided into two classes, the internal and the external monooxygenases. Internal monooxygenases extract two reducing equivalents from the substrate to reduce one atom of dioxygen to water, whereas external monooxygenases utilize an external reductase. Most of the CYP450 are external monooxygenases. They exist as three different types. Most of the eukaryotes contain two-component system, comprising a CYP450 and flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN) containing NADPH-cytochrome P450 reductase (CPR) which are membrane-bound. Bacteria contain soluble three component system, FAD-containing reductase, an iron-sulfur protein, and CYP450. Whereas in mitochondria, reductase and CYP450 are membrane-bound. In few cases, CYP450 and CPR exist as fused system[5,6].
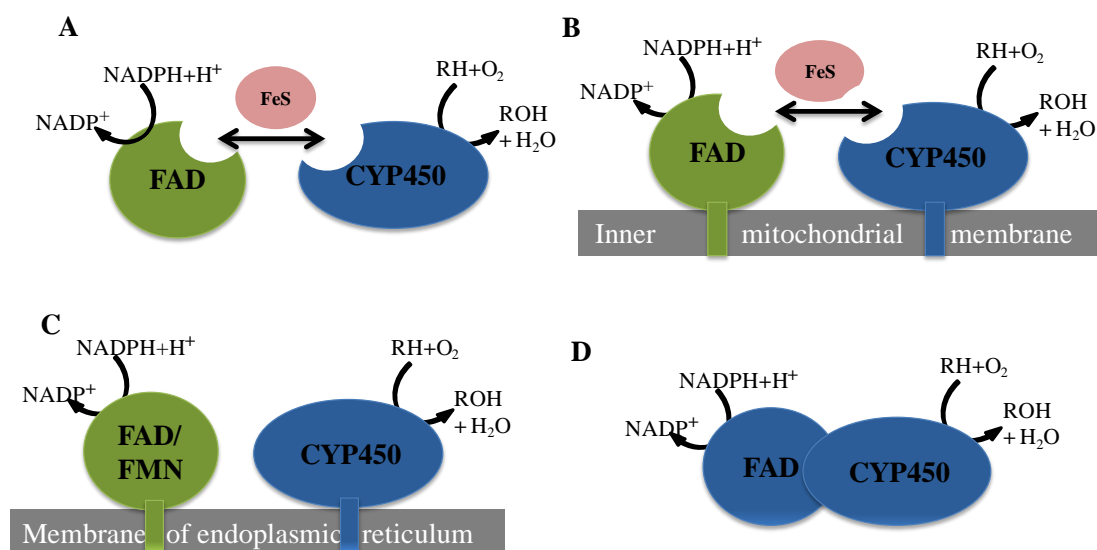
Figure 5. 2 Schematic Representation of CYP450 Systems.
A) Bacterial three component system, B) Mitochondrial three component system, C) Eukaryotic two-component system and D) Fused/self-sufficient system.

In Plants, CPR catalyzes electron transfer from NADPH through FAD / FMN to CYP450. Apart from this, CPR transfers electrons to cytochrome *b5,* fatty acid and sterol desaturase[7]. Mammals and yeast have only one CPR, but in plant species, such as cotton[8], *Arabidopsis thaliana*[9], hybrid poplar[7] and others have two to three CPR. These CPR isoforms have different responses under stress conditions. For example, in *A. thaliana AtCPR1* expresses constitutively and *AtCPR2* is inducible by environmental stimuli. Hence, CPR plays a key role not only in secondary metabolism

but also in electron transport chain and in stress conditions[10]. The CPR is classified into two classes, based on N-terminal hydrophobic region. Class I CPRs contain N-terminal hydrophobic region and class II CPRs contain an extended N-terminal ends with high Ser/Thr content[11].

Triterpene cyclic product catalyzed by triterpene cyclases undergoes site-specific oxidization by cytochrome P450s to produce triterpenoids. β-amyrin-24-hydroxylase (CYP93E1) from *Glycine max*[12] was the first characterized CYP450 related to triterpenoid biosynthesis. Since then, more than 35 CYP450s related to triterpenoid biosynthesis have been identified (Figure 5.3). This chapter deals with the identification and characterization of CPR and CYP450s involved in limonoid biosynthesis.

5.2 Neem CYP450 Systems

NADPH-cytochrome P450 reductases are the enzymes which transfer electrons from NADPH to cytochrome P450 and plays a key role in different biosynthetic pathways. In neem, two CPR were identified from the transcriptome. AiCPR1 (Neem_Transcript_2277/Master_Control_115955; NCBI KM108318) and AiCPR2 (Neem_Transcript_ 1270/Master_Control_117874; NCBI KM108319) are expressed equally in all the tissues. Out of these two, AiCPR2 showed higher expression as compared to AiCPR1. A total of 134 transcripts were predicted as cytochrome P450 monooxygenases. Based on Blast results, with reference to *Arabidopsis thaliana* cytochrome P450, Neem CYP450s were classified into 39 families and 78 subfamilies, out of which most of the CYP450 belonged to the CYP71 family. Based on functional annotation and differential gene expression analysis, fifteen CYPs were predicted to be involved in limonoid biosynthesis. Among these the genes, the ones which showed high expression in kernel and pericarp were selected for functional characterization (Figure 5.4). AiCYP1 (Neem_transcript_34861 /Master_Control_84673) showed functional annotation of β-amyrin 28-oxidase (*Panax ginseng*) which was predicted to be involved in terminal chain modification in protolimonoids. AiCYP2 (Neem_transcript_38933/ Master_Control_57632) was showing functional annotation of β-amyrin 11-oxidase (*Glycyrrhiza uralensis*) which was predicted to be involved in C-ring opening in basic limonoids.

Figure 5. 3 Characterized CYP450 Related to Triterpenoids.
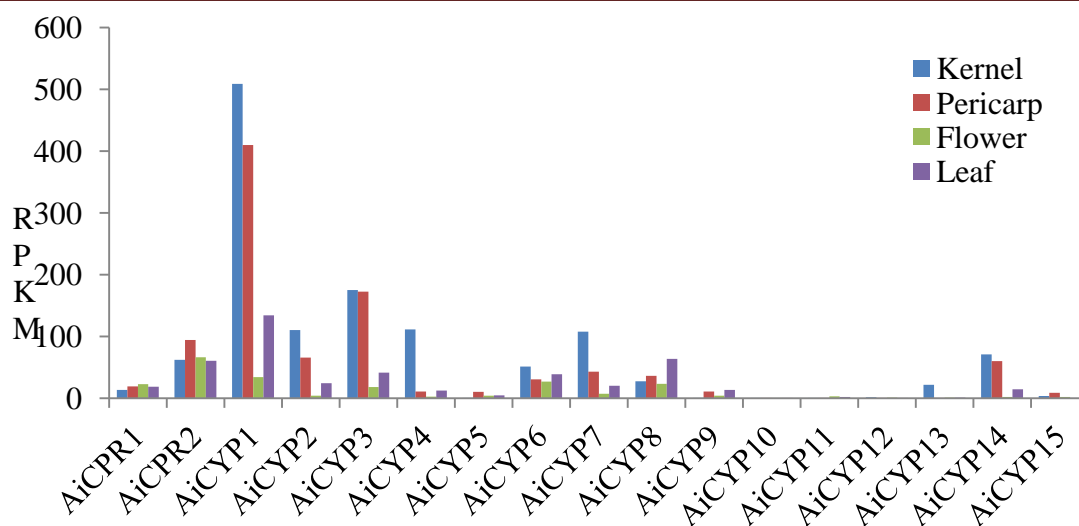
Figure 5. 4 Expression Levels of Neem Cytochrome P450 Systems between Different Tissues of Neem

5.3 Materials and Methods

5.3.1 Materials Used in this Study

5.3.1.1 Bacterial Strains and Plasmids Used in the Study

Escherichia coli Mach1™ T1^R (ThermoFisher Scientific, USA) cloning cells were used for transformation of plasmids or ligation mixtures. pET32a (Novagen, USA) pYES2/CT and pYES3/CT (ThermoFisher Scientific, USA), Pri101-AN vector (Takara Bio, USA) were used of cloning of AiCPRs and AiCYPs. Expression of AiCPRs were done in *Escherichia coli* Rosetta 2 (DE3) cells (Novagen, Addgene, USA). Expression of AiCYPs were done in INVSc1, *S. cerevisiae* yeast Strain (ThermoFisher Scientific, USA). *Agrobacterium* LBA4404 strain was used for transient transformation of AiCYPs in neem

5.3.1.2 Kits and Reagents Used in the Study

SuperScript® III Reverse Transcriptase (ThermoFisher Scientific, USA) was used for cDNA synthesis. JumpStart™ Taq DNA Polymerase (Sigma-Aldrich, USA) and AccuPrime™ *Pfx* DNA Polymerase (ThermoFisher Scientific, USA) were used for amplification of AiCPRs and AiCYPs. PCR products were gel eluted by using PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (ThermoFisher Scientific, USA). Plasmids were isolated by using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, USA). GelRed^{TM} (Biotium Inc., USA) was used for nucleic acid staining. Restriction enzymes (New England Biolabs, USA) and T$_4$ DNA ligase

(Invitrogen/ Life Technologies, USA) used for cloning. pYES2/CT and pYES3/CT Postive clones were transformed into yeast by using *S.c.* EasyComp^{TM} Transformation Kit (ThermoFisher Scientific, USA).

5.3.1.3 Buffers Compositions

5.3.1.3.1 Buffers Used for Characterization of AiCPR1

Lysis buffer

200 mM MOPSO buffer in 10 % glycerol containing NaCl (500 mM), imidazole (10 mM), detergents 1 % CHAPS, 1% Triton X-100, protease inhibitor PMSF (0.5 mM), 1 mg/mL lysozyme and the pH was adjusted to 7.4 with 0.1 M NaOH.

Wash buffer

200 mM MOPSO buffer in 10 % glycerol containing NaCl (500 mM), imidazole (100 mM) and the pH was adjusted to 7.4 with 0.1 M NaOH.

Elution buffer

200 mM MOPSO buffer in 10 % glycerol containing NaCl (500 mM), imidazole (250 mM) and the pH was adjusted to 7.4 with 0.1 M NaOH.

Desalting buffer

50 mM Tris buffer in 10 % glycerol containing KCl (100 mM) and the pH was adjusted to 7.4 with 0.1 M HCl.

Enzyme assay buffer

50 mM Tris buffer in 10 % glycerol containing KCl (100 mM) and co-factors FMN (5 μM), FAD (5 μM), NADPH (20 μM) and the pH was adjusted to 7.4 with 0.1 M HCl.

5.3.1.3.2 Buffers Used for Characterization of AiCPR2

Lysis buffer

200 mM MOPSO buffer in 10 % glycerol containing NaCl (300 mM), imidazole (10 mM), detergents 1 % CHAPS, 1% Triton X-100, protease inhibitor PMSF (0.5 mM), 1 mg/mL lysozyme and the pH was adjusted to 7.4 with 0.1 M NaOH.

Wash buffer

50 mM Tris HCl buffer in 10 % glycerol containing NaCl (300 mM), imidazole (80 mM) and the pH was adjusted to 7.4 with 0.1 M HCl.

Elution buffer

50 mM Tris HCl buffer in 10 % glycerol containing NaCl (300 mM), imidazole (250 mM and the pH was adjusted to 7.4 with 0.1 M HCl.

Desalting buffer

50 mM Tris buffer in 10 % glycerol containing KCl (100 mM) and the pH was adjusted to 7.4 with 0.1 M HCl.

Enzyme assay buffer

50 mM Tris buffer in 10 % glycerol containing KCl (100 mM) and co-factors FMN (5 μM), FAD (5 μM), NADPH (20 μM) and the pH was adjusted to 7.4 with 0.1 M HCl.

5.3.1.4 Primers

5.3.1.4.1 Primers for AiCPRs

Table 5. 1 Primers Used for Cloning of AiCPRs

Primer Name	Primer Sequence
AiCPR1_FP	ATGAGCAACTCTGGTACGGGTAACGATTTG
AiCPR1_RP	TCACCAGACATCTCTAAGATATCGTCCTTC
AiCPR1_pET32_FP	ATCAGGAATTCATGAGCAACTCTGGTACGGGTAAC
AiCPR1_pET32_RP	ATCTAGAAGCTTTCACCAGACATCTCTAAGATATC
AiCPR2 _FP	ATGCAATCATCGTCGTCATCAGCATCATC

AiCPR2 _RP	TCACCACACATCACGTAGATACCTG
AiCPR2_pET32_FP	AAGTCTGAGCTCATGCAATCATCGTCGTCATC
AiCPR2_pET32_RP	AAGTCTCTCGAGTCACCACACATCACGTAGATAC
AiCPR2_pYES3_FP	ATCAGAAGCTTCACACAATGTCCATGCAATCATCGTCGTCATC
AiCPR2_pYES3_RP	ATCTAGCTCGAG TCACCACACATCACGTAGATAC

5.3.1.4.2 Primers for AiCYPs

Table 5. 2 Primers Used FOR Cloning of AiCYPs

Primer Name	Primer Sequence
AiCYP1_FP	ATGGAGCTCTTCTTACTCTCCGTACTTCTC
AiCYP1_RP	TTAATTGTTGTAGGGATAGAGACGAACTGGAAG
AiCYP1_pYES2_FP	ATGCCAGGATCC**AACACAATGTCT**ATGGAGCTCTTCTTACTC
AiCYP1_pYES2_RP	GCATGATCTAGAATTGTTGTAGGGATAG
AiCYP_pRI101_FP	ATCGACCGTCGACATGGAGCTCTTCTTACTC
AiCYP_pRI101_RP	ATACAGGGTACCTCA**ATGATGATGATGATGGTGAGAAGAACC**ATTGTTGTAGGGATAG
AiCYP1_S1_FP	ACTGTCGTCGACATGGAACTGTTCTTACTCTCCGTACTTC
AiCYP1_S1_RP	GCATATAGGTACCCACCAGGATCTAACAAGCTTG
AiCYP1_S2_FP	GCATACGGATCCCACCAGGATCTAACAAGCTTGT
AiCYP1_S2_FP	ATCGACGAGCTCATGGAACTGTTCTTACTCTCCGTACTTC

AiCYP2_FP	ATGGGTTTAGATTTGTTGTGGTTGATTCTTG
AiCYP2_RP	TCATTTGAGCTTAATAATTCTTGCAAGACATTG
AiCYP2_pYES2_FP	GTACCAGGTACCCACACAATGTCCATGGGTTTAGATTTGTTGTG
AiCYP2_pYES2_RP	ACTAATCTCGAGTTTGAGCTTAATAATTC

5.3.2 Cloning and Characterization of AiCPRs

5.3.2.1 Cloning and Characterization of AiCPR1 by Using Bacterial System

Full-length primers for AiCPR1 ORF were designed using the transcript (Neem_Transcripts_2277) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 62 °C for 30 sec, 68 °C for 2.1 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *EcoRI* and *HindIII* cloning sites of pET32a expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter and T7 reverse primers) to identify the positive clones. Cloning was confirmed by analyzing the sequence of positive clones obtained by Sanger sequencing by using T7 promoter and T7 reverse primers.

The expression of the recombinant plasmids containing AiCPR1 was carried out in Rosetta 2 (DE3) cells under 1 mM IPTG inducer at 16 °C incubation for overnight. Bacterial cells were collected by centrifugation at $5000 \times g$ for 10 min. Cell pellet was resuspended in lysis buffer (200 mM MOPSO, 500 mM NaCl, 10 mM imidazole, 1 % CHAPS, 1% Triton X-100, 10 % glycerol, 0.5 mM PMSF, 1 mg/mL lysozyme, pH 7.4) at a ratio of 5 ml per gram of cell pellet. Sonication was done for 10 cycles (30 sec pulse on/ 30 sec pulse off, Amplitude- 75 %) then centrifuged at $10,000 \times g$ for 10 min. The supernatant containing AiCPR1 protein was purified over Ni-NTA (1 mL resin / g cell pellet) affinity chromatography. The column was washed with wash buffer (200 mM MOPSO, 500 mM NaCl, 100 mM imidazole, 10 % glycerol, pH 7.4) till the O.D. reaches to 0.0 at 280 nm. The protein was eluted out

with elution buffer (200 mM MOPSO, 500 mM NaCl, 250 mM imidazole, 10 % glycerol, pH 7.4). Protein was desalted on Hi-Prep^{TM} 26/10 desalting column with desalting buffer (50 mM Tris, 100 mM KCl, 10 % glycerol, pH 7.4) using AKTA Avant (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and checked on 12 % SDS gel.

Enzyme assays for AiCPR1 were performed in assay buffer (50 mM Tris, 100 mM KCl, 10 % glycerol, pH 7.4) with cytochrome C (40 μM) and potassium ferricyanide (40 μM) as substrates and reduction were observed at 550 nm and 424 nm respectively at 25 °C. AiCPR1 activity was confirmed by an increase in absorbance at 550 nm when cytochrome C was reduced and in the same way decrease in absorbance at 424 nm when potassium ferricyanide was reduced.

5.3.2.2 Cloning and Characterization of AiCPR2 by Using Bacterial System

Full-length primers for AiCPR2 ORF were designed using the transcript (Neem_Transcripts_1270) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 60 °C for 30 sec, 68 °C for 2.1 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *SacI* and *XhoI* cloning sites of pET32a expression vector using T$_4$ DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 μg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter and T7 reverse primers) to identify the positive clones. Cloning was confirmed by analysing the sequence of positive clones obtained by Sanger sequencing using T7 promoter forward and T7 reverse primers.

The expression of the recombinant plasmids containing AiCPR2 was carried out in Rosetta 2 (DE3) cells under 1 mM IPTG inducer at 16 °C for overnight. Bacterial cells were collected by centrifugation at 5000 × *g* for 10 min. Cell pellet was resuspended in lysis buffer (50 mM MOPSO, 300 mM NaCl, 10 mM imidazole, 1 % CHAPS, 1% Triton X-100, 10 % glycerol, 0.5 mM PMSF, 1 mg/mL lysozyme, pH 7.4) at a ratio of 5 ml per gram of cell pellet. Sonication was done for 10 cycles (30 sec pulse on/ 30 sec pulse off, Amplitude- 75 %) then centrifuged at 10,000 × *g* for 10 min. The supernatant containing AiCPR1 protein was purified over Ni-NTA (1 mL resin / g cell pellet) affinity chromatography. The column was washed with wash buffer (50 mM

Tris HCl, 300 mM NaCl, 80 mM imidazole, 10 % glycerol, pH 7.4) till the O.D. reaches to 0.0 at 280 nm. The protein was eluted out with elution buffer (50 mM Tris HCl, 300 mM NaCl, 250 mM imidazole, 10 % glycerol, pH 7.4). Protein was desalted on Hi-Prep™ 26/10 desalting column with desalting buffer (50 mM Tris, 100 mM KCl, 10 % glycerol, pH 7.4) using AKTA Avant (GE Healthcare). The desalted proteins were estimated using Bradford reagent (Bio-Rad) and checked on 12 % SDS gel.

Enzyme assays for AiCPR2 were performed in assay buffer (50 mM Tris, 100 mM KCl, 10 % glycerol, 5 µM FMN, 5 µM FAD, 20 µM NADPH, pH 7.4) with cytochrome C (40 µM) and potassium ferricyanide (40 µM) as substrates and reduction was observed at 550 nm and 424 nm respectively at 25 °C. AiCPR2 activity was confirmed by an increase in absorbance at 550 nm when cytochrome C was reduced and in the same way decrease in absorbance at 424 nm when potassium ferricyanide was reduced.

5.3.3 Cloning and Characterization of AiCYPs

5.3.3.1 Cloning and Characterization of AiCYP1 in Yeast by Co-expression of AiCPR2

5.3.3.1.1 Cloning of AiCPR2 into pYES3/CT

Full-length primers for AiCPR2 ORF were designed using the transcript (Neem_Transcripts_1270) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 64 °C for 30 sec, 68 °C for 2.1 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *HindIII* and *XhoI* cloning sites of pYES3/CT expression vector using T$_4$ DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter forward and CYC reverse primers) to identify the positive clones. Cloning was confirmed by analysing the sequence of positive clones obtained by Sanger sequencing using T7 promoter and CYC reverse primers.

5.3.3.1.2 Cloning of AiCYP1 in pYES2/CT

Full-length primers for AiCYP1 ORF were designed using their transcript (Neem_Transcripts_34861) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 63 °C for 30 sec, 68 °C for 2.0 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *BamHI* and *XbaI* cloning sites of pYES2/CT expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter forward and CYC reverse primers) to identify the positive clones. Cloning was confirmed by analysing the Sanger sequence of positive clones using T7 promoter and CYC reverse primers.

5.3.3.1.3 Characterization of AiCYP1 in Yeast

The expression of AiCYP1 was carried out in INVSc1, *S. cerevisiae* Yeast Strain. The cloned AiCPR2 plasmid was transformed by using *S.c.* EasyComp™ Transformation Kit and plated on CSM-URA plates. Then competent cells of INVSc1 cells carrying AiCPR2 was made. Then cloned AiCYP1 plasmid was transformed into the competent cells of INVSc1 carrying AiCPR2 plasmid and plated on CSM-URA-TRPA plates. A single colony was inoculated into 2 % glucose containing CSM-URA-TRP and incubated at 30 °C. The overnight grown culture was induced by transferring to 2 % galactose contain CSM-URA-TRP and incubated at 30 °C for 24 h. Triterpene intermediates such as tirucalla-7,24-dien-3β-ol, euphol and tirucallol (4 mg/ 100 mL) were added to induced culture and incubated at 30 °C for 12 h. Cells were collected by centrifugation at 1500 × *g* for 10 min at 4 °C. Saponification was done with 10 % KOH in 80 % ethanol at 70 °C for 2 h, and then extracted thrice with equal volumes of n-hexane. Metabolite extracts were passed through anhydrous sodium sulphate and concentrated. The extracted samples were analyzed by GC-MS with Restek Rtx-5 (30 m × 0.25 mm × 0.32 µm) capillary columns and the program was 80 °C for 2 min, 5 °C/min till 290 °C and hold for 20 min.

