Major Secondary Metabolites and Their Biosynthesis in Selected Ocimum sp

Thesis submitted to the SPPU
For the degree of
Doctor of Philosophy
in
Biotechnology
by

Atul Anand

Research Supervisor
Dr. A.P. Giri
Research Co-Supervisor
Dr. H.V. Thulasiram
Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune 411008
India

November 2016

CERTIFICATE

This is to certify that the work presented in this thesis entitled, "**Major Secondary Metabolites and Their Biosynthesis in Selected Ocimum sp**" by **Mr. Atul Anand**, for the degree of **Doctor of Philosophy**, was carried out by the candidate under my supervision in the Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune-411008, India. Any material that has been obtained from other sources has been duly acknowledged in the thesis.

Place: Pune

Date:

Dr. A.P. Giri

(Research Guide)

Biochemical Sciences Division

CSIR-National Chemical Laboratory

Pune-411008, India

Dr. H. V. Thulasiram

(Research Co-Guide)

Organic Chemistry Division

CSIR-National Chemical Laboratory

Pune-411008, India

DECLARATION

I, Atul Anand, hereby declare that the work incorporated in the thesis entitled "**Major Secondary Metabolites and Their Biosynthesis in Selected Ocimum sp**" submitted by me to **Savitribai Phule Pune University** for the degree of **Doctor of Philosophy** is original and has not been submitted to this or any other University or Institution for the award of Degree or Diploma. Such material, as has been obtained from other sources, has been duly acknowledged.

Date:

Place: Pune

Atul Anand

Acknowledgement

It gives me immense pleasure to look over my Ph.D. journey and remember all who have helped and supported me enroute. First and foremost, I would like to express my sincere and heartfelt gratitude towards my research supervisor, Dr. A.P. Giri and research co-supervisor Dr. H.V. Thulasiram, for their invaluable advice, constructive support and his extensive discussions around my work, without which it would have been improbable for me to have completed my thesis.

I take this opportunity to sincerely acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi, for providing with the Junior and Senior Research Fellowships which facilitated me to carry out my work. I would like to thank The Director, National Chemical Laboratory (NCL), Pune and HODs of Biochemical Sciences and Division of Organic Chemistry, for infrastructure and support to accomplish my research work.

It is probably the best time to express my obligations towards my M.Sc. professors, Prof. J. K. Pal, Prof. W. N. Gade for kindling my scientific interest and encouraging me to explore various dimensions.

I would like to acknowledge my special friends, Gyan Prakash, Jitendra, Pradeep, Sandeep, Shamim, Abhishek, Manoj and Rajkumar.

I would like to extend a huge thanks to my labmates Prabhakar, Pankaj, Saikat, Dipesh, Harshal, Krithika, Shiva, Soniya, Avinash, Vijayshree, Montu, Vyankatesh, Anurag, Aarthy, Supriya, Fayaz, Pruthviraj, Yashpal, Priyanka Singh, Raviraj, Rahul and Khandekar uncle for providing a healthy and pleasant environment for work.

Last, but not least, I would like to thank God for the strength and courage.

--- Atul Anand

Contents

List of tables

List of figures

Abbreviations

Amp	Ampicillin
AA	Amino acid
BLAST	Basic Local Alignment Search Tool
bp	Base pair
Cam	Chloramphenicol
cDNA	Complementary DNA
C-terminal	Carboxy terminal
DEPC	Diethylpyrocarbonate
DTT	Dithiothritol
EDTA	Ethylene diamine tetra acetic acid disodium salt
FAD	Flavin adenine dinucleotide
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	Kilo dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LA	Luria agar
LB	Luria broth
mg	Milligram
mL	Millilitre
µg	Microgram
µL	Microlitre
µM	Micromolar
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate reduced
NCBI	National Center for Biotechnology Information
ng	Nanogram
N-terminal	Amino terminal
OD	Optical density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction

PDB	Protein Data Bank
pI	Isoelectric point
PMSF	Phenyl methyl sulphonyl fluoride
Pfam	Protein families
PVPP	Polyvinylpolypyrrolidone
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
Rt	Retention time
Sec	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TB	Terrific broth
UTR	Untranslated Region
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactoside

Thesis abstract

Ocimum species present a wide array of diverse secondary metabolites possessing immense medicinal and economic value. The importance of this genus is undisputable and exemplified in the ancient science of Chinese and Indian (Ayurveda) traditional medicine. Unlike several other plant species of *Artemisia, Salvia, Catharanthus, Taxus, Mentha,* etc. that are largely exploited, detailed characterization and identification of important metabolites from *Ocimum* species remained unexplored. This work describes the phenylpropanoid profiling in different tissues from five *Ocimum* species, which revealed significant variations in secondary metabolites including eugenol, eugenol methyl ether, estragole and methyl cinnamate levels. Expression analysis of pathway gene indicated the abundance of eugenol synthase gene in species accumulating more eugenol. Kinetic parameters and *in silico* proved coniferyl acetate to be the preferred substrate. Metabolite diversity of terpenoids was also studied in different tissues of Ocimum species and variations in accumulation pattern in relation to species and class of metabolite was observed. Further, β-caryophyllene synthase gene from *Ocimum kilimandscharicum* Gürke (OkBCS- GenBank accession no. KP226502) was characterized. The open reading frame of 1,629 bp encoded a protein of 542 amino acids with molecular mass of 63.6kDa and pI value of 5.66. The deduced amino acid sequence revealed 50-70% similarity with known sesquiterpene synthases from angiosperms. Recombinant OkBCS converted farnesyl diphosphate to β-caryophyllene as a major product (94%) and 6% α-humulene. Transcript abundance and metabolite variations of this metabolite were studied across different *Ocimum* species. Agro- infiltration based transient expression manipulation with OkBCS over-expression and silencing confirmed its role in β-caryophyllene biosynthesis.

Chapter 1: Introduction

Phenylpropanoids, a group of plant secondary metabolites derived from phenylalanine, typically have propenyl side chains attached to a phenyl ring. These compounds are auto-toxic in nature and perform a vital role in plant communication,

pollinator attraction and defence against herbivores and other pathogens [1-3]. Coniferyl and coumaryl alcohols are monolignol alcohol intermediates of the lignin biosynthetic pathway and serve as the precursors for phenylpropanoid biosynthesis. Biosynthesis of coumaryl alcohol involves coumaroyl-CoA reductase that acts on 4-coumaroyl-CoA to form 4-coumaraldehyde and provides a substrate for the synthesis of chavicol. These alcohols are first converted to acetates by coniferyl- and coumaryl alcohol acyl transferases. The reductive elimination of the acetate moiety which is catalyzed by eugenol synthase (EGS) yields the propenyl side chain. EGS utilizes coniferyl acetate and coumaryl acetate as substrates to produce eugenol and chavicol, respectively [4,5]. Methylation of eugenol and chavicol at the para-OH group of the phenyl ring is catalyzed by eugenol O-methyl transferase (EOMT) and chavicol O-methyl transferase (COMT) to form their methyl ether derivatives.

Terpenes are the largest group of natural products, present ubiquitously with immense diversity in their structure and function [6]. They perform versatile functions including communication and defence in plants [7]. Several terpenes that offer distinct fragrance to flower are crucial in attracting many insects and other pollinators [8]. Plants use different defence strategies for their survival and one of them is to recruit natural enemies of herbivores using induced volatiles. Among herbivore-induced volatiles, terpenoids that comprise mono- and sesqui-terpenes, are the most commonly utilized compounds as tactical arsenal [9]. Like other members of the *Lamiaceae* family, *Ocimum* plant synthesizes and accumulates these volatile compounds in the secretory capitate and peltate trichomes which are located on the surface of aerial parts of plants [10]. Mono- and sesqui-terpenes are the main constituents of volatile components, which impart them unlimited medicinal properties [11] as well as a characteristic flavor and taste of *Ocimum* [12]. Several species from the genus *Ocimum* are known to possess insecticidal and other important bioactive properties [13,14].

Chapter 2: Comparative Functional Characterization of *EGS* from *Ocimum* Species

In the metabolite profiling eugenol, eugenol methyl ether, methyl cinnamate and estragole were identified as the major phenylpropanoids. Young plant tissue such as

leaves and inflorescence had more phenylpropanoids than mature tissues. In mature leaves and flowers, substantial levels of these compounds were detected. There was extensive metabolite diversity, within and across different *Ocimum* species, e.g. selective presence of eugenol and eugenol methyl ether along with varied ratios of different phenylpropanoid in Oba species. The metabolite diversity across different species could be explained by the differential gene expression pattern, e.g. EOMT in OtI and OtII resulting in accumulation of eugenol methyl ether, whereas in other species expression level of this gene was low, thus, no accumulation of these metabolites was observed. In phylogenetic analysis, *Ocimum* EGSs were well separated from those of *Petunia hybrid*a, *Clarkia breweri*, *Vitis vinifera* and *Medicago truncatula* that were grouped in separate sub-cluster. These 10 EGSs obtained from five *Ocimum* species were distributed in two sub-groups. Sequence comparison revealed that EGSs from different *Ocimum* species had high sequence conservation suggesting evolutionary close to EGS from other plants. EGS expression levels were positively correlated with the eugenol levels in different tissues from *Ocimum* species. EGS expression was highest in OgI and OgII, where eugenol was the major phenylpropanoid. Variation in CFAT expression levels, along with EGS and EOMT levels might be crucial for selective eugenol accumulation among these *Ocimum* species. However, further experiments in terms of their in planta over-expression and silencing would be necessary to confirm their exact role. Characterizations and comparisons of four EGSs with coniferyl acetate and coumaryl acetate indicated that the former was preferred substrate based upon its enzyme affinity, turnover number and catalytic efficiency. Additionally, when the substrates were mixed together either in equi-molar ratio or in ratio of their Km, eugenol was the major product, with traces of chavicol detected in ObaI EGS and OtI EGS. A homology model of EGS was made using the reported ObEGS1 structure as a template and interactions of both the substrates with the enzyme were analyzed. The lack of a hydroxymethyl substituent at the 3-position, as in the case of coumaryl acetate, significantly altered its stable interaction with EGS resulting in reduced preference as a substrate. In case of EGS from Og, Ok and Ot, coumaryl acetate form polar contact with catalytic water and Tyr157. In case of ObaI EGS, coumaryl acetate forms contact with Tyr157 but there is absence of polar

interaction between coumaryl acetate and catalytic water, which might be one of the reasons for it inactivity with it.

Chapter3: Metabolic diversity and characterization of synthase involved in monoterpene accumulation

Monoterpenes are the simplest terpenes with two isoprene units. They are synthesized in the oil glands of plants and are widely used in the flavor and fragrance industry. Here, we have analyzed the terpene diversity in six different tissues of five different *Ocimum* species. We could observe specific metabolite distribution between different species and some of them, like β- caryophyllene were present in all the species. We isolated two putative monoterpene synthases from *O. tenuiflorum* and analyzed the sequences. For identification of their product, these two genes were cloned in pRI-101 vector and their over-expression was studied through agro-infiltration in *Ocimum* plants. In case of *OtMTS1*, accumulation of cis-β-terpineol was detected with 2.5 folds increase in its transcript abundance.

Chapter 4: Functional characterization of β-caryophyllene synthase – a sesquiterpene synthase

β-caryophyllene, a bicyclic sesqui-terpene, is widely distributed in plant kingdom including Ocimum species. Like most terpenes, it contributes unique aroma in essential oils from numerous plants. β-caryophyllene might be effective for treating cancer, anxiety and depression and FDA has approved it as food additive. Additionally due to insecticidal nature, it could be potentially useful in plant defence. Though terpenoid biosynthesis from genus Ocimum has received considerable attention, very little is known on β-caryophyllene biosynthesis in Ocimum. In this chapter, we carried molecular characterization of β-caryophyllene synthase from *Ocimum kilimandscharicum* Gürke (OkBCS). Correlation of OkBCS transcript levels with metabolic variations in different tissues from five Ocimum species was performed. Further, transient manipulation of

OkBCS expression in terms of silencing and over expression validated its role in β-caryophyllene biosynthesis.

References:

[1] Gang DR, Wang J, Dudareva N, et al., An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, Plant Physiol. 125 (2001) 539-555.

[2] Louie GV, Baiga TJ, Bowman ME, et al., Structure and reaction mechanism of basil eugenol synthase, PLoS ONE 2 (2007) e993.

[3] Reigosa MJ, Pazos-Malvido E, Phytotoxic effects of 21 plant secondary metabolites on *Arabidopsis thaliana* germination and root growth, J. Chem. Ecol. 33 (2007) 1456-1466.

[4] Koeduka T, Fridman E, Gang DR, et al., Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 10128-10133.

[5] Vassao DG, Gang DR, Koeduka T, et al., Chavicol formation in sweet basil (*Ocimum basilicum*): cleavage of an esterified C9 hydroxyl group with NAD(P)H-dependent reduction, Org. Biomol. Chem. 4 (2006) 2733-2744.

[6] Rock CD, Zeevaart JAD, The ABA mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 7496–7499.

[7] Das A, Lee SH, Hyun TK, et al., Plant volatiles as method of communication, Plant Biotechnol. Rep. 7 (2013) 9–26.

[8] Pichersky E, Gershenzon J, The formation and function of plant volatiles: perfumes for pollinator attraction and defense, Curr. Opin. Plant Biol. 5 (2002) 237–243.

[9] Kessler A, Baldwin IT, Plant responses to insect herbivory: The emerging molecular analysis, Annu. Rev. Plant Biol. 53 (2002) 299–328.

[10] Gang DR, Wang J, Dudareva N, et al., An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, Plant Physiol. 125 (2001) 539–555.

[11] Hakkim FL, Arivazhagan G, Boopathy R, Antioxidant property of selected *Ocimum* species and their secondary metabolite content, J. Med. Plants Res. 2 (2008) 250–257.

[12] Prakash P, Gupta N, Therapeutic uses of *Ocimum sanctum* Linn (Tulsi) with a note on eugenol and its pharmacological actions: a short review, Indian J. Physiol. Pharmacol. 49 (2005) 125–131.

[13] Singh P, Ramesha HJ, Sarate P, et al., Insecticidal potential of defense metabolites from *Ocimum kilimandscharicum* against *Helicoverpa armigera*, PLoS One 9 (2014) e104377.

Chapter 1

Introduction

1. Introduction

Plants produce a diverse array of compounds that can broadly be categorized into primary and secondary metabolites. Primary metabolites are the ones which are required for the normal growth and biological processes and are produced in the pathways that are crucial for plant survival. The other class of metabolites, though generally termed secondary, is also very crucial for plants from the ecological perspective. These are produced by bacteria, fungi, yeast and other organisms and display substantial structural differences with diverse biological functions [1]. As plants are immobile they depend on these metabolites for interaction with its environment, with micro-organisms and other plants [2,3] and in reproduction by acting as pollination attractors [4]. Apart from these, several commercial applications like dyes, drugs, flavors, fragrances and insecticide, are also known for these metabolites [5]. These secondary metabolites are classified on the basis of their biosynthetic pathway and following types are frequently observed – terpenes, phenylpropanoids, alkaloids, saponins and glucosinolates. The availability of carbon, nitrogen and sulfur along with energy from the primary metabolism governs the biosynthesis of these compounds [6].

1.1 Classes of secondary metabolites

1.1.1 Terpenes

Terpenes are the most structurally varied class of secondary metabolites. They are synthesized from five carbon precursors' di-methyl allyl pyrophosphate (DMAPP) and iso-pentenyl diphosphate (IPP). These building blocks of terpenes, also called isoprenes, can condense together in head-to-head-, head-to-tail- and sometimes in head-to-middle fashion to generate an enormous variety of terpenes [7]. Biosynthesis of terpenes can take place either in cytosol (Mevalonic acid pathway or MVA) [8,9] and/or plastids (Methyl erythritol pathway or MEP) [10,11,12]. This compartmentalization for DMAPP/IPP formation is not absolute, because in several cases one metabolite can exchange between these two pathways [13,14]. Terpene synthases catalyze biosynthesis with high regio- and stereo-chemical precision involving intermediates that undergo a sequence of reactions such as cyclizations, alkylations, rearrangements, deprotonations and hydride shifts [15,16,17,18,19,20,21,22]. Terpenes can be classified on the basis of five carbon units

present in the core structure viz. single isoprene unit (C5) are called hemiterpenes, two isoprene units (C10) are called monoterpenes, three isoprene unit (C15) called sesquiterpenes, and so on.

1.1.2 Phenylpropanoids

The successful land adaptation of plants was made possible by the biosynthesis of phenolic compounds by plants that serve to provide the structural support to the plants. While a large fraction of these compounds serves in providing the structural integrity, non-structural molecules are also formed that function in defending the plant, color of the flowers, flavours etc. Phenylpropanoids are derived from amino acid phenylalanine, and have propenyl side chains attached to a phenyl ring. These compounds are auto-toxic in nature and perform a vital role in plant communication, pollinator attraction and defense against herbivores and other pathogens [23,24,25]. The general phenylpropanoid pathway constitutes the first three steps, where phenylalanine is converted to 4-coumaroyl CoA. This molecule is then used as the precursor for biosynthesis of various phenylpropanoid derivatives [26]. Coniferyl and coumaryl alcohols, the precursors for phenylpropanoid biosynthesis, are monolignol alcohol intermediates of the lignin biosynthetic pathway. Biosynthesis of coumaryl alcohol involves coumaroyl-CoA reductase that acts on 4-coumaroyl-CoA to form 4-coumaraldehyde and provides a substrate for the synthesis of these alcohols.