5.3.3.2 Cloning and Characterization of AiCYP1 in Neem

5.3.3.2.1 Cloning of AiCYP1 for Overexpression in Neem

To overexpress AiCYP1 in neem, full-length primers for AiCYP1 ORF were designed using the transcript (Neem_Transcripts_34861) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 57 °C for 30 sec, 68 °C for 1.4 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *SalI* and *KpnI* cloning sites of pRI101-AN expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of kanamycin and incubated overnight at 37 °C. Then colony PCR was carried out (with pRI101 forward and reverse primers) to identify the positive clones. Cloning was confirmed by analyzing the sequence obtained by Sanger sequencing of positive clones using pRI101 forward and reverse primers.

5.3.3.2.2 Cloning of AiCYP1 for its Silencing in Neem

To silence AiCYP1 gene in neem, sense primers for first 300 bp of AiCYP1 ORF were designed using their transcript (Neem_Transcripts_34861) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 54 °C for 30 sec, 68 °C for 30 sec followed by final extension at 68 °C for 5 min. PCR product was cloned into *SalI* and *KpnI* cloning sites of pRI101-AN (contains wheat starch branching intron between *KpnI* and *BamHI*) expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of kanamycin and incubated overnight at 37 °C. Then colony PCR was carried out (with pRI101 forward and reverse primers) to identify the positive clones. Cloning was confirmed by analysing the sequence obtained by Sanger sequencing of positive clones using the pRI101 forward and reverse primers.

Antisense primers for first 300 bp of AiCYP1 ORF were designed using their transcript (Neem_Transcripts_34861) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 55 °C for 30 sec, 68 °C for 30 sec followed by final extension at 68 °C for 5 min. PCR product was cloned into *BamHI*

and *SacI* cloning sites of pRI101-AN (contains AiCYP1 first 300 bp and wheat starch branching intron) expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 µg/mL of kanamycin and incubated overnight at 37 °C. Then colony PCR was carried out (with pRI101 forward and reverse primers) to identify the positive clones. Cloning was confirmed by analyzing the sequence of positive clones obtained by Sanger sequencing using the pRI101 primers.

5.3.3.2.3 Characterization of AiCYP1 in Neem

Cloned AiCYP1 overexpression and silencing constructs were transformed into *Agrobacterium tumefaciens* (LBA4404) chemically competent cells and plated on LA containing 50 µg/mL kanamycin and 25 µg/mL rifampicin. *A. tumefaciens* (LBA4404) cells carrying AiCYP1 constructs were grown in 2 mL LB media with the antibiotics at 28 °C and 180 rpm for 2 days. The starter culture was used to inoculate 15 mL LB with the antibiotics and incubated overnight at 28 °C and 180 rpm. After incubation, cells were centrifuged at $2500 \times g$ for 10 mins at 4°C. Cells were washed with half strength MS media, pH 5.4 and centrifuged again. The cell pellet was re-suspended in half strength MS media and maintained $O.D._{600}$ at 0.3-0.5. Acetosyringone of 10 µM was added and incubated at 24 °C and 180 rpm for 4 h. This suspension was used for syringe assisted infiltration of the adaxial surface of neem leaves. Three plants each were taken for over-expression and silencing. *A. tumefaciens* (LBA4404) cells harbouring empty pRI101-AN vector was used as a control. The plants were kept at 25°C, 18 h light and 6 h dark for acclimatization (3 to 4 days) before infection, and maintained at the same conditions for the entire course of the experiment. Leaves were collected in alternate days and stored in -80 °C until further use. Neem leaves (0.1 g) were extracted with methanol (5 mL $\times$ 3), by overnight extraction for 16 h followed by 1 h extraction two times. The pooled methanol layer after concentration under reduced pressure. The extract was partitioned between ethyl acetate (5 mL) and water (5 mL). The organic layer was separated, passed through anhydrous sodium sulphate and concentrated under similar conditions to obtain the crude triterpenoid extract. Extraction of individual tissues was performed in triplicates.

UPLC-ESI(+)-HRMS runs were performed on Q Exactive Orbitrap associated with Accela 1250 pump (Thermo Scientific, MA, USA). Samples were resolved through Acquity BEH C18 UPLC column (2.1×100 mm) of particle size 1.7 μM with a flow rate of 0.2 mL/min and gradient solvent program of 40 min (0.0 min, 40% methanol/water; 5.0 min, 50.0% methanol/water; 10.0 min, 60% methanol/water; 25.0 min, 65% methanol/water; 30.0 min, 90% methanol/water; 32.0 min, 90% methanol/water; 34.0 min, 40% methanol/water; 40.0 min, 40% methanol/water). 0.1% LC-MS grade formic acid was added to water (mobile phase). Profiling experiments were performed in ESI-positive ion mode using the tuning method as follows: sheath gas (nitrogen) flow rate 45 units, auxiliary gas (nitrogen) flow rate 10 units, sweep gas (nitrogen) flow rate 2 units, spray voltage (|KV|) 3.60, spray current (μA) 3.70, capillary temperature 320 °C, s-lens RF level 50, heater temperature 350 °C. ESI(+)-HRMS data were recorded in full scan mode within the mass range *m/z* 100 to 1000. Profiling data were analyzed through Thermo Xcalibur software.

5.3.3.3 Cloning and Characterization of AiCYP2

Full-length primers for AiCYP2 ORF were designed using the transcript (Neem_Transcripts_28933) as a template. cDNA was used for PCR reaction using AccuPrime *Pfx* Supermix (Invitrogen). The program for PCR was 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 sec, 61 °C for 30 sec, 68 °C for 2.0 min followed by final extension at 68 °C for 5 min. PCR product was cloned into *SalI* and *KpnI* cloning sites of pYES2/CT expression vector using T_4 DNA ligase. The ligation mixture was transformed into TOP10 competent cells and plated on LA containing 100 μg/mL of ampicillin and incubated overnight at 37 °C. Then colony PCR was carried out (with T7 promoter and CYC reverse primers) to identify the positive clones. Cloning was confirmed by analyzing the sequence of positive clones obtained by Sanger sequencing using T7 promoter and CYC reverse primers.

The expression of AiCYP2 was carried out in INVSc1, *S. cerevisiae* yeast Strain. The cloned AiCPR2 plasmid was transformed by using *S.c.* EasyComp™ Transformation Kit and plated on CSM-URA plates. Then competent cells were made with an INVSc1 cell harbouring AiCPR2. Cloned AiCYP2 plasmid was transformed into INVSc1+AiCPR2 cells and plated on CSM-URA-TRPA plates. A single colony was inoculated into 2 % glucose contain CSM-URA-TRP and incubated at 30 °C. The

overnight grown culture was induced by transfer to 2 % galactose contain CSM-URA-TRP and incubated at 30 °C for 24 h. Then tirucalla-7,24-dien-3β-ol, euphol and tirucallol (4 mg/ 100ml) was added to induced cells and incubated at 30 °C for 12 h. Cells were collected by centrifugation at 1500 × g for 10 min at 4 °C. Saponification was done with 10 % KOH in 80 % ethanol at 70 °C for 2 h and then extracted thrice with equal volumes of n-hexane. Metabolite extracts were passed through anhydrous sodium sulphate and concentrated. The extracted samples were analyzed by GC-MS with Restek Rtx-5 (30 m × 0.25 mm × 0.32 μm) capillary columns and the program was 80 °C for 2 min, 5 °C/min till 290 °C and hold for 20 min.

5.4 Results and Discussion

5.4.1 Cloning and Characterization of AiCPRs

5.4.1.1 Cloning and Characterization of AiCPR1

AiCPR1 [GenBank: KM108318] ORF of 2088 bp length was found encode for a protein of 695 amino acids. The theoretical molecular weight and pI for this polypeptide were 77.2 kDa and 5.38, respectively. The sequence comparison of AiCPR1 exhibited 83 % identity with CPR from *Gossypium hirsutum* [GenBank: NP_001313876][8], 77 % identity with that from *Fallopia sachalinensis* [GenBank: BAU59414][13] and *Santalum album* [GenBank: ANQ46483][14]. Amino acid sequence comparison between neem AiCPRs showed 70 % identity with each other. All the functional domains involved in the binding of cytochrome P450 monooxygenase, cytochrome C, and other cofactors NADPH, FAD and FMN were identified for both AiCPRs (Figure 5.5). AiCPR1 belongs to Class I family of dicot cytochrome P450 reductases classification as it shows a close relation to *Arabidopsis thaliana* CPR1 (Figure 5.6).

```
AiCPR1   1   MSNSGTG-----------NDIVKFVESA------------------LGVSLG-SSVTDTVIVIATTLFAVVIGLLVLAWK   50
AiCPR2   1   MQSSSSSASSSNTMKVSPFDIMSAIIKG----KTDPSNVSSSG-SGLEVASIVLENKEFVMILTTSIAVLIGCVVVLIWR   75
AtCPR1   1   MTSALYA-----------SDLFKQLKSI------------------MGTDSL---SDDVVLVIATTSIALVAGFVVVLLWK   48
AtCPR2   1   MSSSSS-------SSTSMIDLMAAIIKGEPVIVSDPANASAYESVAAEISSMLIENRQFAMIVTISIAVLIGCIVMLVWR   73
SaCPR    1   MSSSSE------------IWKSIGSA------------------LGVSPPPAWEEWAAVIVTTSAALIVGFVMFMWR      48
AaCPR    1   MQSTTS-------VKLSPFDIMTALLNG----KVSFDTSNTSD-TNIPLA-VFMENRELLMILTTSVAVLIGCVVLVWR   67
VsCPR    1   MTSSNS-------------DLVRTIESA------------------LGISLG-DSVSDSVVIIATTSAAVIIGLLVFLWR  48
```

FMN-Pyrophosphate

```
AiCPR1   51  KS-SDRSKEVKPVVPLKSPVLKEDHDEADIVWGKTKVTIFYGTQTGTAEGFAKALAEEIKARYEKAAVKVVDLDDYAAD   129
AiCPR2   76  RS---SSQKPKKIEPLKPLVVKEP---PVEVDDGKQKJTIFFGTQTGTAEGFAKALADAAKARYDKAIFKVVDLDDYAAD  149
AtCPR1   49  KTTADRSGELKPLMIPKSLMAKDED-DDLDLGSGKTRVSIFFGTQTGTAEGFAKAISEEIKARYEKAAVKVIDLDDYAAD  127
AtCPR2   74  RS---GSGNSKRVEPLKPLVIKPR---EEEIDDGRKKVTIFFGTQTGTAEGFAKALGEEAKARYEKTREKIVDLDDYAAD  147
SaCPR    49  RS-GEKSKELRPVVALKAAPIEAEE-DDGEVDSGKTKVTVFFGTQTGTAEGFAKALAEAIKARYEKAVVKVVDLDDYAAD  126
AaCPR    68  RS---SSAAKKAAESPVIVVPKKVT--EDEVDDGRKTKVFFGTQTGTAEGFAKALVEEAKARYEKAVFKVIDLDDYAAE  142
VsCPR    49  KS-PDRSRELRPVIVPKFTVKHED--DEVEVDRGKTKVTVFYGTQTGTAEGFAKALAEEIKARYEKAVVKVVDYDDYAID  125
```

FMN-isoalloxazine-ring

```
AiCPR1   130 DEQYGEKLKKETLIFFMVATYGDGEPTDNAARFYKWFTE-ENDRGVWLQQLRYGVFGLGNRQYEHFNKICKVLDEELCKQ  208
AiCPR2   150 DEEYEEKLKKESIAFFFLATYGDGEPTDNAARFYKWFTE-GKERGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDDVLAEQ  228
AtCPR1   128 DDQYEEKLKKETLAFFCVATYGDGEPTDNAARFSKWFTE-ENERDIKLQQLAYGVFALGNRQYEHFNKIGIVLDEELCKK  206
AtCPR2   148 DDEYEEKLKKEDVAFFFLATYGDGEPTDNAARFYKWFTE-GNDRGEWLKNLKYGVFGLGNRQYEHFNKVAKVVDDILVEQ  226
SaCPR    127 DNQYGEKLKNETLIFFMVATYGDGEPTDNAARFYKWFTE-EKEREAWLQQLTYGIFGLGNRQYEHFNKIAKVLDEQLTEQ  205
AaCPR    143 DEEYEEKLKKESLAFFFLATYGDGEPTDNAARFYKWFTE-GEEKGEWLDKLQYAVFGLGNRQYEHFNKIAKVVDEKLVEQ  221
VsCPR    126 DDQYEEKLKKETLVFFMLATYGDGEPTDNAARFYKWFTEGKEERGTWLQQLTYGVFALGNRQYEHFNKICKIVDEDLTEQ  205
```

CYP450, Cytochrome C

```
AiCPR1   209 GGKRLVPVGLGDDDQCIEDDFAAWRELVWPELDQLLRDEDDANTVSTPYTAAIPEYRVVIHDPTATSFVDKYSNIPNGNA  288
AiCPR2   229 GGKRLVPVGLGDDDQCIEDDFSSWRELVWPELDKLLRDDDDPTTVSTPYTAAISEYRVVFYDSADASIGENNWSNANGHA  308
AtCPR1   207 GAKRLIEVGLGDDDQSIEDDFNAWKESLWSELDKLLKDEDDK-SVATPYTAVIPEYRVVTHDPRFTTQKSMESNVANGNT  285
AtCPR2   227 GAQRLVQVGLGDDDQCIEDDFNAWREALWPELDTILREEGD-TAVATPYTAAVLEYRVSIHDSEDAKFNDITLANGNGYT  305
SaCPR    206 GAKRLIQVGLGDDDQCIEDDFSAWRELLWPELDQLLRGDDGANSVSTPYTAAVPEYRVVIHDPTITSSEDKSLAMANGTA  285
AaCPR    222 GAKRLVPVGMGDDDQSIEDDFTAWKELVWPELDQLLRDEDD-TSVATPYTAAVAEYRVVHDKPETYDQDQ---LTNGHA  297
VsCPR    206 GAKRLVPVGLGDDDQSIEDDFNAWKETLWPELDQLLRDEDDVNTASTPYTAAISEYRVVIHDPTVSPSYENHFNVANGCA  285
```

FAD-Pyrophosphate

```
AiCPR1   289 SFDIHHPCRVNVAVQRELHKPESDRSCIHLEFDISGTGITYETGDHVGIYAENCDETVEQAGKLLGQPLELLFSIHADNE  368
AiCPR2   309 VYDAQHPCRSNVAVRKELHTPLSDRSCTHLEFDIAGTGLMYETGDHVGVFCENLTETVEEALSLLGLSPDTYFSVHTDKE  388
AtCPR1   286 TIDIHHPCRVDVAVQKELHTHESDRSCIHLEFDISRTGITYETGDHVGVYAENHVEIVEEAGKLGHSLDLVFSIHADKE  365
AtCPR2   306 VFDAQHPYKANVAVKRELHTPESDRSCIHLEFDIAGSGLIMKLGDHVGVLCDNLSETVDEALRLLDMSPDTYFSLHAEKE  385
SaCPR    286 LFDIHHPCRVKVAVQRELHKADSDRSCIHLEFDISGTGLMYETGDHVGVYAENCVETVEEAGKLLGQPLDLLFSVHTDKD  365
AaCPR    298 VHDAQHPCRSNVAVRRELHKSPLSDRSCTHLEFDINTGLSYETGDHVGVYVENLSEVVDEAEKIIGLPPHTYFSVHADNE  377
VsCPR    286 VFDIHHPCRVNVAVRRELHKPQSDRSCIHLEFDLSGTGVTYETGDHVGVYAENCDETVEEAGKLLGQSLDLLFSLHTDKE  365
```

```
AiCPR1   369 DGTPIG-SSLPPPFPGPCTLRSAIAQYADLLNPPRKAALTALAAHATEPSEAERLKFLSSPQGKDEYSQWIVASQRSLIE  447
AiCPR2   389 DGTPLGGSSLPPPFP-PCSLRTALAQYADLLSSPKKSALLALAAHASDPTEADRLRHLASPAGKDEYAQWIVASQRSLLE  467
AtCPR1   366 DGSPLE-SAVPPPFPGPCTLGTGLARYADLLNPPRKSALVALAAMATEPSEAEKLKHLTSPDGKDEYSQWIVASQRSLLE  444
AtCPR2   386 DGTPIS-SSLPPPFP-PCNLRTALTRYACLLSSPKKSALVALAAHASDPTEAERLKHLASPAGKDEYSKWVVESQRSLLE  463
SaCPR    366 DGTSLE-SSLPPPFPGPCTLRTALFQYADLLNPPRKAALVALAAHAVEPSEADRLKFLSSPQGKDEYAKWIVVGSQRSLLE  444
AaCPR    378 DGTPLGGASLPPPFP-PCTLRKALASYADVLSSPKKSALLALAAHATDSTEADRLKFLASPAGKDEYAQWIVASHRSLLE  456
VsCPR    366 DGTSLG-GSLLPPFPGPCTVRTALAQYADLLNPPRKAAIVALAAHASEPSEAERLKFLSSPQGKDEYSKWVVGSQRSLLE  444
```

FAD-isoalloxazine-ring

```
AiCPR1   448 VMAEFQSAKPPIGVFFAAVAPRLQPRYYSISSSPRFGPNRVHVTCALVYGPTPTGRIHKGVCSTWMKNAIPLEKSQDCSW  527
AiCPR2   468 VMACFPSAKPPLGVFFAAVAPRLQPRYYSISSSPRVAPSRIHVTCALVYEKTPTGRIHKGVCSTWMKNCVPMEKSNDCGW  547
AtCPR1   445 VMAAFPSAKPPLGVFFAAIAPRLQPRYYSISSCQDWAPSRVHVTSALVYGPTPTGRIHKGVCSTWMKNAVPAEKSHECSG  524
AtCPR2   464 VMAEFPSIKVPLGVFFAAVAPRLQPRFYSISSSPKIAETRIHVTCALVYEKMPTGRIHKGVCSTWMKNAVPYEKSEKLFL  543
SaCPR    445 VMAEFPSIKVPLGVFFAAVAPRLQPRYYSISSSPRFSSDRVHVTCALVYGFSPTGRIHRGVCSTWMKNSVPLEESRECSW  524
AaCPR    457 VMEAFPSAKPPLGVFFASVAPRLQPRYYSISSSPREAPNRIHVTCALVYEQTPSGRVHKGVCSTWMKNAVPMTESQDCSW  536
VsCPR    445 VMADFPSAKPPLGVFFAAIAPRLQPRYYSISSSPREAPQRVHVTCALVEGPTPTGRIHKGVCSTWMKSATPLEKSHDCSR  524
```

NADPH-ribose, Pyrophosphate

```
AiCPR1   528 A-PVFIRPSNFKLPTNPSVPVIMIGPGTGLAPFRGFLQERVALKLDGNDLGPALLFFGCRNRRMDFIYKDELYNFVDQGV  606
AiCPR2   548 A-PIFVRQSNFRLPADPKVPVIMIGPGTGLAPFRGFLQERFALKEAGAELGPSVLFFGCRNRMDYIYEDELNNFVQSGA  626
AtCPR1   525 A-PIFIRASNFKLPSNPSTPIVMVGPGTGLAPFRGFLQERVALKDDGEELGSSLLFFGCRNRQMDFIYEDELNNFVDQGV  603
AtCPR2   544 GRPIFVRQSNFKLPSDSKVPIIMIGPGTGLAPFRGFLQERIALVESGVELGPSVLFFGCRNRRMDFIYEEELQRFVESGA  623
SaCPR    525 A-PIFIRTSNFKLPSDPKVPVIMVGPGTGLAPFRGFLQERVALLEGAQLGPALLFFGCRNRQMDFIYEDELNNFVEQGV  603
AaCPR    537 A-PIYVRTSNFRLPSDPKVPVIMIGPGTGLAPFRGFLQERLAQKEAGTELGTAILFFGCRNRKVDFIYEDELNNFVETGA  615
VsCPR    525 A-PIFIRPSNFKLPADHSIPIIMVGPGTGLAPFRGFLQERVALKDGVQLGPALLFFGCRNRQMDFIYEDELNNFVQQGA  603
```

NADPH-nicotinamide

```
AiCPR1   607 ISELIVAFSREGPQKEYVQHKMMDKAAYLWSLLSKDGYLYVCGDAKGMARDVHRTLHTIIQEQENVDSSTAEAIVKKLQM  686
AiCPR2   627 ISELVVAFSRQGPTKEYVQHKMMEKASDIWNMISQGGYLYVCGDAKGMARDVHRTLHTIVQEQGSVDSSKAESIVKNLQM  706
AtCPR1   604 ISELIMAFSREGAQKEYVQHKMMEKAAQVWDLIKEEGYLYVCGDAKGMARDVHRTLHTIVQEQGCVSSSSEAEAIVKKLQT  683
AtCPR2   624 LAEELSVAFSREGPTKEYVQHKMMDKASDIWNMISQGAYLYVCGDAKGMARDVHRSLHTIAQEQGSMDSLKAEGFVKNLQT  703
SaCPR    604 ISELIVAFSRDGPTKEYVQHKMMDKAAYIWSLISQGAYLYVCGDAKGMARDVHRTLHTLVQQQESVDSSKAESIVKKLQM  683
AaCPR    616 ISELVTAFSREGATKEYVQHKMTQKASDIWNLISEGAYLYVCGDAKGMARDVHRTLHTIVQEQGSIDSSKAELYVKNLQM  695
VsCPR    604 ISELIVAFSREGPEKEYVQHKMMDKAEYLWSLISQGGYLYVCGDAKGMARDVHRSLHTIVQQQENADSSKAEATVKKLQM  683
```

```
AiCPR1   687 EGRYLRDVW   695
AiCPR2   707 TGRYLRDVW   715
AtCPR1   684 EGRYLRDVW   692
AtCPR2   704 SGRYLRDVW   712
SaCPR    684 DGRYLRDVW   692
AaCPR    696 AGRYLRDVW   704
VsCPR    684 DGRYLRDVW   692
```

Figure 5. 5 Multiple Sequence Alignment of *A. indica* NADPH-Cytochrome P450 Reductases.