1.1.3 Alkaloids

Alkaloids are nitrogen containing, low molecular weight compound derived from amino acids found in plants and have been reported to be present in numerous animals [7]. The nitrogen atom may be present on the ring structure (true alkaloids) or on the side chains (pseudo-alkaloids) [27]. Owing to their potent biological activity, alkaloids are often used as pharmaceuticals, stimulants and narcotics [28]. The classification of alkaloids is based on the basis of their structural or chemical feature or its biological origin e.g. terpenoid indole alkaloids, benzylisoquinoline alkaloids, tropane alkaloids etc. The biosynthesis of different type of alkaloid starts from different class of precursors [28,29].

1.1.4 Saponins

Saponins are secondary metabolites that consists of tri-terpene or steroidal aglycones that are linked to oligosaccharide moieties. This combination of hydrophilic and

hydrophobic groups gives them foaming and emulsifying properties, and hence used as foaming agent in carbonated drinks and cosmetics and flavoring agent in baked foods etc [30]. Saponins have varied distribution among different plant species and their accumulation in plants also varies according to season and developmental stages. The well known activity of these compounds includes anti-carcinogenic, anti-microbial, anti-oxidant, insecticidal and anti-feedant properties.

1.1.5 Glucosinolates

Glucosinolates are commonly found in the *Brassicaceae* family and contain nitrogen and sulfur molecules. They are classified on the basis of nature of amino acid they are derived from, accordingly glucosinolates are of three types – a) aliphatic glucosinolates - from methionine, leucine, isoleucine, or valine, b) aromatic glucosinolates - from phenylalanine or tyrosine, and c) indole glucosinolates - from tryptophan. The side chain elongation of these amino acid precursors, along with other secondary modifications like oxidation, desaturation, hydroxylation, methoxylation, sulfation and glucosylation result in structural diversity of these metabolites [31]. Their activity as cancer prevention agent, crop protection compounds and as bio-fumigants has already been established [32].

1.2 Biosynthesis of secondary metabolites

1.2.1 Terpenes

Terpenes, biosynthesized from two simple five-carbon building blocks IPP and DMAPP, constitute the largest and diverse class of naturally occurring organic compounds that is found to be present in all life forms. Over 70,000 individual structures of this class of secondary metabolite has been reported with variations in carbon skeletons and functional group [33,34]. Plants serve as an excellent source of terpenes with diverse chemical structures, from universal primary metabolites such as sterols (components of bio-membrane) [35,36], carotenoids and chlorophyll (photosynthetic pigments) [37,38], ubiquinones (electron transport chain) [39,40], vitamins and hormones [41,42] to more unique and species-specific secondary metabolites where they find role in plant defense and communication [43]. Their derivatives are used for the production of drugs like taxol [44], artemisinin [45], etc.

1.2.1.1 Mevalonate pathway (MVA) for biosynthesis of IPP and DMAPP

This pathway was first discovered in yeast and animals in 1950 and leads to biosynthesis of IPP and DMAPP in six enzymatic steps starting from Acetyl-CoA [46,47]. This pathway operates in cytosol and is present in animals, plants, fungi, archaea and some bacteria. The pathway starts by the condensation of two molecules of acetyl-Co-A, to form acetoacetyl-CoA, the reaction being catalyzed by acetoacetyl-CoA transferase. This acetoacetyl-CoA is converted into 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by HMG synthase. The latter step involves conversion of HMG-CoA to mevalonic acid (MVA) catalyzed by 3-hydroxy-3-methylglutaryl-CoA reductase, which utilizes NADPH as co-factor. Two successive phosphorylations catalyzed respectively by, MVA kinase and phospho-MVA kinase, leads to formation of MVA 5-diphosphate. In the last step of IPP biosynthesis undergoes An ATP-dependent decarboxylation of MVA 5-diphosphate catalyzed by diphospho-MVA decarboxylase, constitute the last step of IPP biosynthesis. IPP undergoes isomerization to form DMAPP in a reaction catalyzed by IPP isomerase (IDI). The requirement of cofactor led to the identification of two different types of IDI. Type 1 IDI belongs to zinc metallo-proteins and is widely distributed in fungi, mammals and plants [48]. Type 2 IDI requires flavin mononucleotide (FMN) and NADH for its activity [49,50] and is restricted to archaea and bacteria.

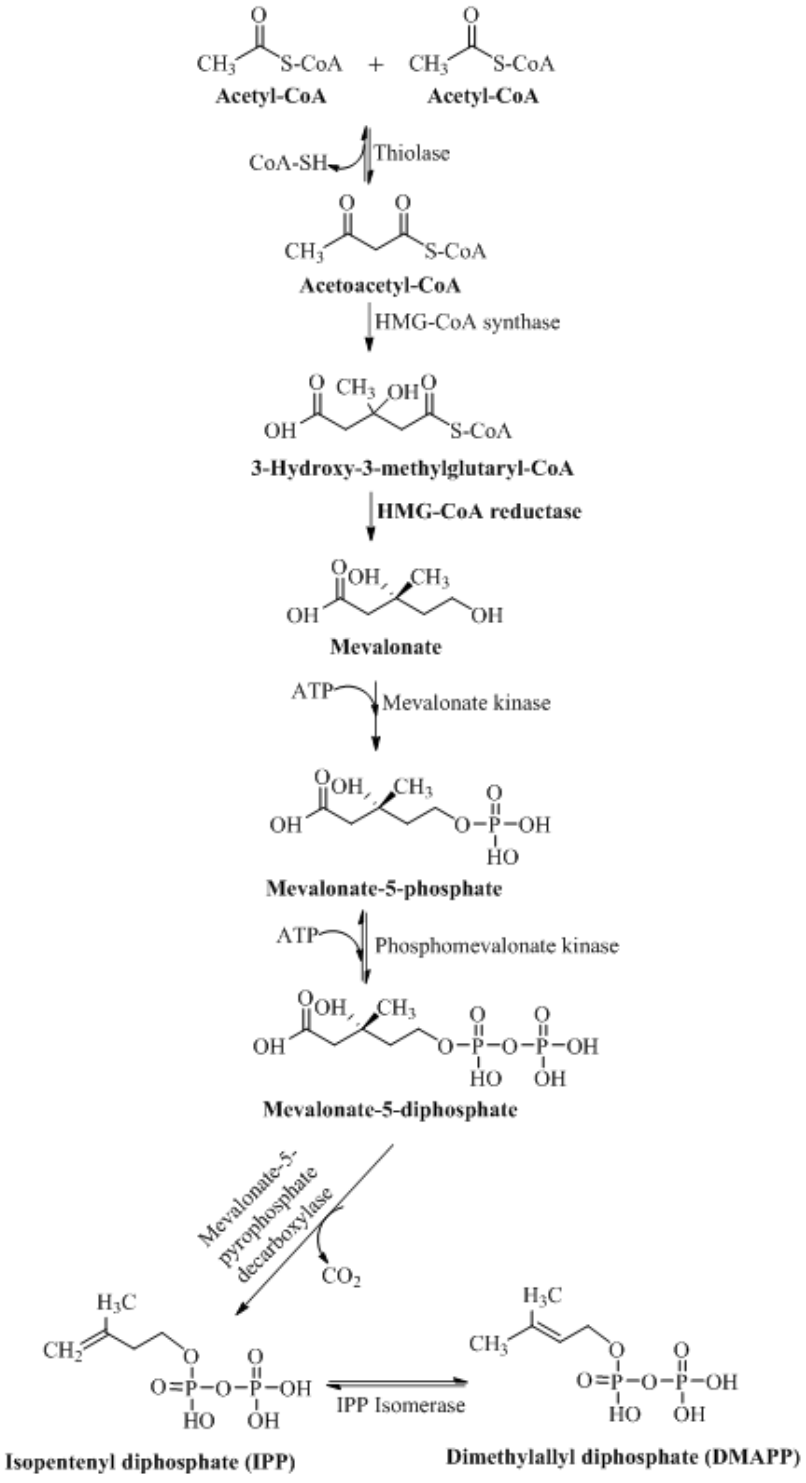


Figure 1: Schematic representation of mevalonic acid pathway

1.2.1.2 Methyl Erythritol Pathway (MEP) / 1-deoxy-D-xylulose 5-phosphate (DXP) pathway

Rohmer and co-workers established the existence of an alternative pathway (the MEP or DXP pathway) [51-53] , present in many eubacteria and plant plastids but absent in humans [43] which makes it an excellent target for developing broad spectrum antibiotics [54-56], as many human pathogenic bacteria rely on MEP pathway for the biosynthesis of essential terpenes. In plants, carotenoids, chlorophylls and prenylquinones are synthesized through the MEP pathway operated in plastids. The pathway comprises of seven steps, starting from glyceraldehyde-3-phosphate and pyruvic acid and leads to the synthesis of IPP and DMAPP in a ratio of 5:1. The first step in the pathway is the synthesis of 1-deoxy-D-xylulose-5-phosphate (DOXP), which is catalyzed by DOXP synthase (DXS) in a trans-ketolase type decarboxylation from glyceraldehyde-3-phosphate and pyruvate. In the second step, DOXP is transformed into MEP catalyzed by the NADPH-dependent DOXP reducto-isomerase. MEP is then converted into 4-(diphosphocytidyl)-2-C-methyl-D-erythritol (CDP-ME) in a CTP (cytidine triphosphate) dependent reaction catalyzed by 2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase. The CDP-ME in the next step undergoes phosphorylation and gets converted into 4-(diphosphocitidyl)-2-C-methyl-D-erythritol-2-phosphate (CDP-ME2P), catalyzed by CDP-ME kinase. CDP-ME2P is then converted into a cyclic 2-C-methyl-D-erythritol 2,4 cyclodiphosphate (MEcPP) mediated by MECP synthase. In the last step, MEcPP is converted into IPP and DMAPP via an intermediate step 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP) catalyzed by HMBPP synthase and HMBPP reductase, both of these steps require NADPH as cofactor.

Figure 2: Schematic representation of Methyl erythritol pathway

1.2.1.3 Major branch points and downstream products

In plants, IPP and dimethylallyl diphosphate DMAPP synthesized from MVA pathway are utilized for the synthesis of biosynthesis in cytosol and mitochondria, whereas IPP and DMAPP synthesized from MEP/DOXP pathway are used in the plastid. DMAPP acts as the primary active substrate for the synthesis of prenyl diphosphates: geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) by head to tail condensation with its isomer IPP. It also acts as a source for synthesis of cytokines and hemiterpenes whereas GPP, FPP and GGPP are the branch points for synthesis of terpenes. GPP synthase, localized in plastids, produces GPP as a precursor of monoterpenes biosynthesis and functions as homomeric and heteromeric enzymes [57-59]. FPP is synthesized mainly in cytosol and mitochondria by homo-dimeric enzyme (FPP synthase) and act as a main branching point for the biosynthesis of sterols, brassinosteroids, dolichols, polyprenols and for protein prenylation [60]. GGPP produced by GGPP synthase act as a precursor for the synthesis of diterpenoids, gibberellins, chlorophylls, phylloquinone, plastoquinone, tocopherols, carotenoids, abscisic acid and oligoprenols. GGPP synthase functions as homodimeric enzyme localized in plastids, mitochondria, and endoplasmic reticulum and utilizes all three allylic prenyl diphosphates (DMAPP, GPP, and FPP) [61].

1.2.1.4 Cross talk between MVA and MEP

In higher plants, these pathways function simultaneously with a compartmental segregation. Experiments with stable isotopes indicate this separation to not be absolute, because in several cases one metabolite can exchange between these two pathways [57-62]. Biosynthetic studies of diterpene ginkgolide using ^{13}C glucose indicated that three IPP are formed through MVA pathway and one from DXP pathway is utilized. Similar studies in liverwort (*H. planus*) and hornwort (*A. punctatus*) shows FPP portion (end C15 unit) of phytol is derived from MVA pathway whereas, terminal IPP portion (C5 unit) is derived from the DXP pathway [63-65]. These results indicate that there is a potential exchange of terpene building blocks between both MVA and DXP pathways which plays a crucial role in regulation of terpene biosynthesis [66-68]. The presence of MVA and MEP pathway enzymes in different compartments of plants suggest a mixed biosynthetic origin. In *A. thaliana*, GGPP synthase which synthesizes GGPP as a precursor for di-terpenoid, carotenoids

and chlorophylls was found to be present in endoplasmic reticulum and mitochondria in addition to plastids [61].

1.2.1.5 Regulation of MVA and MEP

Terpene biosynthesis is regulated differently and at multiple steps in organisms. In yeast and mammalian cells, regulation occurs at the levels of transcription, translation, posttranslational and protein degradation [69], whereas in bacteria, regulation occurs mainly at transcriptional level [70]. The regulation of both these pathways in plants is more complex as compared to other organisms. The presence of both the pathways in different compartments of one cell, existence of isozymes and different environmental stimuli add to the complexity of regulation of terpene biosynthesis. Regulation of MVA and MEP pathways in plants occurs mainly at transcriptional level; transcription of genes for enzymes of these pathways is not very tightly co-ordinated. Transcription of individual enzymes or isozymes can vary significantly in different tissues and developmental stages. MVA pathway genes are mainly expressed in roots, flowers and seeds, whereas MEP pathway genes are predominantly expressed in photosynthetic tissues [71]. Light also plays critical part in regulation of terpene biosynthesis from MEP pathway. The pigment protein complexes involved in photosynthesis like chlorophylls, carotenoids, xanthophylls, side chains of phylloquinone and plastoquinone are derived from the MEP pathway. Exposure of dark-grown *Arabidopsis* seedlings to low frequency red light results in photomorphogenesis and etioplast development in chloroplast. On the other hand, plants exposed to far red light in turn accumulate lutein and α-tocopherol without changes in transcript abundance. These results suggest that MEP pathway is regulated at transcriptional and post transcriptional levels during photomorphogenesis [72]. The main rate determining enzyme of MVA pathway in fungi, mammals, insects and plants is hydroxymethylglutaryl CoA reductase (HMGR), which catalyzes the formation of Mevalonate from 3-hydroxy-3-methylglutaryl-CoA [73-77] regulated by feedback-inhibition. HMGR is anchored to endoplasmic reticulum (ER) whereas all other enzymes of MVA pathway have been found in cytoplasm and peroxisome [78,79].

In vivo feeding experiments suggested that DXS in MEP pathway is a rate limiting step in the biosynthesis of IPP and DMAPP [80,81]. Analysis of *Arabidopsis* transgenic lines [82], tomato [83,84], potato [85] and *Ginkoba biloba* [86] indicates

that changes in the level of DXS changes the formation of final product of terpenes including chlorophyll, carotenoids, tocopherols and abscisic acid (ABA). In addition to DXS, DXR (DXR reductoisomerase) and HDR (4-hydroxy-3-methylbut-2-enyl diphosphate reductase) also have rate limiting roles in MEP pathway for the biosynthesis of IPP and DMAPP.

1.2.1.6 Structural classification of terpenes

Terpenes can be classified into different groups according to the number of five carbon units present in the core structure.

1.2.1.6.1 Hemiterpene (C5)

Hemiterpenes consist of a single isoprene unit (C5), and their oxygenated derivatives are called as hemi-terpenoids eg: prenol, isovaleric acid. Several plant species from mosses, fern, and trees emit isoprenes in very large amounts. Isoprenes are synthesized by chloroplastic enzymes directly from DMAPP by diphosphate elimination, which is produced from MEP pathway [87-89].

Figure 3: Biosynthesis of hemiterpenes

Isoprene emission plays an important role in protecting leaves from abiotic stress (short high temperature condition). The evolution of isoprene emission in plants may have been important in allowing plants to survive during rapid temperature changes.

1.2.1.6.2 Monoterpene (C10)

Monoterpenes consist of two isoprene units with the molecular formula $C_{10}H_{16}$ and are present in secretory tissues such as oil glands of higher plants, insects, fungi and marine organisms. These molecules are volatile in nature (boiling point in the range of 150-185 °C), and are less dense than water. These compounds are widely used in flavour and fragrance industry, due to their characteristic odour. Oxygenated derivatives of monoterpenes are more widespread in nature with greater importance. Monoterpenes can be divided into three subgroups: acyclic (myrcene, geraniol,

linalool), monocyclic (limonene, α-terpineol and terpinolene) and bicyclic (α-pinene, sabinene and camphor). Monoterpenes are biosynthesized from geranyl pyrophosphate (GPP) catalyzed by monoterpene synthase.

Figure 4: Biosynthesis of monoterpenes

Several monoterpenes possess various pharmacological properties including antibacterial antifungal, antioxidant, and anti-cancerous [90-92]. In plants, monoterpenes are synthesized in plastids through the DXP pathway, whereas in other higher organisms and in yeast, they are synthesized through the MVA pathway [93].