Amino acid sequences of AiCPR1 (AIG15451, *A. indica*), AiCPR2 (AIG15452, *A. indica*), AtCPR1 (CAA46814, *A. thaliana*), AtCPR (CAA46815, *A. thaliana*), SaCPR (ANQ46483,

Santalum album), AaCPR (ABI98819, *Artemisia annua*) and VsCPR (CAA81211, *Vicia sativa*) are used for multiple sequence alignment. The highly conserved DCTAE motifs are indicated in blue colour letters.
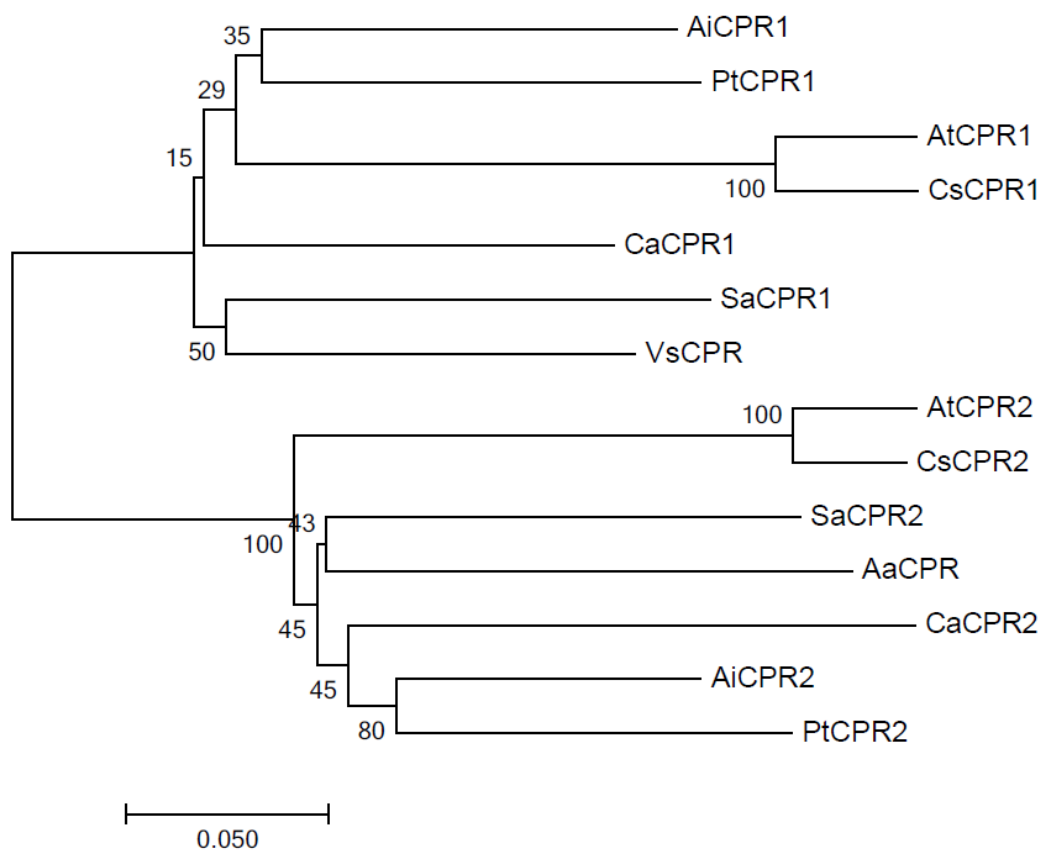

Figure 5. 6 Phylogenetic Analysis of Neem Cytochrome P450 Reductases.

GeneBank Accession numbers used for construction of phylogenetic analysis are AiCPR1 (AIG15451, *Azadirachta indica*), AiCPR2 (AIG15452, *A. indica*), AtCPR1 (CAA46814, *A. thaliana*), AtCPR2 (CAA46815, *A. thaliana*), SaCPR1 (ANQ46483, *Santalum album*), AaCPR (ABI98819, *Artemisia annua*), VsCPR (CAA81211, *Vicia sativa*), CsCPR2 (XP_010433014, *Camelina sativa*), CsCPR1 (XP_010433757, *Camelina sativa*), CaCPR1 (CDI59405, *Capsicum annuum*), CaCPR2 (CDI59406, *Capsicum annuum*), SaCPR2 (AHB33950, *Santalum album*), PtCPR1 (AAK15259, *Populus trichocarpa* x *P. deltoids*), PtCPR2 (AAK15260, *Populus trichocarpa* x *P. deltoids*).
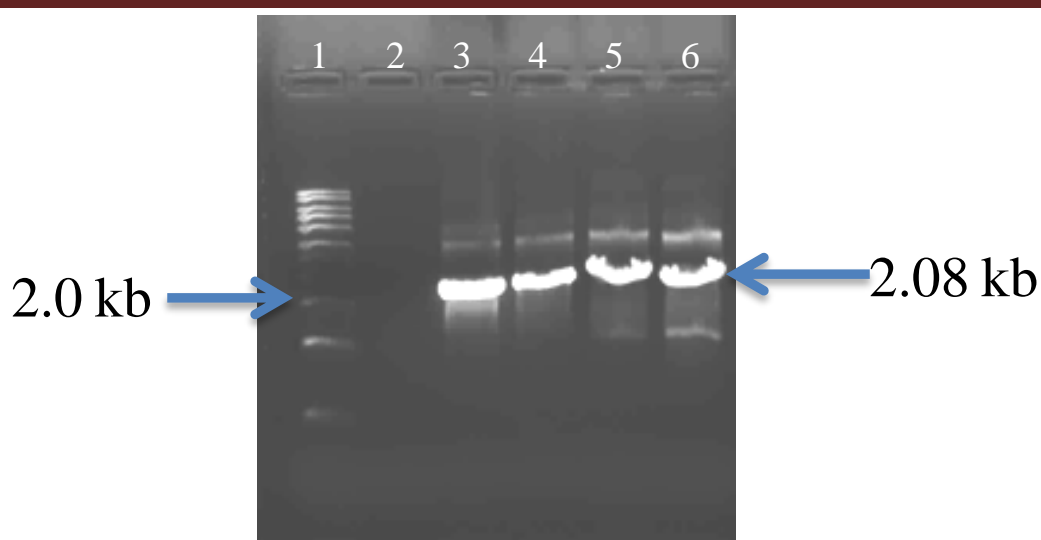
Figure 5. 7 AiCPR1 ORF Amplification.
Lane 1: 1 kb DNA ladder Invitrogen (Addendum Figure A1.B), **Lane 2:** Negative control, **Lane 3:** AiCPR1 PCR product at 61 °C, **Lane 4:** AiCPR1 PCR product at 62 °C, **Lane 5:** AiCPR1 PCR product at 63 °C and **Lane 6:** AiCPR1 PCR product at 64 °C.
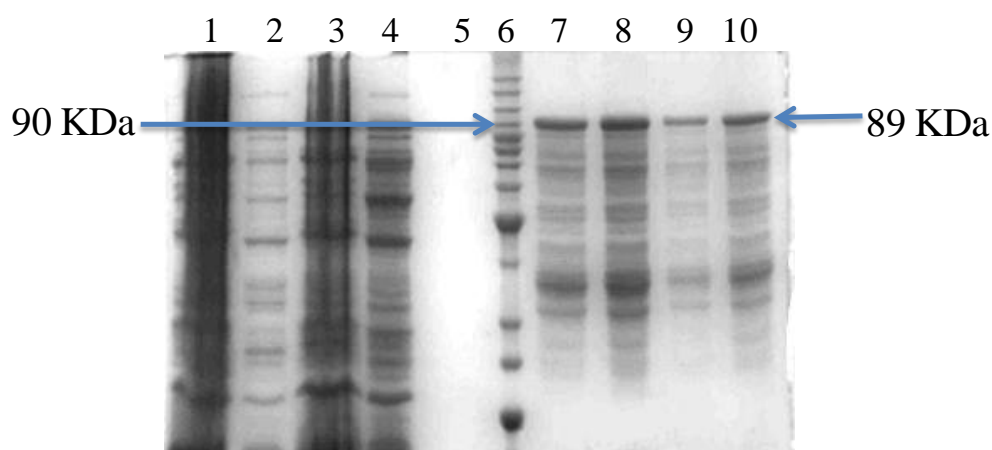

Figure 5. 8 SDS-PAGE for AiCPR1 Protein Purification in pET32a.
Lane1: Un-induced fraction, **Lane 2:** Supernatant fraction, **Lane 3:** Pellet fraction, **Lane 4:** Unbound fraction, **Lane 5:** Wash fraction 1, **Lane 6:** BenchMark™ Protein Ladder (Addendum Figure A2.B), **Lane 7, 8, 9 and 10:** Elution fractions.

AiCPR1 was cloned into a pET32a expression vector (Figure 5.7). The cloned construct was transformed into Rosetta 2 (DE3) cells and expressed. AiCPR1 was obtained as soluble form and purified by Ni-NTA affinity column chromatography. The recombinant protein was approximately 50 % pure as analyzed by SDS-PAGE (Figure 5.8). The partially purified AiCPR1 was incubated with cytochrome C and potassium ferricyanide. Spectrophotometric analyses of the assay indicated the

reduction of cytochrome C and potassium ferricyanide, which confirmes AiCPR1 was NADPH-cytochrome P450 reductase (Figure 5.9).
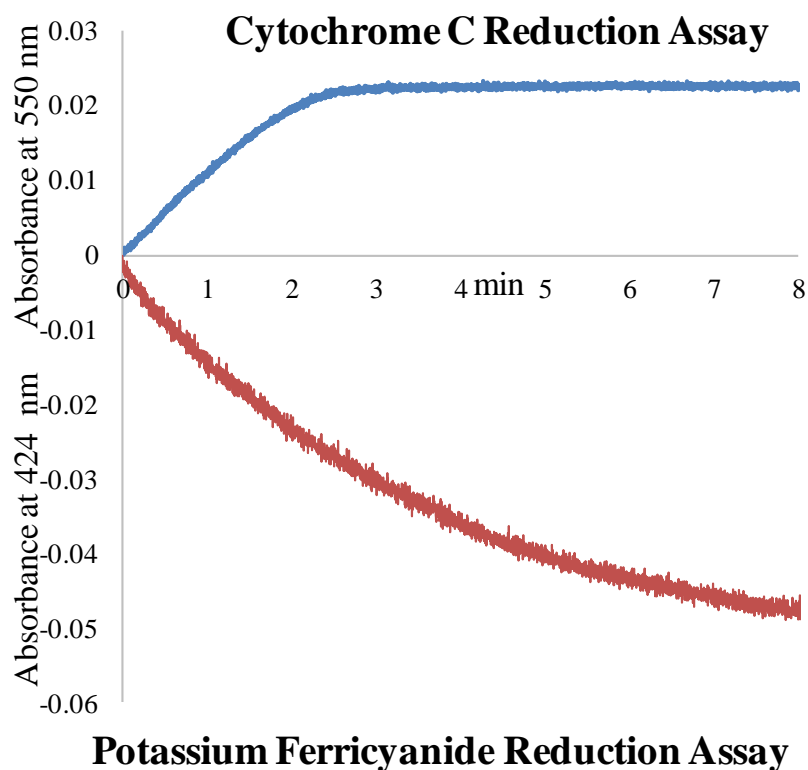
Figure 5. 9 Spectrophotometric Absorbances of Cytochrome C and Potassium Ferricyanide Reduction by AiCPR1.

5.4.1.2 Cloning and Characterization of AiCPR2

AiCPR2 [GenBank: KM108319] ORF of 2088 bp length was found to be encoded for a protein of 695 amino acids. The theoretical molecular weight and pI for this polypeptide were 78.9 kDa and 5.35, respectively. The sequence comparison of AiCPR2 exhibited 82 % identity with CPR3 from *Populus trichocarpa x Populus deltoides* [GenBank: AAK15261][7], 79 % identity with that from *Camptotheca acuminate* [GenBank: AJW67229][15] and 78% identity with *Artemisia annua* [GenBank: ABI98819][14,16]. All the functional domains involved in the binding of cytochrome P450 monooxygenase, cytochrome C, and other cofactors NADPH, FAD and FMN were identified for both AiCPR2 (Figure 5.5). AiCPR2 belongs to Class 2 family of dicot cytochrome P450 reductases classification as it shows a close relation to *Arabidopsis thaliana* CPR2 (Figure 5.6).
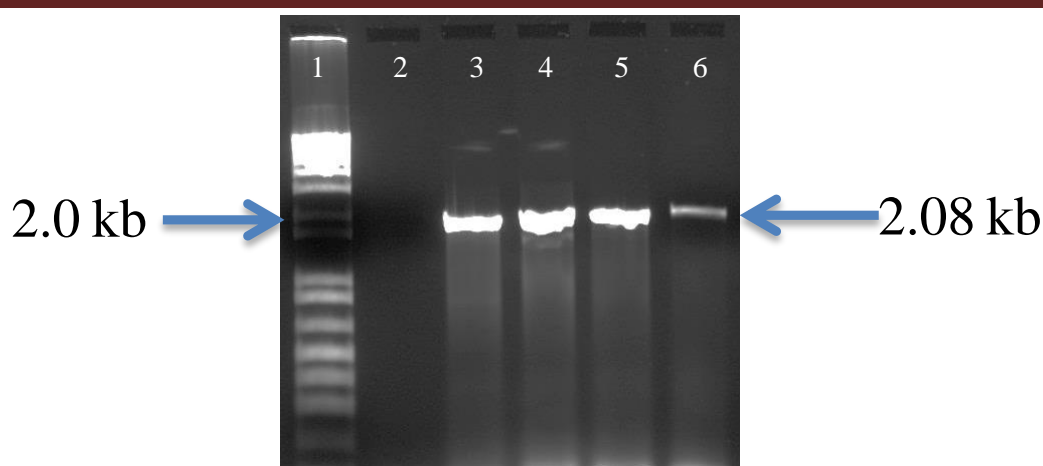
Figure 5. 10 AiCPR2 ORF Amplification.
Lane 1: 1 kb DNA ladder Invitrogen (Addendum Figure A1.2), **Lane 2:** Negative control, **Lane 3:** AiCPR2 PCR product at 56 °C, **Lane 4:** AiCPR2 PCR product at 58 °C, **Lane 5:** AiCPR1 PCR product at 60 °C and **Lane 6:** AiCPR1 PCR product at 62 °C.
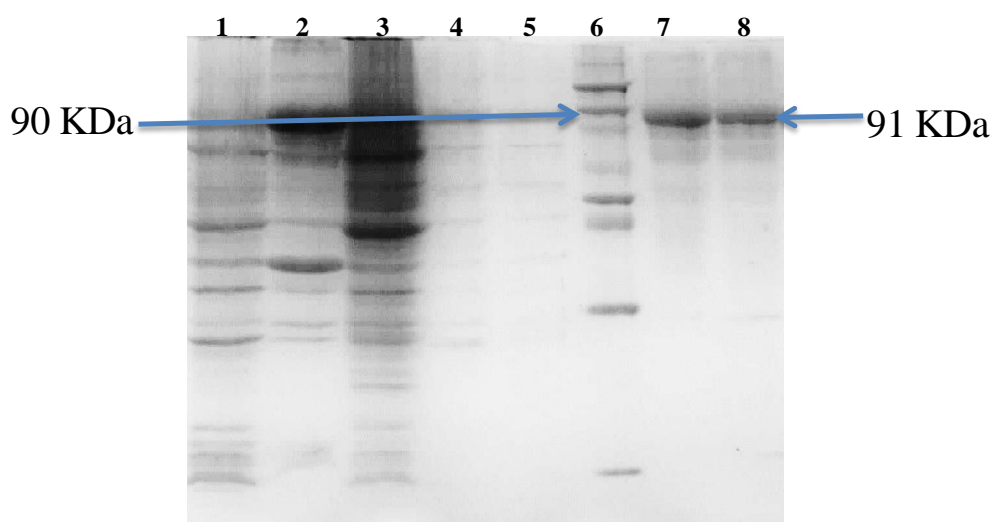

Figure 5. 11 SDS-PAGE for AiCPR2 Protein Purification in pET32a.

Lane1: Un-induced fraction, **Lane 2:** Supernatant fraction, **Lane 3:** Pellet fraction, **Lane 4:** Unbound fraction, **Lane 5:** Wash fraction 1, **Lane 6:** BenchMark™ Protein Ladder (Addendum Figure A2.B), **Lane 7 and 8:** Elution fractions.

AiCPR2 was cloned into a pET32a expression vector (Figure 5.10). The cloned construct was transformed into Rosetta 2 (DE3) cells and expressed. AiCPR2 was obtained as soluble form and purified by Ni-NTA affinity column chromatography. The recombinant protein was approximate 98 % pure as analyzed by SDS-PAGE (Figure 5.11). The purified AiCPR1 was incubated cytochrome C and potassium ferricyanide. Spectrophotometric analyses of the assay indicated the

reduction of cytochrome C and potassium ferricyanide which confirmes AiCPR1 was NADPH-cytochrome P450 reductase (Figure 5.12).
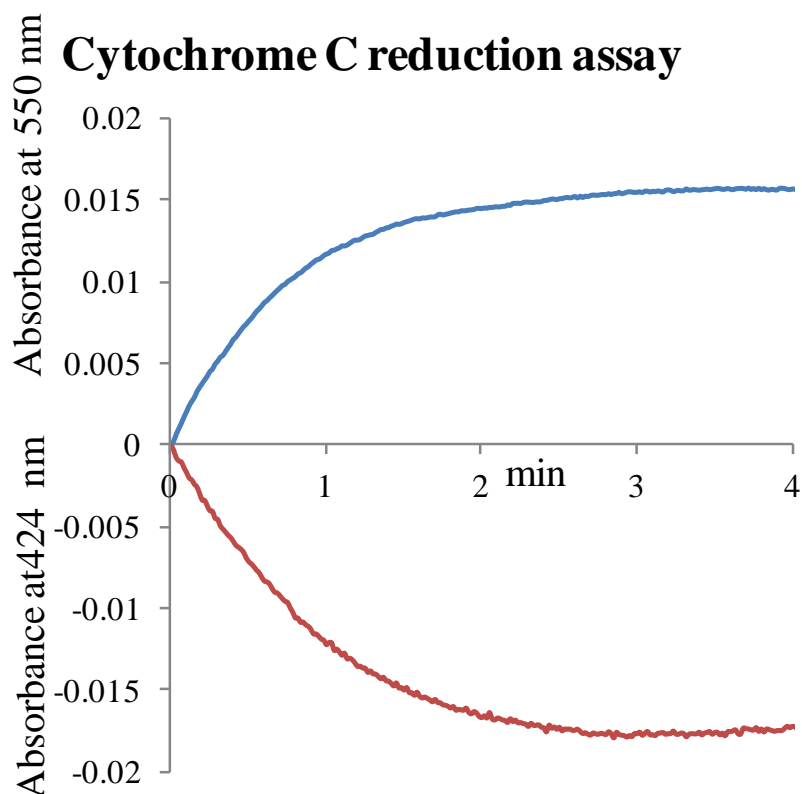
Figure 5. 12 Spectrophotometric Absorbances of Cytochrome C and Potassium Ferricyanide Reduction by AiCPR2.

5.4.2 Cloning and Characterization of AiCYPs

5.4.2.1 Cloning of AiCPR2 in pYES3/CT Vector

AiCPR2 was cloned into a pETYES3/CT expression vector and the construct was expressed in INVSc1 yeast strain. Then competent cells were made with INVSc1 cells having AiCPR2 in pYES3/CT vector.

Figure 5. 13 AiCPR2 ORF Amplification.
Lane 1: 1 kb DNA ladder Invitrogen (Addendum Figure A1.B), **Lane 2:** Negative control, **Lane 3:** AiCPR2 PCR product at 60 °C, **Lane 4:** AiCPR2 PCR product at 62 °C, **Lane 5:** AiCPR1 PCR product at 64 °C and **Lane 6:** AiCPR1 PCR product at 66 °C

5.4.2.2 Cloning and Characterization of AiCYP1

5.4.2.2.1 Cloning and Characterization of AiCYP1 in Yeast

AiCYP1 [Neem_Transcript_34861] ORF of 1437 bp length was found to be encoded for a protein of 478 amino acids. The theoretical molecular weight and pI for this polypeptide were 55.2 kDa and 9.21, respectively. The sequence comparison of AiCPR1 exhibited 43 % identity with protopanaxadiol 6-hydroxylase from *Panax ginseng* [GenBank: I7CT85][17], 42 % identity with dammarenediol 12-hydroxylase [GenBank: H2DH16][18] and 41% identity with β-amyrin 28-oxidase [GenBank: I7C6E8][14,17]. N terminal transmembrane domain, Proline cluster, oxygen binding, Heme binding motif and E-R-R triad for locking the heme pocket into position and to assure stabilization of the conserved core structure are found in AiCYP1[19,20] (Figure 5.14). AiCYP1 belongs to CYP716 family and may be involved in the modification of phytol chain in protolimonoids (Figure 5.15).
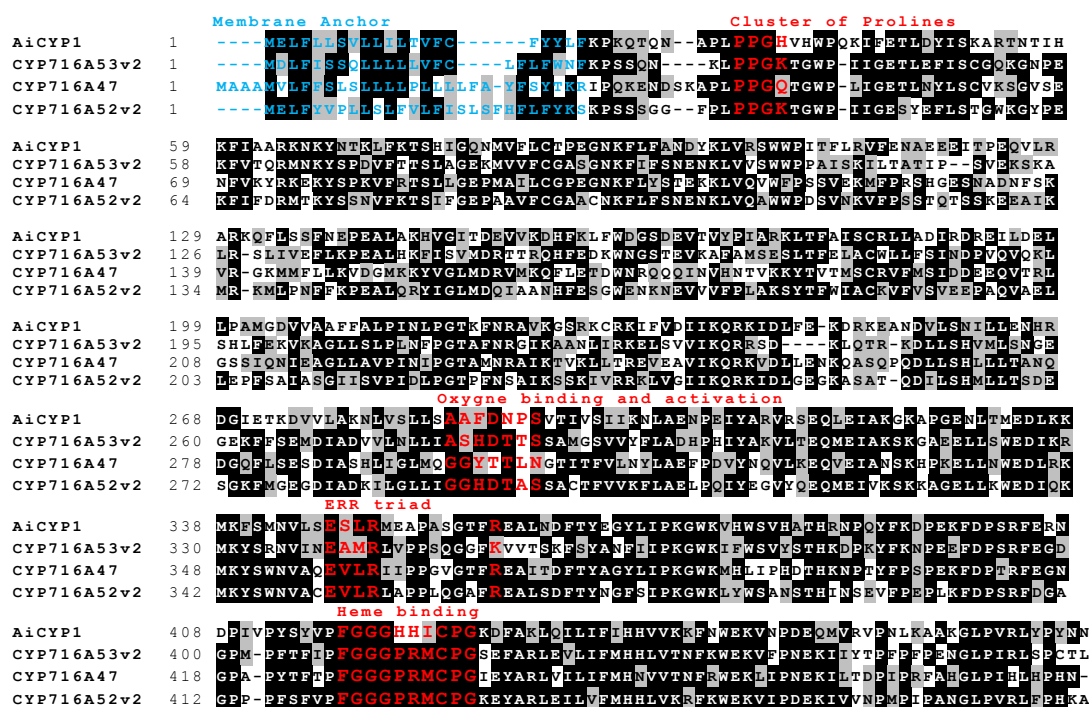
Figure 5. 14 Multiple Sequence Alignment of *A. indica* Cytochrome P450 1 (AiCYP1)
Amino acid sequences of AiCYP1 (*A. indica*), CYP716A53v2 (*Panax ginseng*, I7CT85), CYP716A47 (*Panax ginseng*, H2DH16) and CYP716A52v2 (*Panax ginseng*, I7C6E8) are used for multiple sequence alignment. The transmembrane domain is indicated in blue colour. The highly conserved cluster of proline, oxygen binding, EER triad and heme binding motifs is indicated in red colour letters.
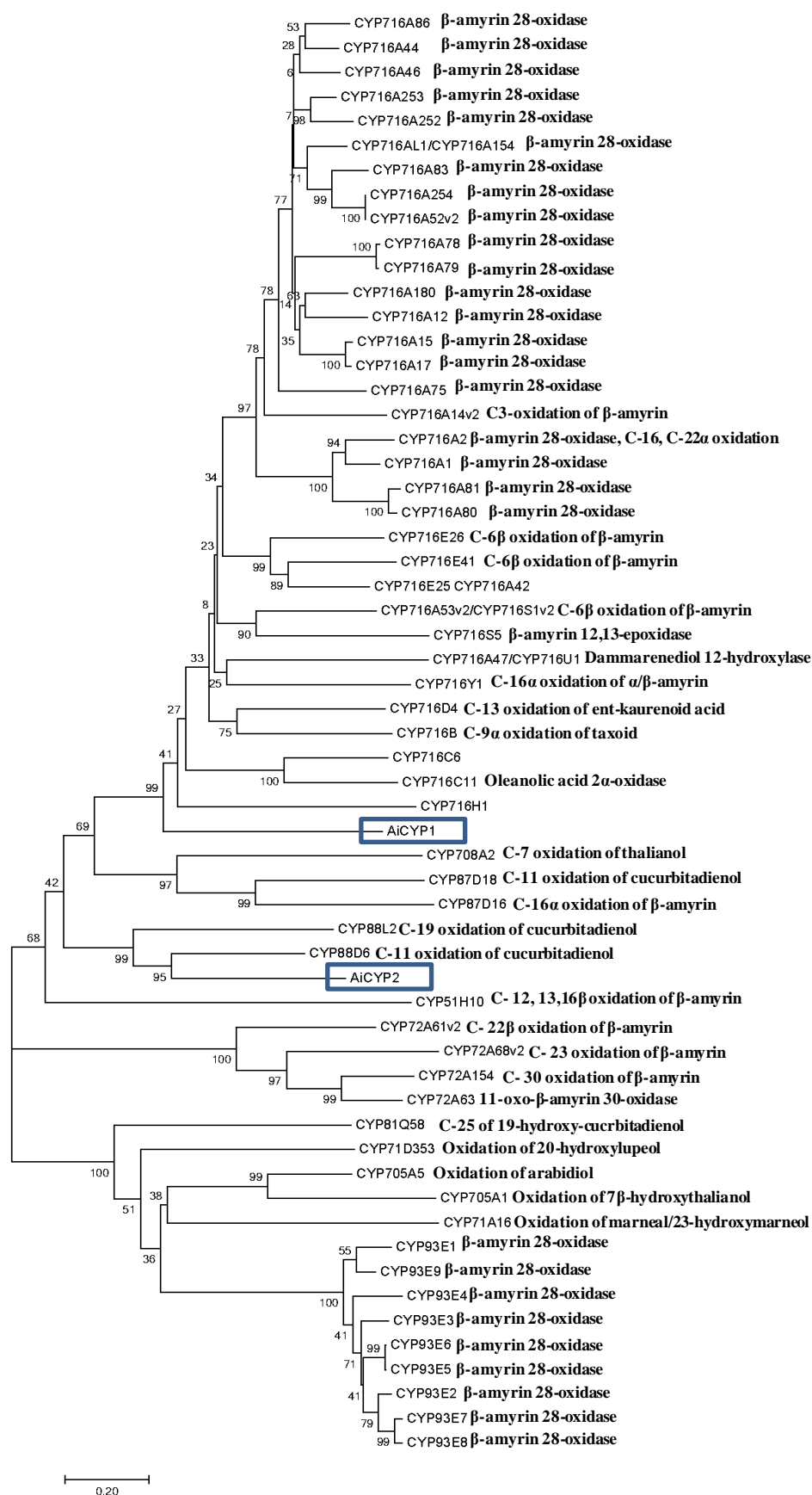
Figure 5. 15 Phylogenetic Analysis of AiCYP1 and AiCYP2.

AiCYP1 was cloned into a pYES2/ct expression vector (Figure 5.16). The cloned construct was transformed into INVSc1 cells containing AiCPR2. Expression was done under 2% galactose. It was then incubated with tirucalla-7,24-dien-3β-ol, euphol and tirucallol (4 mg/ 100 mL) incubated for 12 h. Crude n-hexane metabolite extracts of saponified AiCYP1 yeast cells were analyzed on GC-MS. Hydroxylated triterpene cyclic product was not observed from GC-MS data analysis (Figure 5.17). Hence, tirucalla-7,24-dien-3β-ol, euphol and tirucallol may not be the substrates for AiCYP1.
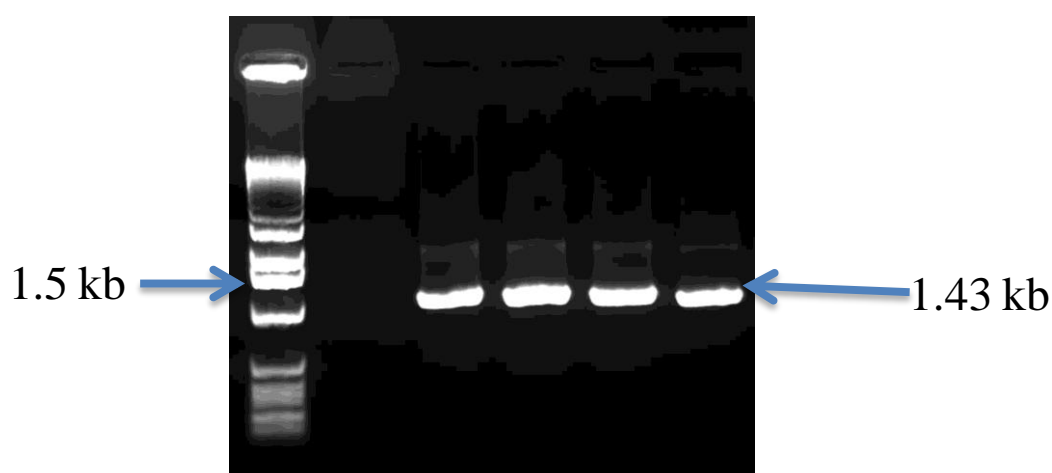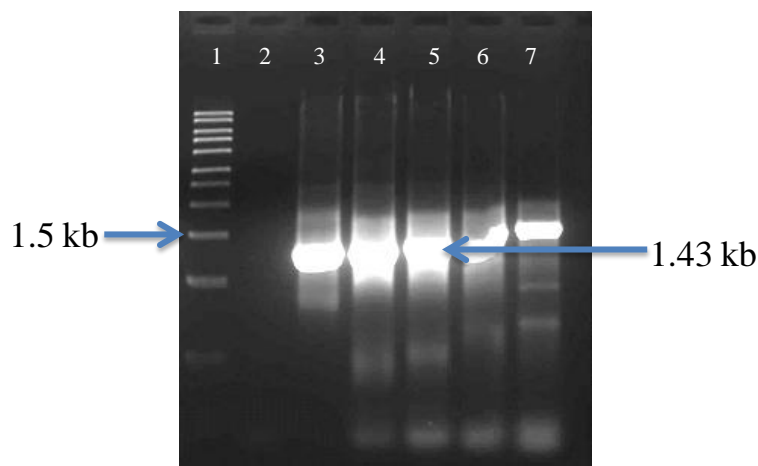

Figure 5. 16 AiCYP1 ORF Amplification.
Lane 1: 1 kb DNA ladder Invitrogen (Addendum Figure A1.B), **Lane 2:** Negative control, **Lane 3:** AiCYP1 PCR product at 60 °C, **Lane 4:** AiCYP1 PCR product at 61 °C, **Lane 5:** AiCYP1 PCR product at 62 °C and **Lane 6:** AiCYP1 PCR product at 63 °C.
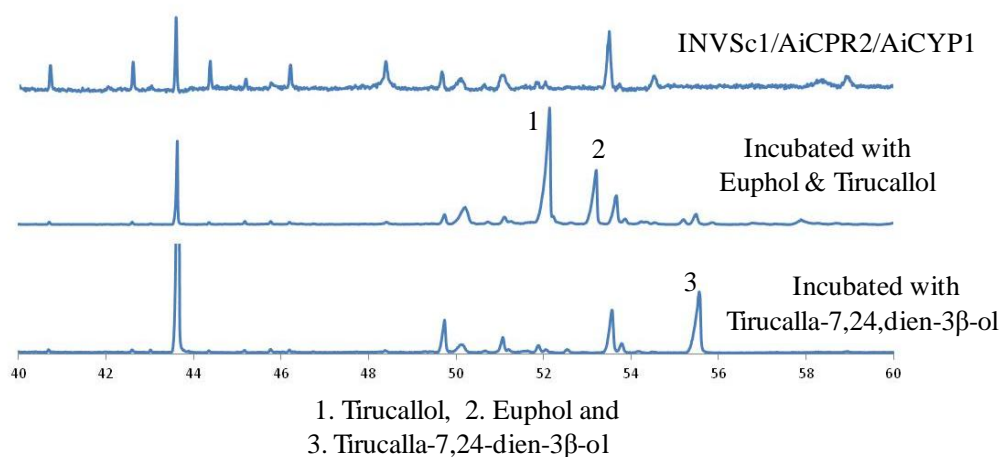
Figure 5. 17 Total Ion Chromatograms of AiCYP1/AiCPR2-INVSc1 Metabolite Extract.

5.4.2.2.2 Cloning and Characterization of AiCYP1 in Neem

AiCYP1 ORF was cloned into a pRI101-AN expression vector (Figure 5.18). Sence (Figure 5.19) and anti-sense (Figure 5.20) primers are used to clone AiCYP1 first 300 bp in PRI101-AN which contains wheat starch branching intron. The cloned construct was transformed into *Agrobacterium tumefaciens* (LBA4404).

Figure 5. 18 AiCYP1 ORF Amplification for Cloning into pRI101-AN.
Lane 1: 1 kb DNA ladder Sigma (Addendum Figure A1.A), **Lane 2:** Negative control, **Lane 3:** AiCYP1 PCR product at 55 °C, **Lane 4:** AiCYP1 PCR product at 57 °C, **Lane 5:** AiCYP1 PCR product at 59 °C, **Lane 6:** AiCYP1 PCR product at 61 °C and **Lane 7:** AiCYP1 PCR product at 63 °C.
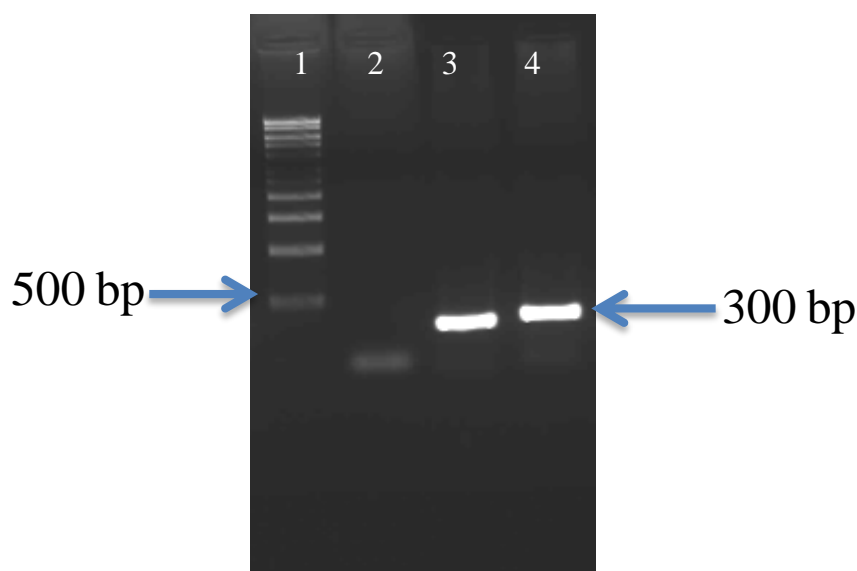
**Figure 5. 19 AiCYP1 Sence 300 bp Amplification.**

Lane 1: 1 kb DNA ladder Sigma (Addendum Figure A1.A), **Lane 2:** Negative control, **Lane 3:** Sence PCR product at 58 °C **and Lane 4:** Sence AiCYP1 PCR product at 60 °C.
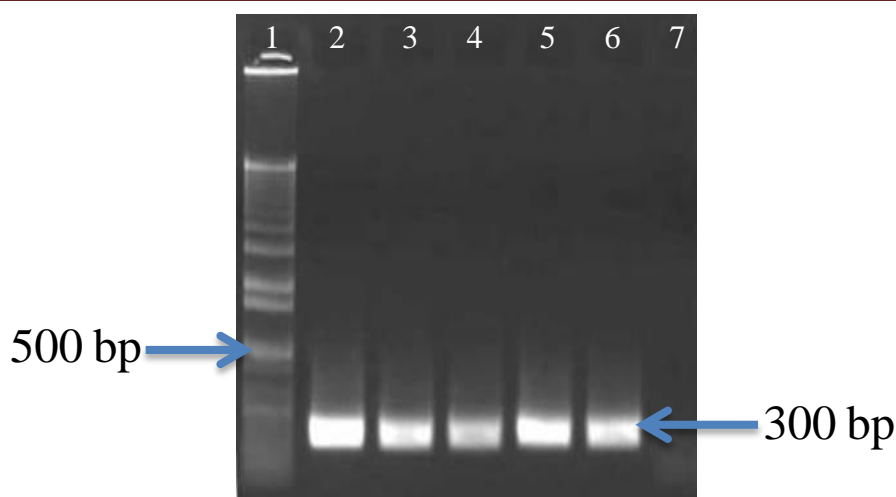
Figure 5. 20 AiCYP1 Antisense 300 bp Amplification.
Lane 1: 1 kb Plus DNA ladder Thermo Fisher (Addendum Figure A1.C), **Lane 2:** Antisense PCR product at 55 °C, **Lane 3:** Antisense PCR product at 57 °C, **Lane 4:** Antisense PCR product at 59 °C, **Lane 5:** Antisense PCR product at 61 °C, **Lane 6:** Antisense PCR product at 63 °C and **Lane 7:** Negative control.

A. tumefaciens (LBA4404) harbouring AiCYP1 constructs were transiently transfected into neem leaves through syringe agroinfiltration. Neem leaves were collected on alternate days for six days. Methanol extracts of leaves are analyzed through Q Exactive Orbitrap associated with Accela 1250 pump (Thermo Scientific, MA, USA). Quantitative profiling of azadirachtin A was analyzed through Thermo Xcalibur software by considering azadirachtin A standard graph[21]. In AiCYP1 overexpression, azadirachtin A production was observed maximum in day 2 (3.5 folds) and day 4 (2.09 folds) as compared to pRI101 vector control. In AiCYP1 silencing, day 4 showed the highest effect i.e., 5 fold lesser was observed in AiCYP1 silenced neem leaves as compared to vector control (Figure 5.21). This analysis confirms the AiCYP1 involves in neem limonoid biosynthesis.
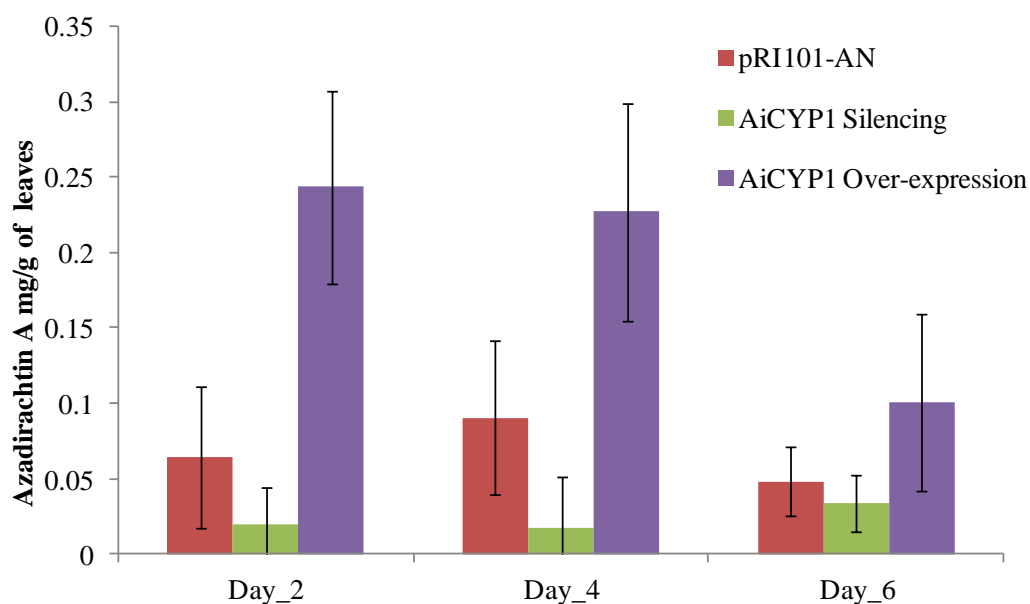
Figure 5. 21 Neem Transient Transformation of AiCYP1.

5.4.2.3 Cloning and Characterization of AiCYP2 in Yeast

AiCYP2 [Neem_Transcript_38933] ORF of 1473 bp length was found to be encoded for a protein of 490 amino acids. The theoretical molecular weight and pI for this polypeptide were 56.1 kDa and 8.38, respectively. The sequence comparison of AiCYP2 exhibited 51 % identity with β-amyrin 11-oxidase from *Glycyrrhiza uralensis* [GenBank: B5BSX1][22], 49 % identity with ent-kaurenoic acid oxidase 2 from *Arabidopsis thaliana* [GenBank: Q9C5Y2][23] and 47% identity with ent-kaurenoic acid oxidase 2 from *A. thaliana* [GenBank: O23051][14,24]. N terminal trans-membrane domain, proline cluster, oxygen binding, heme binding motif and E-R-R triad for locking the heme pocket into position, to assure stabilization of the conserved core structure are found in AiCYP2[19,20] (Figure 5.22). AiCYP2 belongs to the CYP88 family and may be involved in hydroxylation at C-11 or 12 (Figure 5.15).

Figure 5. 22 Multiple Sequence Alignment of *A. indica* Cytochrome P450 2 (AiCYP2). Amino acid sequences of AiCYP1 (*A. indica*), CYPP45088D6 (*Glycyrrhiza uralensis*, B5BSX1), CYPP45088A4 (*Arabidopsis thaliana*, Q9C5Y2), CYPP45088A5 (*A. thaliana*, Q5VRM7) and CYPP45088A3 (*A. thaliana*, O23051) are used for multiple sequence alignment. The transmembrane domain is indicated in blue colour. The highly conserved cluster of proline, oxygen binding, EER triad and heme binding motifs are indicated in red colour letters.

AiCYP2 was cloned into a pYES2/ct expression vector (Figure 5.23). The cloned construct was transformed into INVSc1 cells containing AiCPR2. Expression was done under 2% galactose. It was then incubated with tirucalla-7,24-dien-3β-ol, euphol and tirucallol (4 mg/ 100 mL) for 12 h. Crude n-hexane metabolite extracts of saponified AiCYP2 yeast cells were analyzed on GC-MS. No hydroxylated triterpene cyclic product was observed from GC-MS data analysis (Figure 5.24). Triucalla-7,24-dien-3β-ol, euphol and tirucallol did not act as substrates for AiCYP2.