1.2.1.6.3 Sesquiterpene (C15)

Sesquiterpenes are the most diverse group of isoprenoids consisting of three isoprene units with the molecular formula $C_{15}H_{24}$. This abundant group of isoprenoids consists of over 7000 molecules with more than 300 stereo chemically distinct hydrocarbon skeletons. Sesquiterpenes are derived from the linear substrate, farnesyl pyrophosphate (C15) which is biosynthesized by FDS by head to tail condensation of IPP and GPP or IPP and DMAPP. They represent one of the most important components of plant essential oils along with monoterpenes. The large carbon skeleton of farnesyl diphosphate (FPP) and the presence of three double bonds greatly increase the structural diversity of the resulting products. Sesquiterpenes can be divided into four subgroups: acyclic (farnesene, farnesol, nerolidol), monocyclic (bisabolene, curcumene, zingiberene), bicyclic (bergamotene, β-santalene, aristolochene, cadinene, caryophyllene) and tricyclic (α-santalene, longifolene). Sesquiterpenes are biosynthesized from MVA pathway and targeted in the cytosol [94].

Figure 5: Biosynthesis of sesquiterpenes

Sesquiterpenes are traditionally used as flavours and fragrances but they have potential to serve as anticancer [95] and anti-malarial [96]. In recent years, sesquiterpenes of farnesene and bisabolene skeletons have been recognized as replacements for petroleum-derived fuels [97].

1.2.1.6.4 Diterpene (C20)

Diterpenes consist of four isoprene units with the molecular formula $C_{20}H_{32}$ and are widely distributed in nature. The diterpene compounds are derived from geranylgeranyl diphosphate (GGPP) and divided into the following subgroups: linear (geranylgeraniol, geranyl linalool), bicyclic (casbene, sclareol), tricyclic (abietadiene, taxadiene), tetracyclic (kaurene aphidicolanes, gibberellenes) and macrocyclic (cembranes, daphnanes, tiglianes). They are mainly present in polyoxygenated form with keto and hydroxyl group.

Figure 6: Biosynthesis of diterpenes

Diterpenes have attracted attention because of their biological and pharmacological activities such as anti-bacterial, anti-fungal, anti-inflammatory and anti-leishmanial

activity. Taxol is one of the most effective diterpenes which possesses anticancer properties against a wide range of cancers.

1.2.1.6.5 Triterpene (C30)

Triterpenes consist of six isoprene units and are derived from squalene which is synthesized by head to head condensation of two units of farnesyl diphosphate. Triterpenes constitute the most important class of terpenoids and exhibit a wide range of structural diversity and regulate membrane fluidity and permeability with lipids and proteins [98]. Sterols are precursors for a vast array of compounds involved in important cellular processes in animals whereas plant sterols are linked to brassinosteroids synthesis [99]. Brassinosteroids are polyhydroxy plant steroids essential for normal growth and development and present in all parts of plants.

Figure 7: Biosynthesis of triterpenes

1.2.2 Phenylpropanoids

Phenylpropanoid biosynthesis involves a series of branching enzymatic reactions, to provide plants with a vast array of phenolic products. These metabolites are generally species and tissue specific, and perform a variety of different function like structural components (lignin), protection of plants from biotic and abiotic stress, pigments and signaling molecules. The pathway starts from amino acid phenylalanine, which after three steps gives rise to coumaryl-CoA. Phenylalanine ammonia lyase (PAL) catalyzes non-oxidative de-amination and directs the carbon flow from shikimate pathway towards the phenylpropanoid biosynthesis. Different number of copies of PAL is found across species e.g. four in *Arabidopsis*, five in poplar etc. These different gene copies respond differently to various stresses that the plant encounter

[100]. In poplar, out of the five copies, three shows organ specific expression. The first three steps catalyzed by PAL, coumarate-4-hydroxylase (C4H) and 4-coumarate ligase (4-CL) is required for the pathway, as it dedicated phenylalanine towards the production of monolignols and other phenylpropanoids.

In further steps, 4CL mediates the conversion of coumarate to coumaroyl-CoA. This step represents a very important point of the pathways as it serves as a direct precursor for flavonol or H-liginin, or is involved in the production of G- and S-monolignols. These conversions are brought about by the subsequent activity of coumaroyl-CoA/feruloyl-CoA reductase (CCR) and coumaryl/coniferyl aldehyde dehydrogenase (CAD). CCR, in general, does not exhibit preference for anyone CoA substrates, however in several studies, feruloyl CoA is reported to be the substrate of choice [101]. On the other hand, CAD, because of its role at towards the end of pathway, is often considered to be the indicator of lignin biosynthesis. However, it is also found to be the expressed under stress conditions [101]. The substrate selectivity of CAD differs between angiosperms and gymnosperms, and that governs the type of lignin that is formed. The monolignols that are formed may continue in the lignin biosynthesis pathway to form H- (coumaryl alcohol), S- (Sinapyl alcohol) or G-(coniferyl alcohol) lignin. These alcohols can also be diverted away from the pathway by alcohol acyl transferases. These transferases belong to the BADH family of acyl transferases and are reported to catalyze the acylation of several plants secondary metabolites [102]. Acetyl group from acetyl CoA is transferred to alcohols to form ester. These esters are then acted upon by NADPH dependent reductase, eugenol synthase (EGS) to form eugenol and chavicol. There also exists enzyme in the pathway that utilizes the same substrates and forms the isomers of the products, isoeugenol and anol.

Phenylpropanoids, owing to their impact on various processes in plants, is regulated at multiple levels. The spatial and temporal regulation of pathway genes by endogenous (metabolic and hormonal) and environmental (light and pathogen attack) stimuli brings about different accumulation patterns. Also, transcription factors like R2R3-MYB (EOBII, PAP1 etc.), TCP and bHLH direct the tissue and plant specific expression of pathway genes [103,104,105].

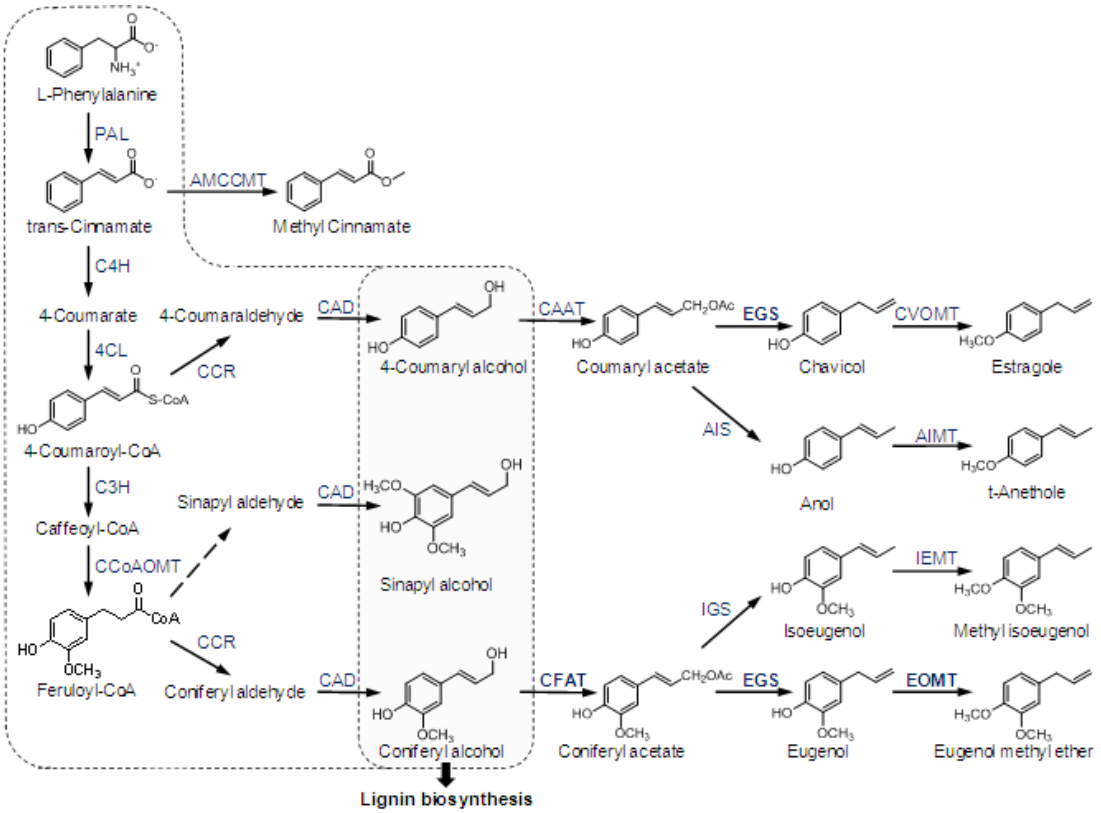


Figure 8: General scheme of the phenylpropanoid biosynthesis

1.3 *Ocimum*

The genus *Ocimum* belongs to the family *Lamiaceae*, with over 30 species being reported till date. [106]. This genus is famous for the diversity of secondary metabolites including terpenes and phenylpropanoid synthesized and stored in different plant parts. These plants are typically found in the semitropical and tropical regions of Africa, South America and India [106]. They are usually grown near temples and houses because of its air freshener properties. Different *Ocimum* species can be identified by the exclusive presence of particular metabolites, that are either not present in other species, or are observed in trace amounts. The diversity of metabolites produced by *Ocimum* plants is indeed enormous, though the purpose of this metabolite diversity is not properly understood. Different parts of *Ocimum* plants are traditionally used in treatment of skin diseases, hepatic disorders, for releasing stress, as an antidote for snake bite, and scorpion sting. The antioxidant properties of the alcoholic extracts responsible for wound healing have been studied using albino rat as model system. The other properties that are already established include anti-

inflammatory, analgesic, antipyretic, radiation protection, anti microbial properties [106].

1.3.1 Classification

Kingdom:	Plantae
Phylum	Steptophyta
Class	Eudicots
Order:	Lamiales
Family:	*Lamiaceae*
Genus:	*Ocimum*

India has a vast and inexhaustible resource of medicinal and aromatic plants. Over 45,000 plant species are known in India, of which about six thousand are known to possess medicinal and organoleptic properties and very few plants have been explored for the natural products which possess medicinal properties [107].

1.4 Scope of thesis

There were reports available at the start of this work that provided information about the metabolite diversity of different *Ocimum* species. These studies took into account the alcoholic extracts of the complete plant and analysis was done. However, very limited information existed about the accumulation of different metabolites in different plant parts across the species. The primary aim of this work was to analyze this diversity in different species of *Ocimum.* We did volatile analysis of six different plant parts and compared the profile with other species to gain further insights. The next step was to understand how this change is brought about, so we did characterization of genes involved in two major secondary metabolite pathway i.e phenylpropanoid and terpene biosynthesis. These studies led us to a better understanding of the chemical diversity and how its brought about in the species that were studied.

The thesis is divided into four chapters

Chapter 1: Introduction

Chapter 2: Comparative functional characterization of *EGS* from *Ocimum* species

Chapter 3: Metabolic diversity and characterization of synthase involved in monoterpene accumulation

Chapter 4: Functional characterization of β-caryophyllene synthase – a sesquiterpene synthase

1.5 References

[1] O'Connor SE, Engineering of secondary metabolism, Annu. Rev. Genet. 49 (2015) 5.1-5.24.

[2] Verpoorte R and Memelink J, Engineering of secondary metabolite production in plants, Curr. Opin. Biotech. 13 (2002) 181-187.

[3] Harborne JB, Twenty five years of chemical ecology, Nat. Prod. Rep. 18 (2001) 361-379.

[4] Dixon RA, Plant natural products: the molecular genetic basis of biosynthetic diversity, Curr. Opin. Biotech. 10 (1999) 192-197.

[5] Verpoorte R, Heijden R van der and Memelink J, Engineering the plant cell factories for secondary metabolite production, Trans. Res. 9 (2000) 323-343.

[6] Dudareva N, Klempien A, Muhlemann JK, et. al., Biosynthesis, function and metabolic engineering of plant organic volatile compounds, New Phytol. 198 (2013) 16-32.

[7] Croteau R, Kutchan TM and Lewis NG, In Natural products, (2000) pg. 1250-1318

[8] McGarvey DJ and Croteau R, Terpenoid metabolism, Plant Cell 7 (1995) 1015-26.

[9] Bach TJ, Some new aspects of isoprenoid biosynthesis in plants—A review, Lipids. 30 (1995) 191-202.

[10] Rohmer M, Knani M, Simonin P, et. al., Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate, Biochem. J 295 (1993) 517-524.

[11] Sprenger GA, Schorken U, Wiegert T, et. al., Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol, Proc. Natl. Acad. Sci. USA 94 (1997) 12857-12862.

[12] Eisenreich W, Schwarz M, Cartayrade A, et. al., The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms, Chem. Biol. 5 (1998) R221-R233.

[13] Burke C and Croteau R, Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate, J. Biol Chem 277 (2002) 3141-3149.

[14] Thiel R and Adam KP, An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants, Phytochem. 59 (2002) 269-274.

[15] Cane DE, Isoprenoid biosynthesis: Stereochemistry of the cyclization of allylic pyrophosphates, Accounts Chem. Res. 18 (1985) 220-226.

[16] Croteau R, Biosynthesis and catabolism of monoterpenoids, Chem. Rev. 87 (1987) 929-954.

[17] Cane DE, Enzymic formation of sesquiterpenes, Chem. Rev. 90 (1990) 1089-1103.

[18] Lesburg CA, Caruthers JM, Paschall CM, et. al., Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism, Curr. Opin. Struc. Biol. 8 (1998) 695-703.

[19] Davis EM and Croteau R, Cyclization Enzymes in the biosynthesis of monoterpenes, sesquiterpenes and diterpenes, Springer, Vol. 209 (2000) p 53-95.

[20] Poulter CD, Farnesyl diphosphate synthase: A paradigm for understanding structure and function relationships in E-polyprenyl diphosphate synthases, Phytochem. Rev. 5 (2006)17-26.

[21] Thulasiram HV and Poulter CD, Farnesyl Diphosphate Synthase: The art of compromise between substrate selectivity and stereo-selectivity, J. Am. Chem. Soc. 128 (2006)15819-15823.

[22] Tantillo DJ, Biosynthesis via carbocations: Theoretical studies on terpene formation, Nat. Prod. Rep. 28 (2011)1035-1053.

[23] Xie Z, Kapteyn J, Gang DR, et al., A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes, Plant J. 54 (2008) 349-361.

[24] Singh P, Kalunke RM, Giri AP, et al., Towards comprehension of complex chemical evolution and diversification of terpenoids and phenylpropanoid pathways in Ocimum species, RSC Adv, 5 (2015) 106886-106904.

[25] Zvi MMB, Shklarman E, Masci T, et al., PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers, New Phytol. 195 (2012) 335-345.

[26] Liu J, Osbourne A and Ma P, MYB transcription factors as regulators of phenylpropanoid metabolism in plants, Mol. Plant 8 (2015) 689-708.

[27] Wink M, In Genetics of Plants, 4 (1987) 17-42.

[28] Facchini PJ, Alkaloid biosynthesis in plants-Biochemistry, cell biology, molecular regulation and metabolic engineering, Ann. Rev. Plant Physiol. Plant Mol. Bio. 52 (2001) 29-66.

[29] Kutchan TM, Alkaloid biosynthesis-The basis for metabolic engineering of metabolic plants, The Plant Cell 7 (1995) 1059-1070.

[30] Osbourne A, Goss RJM and Field RA, The Saponins-Polar isoprenoids with important and diverse biological activities, Nat. Prod. Rep. 28 (2011) 1261-1268.

[31] Ishida M, Hora M, Fukino N, et. al., Glucosinolate metabolism, function and breeding for the improvement of *Brassicaceae* vegetables, Breeding Science 64 (2014) 48-59.

[32] Halkier BA and Gershenzon J, Biology and biochemistry of glucosinolates, Annu. Rev. Plant Biol. 57 (2006) 303-33.

[33] Thulasiram HV, Erickson HK, Poulter CD, Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis, Science 316 (2007) 73-76.

[34] Reiling KK, Yoshikuni Y, Martin VJJ, et. al., Mono and diterpene production in *E. coli*., Biotechnol. Bioeng. 87 (2004) 200-212.

[35] Ding VDH, Sheares BT, Bergstrom JD, et. al., Purification and characterization of recombinant human farnesyl diphosphate synthase expressed in *E. coli*, Biochem. J. 275 (1991) 61-65.

[36] Anderson MS, Yarger JG, Burck CL, et. al., Farnesyl diphosphate synthetase. Molecular cloning, sequence, and expression of an essential gene from *Saccharomyces cerevisiae*, J. Biol. Chem. 264 (1989) 19176-19184.

[37] Beyer P, Mayer M, Kleinig H, Molecular oxygen and the state of geometric isomerism of intermediates are essential in the carotene desaturation and cyclization reactions in daffodil chromoplasts, Eur. J. Biochem. 184 (1989) 141-150.

[38] Krinsky N, Mathews-Roth M, Taylor R, et. al., In Carotenoids; Springer US (1989) p 167-184.

[39] Momose K and Rudney HJ, et. al., 3-Polyprenyl-4-hydroxybenzoate synthesis in the inner membrane of mitochondria from p-hydroxybenzoate and isopentenylpyrophosphate. A demonstration of isoprenoid synthesis in rat liver mitochondria, J. Biol. Chem. 247 (1972) 3930-3940.

[40] Trumpower BL, Houser RM and Olson REJ, Studies on ubiquinone. Demonstration of the total biosynthesis of ubiquinone-9 in rat liver mitochondria, J. Biol. Chem. 249 (1974) 3041-3048.

[41] Rock CD and Zeevaart JA, The aba mutant of Arabidopsis thaliana is impaired in epoxy-carotenoid biosynthesis, P. Natl. Acad. Sci. USA 88 (1991) 7496-7499.