Figure 5. 23 AiCYP2 ORF Amplification.
Lane 1: 1 kb DNA ladder Thermo Fisher (Addendum Figure A1.B), **Lane 7:** Negative control, **Lane 8:** AiCYP2 PCR product at 60 °C, **Lane 9:** AiCYP2 PCR product at 57 °C, **Lane 10:** AiCYP2 PCR product at 62 °C and **Lane 7:** AiCYP2 PCR product at 63 °C

1. Tirucallol, 2. Euphol and
3. Tirucalla-7,24-dien-3β-ol
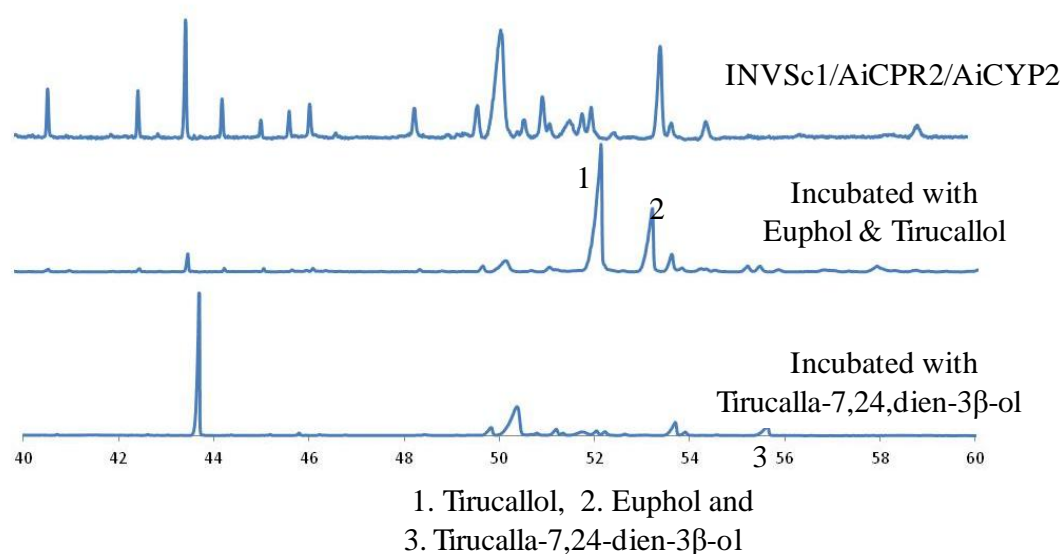
Figure 5. 24 Total Ion Chromatograms of AiCYP2/AiCPR2-INVSc1 Metabolite Extract.

5.5 Conclusion

In order to identify the cytochrome P450 involved in limonoid biosynthesis in neem, a total of two cytochrome P450 reductases and fifteen CYPs were selected based on functional annotation and differential expression data. The two CPR genes ORF was amplified and cloned into pET32a vector. Then expressed in Rosetta 2 (DE3) cells and protein purification was done. The activity of AiCPRs was confirmed by reduction of cytochrome C and potassium ferricyanide. Two CYPs were selected for functional characterization. AiCYP1 was predicted to modify the terminal side chain of protolimonoids and AiCYP2 was predicted to hydroxylate C12 or C13 in basic limonoids/protolimonoids based on BLAST results. AiCPR2 ORF was amplified and

cloned into pYES3/CT vectors and expressed in yeast along with AiCPRs by providing tirucalla-7,24-dien-3β-ol, euphol and tirucallol as substrates. However, the GC-MS analysis indicates that the expressed AiCYPs failed to carry out hydroxylation. To overexpress AiCYP1, the ORF was cloned into pRI101-AN and to silence, cloning of sense and antisense first 300 bp fragments into pRI101-AI harbouring wheat for overexpression. *Agrobacterium*-based syringe infiltration method was used for transient transformation of AiCYP1 constructs into neem leaves. In AiCYP1 overexpression, Azadirachtin A production was observed maximum in day 2 (3.5 folds) and Day 4 (2.09 folds) as compared to vector control. In AiCYP1 silencing, Day 4 showed the highest effect i.e., 5 fold lesser was observed in AiCYP1 silenced neem leaves as compared to vector control. Based on metabolite extraction and HRMS analysis of different time intervals of neem leaves, azadirachtin A levels states that AiCYP1 involves in limonoids biosynthesis. This analysis confirms the AiCYP1 involves in neem limonoids biosynthesis. Further analysis is needed to characterize the genes involved in limonoids biosynthesis in neem.

5.6 Appendix: Agarose Gel Electrophoresis for Colony PCR Screening

5.6.1 Cloning of AiCPR1 into pET32a Vector

Colony PCR with T7 forward and reverse primers for the screening of AiCPR1 cloned into pET32a vector.
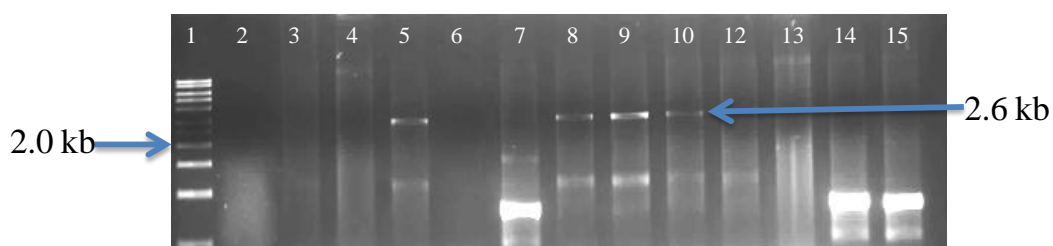
Figure 5. 25 Colony PCR Screening for AiCPR1 Cloned into pET32a on an Agarose Gel. Lane 1: 1 kb DNA ladder Sigma (Addendum Figure A1.A), **Lane 2**: negative control and **Lanes 3-15:** PCR with T7 forward and reverse primers.

5.6.2 Cloning of AiCPR2 into pET32a Vector

Colony PCR with T7 forward and reverse primers for the screening of AiCPR2 cloned into pET32a vector.

Figure 5. 26 Colony PCR Screening for AiCPR2 Cloned into pET32a on an Agarose Gel.
Lane 1: 1 kb DNA ladder Thermo Fisher (Addendum Figure A1.B) and **Lanes 2-7:** PCR with T7 forward and reverse primers.

5.6.3 Cloning of AiCPR2 into pYES3/CT Vector

Colony PCR with T7 forward and CYC reverse primers for the screening of AiCPR2 cloned into pYES3/CT vector.

Figure 5. 27 Colony PCR Screening for AiCPR2 Cloned into pYES3/CT on an Agarose Gel.
Lane 1: 1 kb DNA ladder Sigma (Addendum Figure A1.A), **Lane 2**: negative control and **Lanes 3-7:** PCR with T7 forward and CYC reverse primers.

5.6.4 Cloning of AiCYP1 into pYES2/CT Vector

Colony PCR with T7 forward and CYC reverse primers for the screening of AiCYP1 cloned into pYES2/CT vector.

Figure 5. 28 Colony PCR Screening for AiCYP1 Cloned into pYES2/CT on an Agarose Gel.
Lane 1: 1 kb Plus DNA ladder Thermo Fisher (Addendum Figure A1.C) and **Lanes 2-5:** PCR with T7 forward and CYC reverse primers.

5.6.5 Cloning of AiCYP1 ORF into pRI101-AN Vector

Colony PCR with pRI101 forward and reverse primers for the screening of AiCYP2 cloned into pRI101-AN vector.

Figure 5. 29 Colony PCR Screening for AiCYP1 ORF Cloned into pRI101-AN on an Agarose Gel.
Lane 1: 1 kb DNA ladder Thermo Fisher (Addendum Figure A1.B) and **Lanes 2-15:** PCR with pRI101-AN forward and reverse primers.

5.6.6 Cloning of AiCYP1 Sence 300 bp into pRI101-AN Vector

Colony PCR with pRI101 forward and reverse primers for the screening of AiCYP2 cloned into pRI101-AN vector.

Figure 5. 30 Colony PCR Screening for AiCYP1 Sence 300 bp into pRI101-AN on an Agarose Gel.
Lane 1: 1 kb DNA ladder Thermo Fisher (Addendum Figure A1.B) and **Lanes 2-14:** PCR with pRI101-AN forward and reverse primers.

5.6.7 Restriction Digestion of AiCYP1 Antisense 300 bp into pRI101-AN Vectors

BamHI and *SacI* restriction digestion of AiCYP1 antisense 300 bp into pRI101-AN vectors.
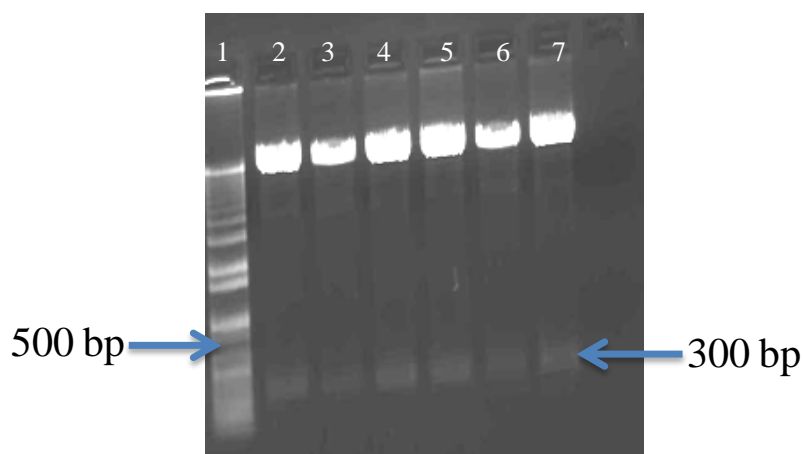
Figure 5. 31 Restriction Digestion for AiCYP1 Antisense 300 bp Cloned into pRI101-AN on an Agarose Gel.
Lane 1: 1 kb Plus DNA Ladder Thermo Fisher (Addendum Figure A1.C) and **Lanes 2-7:** Restriction Digested AiCYP1 antisense 300 bp cloned in pRI101-AN.
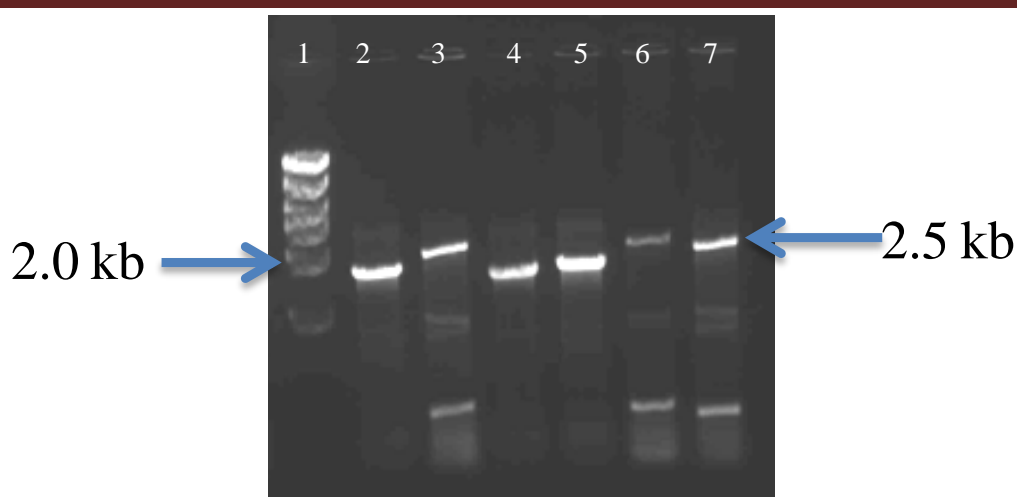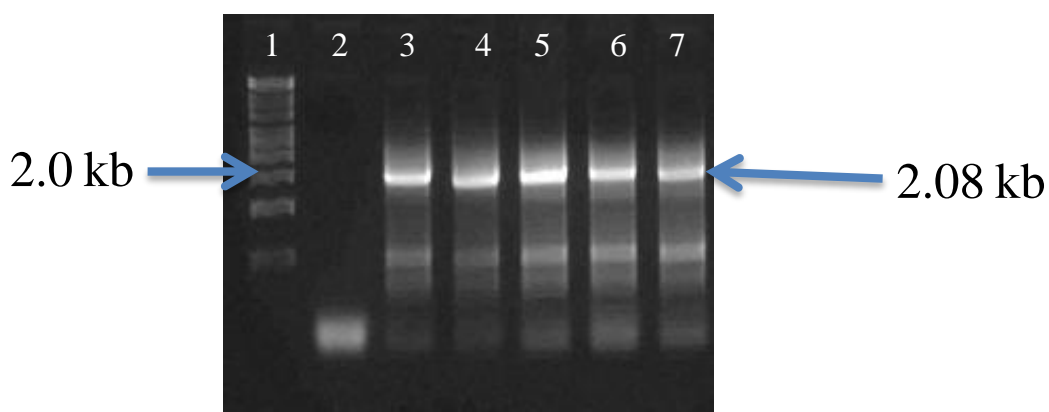
5.6.8 Cloning of AiCYP2 into pYES2/CT Vector

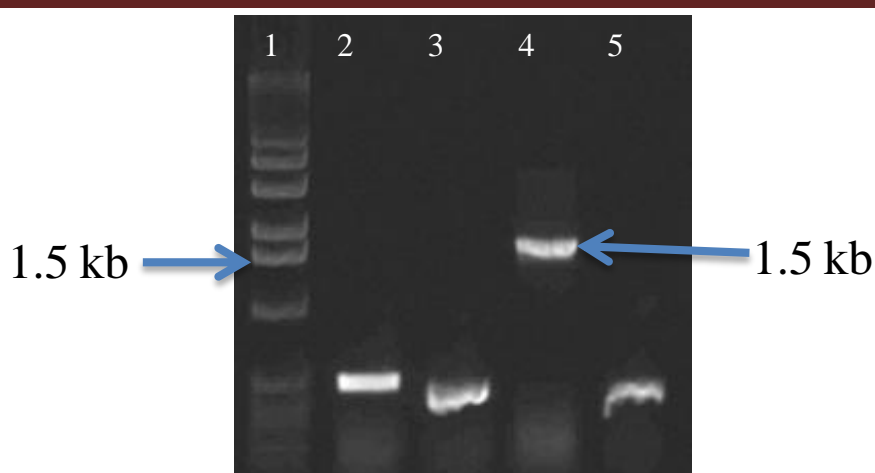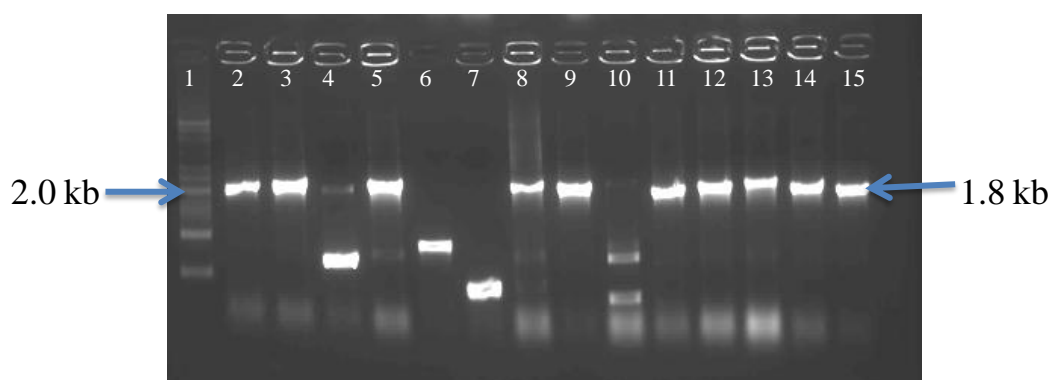Colony PCR with T7 forward and CYC reverse primers for screening of AiCYP2 cloned into pYES2/CT vector.
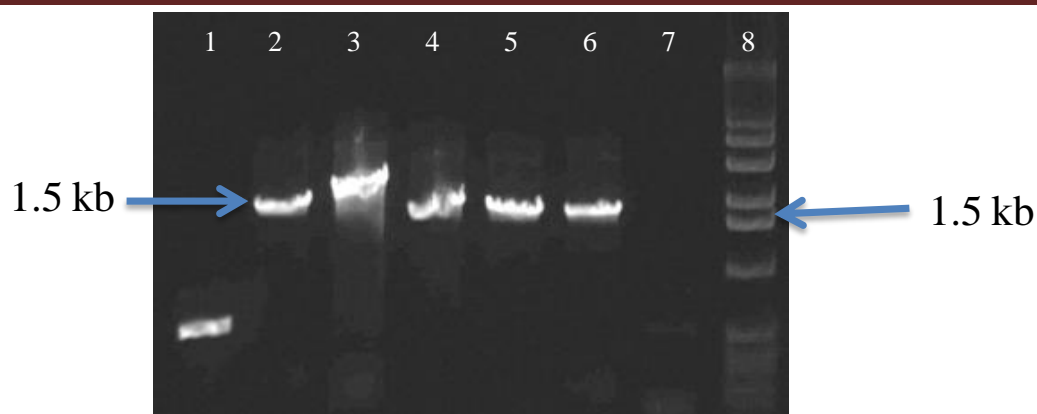
Figure 5. 32 Colony PCR Screening for AiCYP2 Cloned into pYES2/CT on Agarose Gel. Lanes 1-6: PCR with T7 forward and CYC reverse primers, **Lane 7:** 1 kb Plus DNA Ladder Thermo Fisher (Addendum Figure A1.C) and **Lane 8**: negative control.

5.7 References

1. Klingenberg, M. Pigments of rat liver microsomes. *Archives of Biochemistry and Biophysics* **75**, 376-386 (1958).
2. Omura, T. & Sato, R. A new cytochrome in liver microsomes. *The Journal of Biological Chemistry* **237**, 1375-1376 (1962).
3. Nelson, D.R. *et al*. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6**, 1-42 (1996).
4. Schuckel, J. University of York (2012).
5. Bernhardt, R. Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology* **124**, 128-145 (2006).
6. Roberts, G.A., Grogan, G., Greter, A., Flitsch, S.L. & Turner, N.J. Identification of a New Class of Cytochrome P450 from a *Rhodococcus* sp. *Journal of Bacteriology* **184**, 3898-3908 (2002).
7. Ro, D.K., Ehlting, J.r. & Douglas, C.J. Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. *Plant Physiology* **130**, 1837-1851 (2002).
8. Yang, C.Q., Lu, S., Mao, Y.B., Wang, L.J. & Chen, X.Y. Characterization of two NADPH: cytochrome P450 reductases from cotton (*Gossypium hirsutum*). *Phytochemistry* **71**, 27-35 (2010).
9. Urban, P., Mignotte, C., Kazmaier, M., Delorme, F. & Pompon, D. Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *Journal of Biological Chemistry* **272**, 19176-19186 (1997).
10. Mizutani, M. & Ohta, D. Two Isoforms of NADPH: Cytochrome P450 Reductase in *Arabidopsis thaliana*: Gene Structure, Heterologous Expression in Insect Cells, and Differential Regulation. *Plant Physiology* **116**, 357-367 (1998).

11. Huang, F.C., Sung, P.-H., Do, Y.Y. & Huang, P.L. Differential expression and functional characterization of the NADPH cytochrome P450 reductase genes from *Nothapodytes foetida*. *Plant Science* **190**, 16-23 (2012).

12. Shibuya, M. *et al*. Identification of β–amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. *The FEBS Journal* **273**, 948-959 (2006).

13. Yamaguchi, T., Noge, K. & Asano, Y. Cytochrome P450 CYP71AT96 catalyses the final step of herbivore-induced phenylacetonitrile biosynthesis in the giant knotweed, *Fallopia sachalinensis*. *Plant Molecular Biology* **91**, 229-239 (2016).

14. Thabet, I. *et al*. The subcellular localization of periwinkle farnesyl diphosphate synthase provides insight into the role of peroxisome in isoprenoid biosynthesis. *J. Plant Physiol.* **168**, 2110-6 (2011).

15. Qu, X. *et al*. Molecular cloning, heterologous expression, and functional characterization of an NADPH-cytochrome P450 reductase gene from *Camptotheca acuminata*, a camptothecin-producing plant. *PloS one* **10**, e0135397 (2015).

16. Yang, R.-Y. *et al*. Quantitative transcript profiling reveals down-regulation of a sterol pathway relevant gene and overexpression of artemisinin biogenetic genes in transgenic *Artemisia annua* plants. *Planta Medica* **74**, 1510 (2008).

17. Han, J-Y., Hwang, H.S., Choi, S.-W., Kim, H.J. & Choi, Y.E. Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. *Plant and Cell Physiology* **53**, 1535-1545 (2012).

18. Han, J.Y., Kim, H.J., Kwon, Y.-S. & Choi, Y.E. The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant and Cell Physiology* **52**, 2062-2073 (2011).

19. Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A. & Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **3**, 41-62 (1995).

20. Werck-Reichhart, D., Bak, S. & Paquette, S. Cytochromes P450. *The Arabidopsis Book*, e0028 (2002).

21. Pandreka, A. *et al*. Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*). *BMC Plant Biology* **15**, 214 (2015).

22. Seki, H. *et al*. Licorice β-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. *Proceedings of the National Academy of Sciences* **105**, 14204-14209 (2008).

23. Helliwell, C.A., Chandler, P.M., Poole, A., Dennis, E.S. & Peacock, W.J. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of the National Academy of Sciences* **98**, 2065-2070 (2001).

24.   Helliwell, C.A. *et al*. A plastid envelope location of *Arabidopsis* ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway. *The Plant Journal* **28**, 201-208 (2001).