[42] Parry AD and Horgan R, Carotenoid metabolism and the biosynthesis of abscisic acid, Phytochem. 30 (1991) 815-821.

[43] Lange BM, Rujan T, Martin W, et. al., soprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes, P. Natl. Acad. Sci. USA 97 (2000) 13172-13177.

[44] Jennewein S and Croteau R, Taxol: biosynthesis, molecular genetics, and biotechnological applications, Appl. Microbiol. Biot. 57 (2001) 13-19.

[45] Rodriguez-Concepion M, The MEP pathway: a new target for the development of herbicides, antibiotics and antimalarial drugs, Curr. Pharm. Design 10 (2004) 2391-2400.

[46] McGarvey DJ and Croteau R, Terpenoid metabolism, Plant Cell 7 (1995) 1015-26.

[47] Bach TJ, Some new aspects of isoprenoid biosynthesis in plants--a review, Lipids 30 (1995) 191-202.

[48] Wouters J, Oudjama Y, Barkley SJ, 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, J. Biol. Chem. 278 (2003) 11903-11908.

[49] Kaneda K, Kuzuyama T, Takagi M, et. al., An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190, P. Natl. Acad.Sci. USA 98 (2001) 932-937.

[50] Hemmi H, Ikeda Y, Yamashita S, et. al., 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, Biochem. Bioph. Res. Comm. 322 (2004) 905-910.

[51] Rohmer M, Knani M, Simonin P, et. al., Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate, Biochem. J. 295 (1993) 517-524.

[52] Sprenger GA, Schorken U, Wiegert T, et. al., Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol, P. Natl. Acad. Sci. USA 94 (1997) 12857-12862.

[53] Eisenreich W, Schwarz M, Cartayrade A, et. al., The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms, Chem. Biol. 5 (1998) R221-R233.

[54] Zeidler J, Schwender J, Mueller C, et. al., Properties and inhibition of the first two enzymes of the non-mevalonate pathway of isoprenoid biosynthesis, Biochem. Soc. Trans. 28 (2000) 796-798.

[55] Rodriguez-Concepcion M, Fores O, Martinez-Garcia JF, et. al., Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during *Arabidopsis* seedling development, Plant Cell 16 (2004) 144-156.

[56] Rohdich F, Bacher A and Eisenreich W, Isoprenoid biosynthetic pathways as anti-infective drug targets, Biochem. Soc. Trans. 33 (2005) 785-791.

[57] Burke CC, Wildung MR and Croteau R, Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer, P. Natl. Acad. Sci. USA 96 (1999) 13062-13067.

[58] Burke C and Croteau R, Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate, J. Biol. Chem. (2002) 277 3141-3149.

[59] Delourme D, Lacroute FO and Karst F, Cloning of an *Arabidopsis thaliana* cDNA coding for farnesyl diphosphate synthase by functional complementation in yeast, Plant Mol. Biol. 26 (1994) 1867-1873.

[60] Okada K, Saito T, Nakagawa T, et. al., Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three sub-cellular compartments in *Arabidopsis*, Plant Physiol. 122 (2000) 1045-1056.

[61] Thiel R and Adam KP, An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants, Phytochem. 59 (2002) 269-274.

[62] Itoh D, Karunagoda RP, Fushie T, et. al., Nonequivalent labeling of the phytyl side chain of chlorophyll a in callus of the hornwort *Anthoceros punctatus*, J. Nat. Prod. 63 (2000) 1090-1093.

[63] Itoh D, Kawano K and Nabeta K, Biosynthesis of chloroplastidic and extrachloroplastidic terpenoids in liverwort cultured cells: ^{13}C serine as a probe of terpene biosynthesis via mevalonate and non-mevalonate pathways, J. Nat. Prod. 66 (2003) 332-336.

[64] Nabeta K and Komuro K, Biosynthesis of phytyl side-chain of chlorophyll a: apparent reutilization of carbon dioxide evolved during acetate assimilation in biosynthesis of chloroplastidic isoprenoid, Chem. Comm. (1998) 671-672.

[65] Lichtenthaler HK, The 1-Deoxy-D-Xylulose-5-Phosphate pathway of isoprenoid biosynthesis in plants, Annu. Rev. Plant Phys. 50 (1999) 47-65.

[66] Rohmer M, The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, Nat. Prod. Rep. 16 (1999) 565-574.

[67] Eisenreich W, Rohdich F and Bacher A, Deoxyxylulose phosphate pathway to terpenoids, Trends Plant Sci. 6 (2001) 78-84.

[68] Burg JS and Espenshade PJ, Regulation of HMG-CoA reductase in mammals and yeast, Prog. Lipid Res. 50 (2011) 403-410.

[69] Bach TJ, Rohmer M, Rodriguez-Concepcion, M, et. al., In Isoprenoid Synthesis in Plants and Microorganisms, Springer New York (2013) 1-16.

[70] Schmid M, Davison TS, Henz SR, et. al., A gene expression map of *Arabidopsis thaliana* development, Nat. Genet. 37 (2005) 501-506.

[71] Ghassemian M, Lutes J, Tepperman JM, et. al., Integrative analysis of transcript and metabolite profiling data sets to evaluate the regulation of biochemical pathways during photomorphogenesis, Arch. Biochem. Biophys. 448 (2006) 45-59.

[72] Basson ME, Thorsness M and Rine J, *Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase, P. Natl. Acad. Sci. USA 83 (1986) 5563-5567.

[73] Goldstein JL and Brown MS, Regulation of the mevalonate pathway, Nature 343 (1990) 425-430.

[74] Monger DJ, Lim WA, Kozdy FJ, et. al., Compactin inhibits insect HMG-CoA reductase and juvenile hormone biosynthesis, Biochem. Bioph. Res. Co. 105 (1982) 1374-1380.

[75] Chappell J, Wolf F, Proulx J, et. al., Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants?, Plant Physiol. 109 (1995) 1337-1343.

[76] Bach TJ, Synthesis and metabolism of mevalonic acid in plants, Plant Physiol. Biochem. 25 (1987) 163-178.

[77] Sapir-Mir M, Mett A, Belausov E, et. al., Peroxisomal localization of *Arabidopsis* isopentenyl diphosphate isomerases suggests that part of the plant isoprenoid mevalonic acid pathway is compartmentalized to peroxisomes, Plant Physiol. 148 (2008) 1219-1228.

[78] Simkin A, Guirimand G, Papon N, et. al., Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta, Planta 234 (2011) 903-914.

[79] Guevara-Garcia A, San Roman C, Arroyo A, et. al., Characterization of the Arabidopsis clb6 mutant illustrates the importance of posttranscriptional regulation of the methyl-D-erythritol 4-phosphate pathway, Plant Cell 17 (2005) 628-643.

[80] Wolfertz M, Sharkey TD, Boland W, et. al., Rapid regulation of the methylerythritol 4-phosphate pathway during isoprene synthesis, Plant Physiol. 135 (2004) 1939-1945.

[81] Estevez JM, Cantero A, Reindl A, et. al., 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants, J. Biol. Chem. 276 (2001) 22901-22909.

[82] Lois LM, Campos N, Putra SR, et. al., Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis, P. Natl. Acad. Sci. USA 95 (1998) 2105-2110.

[83] Enfissi EMA, Fraser PD, Lois LM, et. al., Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato, Plant Biotechnol. J. 3 (2005) 17-27.

[84] Morris WL, Ducreux LJM, Hedden P, et. al., Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life cycle, J. Exp. Bot. 57 (2006) 3007-3018.

[85] Gong YF, Liao ZH, Guo BH, et. al., Molecular cloning and expression profile analysis of Ginkgo biloba DXS gene encoding 1-deoxy-D-xylulose 5-phosphate synthase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway, Planta Med. 72 (2006) 329-335.

[86] Sharkey TD and Yeh S, Isoprene emission from plants, Annu. Rev. Plant Phys. 52 (2001) 407-436.

[87] Silver GM and Fall R, Enzymatic synthesis of isoprene from dimethylallyl diphosphate in aspen leaf extracts, Plant Physiol. 97 (1991) 1588-1591.

[88] Kuzma J and Fall R, Leaf Isoprene Emission Rate Is Dependent on Leaf Development and the Level of Isoprene Synthase, Plant Physiol. 101 (1993) 435-440.

[89] Garcia R, Alves ESS, Santos MP, et. al., Antimicrobial activity and potential use of monoterpenes as tropical fruits preservatives, Braz. J. Microbiol. 39 (2008) 163-168.

[90] Singh P, Shukla R, Prakash B, et. al., Chemical profile, antifungal, antiaflatoxigenic and antioxidant activity of *Citrus maxima* Burm. and *Citrus sinensis* (L.) Osbeck essential oils and their cyclic monoterpene, DL-limonene, Food Chem. Toxicol. 48 (2010) 1734-1740.

[91] Karkabounas S, Kostoula OK, Daskalou T, et. al., Anticarcinogenic and antiplatelet effects of carvacrol, Exp. Oncol. 28 (2006) 121-5.

[92] Hampel D, Mosandl A and Wust M, Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways, Phytochem. 66 (2005) 305-311.

[93] Yu F and Utsumi R, Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis, Cell. Mol. Life Sci. 66 (2009) 3043-3052.

[94] Bommareddy A, Rule, B, VanWert AL, et. al., α-Santalol, a derivative of sandalwood oil, induces apoptosis in human prostate cancer cells by causing caspase-3 activation, Phytomed. 19 (2012) 804-811.

[95] Klayman DL, Qinghaosu (artemisinin): an anti-malarial drug from China, Science 228 (1985) 1049-1055.

[96] Tippmann S, Chen Y, Siewers V, et. al., From flavors and pharmaceuticals to advanced biofuels: production of isoprenoids in *Saccharomyces cerevisiae*, Biotechnol. J. 8 (2013) 1435-1444.

[97] Ohvo-Rekila H, Ramstedt B, Leppimaki P, et. al., Cholesterol interactions with phospholipids in membranes, Prog. Lipid Res. 41 (2002) 66-97.

[98] Sakurai A, Brassinosteroid biosynthesis, Plant Physiol. Biochem. 37 (1999) 351-361.

[99] Bishop GJ and Yokota T, Plants steroid hormones, brassinosteroids: current highlights of molecular aspects on their synthesis/metabolism, transport, perception and response, Plant Cell Physiol. 42 (2001) 114-120.

[100] Hemberger B, Ellis M, Friedmann M, et. al., Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the populus lignin toolbox and conservation and diversification of angiosperm gene families, Can. J. Bot. 85 (2007) 1182-1201.

[101] Whetter R and Sederoff R, Lignin biosynthesis, The Plant Cell 7 (1995) 1001-1013.

[102] Dexter R, Qualley A, Kish CM, et. al., Characterization of a petunia acyl transferase involved in the biosynthesis of floral volatile isoeugenol, Plant J. 49 (2007) 265-275.

[103] Spitzer-Rimon B, Marhevka E, Barakai O, et. al., EOBII- a gene encoding a flower specific regulator of phenylpropanoid volatile biosynthesis in petunia, The Plant Cell 22 (2010) 1961-1976.

[104] Zvi MMB, Shklarman E, Masci T, et. al., PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers, New Phytol. 195 (2012) 335-345.

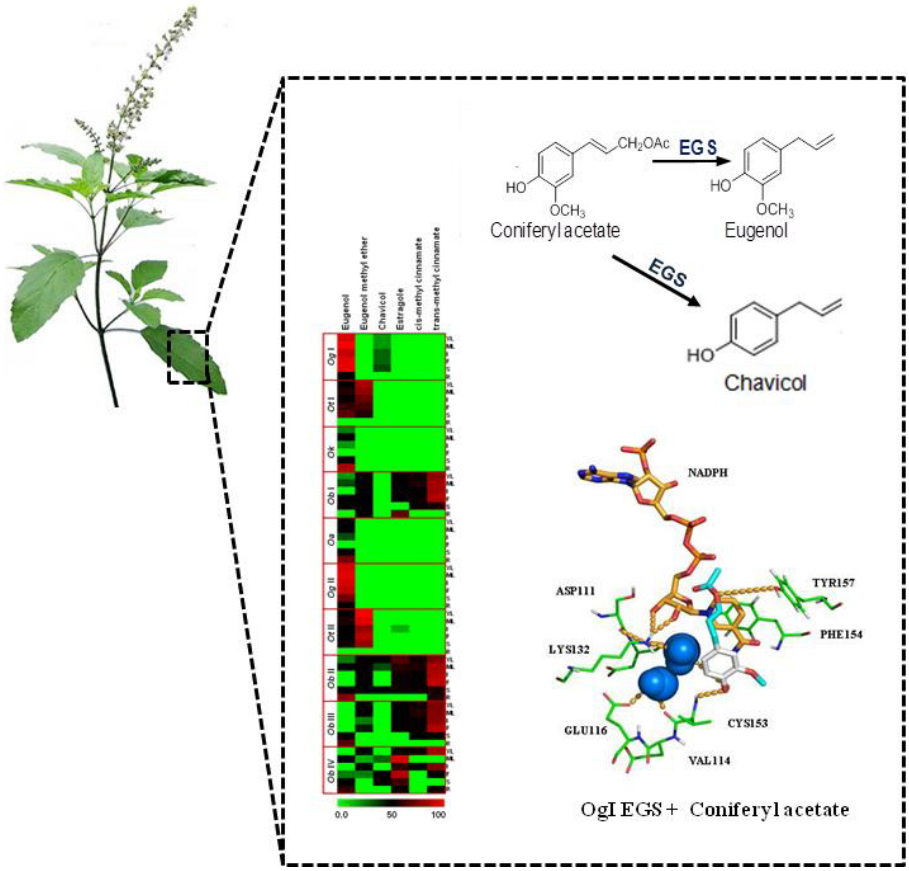
[105] Liu J, Osbourn A and Ma P, MYB transcription factors as regulators of phenylpropanoid metabolism in plants, Mol. Plant 8 (2015) 689-708.

[106] Singh P, Kalunke RM, Giri AP, et al., Towards comprehension of complex chemical evolution and diversification of terpenoids and phenylpropanoid pathways in *Ocimum* species, RSC *Adv*, 5 (2015) 106886-106904.

[107] Grover JK, Yadav S and Vats VJ, Medicinal plants of India with anti-diabetic potential, Ethnopharmacol. 81 (2002) 81-100.

Chapter 2

Comparative Functional Characterization of *EGS* from *Ocimum* Species



Chapter 2

Comparative Functional Characterization of *EGS* from *Ocimum* Species

Isoprenoids and phenylpropanoids are the major secondary metabolite constituents in *Ocimum* genus. Though enzymes from phenylpropanoid pathway have been characterized from few plants, limited information exists on how they modulate levels of secondary metabolites. Phenylpropanoid profiling in different tissues from five *Ocimum* species was performed which revealed significant variations in secondary metabolites including eugenol, eugenol methyl ether, estragole and methyl cinnamate levels. Expression analysis of eugenol synthase (*EGS*) gene showed higher transcript levels especially in young leaves and inflorescence; and were positively correlated with eugenol contents. Additionally, transcript levels of coniferyl alcohol acyl transferase, a key enzyme diverting pool of substrate to phenylpropanoids, were in accordance with their abundance in respective species. In particular, eugenol methyl transferase expression positively correlated with higher levels of eugenol methyl ether in *Ocimum tenuiflorum*. Further, EGSs were functionally characterized from four *Ocimum* species varying in their eugenol contents. Kinetic and expression analyses indicated, higher enzyme turnover and transcripts levels, in species accumulating more eugenol. Moreover, biochemical and bioinformatics studies demonstrated that coniferyl acetate was the preferred substrate over coumaryl acetate when used, individually or together in the enzyme assay.

2.1 Introduction

Phenylpropanoids, a group of plant secondary metabolites derived from phenylalanine, typically have propenyl side chains attached to a phenyl ring. These compounds are auto-toxic in nature and perform a vital role in plant communication, pollinator attraction and defense against herbivores and other pathogens [1-3]. Coniferyl and coumaryl alcohols are monolignol alcohol intermediates of the lignin biosynthetic pathway and serve as the precursors for phenylpropanoid biosynthesis. Biosynthesis of coumaryl alcohol involves coumaroyl-CoA reductase that acts on 4-coumaroyl-CoA to form 4-coumaraldehyde and provides a substrate for the synthesis of chavicol (Fig. 1). These alcohols are first converted to acetates by coniferyl- and coumaryl alcohol acyl transferases. The reductive elimination of the acetate moiety, which is catalyzed by eugenol synthase (EGS), yields the propenyl side chain. EGS utilizes coniferyl acetate and coumaryl acetate as substrates to produce eugenol and chavicol, respectively [4,5]. Methylation of eugenol and chavicol at the para-OH group of the phenyl ring is catalyzed by eugenol O-methyl transferase (EOMT) and chavicol O-methyl transferase (COMT) to form their methyl ether derivatives.