Addendum
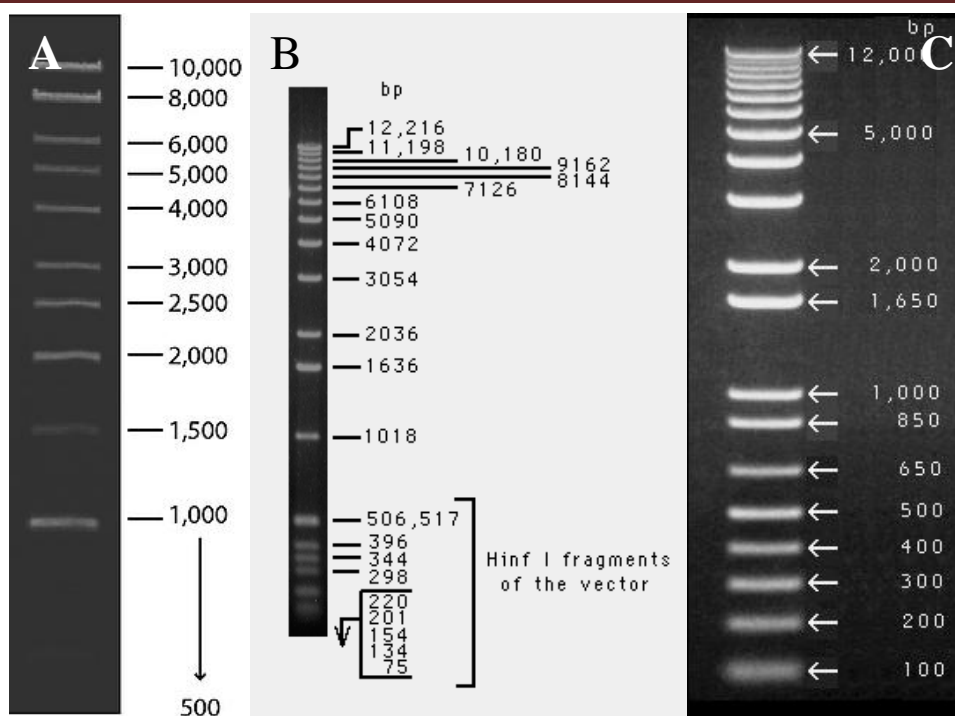
Figure A1: Band distribution of A) 1 Kb DNA Ladder from Sigma Aldrich, B) 1 Kb DNA Ladder and C) 1 Kb Plus DNA Ladder from Thermo Fischer Scientific, USA on a 0.9 % agarose gel stained with Ethidium bromide.
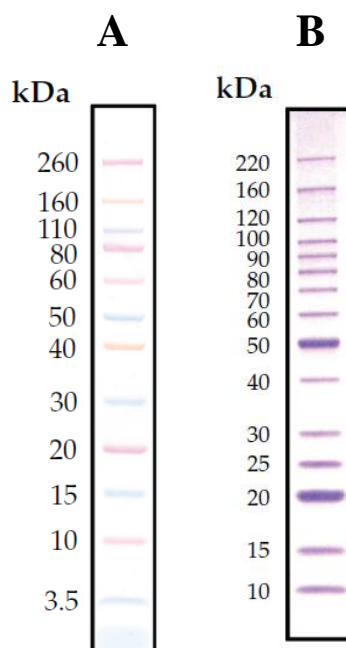
Figure A2: Band distribution of A) Novex® Sharp Pre-Stained Protein Standard and B) BenchMark™ Protein Ladder from Thermo Fischer Scientific, USA on a 0.9 % agarose gel stained with Ethidium bromide.

Sequences of Genes Isolated from *Azadirachta indica* (Neem)

1)   Geranyl Diphospahte Synthase (AiGDS; Accession number: KM108315)

>**Neem_transcript_10912**

```
ATGTTATTTTCTCGTGGACTTTCTCGGATTTCTAGAATCCCGAGAAATAGTTTGATT
GGCTGCCGTTGGCTTGTTTCTTACCGACCCGATACCATTCTTTCCGGTAGTTCACAC
TCTGTTGGAGATTCTACTCAAAAGGTTTTGGGTTGCAGAGAAGCTTATTTATGGAGT
TTGCCTGCCCTGCATGGAATTAGGCATCAGATTCATCAACAGAGCAGCTCCCTAATT
GAGGAGGAACTTGATCCATTTTCCCTTGTTGCTGATGAACTATCACTAGTTGCTAAC
AGGTTGCGCTCCATGGTTGTTGCTGAGGTGCCTAAACTTGCCTCAGCTGCTGAATAT
TTCTTCAAAATGGGGGTGGAAGGAAAGAGGTTTCGTCCTACGGTTTTATTGTTGATG
GCATCGGCTTTGAATGTGCAAGTACCTCAGCCTCTTTCTGATGGAGTAGGAGATGCT
TTGACAACTGAACTCCGTACAAGACAACAATGTATAGCTGAGATTACAGAGATGATC
CATGTAGCTAGCCTTCTTCATGATGATGTCTTGGATGATGCAGATACCAGACGTGGT
ATTGGTTCATTGAATTTTGTAATGGGCAATAAGTTAGCTGTATTAGCTGGAGATTTT
CTGCTTTCTCGAGCTTGTGTTGCCCTTGCTTCTTTAAAAAACACTGAGGTTGTGTCA
TTACTGGCAACTGTTGTAGAGCATCTTGTTACTGGTGAAACCATGCAAATGACTACT
ACAGCAGAACAACGTCGTAGCATGGATTATTATATGCAAAAAACATACTACAAGACT
GCATCATTGATATCAAATAGCTGCAAGGCCATTGCCCTTCTTGCTGGACAAACAACT
GAAGTTGCAATGTTGGCTTTTGATTATGGCAAAAATCTGGGTCTGGCGTTTCAATTA
ATAGATGATGTTCTCGATTTCACAGGCACATCAGCTTCCCTTGGAAAGGGTTCTTTG
TCTGACATCCGACATGGAATTGTAACAGCTCCAATTCTGTTTGCAATGGAAGAGTTC
CCTGAGTTGCGTAAAGTAGTTGACAAGGGCTTTGACGATCCCTCAAATGTCGATATT
GCCTTGGAGTACCTTGGAAAGAGTCGAGGAATACAAAGAACGAGAGAACTAGCTCAG
AAGCATGCTAACCTAGCCACAGTTGCACTTGACTCTCTACCAGAAAGCAACGATGAC
GATGTGAAAAAGTCGAGGCGTGCACTTTTAGATCTCGCTCAAAGAGTCATCACAAGA
AATAAATAG
```

>**AiGDS_Protein_Sequence**

```
MLFSRGLSRISRIPRNSLIGCRWLVSYRPDTILSGSSHSVGDSTQKVLGCREAYLWS
LPALHGIRHQIHQQSSSLIEEELDPFSLVADELSLVANRLRSMVVAEVPKLASAAEY
FFKMGVEGKRFRPTVLLLMASALNVQVPQPLSDGVGDALTTELRTRQQCIAEITEMI
HVASLLHDDVLDDADTRRGIGSLNFVMGNKLAVLAGDFLLSRACVALASLKNTEVVS
LLATVVEHLVTGETMQMTTTAEQRRSMDYYMQKTYYKTASLISNSCKAIALLAGQTT
EVAMLAFDYGKNLGLAFQLIDDVLDFTGTSASLGKGSLSDIRHGIVTAPILFAMEEF
PELRKVVDKGFDDPSNVDIALEYLGKSRGIQRTRELAQKHANLATVALDSLPESNDD
DVKKSRRALLDLAQRVITRNK
```

2)   Farnesyl Diphospahte Synthase (AiGDS; Accession number: KM10831)

>**Neem_transcript_25722**

```
ATGAGTGATCTGCATTCCAAATTCCGGGAGGCTTACCCTGTCTTAAAAGAAGAGCTC
CTCAATGACCCTGCATTTGAGTACGATGATGCTTCTCGTCAATGGGTCGAGCGTATG
CTGGACTACAATGTGCCTGGAGGAAAGCTGAACCGGGGACTATCTGTTCTTGACAGC
TACAAGTTACTAAAAGAAGGAAAGAACTAACGGATGATGAACTTTTTCTTGCATCT
GCACTAGGCTGGTGCATCGAATGGCTTCAAGCATATTTTCTTGTTCTTGACGATATC
```

```
ATGGACAACTCTGTGACACGCCGTGGTAACCCTTGCTGGTTCAGACTCCCCAAGGTT
GGCATAATTGCTGTAAATGATGGTGTAATTCTACGCAACCATGTGCACAGAATTCTT
AAGAAACATTTTAGAGAAAAGCCTTACTATGTGGAACTGGTGGATTTATTTAATGAG
GTGGAATTTCAAACAGCTTCAGGACAATTGATAGATTTAATTACCACACATGAAGGA
GAGAAAGATCTATCAAAGTACTCATTGACAGTTCATCGTCGCATTGTTCAGTACAAA
ACCGCTTACTATTCCTTTTACCTTTCGGTTGCTTGTGCATTGCTTATGTCAGGTGAG
AATCTTGACAATCATGTTAATGTCAAGAACATCCTTGTTGAAATGGGTGTCTATTTT
CAAGTACAGGATGATTATCTTGATTGTTTTGGTGCTCCAGAAGTGACTGGTAAGATT
GGAACTGATATCGAAGACTTTAAGTGTTCTTGGTTGGTCGTGAAAGCTCTGGAACAT
TGTAACGAGGAACAGAAAAAATTGCTATATGAGAACTACGGGAAGGCAGATGCAGCT
TGTATTTCAAAAGTGAAAGAGCTTTATAAAACTCTTGATATTCAGGGTGCATTTGAG
GAGTACGAGAGGGAGAGTTATGAAAAGCTGACAAAATCCATTGAAGCTCATCCAAGT
AAGGCAGTTCAAGCTGTGTTAAAATCATTCTTGGCAAAGATATACAAGAGGCAGAAG
TAA
```

>AiFDS_Protein_Sequence

```
MSDLHSKFREAYPVLKEELLNDPAFEYDDASRQWVERMLDYNVPGGKLNRGLSVLDS
YKLLKEGKELTDDELFLASALGWCIEWLQAYFLVLDDIMDNSVTRRGNPCWFRLPKV
GIIAVNDGVILRNHVHRILKKHFREKPYYVELVDLFNEVEFQTASGQLIDLITTHEG
EKDLSKYSLTVHRRIVQYKTAYYSFYLSVACALLMSGENLDNHVNVKNILVEMGVYF
QVQDDYLDCFGAPEVTGKIGTDIEDFKCSWLVVKALEHCNEEQKKLLYENYGKADAA
CISKVKELYKTLDIQGAFEEYERESYEKLTKSIEAHPSKAVQAVLKSFLAKIYKRQK
```

3)   **Squalene Epoxidase 1 (AiSQE1; Accession number: JX997152)**

>Neem_transcript_11067/ Master_Control_31859

```
ATGGCGGCTGTGATTGATCAATACATTGTTTGGACTTTCTGTGCCTCTCTCATTGGG
TTTCTTCTTCTCTTCATTTTGCGACTCAATGATAAGAACGAGTCAAAGGAGGATACC
GGAAACTGCAATATCCAAAATGACATTGTTAAGAGCTCTTCTAACGATGGAGATTCC
TCGCCGGAGCATGGCTCCCGCACTGACGTCATCATTGTTGGCGCCGGCGTTGCTGGT
ACTGCCTTGGCTCACACACTTGGCAAGGATGGAAGGCGAGTTCATGTGATAGAAAGA
GACTTGACAGAGCCTGACAGAATTGTTGGTGAACTGCTTCAGCCTGGGGAATACCTG
AAACTAGTTGAGTTGGGCCTTGAGGACTGTGTGGAAGAAATTGATGCCCAGCAAGTA
CTTGGTTATGCTCTTTTCAAGGATGGAAAAAATACTAGACTGTCATATCCCTTGGAG
AAATTTCATGCGGACGTCTCTGGGAGGAGCTTCCACAATGGGCGTTTCATTCAGCGG
ATGAGAGAAAAAGCAGCTACTCTTCCCAATGTACAGCTAGATCAAGGAACTGTAACA
TCCCTACTTGAAGAAATGGGACTATAAAGGGGTTCAGTACAAAACTAAGGATGGT
CAAGAACTTCAATCTTTTGCCCCTCTTACAATTGTGTGTGATGGTTGCTTCTCAAAT
CTGCGCCGTTCTCTTTGCAAACCAAAGGTTGATGTGCCCTCTTGTTTTGTGGGTTTA
GTCTTAGAGAACTGCCAGCTTCCATTTGCGAACCACGGGCATGTCATTCTGGCAGAT
CCTTCCCCCATCTTATTTTATCCAATCAGTAGCACAGAGATTCGATGCTTGGTTGAT
GTTCCTGGTCAGAAGATACCACCAATAGCTAATGGTGAAATGGCCAATTATTTGAAG
ACTGTGGTGGCACCGCAGGTTCCCCCTGAGCTTCATGATGCCTTCATTTCTGCAGTT
GAGAAAGGAAATATTAGAACTATGCCAAATAGAAGCATGCCAGCTGATCCCCAGCCT
ACTCCTGGAGCCCTTCTGATGGGTGATGCTTTCAACATGCGTCATCCTTTAACTGGC
GGCGGAATGACTGTGGCATTATCTGATATTGTTGTACTGCGGAATCTCCTTAAGCCA
```

```
CTCAATGACTTTCATGATTCAGCTTCCCTTACCACATATCTTGAATCCTTTTATACC
TTACGCAAGCCAGTTGCATCGACAATAAATACTCTGGCAGGTGCTCTTTACAAGGTG
TTTTCCTCTTCCCCTGATCAAGCAAGAAAGGAAATGCGCCAAGCATGTTTTGACTAT
TTAAGTCTTGGAGGTGTCTTTTCATCCGGACCAATTGCTTTACTCTCTGGTCTAAAC
CCTCGCCCCTTAAGTTTAGTTCTGCATTTCTTTGCTGTGGCAATTTATGGTGTTGGT
CGTTTGCTACTTCCATTTCCATCGCTTAAGAGGATGTGGATTGGGGCTAGGCTAATA
TCGAGTGCATCAGGTATTATTTTTCCCATCATCAAGGCAGAAGGAGTTAGGCAAATG
TTCTTCCCTGCCACTATTCCTGCAATATACAGAGCTCCTCCTGTTGACGATTGA
```

>AiSQE1_Protein_Sequence

```
MAAVIDQYIVWTFCASLIGFLLLFILRLNDKNESKEDTGNCNIQNDIVKSSSNDGDS
SPEHGSRTDVIIVGAGVAGTALAHTLGKDGRRVHVIERDLTEPDRIVGELLQPGEYL
KLVELGLEDCVEEIDAQQVLGYALFKDGKNTRLSYPLEKFHADVSGRSFHNGRFIQR
MREKAATLPNVQLDQGTVTSLLEENGTIKGVQYKTKDGQELQSFAPLTIVCDGCFSN
LRRSLCKPKVDVPSCFVGLVLENCQLPFANHGHVILADPSPILFYPISSTEIRCLVD
VPGQKIPPIANGEMANYLKTVVAPQVPPELHDAFISAVEKGNIRTMPNRSMPADPQP
TPGALLMGDAFNMRHPLTGGGMTVALSDIVVLRNLLKPLNDFHDSASLTTYLESFYT
LRKPVASTINTLAGALYKVFSSSPDQARKEMRQACFDYLSLGGVFSSGPIALLSGLN
PRPLSLVLHFFAVAIYGVGRLLLPFPSLKRMWIGARLISSASGIIFPIIKAEGVRQM
FFPATIPAIYRAPPVDD
```

4)   Triterpene synthase 1 (AiTTS1)

> Master_Control_24780/ Neem_transcript_28920

```
ATGTGGAAGCTGAAGATTGCAGAGGGTGACAAAAATAGCCCATATATTTCTACAACA
AACAATTTCGTTGGAAGGCAAATATGGGAATTTGATCCGAACGCCGGAACTGCTGAA
GAGCTTGCTGAAGTTGAAGAAGCTCGTCAGAATTTCTACAAGAATCGCCATCAAGTC
AAACCTGCTAGTGATCTTATTTTTCGTCTTCAGTTTCTTAGAGAGAAAAACTTCAAG
CAAACGATTCCTCAAGTGAAGGTTGAAGATGGGGAGGAGATCACATATGACACTGCC
ACAGCAGCAATGAAGAGGGCTGCTCACTACTTCTCAGCAATTCAGGCTAGCGATGGC
CATTGGCCTGCTGAAAATTCTGGCCCTATGTATTTCCTTCCTCCATTTGTATTCTGC
TTGTACATTACAGGACATCTTGATACTGTATTTACAGCTGCTCATCGCAGAGAAGTC
CTTCGTTACTTATACAATCATCAGCATGAAGATGGAGGATGGGGAATACACATAGAA
GCGCCAAGCAGTATGTTTGGTACAGTTTACAGTTATCTTACAATGCGTTTGCTAGGG
TTAGGACCCAACGATGGTGAAAACAATGCCTGCGCCAGAGCTAGAAAATGGATTCGT
GATAATGGTGGTGTCACTTACATTCCCTCTTGGGGAAAGAATTGGCTTTCGATTCTT
GGTTTGTTTGAATGGGCTGGAACACACCCAATGCCCCCAGAGTTCTGGATGCTTCCT
TCTCATTTTCCACTTCATCCAGCCCAAATGTGGTGCTTCTGCCGGCTGGTTTACATG
CCCTTGTGTTATTTATACGGCAAAAGATTTGTTGGTCCAATCACTCCACTTATCAAA
CAACTGAGAGAAGAACTTCATACAGAGCCTTACGATAAATCAACTGGAGGAAAGTT
CGTCATCAATGTGCAAAGACTGATCTCTACTACCCCCATCCATTCGTACAAGAAGTT
CTATGGGATACTCTATACTTTGCTACAGAGCCTCTGCTTACTCGTTGGCCATTGAAC
AAGTATGTCAGAGAGAAGGCTTTGAAACAAACGATGAAGATCATTCATTATGAAGAC
CAAAGCAGTCGATATATTACTATTGGATGCGTCGAGAAGCCGCTGTGTATGCTTGCT
TGTTGGGTGGAGGATCCTGAAGGGGTTGCTTTCAAGAAGCATCTTGAGAGAATTGCT
GATTTTATTTGGATTGGAGAAGATGGAATGAAAGTTCAGACATTTGGCAGTCAAACA
TGGGATACTGCTCTTGGACTTCAAGCTTTGCTTGCTTGCAATATCGTTGATGAAATT
```

```
GGACCTGCACTTGCTAAAGGACACGACTACTTGAAGAAAGCTCAGGTGAGGGATAAT
CCAGTGGGTGATTATACAAGCAATTTCCGACACTTTTCCAAAGGAGCATGGACTTTC
TCTGATCAAGATCATGGTTGGCAAGTTTCAGATTGTACTGCAGAAAGTTTGAAGTGC
TGCCTGCATTTCTCAATGCTGCCTCCAGAAATTGTTGGAGAGAAACATGATCCTGAG
AGATTATATGAAGCTGTCAATTTCATACTCTCTCTTCAGGATAAAAATGGTGGAATA
GCAGTTTGGGAGAAAGCTGGTGCCTCTTTGATGTTAGAGTGGCTCAATCCTGTAGAG
TTTCTGGAGGACCTTATTGTTGAGCATACTTACGTGGAATGCACTGCTTCAGCAATC
GAGGCATTTGTTATGTTCAAGAAATTATACCCACATCATCGCAAGAAGGAGATTGAA
AATTTCCTCGTAAAAGCTGTACAGTACATTGAAAATGAACAAACTGCTGATGGTTCA
TGGTATGGAAACTGGGGAGTTTGCTTCTTATATGGAACATGTTTTGCACTTGGAGGT
TTACATGCTGCTGGAAAGACTTACAACAATTCTCTTGCCATTCGTAGAGCAGTTGAG
TTTCTGCTTCAAGCACAGAGTGATGATGGTGGTTGGGGAGAGAGCTACAAATCTTGC
CCTAGTAAGATATACGTACCTCTTGATGGGAAAGATCAACTGTGGTACACACTGCA
TTGGCTATTCTTGGTTTAATCCATGCTGGGCAGGCTGAAAGAGACCCAACCCCTATT
CATCGTGGTGTAAAATTGCTGATCAACTCTCAATTGGAGAATGGAGACTTCCCTCAA
CAGGAAATTATGGGAGTTTTTATGAGAAACTGTATGTTACACTATGCTCAATACAGG
AATATTTTTCCTTTGTGGGCTTTAGCTGAATATAGAAGAAAGTTCCATTGCCTAAT
TAA
```

>AiTTS1_Protein_Sequence

```
MWKLKIAEGDKNSPYISTTNNFVGRQIWEFDPNAGTAEELAEVEEARQNFYKNRHQV
KPASDLIFRLQFLREKNFKQTIPQVKVEDGEEITYDTATAAMKRAAHYFSAIQASDG
HWPAENSGPMYFLPPFVFCLYITGHLDTVFTAAHRREVLRYLYNHQHEDGGWGIHIE
APSSMFGTVYSYLTMRLLGLGPNDGENNACARARKWIRDNGGVTYIPSWGKNWLSIL
GLFEWAGTHPMPPEFWMLPSHFPLHPAQMWCFCRLVYMPLCYLYGKRFVGPITPLIK
QLREELHTEPYDKINWRKVRHQCAKTDLYYPHPFVQEVLWDTLYFATEPLLTRWPLN
KYVREKALKQTMKIIHYEDQSSRYITIGCVEKPLCMLACWVEDPEGVAFKKHLERIA
DFIWIGEDGMKVQTFGSQTWDTALGLQALLACNIVDEIGPALAKGHDYLKKAQVRDN
PVGDYTSNFRHFSKGAWTFSDQDHGWQVSDCTAESLKCCLHFSMLPPEIVGEKHDPE
RLYEAVNFILSLQDKNGGIAVWEKAGASLMLEWLNPVEFLEDLIVEHTYVECTASAI
EAFVMFKKLYPHHRKKEIENFLVKAVQYIENEQTADGSWYGNWGVCFLYGTCFALGG
LHAAGKTYNNSLAIRRAVEFLLQAQSDDGGWGESYKSCPSKIYVPLDGKRSTVVHTA
LAILGLIHAGQAERDPTPIHRGVKLLINSQLENGDFPQQEIMGVFMRNCMLHYAQYR
NIFPLWALAEYRRKVPLPN
```

5)   Triterpene synthase 2 (AiTTS2)