In this chapter, we have focused on EGS because eugenol has interesting properties like effector molecule in diabetes control [6] and is an important scent compound [7]. Till date, EGS has been characterized from few plant species such as *Ocimum basilicum* (ObEGS1) [4], *Gymnadenia odoratissima* [7], petunia (PhEGS1), *Clarkia breweri* (CbEGS1 and CbEGS2) [8] and *Pimpinella anisum* t-anol/iso-eugenol synthase (PaAIS1) [9]. These enzymes belong to the pinoresinol iso-flavone phenylcoumaran (PIP) family and perform NADPH-dependent reductases activity for the conversion of coniferyl acetate to eugenol [10]. EGS from *Larrea tridentate* (LtCES1) has been also shown to accept both coniferyl acetate and coumaryl acetate [5]. Although these reports describe about the characterization of EGS, very limited knowledge exists about the molecular basis for substrate selection and its effect on the modulation of secondary metabolites by EGS. Therefore, phenylpropanoids profiling was performed in different tissues from five *Ocimum* species and further functionally characterized EGSs from them. The kinetic parameters for these EGSs with coniferyl acetate indicated high turnover number of EGS in species having high eugenol content. Additionally, this was supported by EGS expression analysis indicating significant correlation between the EGS expression and eugenol levels. Though EGS

could utilize two substrates, in vitro and in silico studies confirmed that eugenol was the predominantly detected phenylpropanoid while chavicol was present in trace amounts in assays when both the substrates used together.

2.2 Materials and Methods

2.2.1 Plant material

Authenticated *Ocimum* plants belonging to five species, *Ocimum gratissimum* L. I and II (*Og*I and *Og*II), *O. tenuiflorum* I and II (*Ot*I and *Ot*II), *O. kilimandscharicum* Gürke (*Ok*), *O. americanum* (*Oa*) and *O. basilicum* I, II, III and IV (*Oba*I, *Oba*II, *Oba*III and *Oba*IV) were grown in a greenhouse at 25-28 ºC with ~35-40% humidity and 16 h light and 8 h dark periods. Plant samples of all the species were deposited in the herbarium of Botanical Survey of India, Pune as voucher specimens. All the samples were collected from 2-3 months old plants, immediately frozen in liquid nitrogen and stored at -80 ºC until further use. Fresh tissues were used for metabolite extraction and analysis.

2.2.2 Reagents

All chemicals were purchased from Sigma-Aldrich (Sigma Chemical Co., USA), unless stated. Coniferyl acetate and coumaryl acetate were synthesized from coniferyl alcohol and coumaryl alcohol, respectively, as previously described [2]. The TA cloning kit with pCR 2.1 vector, Generacer kit, Zero Blunt Vector, pET102 directional cloning kit, PCR purification kit, gel extraction kit and Accuprime proof reading polymerase were purchased from Thermo Scientific (USA).

Synthesis of Coumaryl alcohol

para- coumaric acid (247 mg) was put in a flask into which 5 ml methanol & 1 drop of H_2SO_4 were added and refluxed at 65^0C for 6 hours. The crude reaction product was subjected to column chromatography on 60-120 mesh silica gel, and compound was eluted with 300ml of 3% MeOH in DCM. Fractions were collected in 50ml test tubes. Pure fractions were collected and concentrated to get 268 mg of α, β-unsaturated ester as a white solid (100 % yield). α, β-unsaturated ester (236 mg), was kept in an inert atmosphere and 9 ml of anhydrous tetra hydro furan (THF) was added to the reaction mixture that had been cooled to -78 °C using dry ice and acetone as coolants. 5.30 mL DIBAL-H in toluene was added drop-wise. The reaction was left in the same coolant bath without maintaining the temperature for 4 hours, after which the reaction was quenched with 8ml saturated solution of sodium potassium tartarate and stirred for 15 min. THF was then removed under reduced pressure, and the reaction mixture was extracted with DCM (2×25 ml). This DCM layer was

concentrated to obtain 190 mg crude product which was then subjected to column chromatography on 60-120 mesh silica gel and eluted with 500 ml of 4% MeOH in DCM. Fractions were collected in 50 ml test tubes and concentrated to get 123 mg of coumaric alcohol as a white solid (61.85 % yield).

Coumaric ester

^{1}H NMR (CD_3OD, 200 MHz): δ 7.62(d, *J*=15.92 Hz, 1H), 7.45(d, *J*=8.21 Hz, 2H), 6.81 (d, *J*=8.59 Hz, 2H), 6.32 (d, *J*=15.92 Hz, 1H,), 3.76 (s, 3H)

^{13}C NMR (CD_3OD, 50 MHz): δ 52.12, 115.02, 116.93, 127.24, 131.27, 146.65, 161.33, 169.86.

Coumaryl alcohol

^{1}H NMR (CD_3OD, 200 MHz): δ 7.20 (d, *J*=8.59 Hz, 2H), 6.69 (d, *J*=8.46 Hz, 2H), 6.46 (d, *J*=15.79 Hz, 1H), 6.12 (m, 1H), 4.14 (d, *J*=5.68 Hz, 2 H)

^{13}C NMR (CD_3OD, 50 MHz): δ 64.09, 116.46, 126.79, 128.83, 130.13, 132.03, 158.32.

Synthesis of Chavicol

Estragole (180.0 mg, 1.34 mmol) was weighed in round bottom flask and flushed with nitrogen to make inert environment; dry dichloromethane (5 ml) was added in it. Reaction mixture was kept at -78 °C for 20 min. Boron tribromide (402.8 mg, 1.6 mmol) was added slowly into the reaction mixture and kept at same temperature for

30 min. Temperature bath was then removed and reaction kept for 15 min at room temperature till green colour appeared. Reaction was quenched by saline solution till colour disappeared. Reaction mixture was extracted two times with DCM. Organic layer dried over sodium sulphate. After evaporation, purification was carried out by column chromatography on silica gel with (Pet ether/ EtOAc) gradient mixture.

Chavicol

^{1}H NMR (200 MHz, $CDCl_3$, ppm): δ 7.03-7.08 (m, 2H), 6.75-6.79 (m, 2H), 5.88-6.01 (m, 1H), 5.06-5.11 (m, 1H), 4.99-5.02 (m, 1H), 3.30-3.33 (d, *J* = 6.57, 2H).
^{13}C NMR (500 MHz, $CDCl_3$, ppm): δ 153.77, 137.81, 132.25, 130.41, 129.68, 115.44, 115.21, 39.30.

2.2.3 Analysis of phenylpropanoids from different *Ocimum* species

Phenylpropanoid abundance was measured in six different tissues including young leaves (YL; top whorl), mature leaves (ML; third whorl), inflorescence (I), flower (F), stem (S) and root (R) from five Ocimum species. Extractions were performed by dichloromethane (DCM) extraction method, as reported previously [11]. Briefly, fresh tissues (5 g each) from different plant parts were harvested separately and immediately soaked in 50 mL DCM for 20 h at 28ºC. The combined organic phase was cooled to -20°C for 2 h (for lipid precipitation) and filtered. The contents were dried, weighed, re-dissolved in 2 mL DCM and subjected to GC and GC-MS analysis was carried out for 3 biological replicates in triplicates. GC analyses were carried out on an Agilent 7890A instrument equipped with a hydrogen flame ionization detector and HP-5 capillary column (30 m X 0.32 mm X 0.25 µm, J and W Scientific, USA). Nitrogen was used as the carrier gas at a flow rate of 1 mL/min. The column temperature was raised from 70°C to 110°C at 2°C min-1, then raised to 180°C at 3°C min-1 and finally at 10°C min-1 raised to 220°C and held for 2 min. Injector and detector temperatures were 230°C and 235°C, respectively. GC-MS was performed on an Agilent 5975C mass selective detector interfaced with an Agilent 7890A gas chromatograph using an HP-5 MS capillary column with helium as the carrier gas and using above mentioned conditions. Compounds were identified by co-injection studies, comparing the retention time and mass fragmentation pattern with those of

reference compounds and also compared acquired mass spectra and retention indices with those of NIST/NBS and the Wiley mass spectral library (software version 2.0, Dec. 2005).

2.2.4 RNA isolation, semi-quantitative and quantitative RT-PCR

Total RNA was extracted from the different tissues using Spectrum™ Plant Total RNA Isolation Kit (Sigma Chemical Co.). Total RNA was quantified using a spectrophotometer (NanoDrop, Thermo Scientific) by measuring optical density of isolated RNA in 10 mM TE buffer at a wavelengths of 230, 260, and 280 nm and purity was checked by comparing the ratio of 230/260 and 260/280. The integrity of total RNA was assessed by sharpness of rRNA (28S and 18S rRNA) on 1.5 % agarose gels and visualized by GelRed™ (Biotium). The DNase-treated total RNA (4 µg) was used for cDNA preparation in a 20 µL reaction using SuperScript™ III reverse-transcriptase system (Thermo Scientific). Semi-quantitative RT-PCR (sqRT-PCR) was performed for EGS, eugenol o-methyl transferase (EOMT) and coniferyl alcohol acyl transferase (CFAT) in 10 µL reaction consisting of 5 µL of 2X Jumpstart readymix, 1 µL each of 10 µM gene specific forward and reverse primers, 1 µL of cDNA optimized with endogenous control (18S rRNA) and nuclease free water was added to volume of 10 µL. PCR was carried out for 30 cycles and the amplified products were run on 1% agarose gel. Quantitative RT-PCR (qRT-PCR) was performed by using Taqman chemistry. A typical reaction consisted of 5 µL of Taqman advanced master-mix, 0.5 µL of primer-probe mix and 1 µL of diluted cDNA (1:2) with nuclease-free water added to make up a volume of 10 µL. For qRT-PCR reactions, actin was used as an endogenous control and the reactions were carried out in triplicates for 3 biological replicates in the 7500 Fast Real Time PCR System (Thermo Scientific). Annealing temperature was kept at 60°C and cycling conditions were kept as per the manufacture's instruction. Primer sequences, and assay identification number for Taqman primer –probes were as given in Table 1.

Table 1: List of primer and probe sequences

Primer Code	Primer sequence/ Assay IDs
EGS FF	5'-ATGGAGGAAAAAGGGTCGAAAAGC-3'
EGS FR	5'-TTATGCTGCTGAAGCAGGCGC-3'
3'RACE EGS	5'-GGAGGCCATCAAGGTTGCTGG-3'
5'RACE EGS	5'-TGGTCTGTAGATCACCACACGATTCAATG-3'
18S rRNA F	5'-TCGAAACCTGCAAAGCAGACC-3'
18S rRNA R	5'-GATTCTGCAATTCACACCAAGTATCG-3'
RT- Actin	Assay ID: AIQJB2E
RT- EGS	Assay ID: AIHSOOT
RT-EOMT	Assay ID: AIGJQIL
RT-CFAT-F	5'-TTGGACAATGTTGCTAACGA-3'
RT-CFAT-R	5'-TAGACCAGAGGAGACAACG-3'
EGS-F	5'-GTGCCATCATAGTCAAGGGAGAGT-3'
EGS-R	5'-TCCGGCAGTTCTTTTGTGAGG-3'

2.2.5 Isolation of *EGS* from ten *Ocimum* subtypes and sequence analysis

*Og*I young leaf cDNA was used for PCR with EGS-F and EGS-R primers (Table 1) to generate a 750 bp amplicon, which was cloned in pCR 2.1 vector and positive clones were sequenced. The sequence upon BLAST analysis indicated it to be eugenol synthase. 5' and 3' RACE PCR was performed by using primers (Table 1) designed from the reported ObEGS1 nucleotide sequence (Accession no. DQ372812) and the partial fragment obtained from *Og*I. RACE cDNA was prepared from 5 µg of total RNA from the leaf tissue using the Generacer kit (Thermo Scientific). RACE PCR

products were cloned in the pCR 2.1 vector and transformed into TOP10 chemically competent cells. Sequence verified RACE fragments were used to generate open reading frame for EGS. EGSs from five species were isolated by using terminal primers and similarly cloned, sequenced and their sequences have been submitted to NCBI (GenBank accession no. KU977432- KU977441). For cloning EGSs into pET102 vector for heterologus expression in *E. coli*, 5'-CACC-3' was added to the 5' end of the forward primer, and in reverse primer stop codon was removed to allow for C-terminal 6X his-tag expression. The sequence analyses were carried out on Bioedit software (Ibis Biosciences, USA); alignment was done using the CLUSTALW2 program (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Nucleotide sequences were translated using the ExPASy translate tool (http://web.expasy.org/translate/) and BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html) was used for marking identical and similar amino acid residues. Multiple alignments of EGS sequences were done using CLUSTALW2, and the neighbor-joining tree was constructed (with 1000 bootstrap value) using MEGA6 [12].

2.2.6 Expression, purification and enzyme assay of recombinant EGS

Plasmid of the pET102 harboring full-length transcripts of EGS was introduced into Rosetta 2 (DE3) chemically competent cells. A single colony was inoculated in 5 mL of Luria Bertani broth (LB) and incubated on a rotary shaker (200 rpm) at 37ºC for 12 h. This culture was used for inoculating 50 mL "terrific broth" (TB) in a 250 mL Erlenmeyer flask and incubated on a rotary shaker (200 rpm) at 37ºC till optical density reached 0.6-0.8. The culture was induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated on a rotary shaker for 12 h at 18ºC. The cell pellet (7 grams) was re-suspended in lysis buffer (50 mM MOPSO, pH 7.4, 300 mM NaCl, 1mg/mL lysozyme, 1mM PMSF and 0.5% CHAPS; 5mL/gram of pellet) and incubated on ice for 30 minutes, followed by sonication using a probe sonicator for 10 cycles of 30 sec burst and 1 minute cooling with amplitude set at 85%. Lysed suspension was centrifuged at 10, 000 x g for 15 minutes. The supernatant was mixed with 2 mL of washed Ni-NTA slurry and kept on a rocker for 1 hour at 4 °C. The resin was washed with wash buffer (50 mM MOPSO, pH 7.4, 300 mM NaCl and 20 mM imidazole) and bound protein was eluted using elution buffer (wash buffer supplemented with 230 mM imidazole), fractions run on 12% poly acryl amide-SDS gel. The fractions containing pure protein were pooled together and desalted in buffer

(50 mM MOPSO, pH 7.4) using Sephadex G-25 (Hi-prep 26/10 desalting column). The protein concentration of each fraction was estimated by Bradford's method [13] and run on 12% poly acryl amide-SDS gel to check for purity and homogeneity. Enzyme activity was measured by GC-FID as described previously [2,4]. The assay mixture (200 µL) consisted of 50 mM MES buffer (pH 6.5), 1 mM coniferyl acetate/ coumaryl acetate/cinnamyl acetate and 20 µg of purified protein. The enzyme reaction was initiated by adding 1 mM NADPH. The mixture was incubated at 25ºC for 30 mins. After this incubation period, the assay contents were cooled to 4°C on ice and then linalool (1 µg) as an internal standard and NaCl (~ 200 mg) were added. The assay contents were extracted with DCM (1 mL X 3). The combined organic phase was dried over anhydrous sodium sulphate and concentrated to nearly 50 µL by flushing it with nitrogen. This extract (1 µL) was subjected to GC-FID and GC-MS analyses. For kinetic experiments, substrate concentrations ranged from 0.1 to 5 mM.

2.2.7 Modeling and docking analysis

Homology models for OgI EGS, ObaI EGS, OtI EGS and Ok EGS were constructed with SwissModel server (http://swissmodel.expasy.org/) using the structure of EGS-NADP+EMDF complex (PDB ID: 2QZZ). Energy minimization of predicted models was done with GROMOS96 force field using Swiss PDB viewer (http://spdbv.vital-it.ch/). Modelled structures were validated by RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), Molprobity (http://molprobity.biochem.duke.edu/) and ProSA analysis (https://prosa.services.came.sbg.ac.at/prosa.php). AutoDock 4.2 software (http://autodock.scripps.edu/) was used to convert receptor and ligand from *.pdb to *.pdbqt format and to set other docking parameters. The grid map was set around the Lys132 residue of EGS. Initially, two catalytic water molecules were docked in the active site. EGS with catalytic water molecules was taken for further docking with substrates i.e. cinnamyl, coumaryl and coniferyl acetates. The structure of the EGS + NADP + substrate complex was further optimized by energy minimization in Swiss PDB viewer (http://spdbv.vital-it.ch/). Complexes were analyzed for polar contact in the binding pocket using PyMol (https://www.pymol.org/).

2.2.8 Statistical analysis

All the metabolite and gene expression analysis were carried out with three biological and three technical replicates each. Data sets were represented as mean ± standard deviation. Significant differences between samples for gene expression analysis were performed with two-way ANOVA followed by Bonferroni's multiple comparisons test.

2.3 Results

2.3.1 Distribution of phenylpropanoids in different tissues of *Ocimum* species

The GC-FID and GC-MS analyses of the metabolite extracts from different tissues of various *Ocimum* species indicated the presence of at least major six phenylpropanoids with significant variation in their distribution across five species (Table 2). Predominantly, we detected end products of the phenylpropanoid pathway that were active in respective *Ocimum* species (Fig. 1). In *Og*I, eugenol content varied from 15106 ng/g of FW in young leaves to 1414 ng/g of fresh weight (FW) in stem (Table 2). In case of *Og*II, eugenol content was highest in young leaves (14942 ng/g of FW) which showed lower levels in mature leaves (5192 ng/g of FW). Eugenol methyl-ether was detected only in *Ot*I and *Ot*II and was most abundant in young leaves with 5653 and 7320 ng/g in *Ot*I and II, respectively. Eugenol was also detected in *Ot*I young and mature leaves, accounting for 213 ng/g to 490 ng/g of FW in *Ot*I, and 41 to 490 ng/g of FW in *Ot*II. Methyl cinnamate was a major phenylpropanoid in *Ob* subtypes, with its *trans* isomer being more abundant than *cis*. In *Oba*I, t-methyl cinnamate was present in varied concentrations among the different tissues ranging from 28ng/g in stem to 7325 ng/g in inflorescence. In *Oba*III, *t*-methyl cinnamate was the principal component in flower (6221 ng/g of FW) and inflorescence in *Oba*II (7645 ng/g FW). Root tissues had trace levels of these compounds in most of these species.