> Master_Control_74892

```
ATGTGGAAGCTTAAGGTTGCAGATGGAGGAAACGACCCGTACATTTACAGCACAAAT
AACTTTGTTGGGAGGCAAATATGGGAGTTTGATCCTGATGCTGGAACTCCTGAAGAA
AGAGCCGAGGTCGAAGCCGCTCGTCAAAATTTCTGCAAAAATCGTTTCCAAGTCAAG
CCTAGTGGTGATCTCCTCTGGCGAATGCAGTTTTTGAAGGAGAAAGTTTCAAACAA
ACAATCCCACAAGTGAAGGTCAAGCTGGGGAGGAAATCACATATGAAACTGCGACA
ACCTCATTGAGGAGAGCCGTTCATTTCTTCTCAGCCTTGCAAGCCAGTGATGGCCCT
TGGCCTGCCGAAATGCTGGTCCCTTGTTTTTCCTTCCTCCCCTGGTCATGTGCATG
TACATAACCGGTCATCTCGATGCCGTGTTCCCACCCGAGTATCGCAAGGAAATCCTT
CGTTACATTTACTGCCATCAGAATGTAGATGGTGGGTGGGGGTTACATATCGAGGGT
```

```
CACAGCACCATGTTTTGCACAGTTTTTAGCTATATTTGCATGCGTATTCTTGGAGAA
GGACCTGATGGTTGCCAAGACAATACTTGTGCCAGAGCTCGAAAGTGGTTTCTTGAC
CGCGGTGGTGTAATACACATTCCTTCTTGGGGAAAGACTTGGCTTTCGATACTTGGT
GTTTTTGAATGGTCCGGAAGCAACCCAATGCCTCCAGAATTTTGGATCCTTCCTTCG
TTTCTTCCCATGCACCCAGCAAAAATGTGGTGCTATTGCCGCATGGCGTACATGCCG
ATGTCATATTTATATGGTAAGAAGTTTGTCGGATTAATTACACCTCTCATTTTGCAA
TTGAGAGAAGAACTCTACACCCAACCTTATCATGAAGTTAATTGGAGGAAATCACGC
CATTTATGTGCAAAGGAGGATCTTTACTATCCCCATCCTTTAATACAAGACTTGATT
TGGGACAGTCTGTATATATTCACAGAGCCTTTTCTTTCTCGTTGGCCTTTTAACAAG
TTAATCAGAGAGAAAGCACTTCAAGTGACAATGAAGCACATTCATTATGAAGATGAG
AACAGTCGATACATCACAATTGGGTGTGTAGAAAAGGTATTATGCATGCTTTCTTGT
TGGGTCGAAGATCCAAATGGAGAATATTTTAAGAAGCATCTTGCTAGGATTCCGGAT
TATTTATGGGTTGGCGAAGATGGAATGAAGATGCAGAGTTTTGGGAGCCAAGAATGG
GATGCTGGTTTTGCTATTCAAGCTTTGCTCGCTAGTAATCTCGTTGACGAAATTGGA
CCCGTACTTAAGAGAGGACATGAATTCATCAAGGCATCTCAGGTAAAAGACAATCCT
TCTGGCGACTTTAAGGGAATGCACCGTCATATCTCAAAAGGATCTTGGACTTTCTCT
GATCAAGATCATGGATGGCAAGTTTCTGATTGCACTGCTGAAGGATTGAAGTGTTGC
CTGCTCTTTTCAATGATGCCGCCAGAAATTGTTGGCGAGAAGATGGAACCTGAGCGG
TTACATGATTCTGTGAATATCATACTTTCACTCCAGAGTAAAAATGGAGGTTTAGCA
GCCTGGGAGCCGGCCGGTGCTCAAGAATGGTTGGAACTGCTGAATCCCACAGAATTT
TTTGCGGACATTGTCGTTGAACATGAATATGTTGAGTGCACTGCATCGGCTATCCAT
TCCTTAATTTTGTTTAAGAAGTTATATCCAGGACATAGGAAGAAAGAAATCGAAAAT
TTCATAGCAAATGCTGTTCGATACTTGGAAGACGTACAATTGCCTGATGGTTCATGG
TATGGAAATTGGGGAGTTTGCTTCACATACGGTACGTGGTTTGCACTTGGAGGGTTA
GCTGCTGCTGGTAGAACTTATTATAATAGTCCAACCATGCGCAAAGCTGTTGATATT
TTTTTGAAATCACAGAGAGATAATGGTGGCTGGGGAGAGAGCTATCGTTCTTGTCCA
GAGAAGAAAAATATACCACTTGAAGGAAACAGATCGAATTTGGTGCATACTGGATGG
GCTATGATGCTGGACAGGCGGAAAGAGATCCAACTCCTTTTCACCGTGCAGCTAAGT
AGCTAA
```

>AiTTS2_Protein_Sequence

```
MWKLKVADGGNDPYIYSTNNFVGRQIWEFDPDAGTPEERAEVEAARQNFCKNRFQVK
PSGDLLWRMQFLKEKSFKQTIPQVKVKAGEEITYETATTSLRRAVHFFSALQASDGP
WPAENAGPLFFLPPLVMCMYITGHLDAVFPPEYRKEILRYIYCHQNVDGGWGLHIEG
HSTMFCTVFSYICMRILGEGPDGCQDNTCARARKWFLDRGGVIHIPSWGKTWLSILG
VFEWSGSNPMPPEFWILPSFLPMHPAKMWCYCRMAYMPMSYLYGKKFVGLITPLILQ
LREELYTQPYHEVNWRKSRHLCAKEDLYYPHPLIQDLIWDSLYIFTEPFLSRWPFNK
LIREKALQVTMKHIHYEDENSRYITIGCVEKVLCMLSCWVEDPNGEYFKKHLARIPD
YLWVGEDGMKMQSFGSQEWDAGFAIQALLASNLVDEIGPVLKRGHEFIKASQVKDNP
SGDFKGMHRHISKGSWTFSDQDHGWQVSDCTAEGLKCCLLFSMMPPEIVGEKMEPER
LHDSVNIILSLQSKNGGLAAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTASAIH
SLILFKKLYPGHRKKEIENFIANAVRYLEDVQLPDGSWYGNWGVCFTYGTWFALGGL
AAAGRTYYNSPTMRKAVDIFLKSQRDNGGWGESYRSCPEKKNIPLEGNRSNLVHTGW
AMMLDRRKEIQLLFTVQLSS
```

6)   NADPH Cytochrome P450 reductase 1 (AiCPR1; Accession number: KM108318)

> Neem_Transcript_2277/Master_Control_115955

```
ATGAGCAACTCTGGTACGGGTAACGATTTGGTCAAATTCGTGGAATCAGCGCTAGGG
GTGTCGTTGGGGAGTTCCGTGACGGACACGGTCATCGTGATTGCAACGACGTTGTTT
GCGGTGGTAATTGGTTTGCTGGTGCTGGCTTGGAAGAAATCATCGGATCGCAGTAAA
GAAGTGAAGCCGGTGGTGCCGCTTAGTCGCCGGTGTTGAAGAAGGAGGACCATGAC
GAAGCTGATATTGTTTGGGGCAAGACTAAAGTCACTATATTTTACGGTACTCAGACC
GGAACTGCGGAGGGGTTTGCTAAGGCTTTGGCTGAAGAAATCAAGGCAAGATATGAA
AAAGCAGCTGTCAAAGTTGTTGATCTGGATGATTATGCTGCGGATGATGAGCAATAT
GGAGAAAAACTGAAGAAAGAGACTTTGACATTTTTCATGGTGGCCACTTATGGAGAT
GGAGAGCCTACTGATAATGCAGCAAGATTCTACAAATGGTTTACTGAGGAAAATGAT
CGGGGAGTCTGGCTTCAACAGCTCAGATATGGCGTGTTTGGGCTGGGTAACCGTCAG
TATGAACATTTTAACAAGATAGGGAAGGTGCTTGATGAAGAACTTTGCAAGCAAGGT
GGAAAGCGTCTAGTCCCTGTGGGTCTTGGGGATGATGATCAGTGCATTGAGGATGAT
TTTGCTGCCTGGAGAGAATTAGTGTGGCCTGAGTTAGATCAATTGCTCCGTGATGAA
GATGATGCAAATACTGTATCTACACCATACACAGCAGCTATTCCTGAATATCGAGTA
GTGATTCATGATCCCACTGCTACTTCTTTTGTGGATAAGTACTCAAACATTCCAAAT
GGCAATGCTTCTTTTGACATCCACCATCCATGCAGAGTCAATGTGGCTGTCCAGAGA
GAGCTGCACAAACCTGAATCTGACCGCTCTTGCATACATCTGGAGTTCGATATATCT
GGGACTGGTATCACATATGAAACTGGAGACCATGTGGGCATTTATGCAGAGAATTGT
GATGAAACTGTTGAACAAGCCGGAAAGTTATTGGGTCAACCTCTAGAGCTGCTGTTC
TCTATTCATGCTGACAATGAAGATGGCACTCCAATTGGAAGCTCATTGCCGCCTCCT
TTCCCTGGTCCATGCACACTTCGCTCTGCAATAGCACAATATGCAGATTTATTGAAT
CCTCCTCGTAAGGCTGCTCTGACTGCTTTGGCTGCACATGCCACTGAACCTAGTGAG
GCAGAGAGACTCAAGTTTTTGTCATCACCTCAGGGGAAGGATGAGTACTCACAATGG
ATTGTAGCCAGTCAGAGAAGTCTTATTGAGGTAATGGCTGAGTTTCAGTCAGCAAAA
CCTCCCATTGGTGTATTTTTTGCAGCTGTAGCCCCTCGCTTACAGCCTCGTTACTAC
TCTATCTCTTCTTCACCCAGATTTGGTCCCAATAGAGTTCATGTAACATGTGCTTTA
GTTTATGGTCCTACTCCCACTGGTAGAATACACAAGGGGGTCTGTTCAACCTGGATG
AAGAATGCAATTCCCCTCGAGAAAAGTCAAGACTGTAGCTGGGCTCCTGTTTTCATC
AGACCATCTAATTTCAAGTTACCAACCAATCCTTCTGTTCCTGTTATCATGATAGGA
CCTGGTACTGGATTGGCACCTTTCAGAGGGTTTCTACAGGAAAGAATGGCCTTGAAA
CTGGATGGTAATGACCTTGGTCCTGCCCTACTATTCTTTGGATGTCGAAATCGCCGA
ATGGATTTTATTTACAAGGATGAGCTCTATAATTTTGTGGATCAAGGCGTTATATCT
GAGCTGATTGTTGCATTCTCACGGGAGGGGCCACAGAAGGAATATGTTCAACATAAG
ATGATGGACAAAGCAGCATACTTGTGGAGCCTACTTTCTAAGGATGGATATCTTTAC
GTGTGTGGAGATGCCAAGGGAATGGCAAGAGACGTTCATCGCACTTTGCACACCATT
ATTCAGGAGCAGGAAAATGTGGACTCATCAACGGCAGAAGCCACAGTGAAGAACTC
CAGATGGAAGGACGATATCTTAGAGATGTCTGGTGA
```

>AiCPR1_Protein_Sequence

```
MSNSGTGNDLVKFVESALGVSLGSSVTDTVIVIATTLFAVVIGLLVLAWKKSSDRSK
EVKPVVPLKSPVLKKEDHDEADIVWGKTKVTIFYGTQTGTAEGFAKALAEEIKARYE
KAAVKVVDLDDYAADDEQYGEKLKKETLTFFMVATYGDGEPTDNAARFYKWFTEEND
RGVWLQQLRYGVFGLGNRQYEHFNKIGKVLDEELCKQGGKRLVPVGLGDDDQCIEDD
```

```
FAAWRELVWPELDQLLRDEDDANTVSTPYTAAIPEYRVVIHDPTATSFVDKYSNIPN
GNASFDIHHPCRVNVAVQRELHKPESDRSCIHLEFDISGTGITYETGDHVGIYAENC
DETVEQAGLLGQPLELLFSIHADNEDGTPIGSSLPPPFPGPCTLRSAIAQYADLLN
PPRKAALTALAAHATEPSEAERLKFLSSPQGKDEYSQWIVASQRSLIEVMAEFQSAK
PPIGVFFAAVAPRLQPRYYSISSSPRFGPNRVHVTCALVYGPTPTGRIHKGVCSTWM
KNAIPLEKSQDCSWAPVFIRPSNFKLPTNPSVPVIMIGPGTGLAPFRGFLQERMALK
LDGNDLGPALLFFGCRNRRMDFIYKDELYNFVDQGVISELIVAFSREGPQKEYVQHK
MMDKAAYLWSLLSKDGYLYVCGDAKGMARDVHRTLHTIIQEQENVDSSTAEATVKKL
QMEGRYLRDVW
```

7)   NADPH Cytochrome P450 reductase 2 (AiCPR2; Accession number: KM108319)

>Neem_Transcript_1270/Master_Control_117874

```
ATGCAATCATCGTCGTCATCAGCATCATCTTCCAACACGATGAAAGTATCACCGTTT
GATTTGATGTCCGCAATCATTAAAGGCAAAACGGATCCATCCAACGTGTCGTCGTCG
GGTTCCGGTTTGGAAGTGGCGTCTATTGTTCTTGAGAATAAAGAGTTTGTTATGATC
TTGACTACCTCTATAGCCGTCTTGATTGGCTGCGTTGTTGTTTTGATATGGCGGAGA
TCCAGCAGCCAGAAACCGAAGAAGATTGAGCCACTTAAGCCGTTGGTCGTTAAAGAA
CCAGAAGTGGAGGTTGATGATGGTAAGCAGAAGATTACCATTTTCTTCGGCACTCAA
ACCGGCACAGCTGAAGGTTTTGCTAAGGCTTTGGCTGACGAGGCCAAAGCGCGATAT
GATAAGGCCATTTTTAAAGTTGTTGATTTGGATGATTACGCTGCTGATGATGAGGAA
TATGAGGAGAAACTGAAAAAGAATCTATAGCATTTTTCTTCTTGGCTACATATGGT
GATGGTGAGCCAACCGATAATGCTGCTAGGTTTTATAAATGGTTCACGGAGGGAAAA
GAGAGAGGGGAGTGGTTACAGAACCTTAAGTACGGAGTGTTTGGTCTTGGAAACAGG
CAATATGAACATTTCAACAAGATTGCAAAGGTGGTGGATGATGTTCTTGCTGAACAG
GGTGGGAAACGCCTTGTCCCTGTGGGTCTTGGAGATGATGATCAGTGCATTGAAGAT
GACTTTTCTTCATGGCGAGAATTGGTGTGGCCTGAGCTGGATAAGTTGCTTCGGGAT
GATGATGATCCAACAACTGTTTCTACACCTTACACCGCTGCCATATCAGAATATCGG
GTTGTATTTATGACTCTGCCGATGCATCAATAGGGGAGAACAACTGGAGCAATGCA
AATGGCCATGCTGTTTATGATGCTCAACACCCCTGCAGGTCTAATGTGGCGGTAAGG
AAGGAGCTTCACACTCCATTATCGGATCGCTCTTGCACCCATCTAGAATTTGAAATT
GCTGGAACTGGATTAATGTATGAAACAGGGGATCATGTTGGTGTTTTCTGTGAGAAT
CTAACTGAGACTGTGGAGGAGGCATTGTCATTGTTAGGCTTGTCACCAGATACTTAC
TTCTCCGTCCATACCGATAAAGAGGATGGCACACCACTTGGTGGAAGCTCCTTGCCT
CCTCCTTTCCCACCATGCTCTTTAAGAACAGCTTTGGCTCAATATGCTGATCTTTTG
AGTTCCCCGAAAAAGTCTGCATTGCTTGCTTTAGCAGCTCATGCATCTGACCCCACT
GAAGCTGATCGATTGAGACATCTTGCATCACCTGCTGGAAAGGATGAATATGCACAG
TGGATTGTTGCAAGTCAGAGAAGCCTCCTTGAGGTCATGGCTCAATTTCCATCTGCC
AAGCCCCCACTCGGTGTCTTCTTTGCAGCAGTTGCCCCACGCTTGCAACCAAGATAT
TATTCAATTTCATCTTCCCCAAGGGTGGCACCATCTAGAATTCATGTTACCTGTGCA
CTAGTTTATGAGAAACACCAACAGGTCGCATTCACAAGGAGTGTGCTCTACTTGG
ATGAAGAATTGTGTGCCTATGGAGAAAGCAATGACTGCGGCTGGGCACCCATTTTT
GTTAGGCAATCAAACTTCAGACTCCCAGCAGATCCTAAAGTTCCTGTCATTATGATT
GGCCCTGGCACTGGCTTGGCTCCTTTCAGGGGTTTCCTGCAGGAACGATTTGCTCTG
AAGGAAGCTGGAGCAGAACTGGGACCATCTGTATTATTTTTCGGCTGCCGGAATCGT
CAGATGGATTACATATATGAGGACGAGCTCAACAACTTTGTTCAAAGTGGTGCTCTT
```

```
TCAGAGTTGGTTGTTGCCTTCTCACGTCAGGGACCCACCAAGGAGTATGTGCAGCAT
AAAATGATGGAGAAGGCTTCGGATATCTGGAACATGATATCTCAGGGAGGTTACTTG
TATGTTTGCGGTGATGCCAAAGGCATGGCCAGAGATGTCCACCGAACTCTGCACACC
ATTGTGCAAGAGCAGGGATCTGTGGACAGCTCTAAGGCTGAGAGCATCGTGAAGAAC
TTGCAAATGACTGGCAGGTATCTACGTGATGTGTGGTGA
```

>AiCPR2_Protein_Sequence

```
MQSSSSSASSSNTMKVSPFDLMSAIIKGKTDPSNVSSSGSGLEVASIVLENKEFVMI
LTTSIAVLIGCVVVLIWRRSSSQKPKKIEPLKPLVVKEPEVEVDDGKQKITIFFGTQ
TGTAEGFAKALADEAKARYDKAIFKVVDLDDYAADDEEYEEKLKKESIAFFFLATYG
DGEPTDNAARFYKWFTEGKERGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDDVLAEQ
GGKRLVPVGLGDDDQCIEDDFSSWRELVWPELDKLLRDDDDPTTVSTPYTAAISEYR
VVFYDSADASIGENNWSNANGHAVYDAQHPCRSNVAVRKELHTPLSDRSCTHLEFEI
AGTGLMYETGDHVGVFCENLTETVEEALSLLGLSPDTYFSVHTDKEDGTPLGGSSLP
PPFPPCSLRTALAQYADLLSSPKKSALLALAAHASDPTEADRLRHLASPAGKDEYAQ
WIVASQRSLLEVMAQFPSAKPPLGVFFAAVAPRLQPRYYSISSSPRVAPSRIHVTCA
LVYEKTPTGRIHKGVCSTWMKNCVPMEKSNDCGWAPIFVRQSNFRLPADPKVPVIMI
GPGTGLAPFRGFLQERFALKEAGAELGPSVLFFGCRNRQMDYIYEDELNNFVQSGAL
SELVVAFSRQGPTKEYVQHKMMEKASDIWNMISQGGYLYVCGDAKGMARDVHRTLHT
IVQEQGSVDSSKAESIVKNLQMTGRYLRDVW
```

8)   Cytochrome P450 1 (AiCYP1)

>Neem_transcript_34861 /Master_ Control_84673

```
ATGGAGCTCTTCTTACTCTCCGTACTTCTCATTCTTACTGTCTTCTGCTTCTATTAT
CTCTTCAAACCCAAGCAAACCCAGAATGCTCCTCTCCCACCAGGTCATGTGCATTGG
CCTCAGAAGATTTTTGAAACTTTAGATTACATTTCCAAGGCCAGAACCAACACTATT
CACAAATTCATTGCTGCGAGGAAAAACAAATACAACACCAAGCTTTTCAAGACGTCT
CATATTGGTCAAAACATGGTGTTTTTATGCACTCCTGAGGGGAACAAGTTCTTGTTT
GCTAATGATTACAAGCTTGTTAGATCCTGGTGGCCTATAACTTTTCTCAGAGTGTTT
GAAACGCCGAGGAAGAGATTACTCCGGAGCAAGTTCTCAGAGCTCGCAAGCAGTTT
CTCTCCTCCTTCAATGAGCCTGAGGCTCTTGCAAAACATGTTGGTATTACTGATGAG
GTCGTGAAAGATCACTTCAAGTTGTTCTGGGACGGGTCTGATGAAGTTACTGTTTAC
CCTATTGCAAGGAAGTTGACGTTTGCGATATCTTGTAGATTGCTTGCTGATATCAGG
GATCGGGAAATTCTCGACGAGCTTCTGCCTGCGATGGGAGATGTTGTGGCTGCGTTC
TTTGCGTTGCCTATAAACCTTCCAGGTACAAAATTCAATCGTGCTGTTAAAGGGTCG
AGAAAATGTCGCAAGATATTCGTGGACATCATTAAGCAGAGAAAGATTGATTTGTTT
GAGAAGGATAGAAAAGAAGCAAACGACGTTCTGTCGAACATATTACTGGAGAATCAT
AGGGATGGTATAGAAACAAAGGACGTCGTACTTGCTAAGAACTTGGTATCATTGCTT
TCAGCTGCGTTTGATAATCCCAGTGTAACCATTGTTTCCATTATAAAAAATCTTGCA
GAGAATCCTGAAATCTATGCAAGAGTTCGTTCAGAGCAATTGGAGATAGCGAAAGGA
AAAGCACCAGGAGAGAATTTGACAATGGAGGACCTGAAAAAGATGAAGTTCTCAATG
AATGTGTTGAGTGAATCTCTGAGAATGGAGGCACCTGCTTCTGGGACTTTCAGGGAA
GCCCTCAATGACTTCACCTATGAAGGATATTTGATTCCAAAGGGATGGAAGGTACAC
TGGAGTGTTCATGCAACACACAGGAATCCACAGTACTTCAAAGACCCAGAAAAGTTT
GATCCATCAAGATTCGAACGAAATGACCCAATTGTCCCATACTCATATGTACCATTC
GGAGGAGGCCATCACATTTGCCCTGGAAAAGATTTTGCCAAGCTACAAATTCTTATC
```

```
TTCATTCATCATGTTGTCAAGAAATTCAACTGGGAAAAGGTTAATCCTGACGAGCAA
ATGGTCAGAGTTCCAAACTTAAAGGCAGCCAAGGGGCTTCCAGTTCGTCTCTATCCC
TACAACAATTAA
```