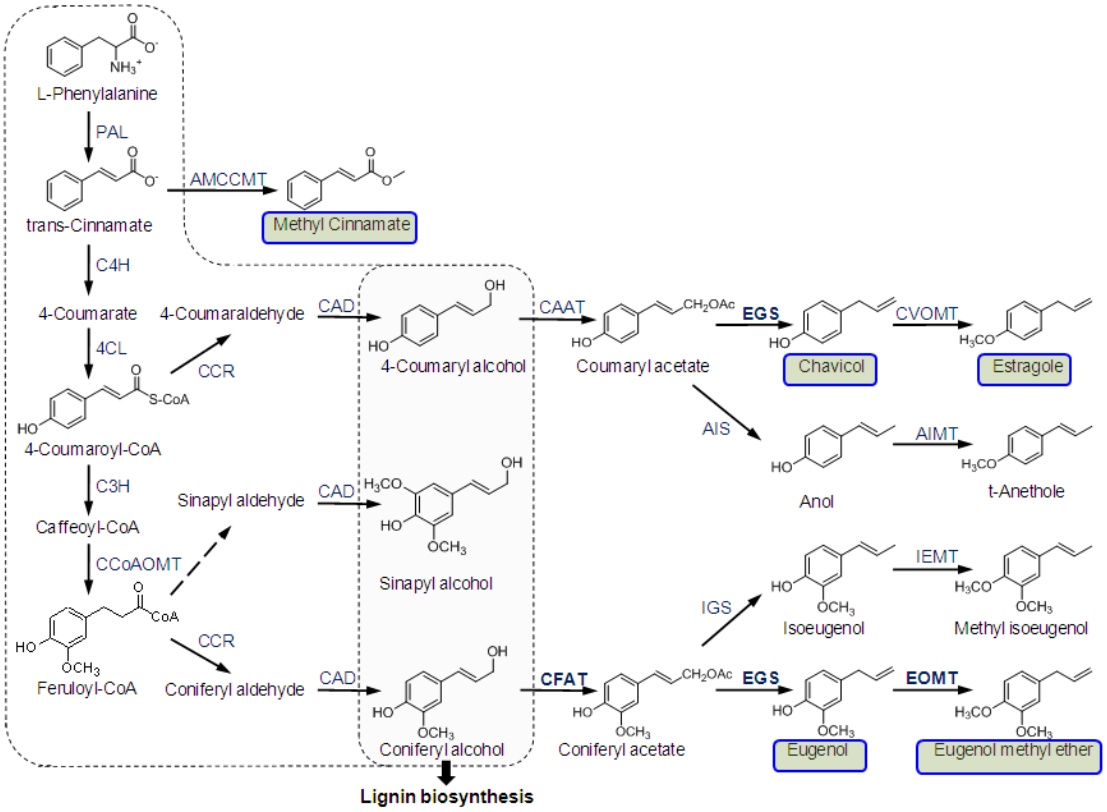


Figure 1: Phenylpropanoid biosynthetic pathway showing identified metabolites (marked in box). In most plants, alternative reactions or enzymes are present. PAL, Phenylalanine lyase; C4H, trans-Cinnamate 4-hydroxylase; 4CL, 4-Coumarate CoA ligase; C3H, 4-Coumarate-3-hydroxylase; CCoAOMT, Caffeoyl-CoA O-methyl transferase; AMCCMT, S-adenosyl L-methionine Cinnamic acid carboxy methyl transferase; CCR, coumaroyl-CoA/feruloyl-CoA reductase; CAD, coumaryl/coniferyl aldehyde dehydrogenase; CAAT, 4-Coumaryl/Coniferyl alcohol acetyl transferase; EGS, Eugenol synthase; AS, anol synthase; CVOMT, chavicol O-methyl transferase; AOMT, anol O-methyl transferase, EOMT, eugenol O-methyl transferase.

Table 2: Levels of phenylpropanoids in different tissues of five *Ocimum* species

Species	Tissues	Level of phenylpropanoids [ng/g of fresh weight (FW) ± SE]					
		Eugenol	**EME**	**Chavicol**	**Estragole**	***c*-MC**	***t*-MC**
***Og* I**	**YL**	15106.43 ± 26.72	ND	7.68 ± 0.33	ND	ND	ND
	ML	11692.18 ± 82.45	ND	5.57 ± 0.04	ND	ND	ND
	I	9229.38 ± 14.92	ND	8.32 ± 0.51	ND	ND	ND
	F	9911.74 ± 8.48	ND	8.53 ± 0.19	ND	ND	ND
	S	1414.12 ± 2.97	ND	1.32 ± 0.11	ND	ND	ND
	R	53.61 ± 1.68	ND	ND	ND	ND	ND
***Og* II**	**YL**	14942.73 ±147.1	ND	ND	ND	ND	ND
	ML	5192.45 ±45	ND	ND	ND	ND	ND
	I	4178.02 ±5.6	ND	ND	ND	ND	ND
	F	3329.96 ±0.04	ND	ND	ND	ND	ND
	S	364.21 ±22.6	ND	ND	ND	ND	ND
	R	53.08 ±21	ND	ND	ND	ND	ND
***Ot* I**	**YL**	212.76 ± 0.54	5653.86 ± 11.52	ND	ND	ND	ND
	ML	489.6 ± 0.41	2037.41 ± 2.17	ND	ND	ND	ND
	I	237.6 ± 3.22	2376.26 ± 31.54	ND	ND	ND	ND
	F	1003.88 ± 8.5	1212.21 ± 8.55	ND	ND	ND	ND
	S	351.75 ± 3.65	143.49 ± 3.88	ND	ND	ND	ND
	R	ND	ND	ND	ND	ND	ND
***Ot* II**	**YL**	41.57 ±2.2	7320.34 ± 30.78	ND	ND	ND	ND
	ML	386.97 ±16	2676.53 ± 23.81	ND	ND	ND	ND
	I	59.52 ±5.4	5818.25 ± 40.72	ND	3.86 ± 0.19	ND	ND
	F	90.39 ± 0.16	3988.1 ± 10.84	ND	ND	ND	ND
	S	119.59 ± 6.45	786.2 ± 10.28	ND	ND	ND	ND
	R	ND	ND	ND	ND	ND	ND
Ok	**YL**	3.39 ± 0.59	ND	ND	ND	ND	ND
	ML	4.11 ± 0.82	ND	ND	ND	ND	ND
	I	2.65 ± 0.98	ND	ND	ND	ND	ND
	F	ND	ND	ND	ND	ND	ND
	S	43.08 ± 13.16	ND	ND	ND	ND	ND
	R	967.88 ± 21.44	ND	ND	ND	ND	ND

Oa	**YL**	13.41 ± 0.6	ND	ND	ND	ND	ND
	ML	37.84 ± 17.81	ND	ND	ND	ND	ND
	I	3.76 ± 0.03	ND	ND	ND	ND	ND
	F	ND	ND	ND	ND	ND	ND
	S	47.16 ± 3.5	ND	ND	ND	ND	ND
	R	855.35 ±	ND	ND	ND	ND	ND
***Oba* I**	**YL**	2.95 ± 0.23	24.6 ± 1.68	ND	50.61 ± 0.7	430.28 ± 2	3250.07 ± 10.95
	ML	2.04 ± 0.43	10.05 ± 0.35	ND	17.62 ± 0.55	259.68 ± 1.28	1518.91 ± 5.19
	I	ND	24.75 ± 6.1	ND	225.49 ± 48.99	232.29 ± 3.73	7325.25 ± 28.57
	F	6.26 ± 0.54	8.28 ± 0.88	ND	37.13 ± 1.34	145.06 ± 3.21	2331.16 ± 10.85
	S	0.65 ± 0.01	1.54 ± 0.06	ND	ND	11.5 ± 0.03	27.93 ± 0.07
	R	ND	9.68 ± 0.1	ND	343.47 ± 3.22	ND	ND
***Oba* II**	**YL**	3.28 ± 0.15	29.19 ± 0.67	10.86 ± 0.27	1735.77±103	358.46 ± 27	2565.36 ± 67
	ML	5.15 ± 0.33	25.61 ± 1.56	3.54 ± 0.17	680.57 ± 72	443.29 ± 16	2769.87 ± 34
	I	ND	23.29 ± 1.42	ND	569.94 ± 18	237.99 ± 4.08	7645.68 ± 39
	F	ND	10 ± 1.39	ND	297.03 ± 18	150.09 ± 3.63	3877.79 ± 80
	S	2.21 ± 0.05	1.92 ± 0.06	2.52 ± 0.19	23.89 ± 0.91	12.11 ± 0.57	39.66 ± 1.66
	R	266.48 ± 3.12	ND	ND	ND	ND	59.62 ± 5.44
***Oba* III**	**YL**	ND	7.91 ± 0.46	ND	26.72 ± 0.26	307.81 ± 4.61	2274.11 ± 35
	ML	ND	9.46 ± 0.53	ND	30.55 ± 2.06	422.5 ± 6.81	1509.08 ± 25
	I	ND	3.31 ± 0.25	ND	36.64 ± 7.01	177.04 ± 19	2772.45 ± 11
	F	ND	9.92 ± 1.06	ND	77.96 ± 8.15	310.73 ± 31	6221.9 ± 11
	S	12.36 ± 0.37	ND	ND	ND	6.2 ± 1.08	22.21 ± 1.08
	R	315.14 ± 5.15	ND	ND	ND	ND	ND
***Oba* IV**	**YL**	ND	19.45 ± 1.75	ND	170.03 ± 1.54	272.09 ± 7.9	2746.56 ± 9.55
	ML	4.84 ± 0.14	1.07 ± 0.06	2.09 ± 0.05	2439.36 ± 7	ND	ND
	I	ND	6.72 ± 0.22	ND	724.76 ± 20	95.59 ± 1.53	2014.48 ± 33
	F	3.03 ± 0.2	2.74 ± 0.22	9.43 ± 0.45	3526.9 ± 11	ND	10.57 ± 0.34
	S	2.54 ± 0.03	ND	22.57 ± 0.33	151.75 ± 2.3	ND	ND
	R	112.58 ± 1.44	ND	ND	32.9 ± 2.92	ND	12.66 ± 1.07

ND = Not detected. The results shown are an average of three biological replicates ± standard errors. EME = eugenol methyl ether, *c*-MC = cis-methyl cinnamate, *t*-MC = trans-methyl cinnamate

2.3.2 Cloning and phylogenetic analysis of *EGS* from various *Ocimum* species

The sequence of the 750 bp amplicon obtained from *Og*I (Fig. 2) along with ObEGS1 (*O. basilicum* EGS; accession No: DQ372812) sequence obtained from NCBI database was used for 5' and 3' RACE, both of which generated 750 bp fragment (Fig. 3). This sequence information was used for OgIEGS cloning, which generated a 945 bp open reading frame (ORF) (Fig. 4) encoding 314 amino acids with an estimated molecular weight of 36 KDa. Similarly, nine EGS ORFs from other *Ocimum* species were also cloned and sequenced. Amino acid sequence alignments of EGSs indicated high sequence conservation among them. The lead residue, Lys132, was conserved in all the EGSs, with variations in residues lining both the active site (Phe85) and NADP+ binding sites (Ser43) (Fig. 5). Phylogenetic analysis based on the neighbor-joining tree clustered these EGSs into four different groups. ObaI EGS and Ok EGS formed a sub-cluster with reported Ob EGS1 indicating evolutionary proximity among them (Fig. 6). The evolutionary relatedness of OgI EGS and OgII EGS (99% identity) and that of Oa EGS, OtI EGS and OtII EGS (97% identity) was also evident as they were in separate sub-cluster of EGSs.

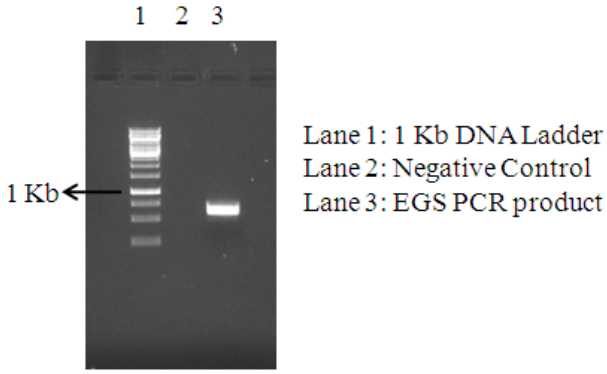


Figure 2: EGS PCR product using EGS-F and EGS-R primers and *Og*I young leaf cDNA.

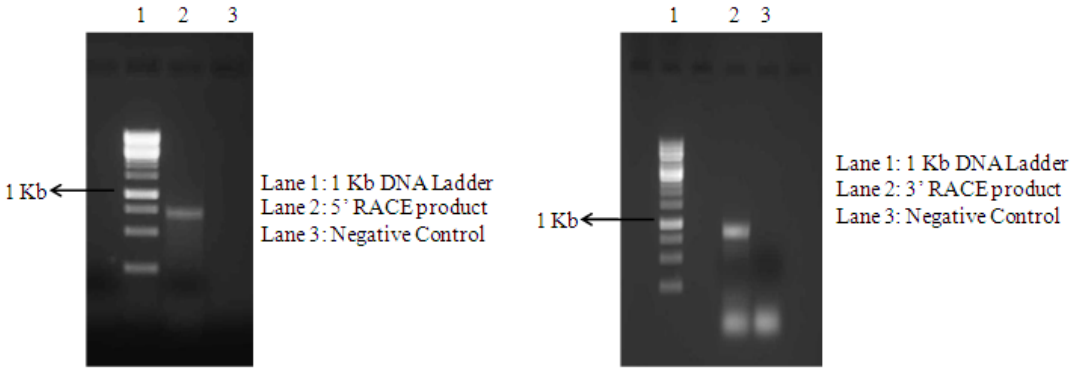


Figure 3: 5' and 3' RACE products of EGS

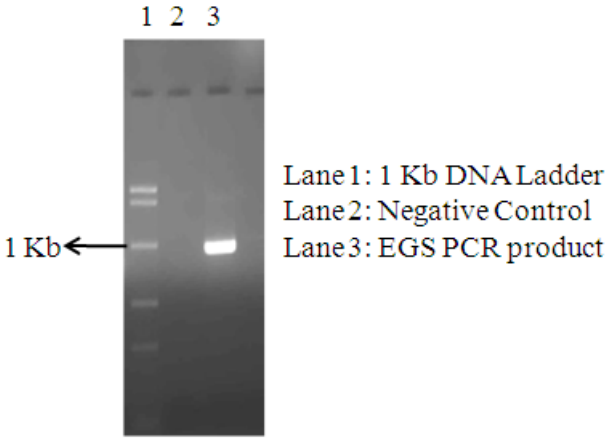


Figure 4: Full length amplicon of EGS from *Og*I

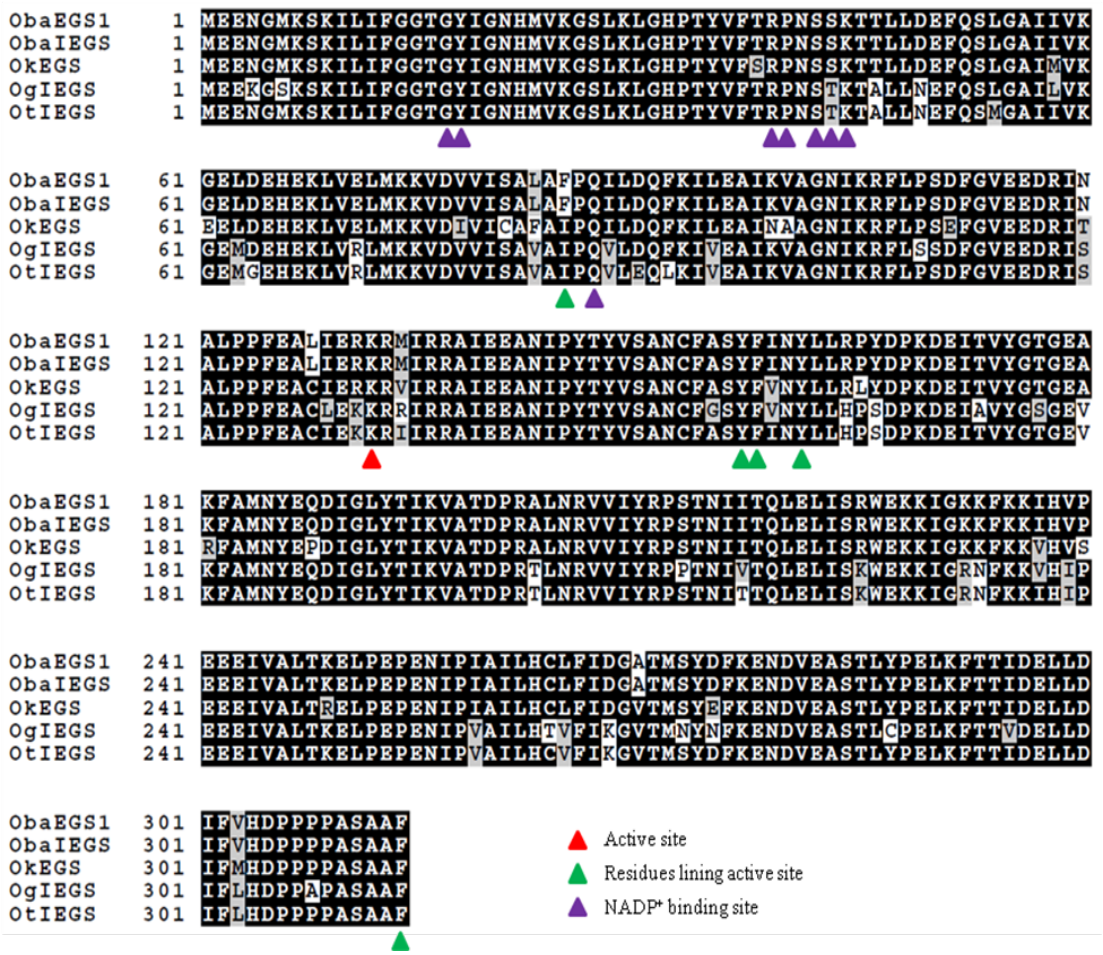


Figure 5: Amino acid sequence alignment showing sequence conservation among eugenol synthases from five *Ocimum* species. Alignment was done by CLUSTALW and shaded using BOXSHADE 3.21. Identical residues are shaded black and similar residues are marked gray. The residues involved in substrate binding, phosphate binding, $NADP^+$ binding and active sites are indicated.