>AiCYP1_Protein_Sequence

```
MELFLLSVLLILTVFCFYYLFKPKQTQNAPLPPGHVHWPQKIFETLDYISKARTNTI
HKFIAARKNKYNTKLFKTSHIGQNMVFLCTPEGNKFLFANDYKLVRSWWPITFLRVF
ENAEEEITPEQVLRARKQFLSSFNEPEALAKHVGITDEVVKDHFKLFWDGSDEVTVY
PIARKLTFAISCRLLADIRDREILDELLPAMGDVVAAFFALPINLPGTKFNRAVKGS
RKCRKIFVDIIKQRKIDLFEKDRKEANDVLSNILLENHRDGIETKDVVLAKNLVSLL
SAAFDNPSVTIVSIIKNLAENPEIYARVRSEQLEIAKGKAPGENLTMEDLKKMKFSM
NVLSESLRMEAPASGTFREALNDFTYEGYLIPKGWKVHWSVHATHRNPQYFKDPEKF
DPSRFERNDPIVPYSYVPFGGGHHICPGKDFAKLQILIFIHHVVKKFNWEKVNPDEQ
MVRVPNLKAAKGLPVRLYPYNN
```

9)   Cytochrome P450 2 (AiCYP2)

>Neem_transcript_38933/ Master_Control_57632

```
ATGGGTTTAGATTTGTTGTGGTTGATTCTTGCAATTGTGGTAGGCACTTATGTTGTT
CTGTTTGGGTTTCTGAGGAGAGCAAATGAGTGGTACTATTCCATAAAATTGGGTGAC
AAGAGCCGTTATCTTCCTCCAGGAGACATGGGATGGCCTATAATTGGTAACATGATA
CCCTATTTCAAAGGCTTCAGATCTGGTGAACCTGAAAGTTTCATCTTCGACTTATTT
GAAAGTATGGTCGAAAAGGGATATACAGAAATCACATATTCGGAAGCCCCAGTATA
ATAGTTTTGGCTCCTGAAGCTTGCAGGCAGGTATTTCTGGACGATGATAACTTTAAA
ATGGGTTATCCTGAATCTACCAATAAGTTAACATTCAGAGGATCGTTCAATACTGCT
TCTAAAGAAGGGCAGAGGCGTATAAGGAAGCTAGCAACAAGTCCGATAAGAGGACAC
AAGGCAATCGCCATTTACATTGACAACATTGAAGACATTGTGGTTAAATCTCTGAAA
GAATGGGCAAGTAAGGATAAACCAATCGAGTTCTTATCCGAAATGAGGAAAGCAACT
TTCAAGGTCATTTCCAACATCTTCCTGGGATCAAGCAGTGATTCTGTCATCGGTTCA
GTGGAGCAATACTACGTTGATTATGCCAACGGGTTGATCTCTCCTCTGGCCATTAAT
TTGCCTGGTTTCGCTTTCCATAAAGCAATGAAGGCAAGAGATATGTTGGGAGAGATC
TTGGAACCAATCCTGAGGGAAAGAAGATCAATGAAGGAAAAGGATCAGCTGAAGGGA
AAGAGGGGATTGGTTGATTTATTGATGGAGGTTGAAGATGAGAATGGTGAAAAGCTG
GAGGATGTCGATATTGTAGATATGCTCATCGCATTTTTGTCAGCTGGTCATGAAAGC
TCAGCGCATATTGCAACATGGGCAATAATCCACCTCCATAGACATCCTGAAATGCTT
CAAAAAGCAATAAAAGAGCAAGAGGAGATTGTAAAGAAAAGACCAGCTTCACAGCAA
GGATTCAGTGTTGAGGATTTTAAACGAATGGAATATATTGCCAAGGTAATAGATGAA
ACATTACGCATAACCAATCTTTCATCATCAAGCTTTCGAGAGGCAGAAGCAGATGTC
AACTTACAAGGTTATATAATACCAAAAGGATGGAAAGTTTTGCTTTACAATAGAGGT
GTTCATCGAAATCCTGAAAATTATCCAAACCCAAAGGAATTCGATCCTTCAAGATGG
GATAATTATGCGAACAGACCAGGATATTTCATTCCCTTTGGAGGAGGACCAAGGATT
TGCCCTGGAGCTGATCTAGCCAAACTTGAAATGTCCATTTTCATTCATTATTTTCTT
CTTAACTATAGACTCGAACCACTGAATCCTGAGTGTCCAACTGAGTACTTACCTGTA
CCAAGGCCTTCAGATCAATGTCTTGCAAGAATTATTAAGCTCAAATGA
```

>AiCYP2_Protein_Sequence

```
MGLDLLWLILAIVVGTYVVLFGFLRRANEWYYSIKLGDKSRYLPPGDMGWPIIGNMI
PYFKGFRSGEPESFIFDLFEKYGRKGIYRNHIFGSPSIIVLAPEACRQVFLDDDNFK
MGYPESTNKLTFRGSFNTASKEGQRRIRKLATSPIRGHKAIAIYIDNIEDIVVKSLK
EWASKDKPIEFLSEMRKATFKVISNIFLGSSSDSVIGSVEQYYVDYANGLISPLAIN
LPGFAFHKAMKARDMLGEILEPILRERRSMKEKDQLKGKRGLVDLLMEVEDENGEKL
EDVDIVDMLIAFLSAGHESSAHIATWAIIHLHRHPEMLQKAIKEQEEIVKKRPASQQ
GFSVEDFKRMEYIAKVIDETLRITNLSSSSFREAEADVNLQGYIIPKGWKVLLYNRG
VHRNPENYPNPKEFDPSRWDNYANRPGYFIPFGGGPRICPGADLAKLEMSIFIHYFL
LNYRLEPLNPECPTEYLPVPRPSDQCLARIIKLK
```

Thesis Summary

Neem (*Azadirachta indica*) is an evergreen tree, native to the Indian sub-continent. It has potential use in medicine, agriculture, environment protection and pest management. Neem has more than 150 different limonoid skeletons and their derivatives, which play a key role in its potential use. These limonoids are synthesized by triterpenoid biosynthetic pathway. Fifteen different limonoids were previously isolated and the metabolic fingerprinting across different tissues of neem was done by my senior in this lab. The resulting analysis of these data showed that basic and C-seco limonoids were abundant in pericarp and kernel, respectively and low in flowers. To understand biosynthesis of limonoids in neem, pooled RNA (fruit, flower and leaves) transcriptome was analyzed. The library preparation, sequencing and assembly were performed at Genotypic Technology Pvt. Ltd, Bengaluru, India. Assembly was done by Velvet_1.1.05 followed by Oasis_0.2.01 with hash length 41. Functional annotation of transcripts was carried out by using BLASTx (nr database), KAAS (KEGG Automatic Annotation Server), Virtual ribosome (translation and ORF finding) and Pfam domain. From the transcriptome analysis, we were able to predict genes related to MVA and MEP pathways viz., prenyltransferases, squalene synthase, squalene epoxidases, CYP450 reductases, three triterpene synthases and CYP450 genes.

The pooled transcriptome analysis identified three triterpene synthases and CYP450s. Tissue-specific (pericarp, kernel, flower and leaves) transcriptome analysis was done to understand the involvement of identified genes in limonoid biosynthesis. Expression profile data (RPKM) of terpenoid biosynthesis (manually curetted) in neem stated that triterpene biosynthetic genes (from MVA pathway rate-limiting genes to triterpene synthases) were highly expressed in seeds (kernel and pericarp), which correlated with limonoid fingerprinting data. Steroid biosynthetic pathway genes showed higher expression in flowers and diterpenoid biosynthesis genes were highly expressed in leaves. This helped us to understand how the terpenoid pathway genes expression is regulated in order for production of limonoids abundantly in seeds as compared to other tissues. From differential gene expression (DESeq2) studies, MVA pathway rate-limiting enzymes, mevalonate kinase, farnesyl

diphosphate synthase (AiFDS), squalene synthase (AiSQS), squalene epoxidases (AiSQE3)and triterpene synthase 1 (AiTTS1) were over-expressed in kernel as compared to flowers. These gene expression analyses predicted the involvement of AiTTS1 in triterpenoid biosynthesis. Further, AiTTS1 involvement in limonoid biosynthesis was confirmed by real-time PCR. fifteen CYP450 genes were predicted to be involved in triterpenoid biosynthesis from annotation and expression analysis of the transcriptome.

Two short-chain prenyltransferases, homomeric AiGDS and AiFDS, were cloned and functionally characterized. The real-time PCR analysis revealed that AiFDS is involved in limonoids biosynthesis. The functional characterization of AiSQE was done by checking the change in sterol production in yeast. When AiSQE1 was expressed in yeast, two to three-fold increase in ergosterol and lanosterol was observed. Based on literature, euphol, tirucallol, butyrospermol, tirucalla-7,24-dien-3β-ol could possibly be involved in limonoids biosynthesis. The triterpene cyclic product formed by AiTTS1 was purified and analysed by NMR studies. The recombinant AiTTS1 produces tirucalla-7,24-dien-3β-ol which confirms its involvement in limonoids biosynthesis. AiTTS2 gene analysis showed a missing C terminal β-sheet which provides key active site amino acid for cyclization of 2,3-oxidosqualene. This was confirmed by 3'-RACE sequencing analysis. AiTTS1 was coexpressed with AiSQE1 resulting in two-fold increase in the production of tirucalla-7,24-dien-3β-ol. Two CYP450 reductase genes, which transfer the electrons from NADPH to CYP450, were cloned and characterized by cytochrome C reduction assay. One cytochrome P450 gene was cloned and transient transformation (overexpression) was done into neem, which resulted in increased production of azadirachtin A as compared to normal plants.

In the active site of triterpene synthases, 2,3-oxidosqualene forced to take up a pre-organized conformation. The protonation of the epoxide ring starts a cascade cyclization through intermediate carbocations. Aromatic amino acids present in the active site stabilize the intermediate carbocations through cation–π interactions. Skeletal rearrangement of carbocation by 1,2 methyl and 1,2 proton shifts (from high to low π electron density) takes place and finally deprotonation occurs to form the

triterpene cyclic product. Phylogenetic analysis of triterpene synthases revealed that synthases, which stabilize complex higher triterpene cyclic carbocations (ex: oleanyl C-19 cation) are evolved from synthases which stabilizes low complex triterpene cyclic carbocation (ex: protosteryl C-20 cation). This hypothesis is also observed in the active site residues of triterpene synthases. To prove this hypothesis, we started mutating AiTTS1 in such a way that it stabilizes oleanyl C-19 cation to form amyrins. Eight active site amino acids showed significant variation with amyrin synthase. AiTTS1 became inactive/showed lowest activity when mutated at V484L, V550T and L553F, this explains their key role in stabilizing dammarenyl C-20 cation. F260Y and Y125F increased the stability of dammarenyl C-20 cation. Further analyses are required for mutating AiTTS1 such that it can stabilize the oleanyl C-19 cation.

Thus the study on "*De novo* Sequencing and Analysis of Transcriptome from *zadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis" is the first to carryout and report the functional characterization of genes involved in limonoid biosynthesis. This work states that tirucalla-7,24-dien-3β-ol is involved in limonoids biosynthesis, which proves to be a novel finding. The overall work helps in identification of the downstream enzymes involved in limonoids biosynthesis.

List of Publications

- **Pandreka, Avinash**, Devdutta S. Dandekar, Saikat Haldar, Vairagkar Uttara, Shinde G. Vijayshree, Fayaj A. Mulani, Thiagarayaselvam Aarthy, and Hirekodathakallu V. Thulasiram. "Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*)." **BMC Plant Biology 15, no. 1 (2015): 214.**

- Srivastava, Prabhakar Lal, Pankaj P. Daramwar, Ramakrishnan Krithika, **Avinash Pandreka**, S. Shiva Shankar, and Hirekodathakallu V. Thulasiram. "Functional characterization of novel sesquiterpene synthases from Indian sandalwood, *Santalum album*." **Scientific reports 5 (2015): 10095**.

- **Pandreka, Avinash**., Chaya, P. S., Bhagyashree D. B., Ashish R. K., Aarthy, T., Mulani, F.A., and Hirekodathakallu V. Thulasiram, "Tirucalla-7,24-dien-3β-ol synthase, a key gene involved in Neem limonoid biosynthesis". (**Manuscript under Preparation**)

- **Pandreka, Avinash**., Chaya, P. S., Pooja D. Sharma., Aarthy, T., Mulani, F.A., and Hirekodathakallu V. Thulasiram. "Characterization of Cytochrome P450 Systams Involves in Neem Limonoids Biosynthesis". (**Manuscript under Preparation**)

- Ashish R. K., Sudha Punnusamy., Pooja D. Sharma., **Pandreka, Avinash.,** and Hirekodathakallu V. Thulasiram. "Cloning and Characterization of Triterpene Synthases from *Euphorbia tirucalli* and *Euphorbia grantii*". (**Manuscript under Preparation**)

- Thiagarayaselvam Aarthy, Fayaj A. Mulani, **Pandreka Avinash,** Ashish Kumar, Sharvani S Nandikol and Hirekodathakallu V. Thulasiram "Tracing the biosynthetic origin of limonoids and their functional groups through stable isotope labeling and pathway inhibition in *Azadirachta indica* cell suspension". **BMC Plant Biology, 2018: (Manuscript No. PBIO-D-17-00754, Under Review)**

AVINASH PANDREKA

CSIR- Institute of Genomics and Integrative Biology, New Delhi - 110025
CSIR- National Chemical Laboratory, Pune - 411008
✉ **Mail ID: 24.avinash@gmail.com**
📱 **Mob: +91 7798898719/9293653790**

Education:

Doctor of Philosophy (PhD) from Academic CSIR, Institute of Genomics and Integrative Biology, New Delhi, India

Title of Thesis: "***De novo* Sequencing and Analysis of Transcriptome from *Azadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis**"
*Expected by July 2018.

Research Interest:

- Elucidation of biosynthetic pathways and regulation
- Bioinformatics
- Molecular biology
- Protein and metabolic engineering

Academic Record:

Degree	Institute	University	Year of passing	Marks % CGPA
PhD (Guide: Dr. H. V. Thulasiram)	CSIR- Institute of Genomics and Integrative Biology	AcSIR, New Delhi, India	2018	
Master of Science (Biotechnology)	SRM University	SRM University, Chennai, India	2010	8.6
Bachelor of Science (Biochem/Biotech/Chem)	A.S.N. Degree College	Achary Nagarjuna University, Andhra Pradesh, India	2008	84

Awards and Honors:

- Cleared national entrance exam **CSIR- Junior Research Fellowship (JRF) and Lectureship award (LS) in Dec 2010 (Rank 212 UGC – JRF**) for a PhD Fellowship conducted by University Grant Commission (UGC) and Council of Scientific & Industrial Research (CSIR), Human Resource and Development Group (HRDM), Government of India, New Delhi.

- Cleared National entrance exam **DBT- JRF in 2011** (**Rank 25 Category B)** for PhD Fellowship conducted by National Center for Cellular Science (NCCS), Department of Biotechnology (DBT) and Ministry of Science and Technology, Government of India, New Delhi.

- Cleared (Graduate Aptitude Test in Engineering) **GATE-2011 with rank 411** in the Life Science field conducted by Department of Higher Education, Ministry of Human Resource Development (MHRD), Government of India.

Research Experience:

- August 2011 onwards (PhD Title "***De novo* Sequencing and Analysis of Transcriptome from *Azadirachta indica* to Characterize the Genes Involved in Limonoid Biosynthesis**")

  - RNA isolation, transcriptome analysis by **BLAST, KO number, Pfam, differential gene expression analysis (DESeq2), co-expression analysis (SOM and WGCNA)** and correlating limonoids profiling to predict genes involved in limonoids biosynthesis in *A. indica*
  - Elucidation of **gene expression profile of terpenoid biosynthesis** in different tissues and correlating with limonoids profiling of *A. indica*
  - Cloning and **characterization** of prenyltransferases (Homomeric geranyl diphosphate synthase - **AiGDS**, farnesyl diphosphate synthase - **AiFDS)**, squalene epoxidase - **AiSQE, tirucalla-7,24-dien-3β-ol synthase - AiTTS1,** inactive triterpene synthase - **AiTTS2,** cytochrome P450 reductases **- AiCPR1 and AiCPR2** from *A. Indica*
  - Cloning and characterization (in progress) of **two cytochrome P450** genes involved in limonoids biosynthesis
  - **Transient transformation** of predicted **cytochrome P450** involved in limonoids biosynthesis in *A. Indica*
  - Attempting **site-directed mutagenesis** in the active site residues of **tirucalla-7,24-dien-3β-ol synthase** for cyclization of 2,3-oxidosqualene into **higher cyclic triterpene (lupeol or amyrin)**
  - Co-expression of **squalene epoxidase and tirucalla-7,24-Dien-3β-ol synthase** in yeast for **metabolic engineering**
  - **Real-time PCR analysis** of AiGDS, AiFDS, AiTTS1 and AiTTS2
  - **Involved in analysing transcriptome** from *Santalum album Linn* (santalene biosynthesis)*, Catharanthus roseus* (Indole alkaloids)*, Euphorbia tirucalli* (Triterpenoid) and *Euphorbia grantii* (Triterpenoid)
  - **Involved in genomic integration or gene knockdown** in yeast for metabolic engineering

Research Expertise:

- **Bioinformatics**

  - **Bioedit and MS Excel** for transcriptome data analysis and handling
  - Functional annotation software: **BLAST2GO**, **KAAS**, **Virtual Ribosome**, **TargetP** and **Pfam** domain
  - Pathway analysis through online databases: **KEGG, BRENDA and METACYC**
  - Hands-on: **RStudio**, DESeq2, SOM and WGCNA
  - **Transcription factor** identification and **CYP450 classification**
  - **EasyModeller 4.0** for protein modeling and **PyMol** for visualization
  - Multiple sequence alignment using **ClustalW**, BoxShade server, **phylogenetic analysis**

- **Molecular Biology**:

    - **RNA isolation** from various tissues of *A. indica* and cDNA synthesis
    - **Gene RACE** (Rapid amplification of cDNA ends) to obtain full-length 5' and 3' ends of transcripts
    - **Cloning techniques** such as primer designing, plasmid isolation, PCR amplification, PCR purification, restriction digestion, gel elution, competent cell preparation.
    - Primer designing for **site-directed mutagenesis** and screening of mutants by analysing Sanger sequences.
    - **Real-time PCR analysis**
    - Designing cassette for **yeast genome integration**

- **Protein Biology**:

    - **Protein expression**, high-speed centrifuge, ultra-centrifuge, microsome preparation, homogenization, **protein solubilization**
    - **Affinity chromatography** from protein purification, **ÄKTA Avant, Western blotting and enzyme characterization**
    - **Spectrophotometric assays** from enzyme characterization

- **Organic Chemistry Techniques**:

    - **Metabolites extraction** from various systems such as enzyme assays, plants and yeast
    - **Thin layer chromatography and silica gel column chromatography** for metabolite separation
    - **GC/FID, GC/MS, GC/Q-TOF, LC/HRMS** analysis for characterization of metabolites
    - Hands-on **1D and 2D NMR** data analysis
    - **ChemDraw**

Research Publications:

- **Pandreka, Avinash**, Devdutta S. Dandekar, Saikat Haldar, Vairagkar Uttara, Shinde G. Vijayshree, Fayaj A. Mulani, Thiagarayaselvam Aarthy, and Hirekodathakallu V. Thulasiram. "Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*)." **BMC Plant Biology 15, no. 1 (2015): 214.**

- Srivastava, Prabhakar Lal, Pankaj P. Daramwar, Ramakrishnan Krithika, **Avinash Pandreka**, S. Shiva Shankar, and Hirekodathakallu V. Thulasiram. "Functional characterization of novel sesquiterpene synthases from Indian sandalwood, *Santalum album*." **Scientific reports 5 (2015): 10095**.

- **Pandreka, Avinash**., Chaya, P. S., Bhagyashree D. B., Ashish R. K., Aarthy, T., Mulani, F.A., and Hirekodathakallu V. Thulasiram, "Characterization of Tirucalla-7,24-dien-3β-ol synthase which involves in limonoids biosynthesis". (**Manuscript under preparation**)

- Thiagarayaselvam Aarthy, Fayaj A. Mulani, **Pandreka Avinash,** Ashish Kumar, Sharvani S Nandikol and Hirekodathakallu V. Thulasiram "Tracing the biosynthetic origin of

limonoids and their functional groups through stable isotope labeling and pathway inhibition in *Azadirachta indica* cell suspension". (**Manuscript under Revision**)

Conferences and Symposia:

- Poster presentation in "Indo-Mexican workshop on Biotechnology: Beyond Borders" organized by CSIR-NCL, Pune, October 2013.
- Poster presentation in "10th International Symposium on Bio-Organic Chemistry-ISBOC-2015", IISER, Pune, January 2015.
- Conducted tutorial for pathway analysis during "Training Course on Methodologies Used in Gene Expression analysis" organized by Department of Botany, Savitribai Phule Pune University, Pune, October 2015.

Personal Details:

Father's Name:	Veera Brahmamau P
Mother's Name:	Padmavati P
Date of Birth:	24th May 1988
Sex:	Male
Marital status:	Unmarried
Nationality:	Indian
Languages known:	English, Hindi, Telugu
Permanent Address:	S/O P. Veera Brahmamu, Balijepalli (Post), Vemuru (Mandal), Guntur (District), Andhra Pradesh, Pin: 522261, India

Declaration:

I hereby declare that the above-mentioned details are true to the best of my knowledge and belief.

Place: Pune   Avinash Pandreka