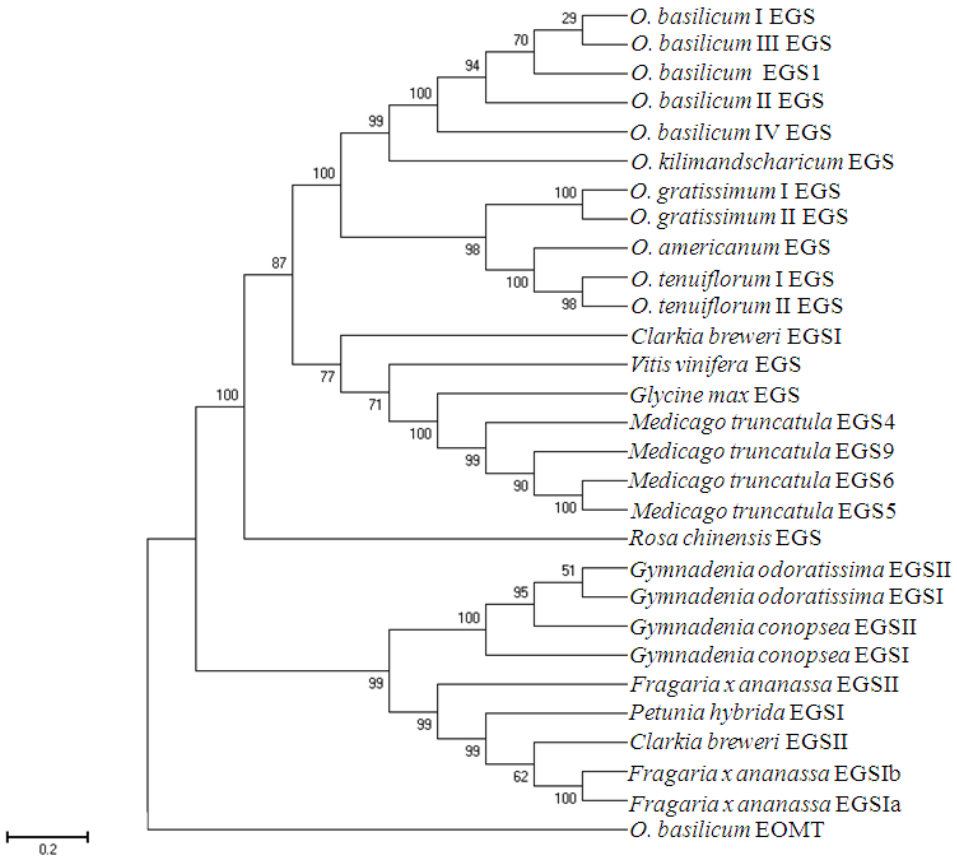


Figure 6: Phylogenetic tree of four eugenol synthases with other reported sequences using MEGA6 with the bootstrap value of 1000 iterations. Total 29 EGS sequences were used to construct NJ tree with *O. basilicum* Eugenol O-methyltransferase (AF435008) as an out group.

2.3.3 Differential expression patterns of *CFAT* and *EOMT* result in varied levels of metabolites

The expression pattern of *CFAT*, responsible for diverting flux towards phenylpropanoid pathway, was analyzed in young leaf tissue of all five *Ocimum* species (Fig. 7). Transcript levels of *CFAT* were highest in species accumulating eugenol (*Og*I) or eugenol methyl ether (*Ot*I), for which eugenol is the precursor molecule. In other species (*Ok*, *Oba*I and *Oa*), where eugenol levels were in traces, low levels of *CFAT* expression were evident. This suggested that expression of *CFAT* could be crucial for accumulation of eugenol or phenylpropanoids. Further, sqRT-PCR and qRT-PCR of *EOMT* indicated increased expression in *Ot*I (Fig. 7) with high abundance in *Ot*I and *Ot*II (Fig. 8), implying that *EOMT* expression might be responsible for the selective accumulation of eugenol methyl ether in *Ot*I and *Ot*II.

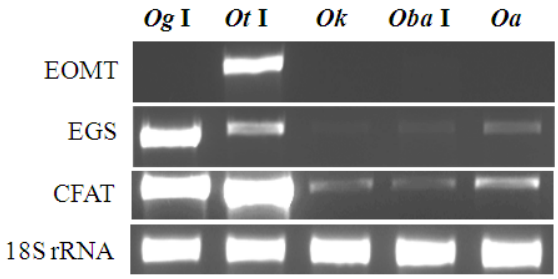


Figure 7: Semi-quantitative real time PCR of *EGS*, *EOMT* and *CFAT* in young leaf tissue of all five species, with 18S rRNA as internal reference gene.

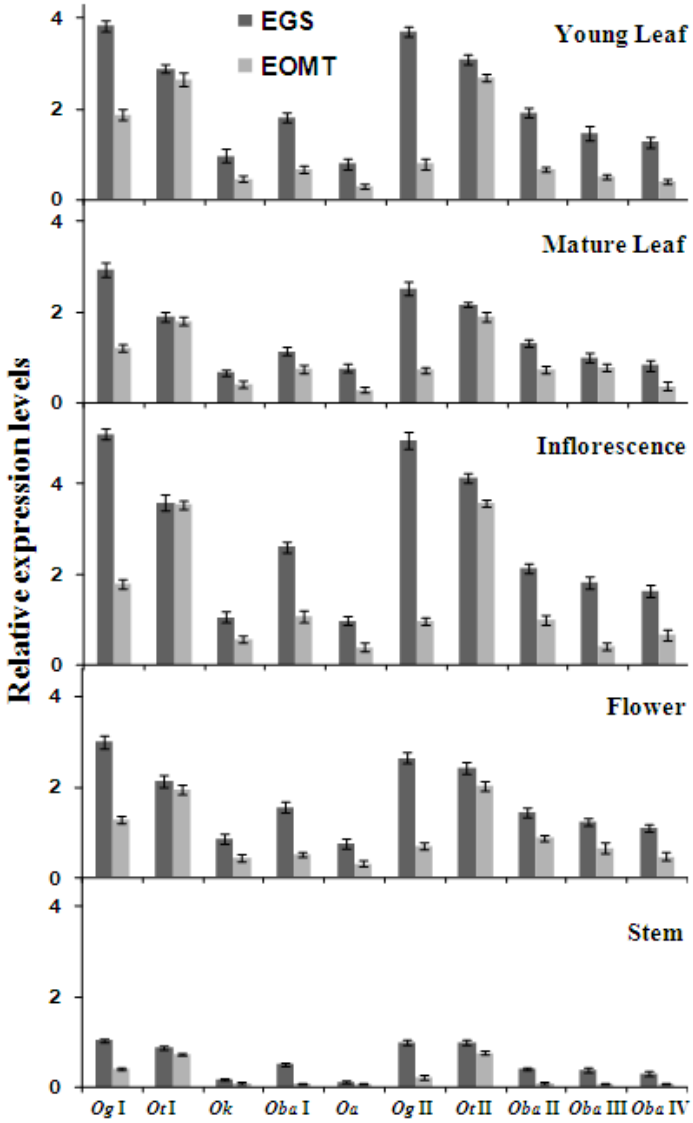


Figure 5: qRT-PCR expression analyses of *EGS* and *EOMT* in different tissues from five *Ocimum* species. Absolute quantification was done by calculating the Ct values and normalizing the results with actin, the endogenous control. Bars represent the relative standard errors.

2.3.4 *EGS* expression pattern correlates with the eugenol levels

The expression level of *EGS* was determined by qRT-PCR in five tissues of different *Ocimum* species (Fig. 8). These analyses showed a positive correlation between gene expression and eugenol levels. *EGS* expression was highest in young leaves and inflorescence tissues across all the *Ocimum* species, while stem had the lowest levels of *EGS*. The expression levels of *EGS* in mature leaves were lower than those in young ones. In *Ot*I and II, *EGS* expression levels were lower than that observed in *Og*I and II, which was consistent with the low levels of eugenol detected in these two species.

2.3.5 Substrate selectivity of recombinant *EGS* leads to high eugenol content

The recombinant EGSs were expressed and purified from four *Ocimum* species (Fig. 9 and 10). Enzyme assays were performed using coniferyl acetate, coumaryl acetate and cinnamyl acetate as substrates with NADPH as a cofactor [1,4,14]. EGS showed catalytic activity towards coniferyl acetate (Fig. 11) and coumaryl acetate with different kinetic parameters (Fig. 12 and Table 3). Assays with buffer as control did not yield any product. For coniferyl acetate, the apparent K_m and k_{cat} values of OgIEGS were 0.5 mM and 0.68 min^{-1}, respectively, while these values were 1.2 mM and 0.2 min^{-1}, respectively for coumaryl acetate (Table 3). The kinetic parameters of OtIEGS, OkEGS and ObIVEGS revealed similar biochemical characteristics. Comparison of coniferyl acetate with coumaryl acetate for their substrate affinity and turnover rate indicated that the former was the preferred substrate. When enzymatic assays were performed by mixing two substrates in the ratio of the K_m values for respective EGS or in equi-molar concentration (1 mM each), a similar trend was observed. This was also in agreement with chemical profiling results confirming eugenol as the main phenylpropanoid over chavicol. However, EGS did not yield any product when cinnamyl acetate was used as a substrate.

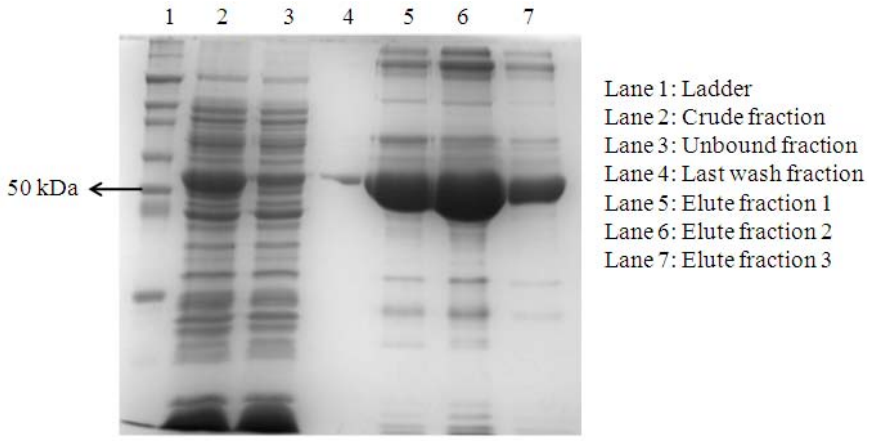


Figure 9: His-tag purification of OgI EGS. Different fractions were run on 12% SDS-PAGE. Band seen in lane 4 is sample loading error from lane 5.

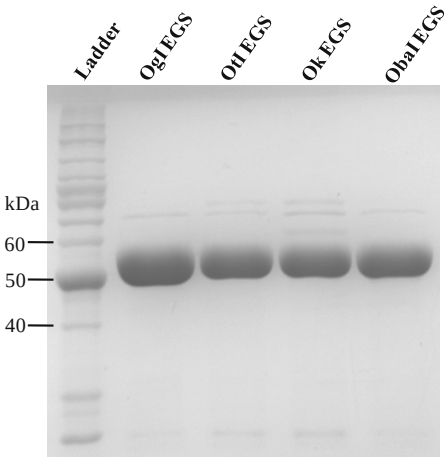


Figure 10: His-tag purified and desalted recombinant EGSs run on 12% SDS-PAGE.

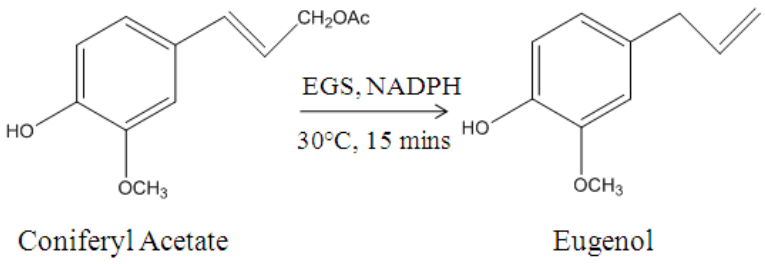


Figure 11: General scheme of reaction catalyzed by EGS.

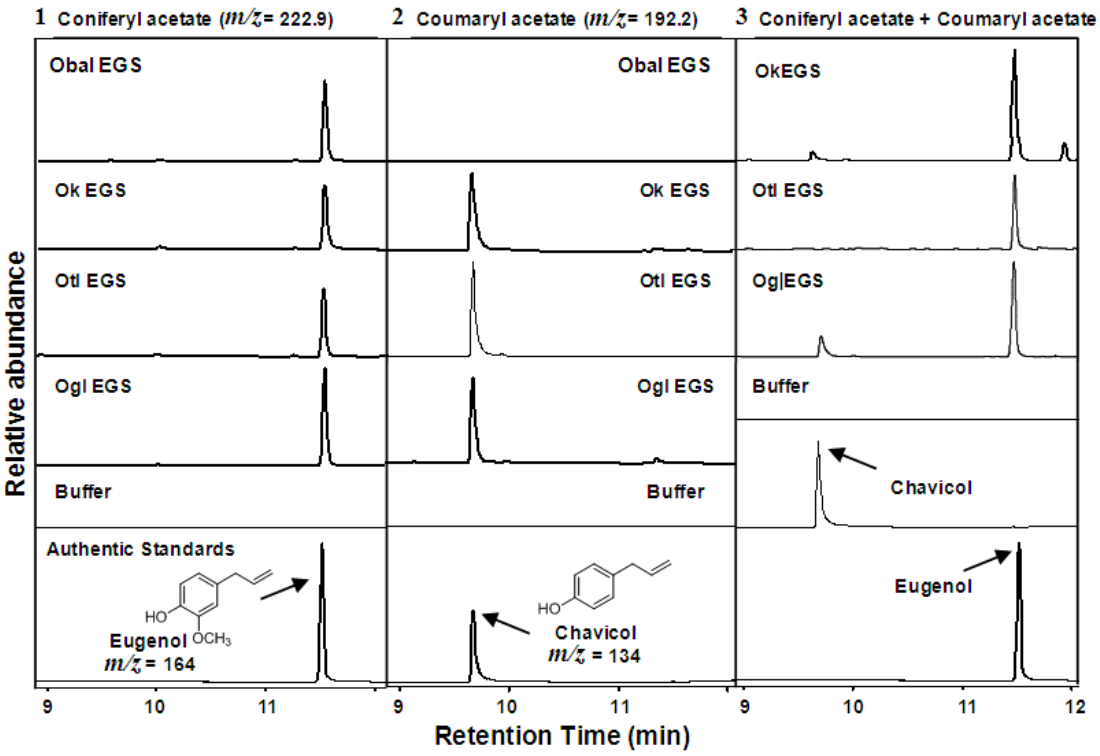


Figure 12: Characterization of recombinant EGS – assay of recombinant *Og*I EGS, *Ok* EGS and *Ot*I EGS with (1) 8 mM coniferyl acetate, (2) 8 mM coumaryl acetate and (3) with substrate mixed in respective K_m ratios.

Table 3: Values of kinetic parameters for four recombinant EGS

Enzyme	Substrate	K_m (mM)	k_{cat} (min^{-1})	k_{cat}/K_m ($mM^{-1}min^{-1}$)
***Og*I EGS**	Coniferyl acetate	0.5 ± 0.04	0.68 ± 0.012	1.36
	Coumaryl acetate	1.25 ± 0.09	0. 2017 ± 0.012	0.16
***Ot*I EGS**	Coniferyl acetate	1.17 ± 0.3	1.68 ± 0.25	1.44
	Coumaryl acetate	1.7 ± 0.2	0.8 ± 0.08	0.47
***Ok* EGS**	Coniferyl acetate	0.41 ± 0.14	0.18 ± 0.013	0.44
	Coumaryl acetate	0.55 ± 0.2	0.63 ± 0.064	1.15
***Oba*I EGS**	Coniferyl acetate	0.7 ± 0.02	0.23 ± 0.003	0.33
	Coumaryl acetate	ND	ND	

ND: Not detected. The results shown are an average of three technical replicates ± standard errors

2.3.6 Stable interaction between EGS and its substrates results in preferential eugenol accumulation

Three-dimensional structures of EGS showed similarity in structural attributes with other PIP-family proteins. EGS showed an N-terminal, Rossman-fold domain, containing a core, six-stranded parallel β-sheet flanked on each face by a helical layer. It had two important regions of catalytic function. First region was the binding pocket of active site, where a substrate's central scaffold attached leading to successful enzymatic reaction. β-sheet of the one core edge served as binding surface for the $NADP^+$ cofactor binding, while C-terminal domain was presumed to function in substrate binding. The examination of complex structure showed that in the binding pocket, the quinone-methide transition state was involved in cleaving the carbon-oxygen bond from the acetate moiety [15-17] and in serving as the actual substrate of the reduction reaction *via* NADPH-mediated hydride transfer (Fig. 13). Analysis of the polar contact network at the e-amino group of Lys132 indicated that the unprotonated $-NH_2$ state of this residue was the key donor in hydrogen-bond interactions with the hydroxyl group of the nicotinamide-ribose and the backbone carbonyl oxygen of Ser110. Furthermore, due to the electronegative nature of Lys132, it served as the acceptor in a hydrogen bond with the bridging water molecule and thus, acted as a general base. The de-protonation of *p*-hydroxyl group of the substrate was facilitated by its interaction with the water molecule (as a hydroxide ion). The second important region of catalytic function was the capping site of EGS active site, which was non-polar in nature. Tyr157 was a single residue available for hydrogen bond interactions with the polar oxygen atoms of the acetate moiety. The presence of aromatic side-chains in the lining region of the active site stabilized the binding of substrates.

All the variants of *Ocimum* EGS had similar three-dimensional structures. By superimposing OgI EGS, Ok EGS, and OtI EGS on ObaI EGS, we observed that the active site was conserved, including the binding (Lys132) and capping regions (Tyr157) [18,19]. A docking study of various substrates (cinnamyl, coumaryl and coniferyl acetates) with OgI EGS, OtI EGS and ObaI EGS, indicated difference in binding energies (Table 4). This might be due to the variation in the residues that line active sites. Taken together, EGS binding to coniferyl acetate was relatively stronger than other substrates, coumaryl or cinnamyl acetate. In the case of coniferyl acetate,

the presence of an additional hydroxymethyl group at the C3 position led to the formation of extra contact *via* attractive van-der-Waals force with lining residues. This might explain the strong binding of coniferyl acetate establishing it as the preferred substrate and thus, leading to more eugenol production than chavicol.

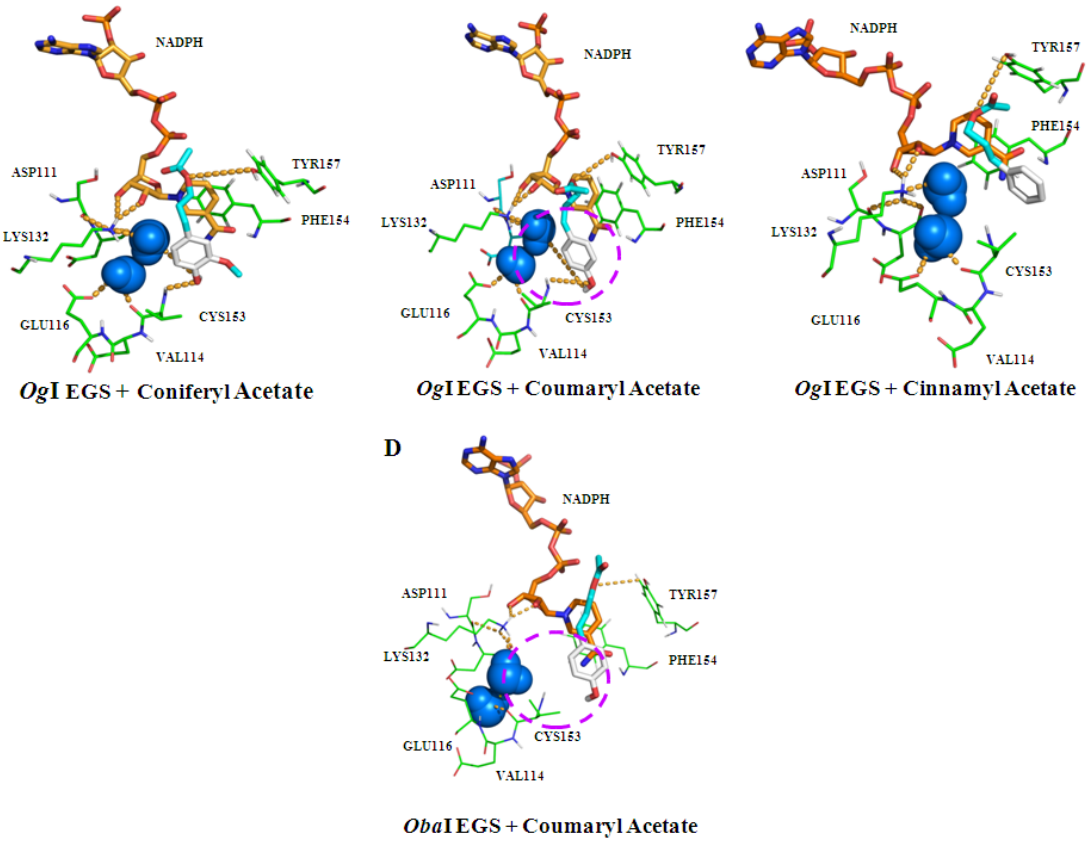


Figure 13: Molecular docking of *Og*I EGS with different substrates. Interaction of active site lining residues of *Og*I EGS with (A) coniferyl acetate, (B) coumaryl acetate and (C) cinnamyl acetate. The standard atom coloring is used, with white carbon atoms for the substrates and orange carbons for NADPH. The close interaction between the binding pocket residues of *Og*I EGS, substrates and the hydride donor of the nicotinamide (C4) is shown as orange dashed line. Variation in the interaction of *Oba*I EGS with (D) coumaryl acetate as compared to *Og*I EGS is highlighted with dashed circle.

Table 4: Binding scores of EGSs with three substrates.

	Binding Score (kcal/mol)			
Substrate	***Og*I EGS**	***Ot*I EGS**	***Ok* EGS**	***Oba*I EGS**
Coniferyl acetate	**-7.6**	**-6.9**	**-8.1**	**-6.7**
Coumaryl acetate	**-7.3**	**-6.5**	**-7.6**	**-6.6**
Cinnamyl acetate	**-7.0**	**-6.3**	**-7.5**	**-6.4**

2.4 Discussion

Ocimum is a member of the mint family (*Lamiaceae*). Different *Ocimum* species have various combinations of phenylpropanoids, with eugenol and eugenol O-methyl ether being the most abundant in *O. gratissimum* and *O. tenuiflorum* species, while methyl cinnamate and estragole are more in *O. basilicum* species [5,20]. These compounds are synthesized in specialized structures (glandular trichomes), located on the leaf surface [21]. Here, eugenol, eugenol methyl ether, methyl cinnamate and estragole were identified as the major phenylpropanoids. Young plant tissue such as leaves and inflorescence had more phenylpropanoids than mature tissues that was consistent with their defensive role against pathogen attack and herbivory [5]. In mature leaves and flowers, substantial levels of these compounds were detected. The root tissue contained trace levels of these compounds, which could be transported through xylem and phloem [22]. There was extensive metabolite diversity, within and across different *Ocimum* species, e.g. selective presence of eugenol and eugenol methyl ether along with varied ratios of different phenylpropanoids in *Oba* species. The metabolite diversity across different species could be explained by the differential gene expression pattern (e.g. *EOMT* in *Ot*I and *Ot*II) resulting in accumulation of eugenol methyl ether. Consequently in other species, *EOMT* expression level was low leading to absence of this metabolite. Additionally, differences in metabolite profile could also be attributed to the post-translational modifications, whereby transcript abundance was not reflected in metabolite formation. For example, in *O. basilicum var* SD, enzyme responsible for methylating chavicol was abundant but methylchavicol was not detected due to post-translation ubiquitylation of this enzyme [23]. Variations in metabolite accumulation within the species could be attributed to a number of factors. Differential gene expression of pathway enzymes and transcription factors could also bring these changes [24,25]. For instance, in two different varieties of *O. basilicum* SD and EMX-1, differential accumulation of phenylpropanoid was observed due to varied expression patterns of entry point enzymes, phenylalanine ammonia-lyase and 4-coumarate-CoA ligase [23].

EGS has been characterized in a number of plant species such as petunia, clarkia and strawberry [8,14]. In phylogenetic analysis, *Ocimum* EGSs were well separated from those of *Petunia hybrida, Clarkia breweri, Vitis vinifera* and *Medicago truncatula* that were grouped in separate sub-cluster. These 10 EGSs

obtained from five *Ocimum* species were distributed in two sub-groups. Sequence comparison revealed that EGSs from different *Ocimum* species had high sequence conservation suggesting evolutionary close to EGS from other plants. *EGS* expression levels were positively correlated with the eugenol levels in different tissues from *Ocimum* species. *EGS* expression was highest in *Og*I and *Og*II, where eugenol was the major phenylpropanoid. Variation in *CFAT* expression levels, along with *EGS* and *EOMT* levels might be crucial for the selective eugenol accumulation among these *Ocimum* species. It has been reported that silencing of *CFAT* led to decreased accumulation of iso-eugenol in petunia [26]. However, further experiments in terms of their *in planta* over-expression and silencing would be necessary to confirm their exact role. Characterizations and comparisons of four EGSs with coniferyl acetate and coumaryl acetate indicated that the former was preferred substrate based upon its enzyme affinity, turnover number and catalytic efficiency. Additionally, when the substrates were mixed together either in equi-molar ratio or in ratio of their K_m, eugenol was the major product, with traces of chavicol detected in *Oba*I EGS and *Ot*I EGS. A homology model of EGS was made using the reported ObEGS1 structure as a template and interactions of both the substrates with the enzyme were analyzed. The lack of a hydroxyl-methyl substituent at the 3-position, as in the case of coumaryl acetate, significantly altered its stable interaction with EGS resulting in reduced preference as a substrate. In case of EGS from *Og*, *Ok* and *Ot*, coumaryl acetate made polar contact with catalytic water and Tyr157. Although with *Oba*I EGS, coumaryl acetate interacted through Tyr157, there was absence of polar interaction between coumaryl acetate and catalytic water, which might be one of the reasons for inactivity of EGS with it.

2.5 Conclusion

The work described in this chapter provides better insight into the phenylpropanoids distribution in different tissues of five *Ocimum* species. Eugenol abundance was positively correlated with the *EGS* and expression levels of *CFAT* along with kinetic parameters of EGS revealed selective and higher eugenol accumulation than chavicol in these *Ocimum* species. Gene expression studies of *EOMT* provided evidence for its accumulation only in *O. tenuiflorum* species. Homology modeling and substrate docking clearly demonstrated the stable binding of coniferyl acetate making it as the preferred substrate compared to others. Taken together, these findings presented crucial insight for higher eugenol content in these *Ocimum* species and could be further utilized for metabolic manipulations in plants.

2.6 Appendix:

2.6.1 Amino acid sequence of all ten eugenol synthases

>OgI EGS

MEEKGSKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSTKTALLNEFQS
LGAILVKGEMDEHEKLVRLMKKVDVVISAVAIPQVLDQFKIVEAIKVAGNIK
RFLSSDFGVEEDRISALPPFEACLEKKRRIRRAIEEANIPYTYVSANCFGSYFVN
YLLHPSDPKDEIAVYGSGEVKFAMNYEQDIGLYTIKVATDPRTLNRVVIYRPP
TNIVTQLELISKWEKKIGRNFKKVHIPEEEIVALTKELPEPENIPVAILHTVFIKG
VTMNYNFKENDVEASTLCPELKFTTVDELLDIFLHDPPAPASAAF

>OgII EGS

MEEKGSKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSTKTALLNEFQS
LGAILVKGEMDEHEKLVRLMKKVDVVISAVAIPQVLDQFKIVEAIKVAGNIK
RFLPSDFGVEEDRISALPPFEACLEKKRRIRRAIEEANIPYTYVSANCFGSYFVN
YLLHPSDPKDEIAVYGSGEVKFAMNYEQDIGLYTIKVATDPRTLNRVVIYRPP
TNIVTQLELISKWEKKIGRNFKKVHIPEEEIVALTKELPEPENIPVAILHTVFIKG
VTMNYNFKENDVEASTLYPELKFTTVDELLDIFLHDPPAPASAAF

>Ok EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFSRPNSSKTTLLDEFQS
LGAIMVKEELDEHEKLVELMKKVDIVICAFAIPQILDQFKILEAINAAGNIKRF
LPSEFGVEEDRITALPPFEACIERKRVIRRAIEEANIPYTYVSANCFASYFVNYL
LRLYDPKDEITVYGTGEARFAMNYEPDIGLYTIKVATDPRALNRVVIYRPSTN
IITQLELISRWEKKIGKKFKKVHVSEEEIVALTRELPEPENIPIAILHCLFIDGVT
MSYEFKENDVEASTLYPELKFTTIDELLDIFMHDPPPPASAAF

>OtI EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSTKTALLNEFQ
SMGAIIVKGEMGEHEKLVRLMKKVDVVISAVAIPQVLEQLKIVEAIKVAGNIK
RFLPSDFGVEEDRISALPPFEACIEKKRIIRRAIEEANIPYTYVSANCFASYFINY
LLHPSDPKDEITVYGTGEVKFAMNYEQDIGLYTIKVATDPRTLNRVVIYRPST
NITTQLELISKWEKKIGRNFKKIHIPEEEIVALTKELPEPENIPVAILHCVFIKGV
TMSYDFKENDVEASTLYPELKFTTIDELLDIFLHDPPPPASAAF

>OtII EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSTKTALLNEFQ
SMGAIIVKGEMDEHEKLVRLMKKVDVVISAVAIPQVLEQLKIVEAIKVAGNIK
RFLPSDFGVEEDRISALPPFEACIEKKRIIRRAIEEANIPYTYVSANCFASYFINY
LLHPSDPKDEITVYGTGEVKFAMNYEQDIGLYTIKVATDPRTLNRVVIYRPST
NITTQLELISKWEKKIGRNFKKIHIPEEEIVALTKELPEPENIPVAILHCVFIKGV
TMSYDFKENDVEASTLYPELKFTTIDELLDIFLHDPPPPASAAF

>ObI EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSSKTTLLDEFQS
LGAIIVKGELDEHEKLVELMKKVDVVISALAFPQILDQFKILEAIKVAGNIKRF
LPSDFGVEEDRINALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYL
LRPYDPKDEITVYGTGEAKFAMNYEQDIGLYTIKVATDPRALNRVVIYRPSTN
IITQLELISRWEKKIGKKFKKIHVPEEEIVALTKELPEPENIPIAILHCLFIDGATM
SYDFKENDVEASTLYPELKFTTIDELLDIFVHDPPPPASAAF

>ObII EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSSKTTLLDEFQS
LGAIIVKGELDEHEKLVELMKKVDVVISALAFPQILDQFKILEAIKVAGNIKRF
LPSDFGVEEDRINALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYL
LRPYDPKDEITVYGTGEAKFAMNYEQDIGLYTIKVATDPRALNRVVIYRPSTN
IITQLELISRWEKKIGKKSKKIHVPEEEIVALTKELPEPENIPIAILHCLFIDGATM
SYDFKENDVEASTLYPELKFTTIDELLDIFVHDPPPPASAAF

>ObIII EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSSKTTLLDEFQS
LGAIIVKGELDEHEKLVELMKKVDVVISALAFPQILDQFKILEAIKVAGNIKRF
LPSDFGVEEDRINALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYL
LRPYDPKDEITVYGTGEAKFAMNYEQDIGLYTIKVATDPRALNRVVIYRPSTN
IITQLELISRWEKKIGKKFKKIHVPEEEIVALTKELPEPENIPIAILHCLFIDGATM
SYDFKENDVEASTLYPELKFTTIDELLDIFVHDPPPPASAAF

>ObIV EGS

MEENGMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSSKTTLLDEFQS
LGAIIVKGELDEHEKLVELMKKVDVVISALAFPQILDQFKILEAIKVAGNIKRF
LPSDFGVEEDRINALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYL
LRPYDPKDEIKVYGTGEAKFAMNYEQDIGLYTIKVATDPRALNRVVIYRPSTN
IITQLELISRWEKKIGKKFKKIHVPEEQIVALTKELPEPENIPIAILHCLFIDGATM
SYEFKENDVEASTLYPELKFTTIDELLDIFVHDPPPPASAAF

>Oa EGS

MEEKGSKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNSTKTALLNEFQS
MGAIIVKGEMGEHEKLVRLMKKVDVVISAIAIPQVLEQLKIVEAIKVAGNIKR
FLPSDFGVEEDRISALPPFEACIEKKRIIRRAIEEANIPYTYVSANCFASYFINYL
LHPSDPKDEITVYGTGEVKFAMNYEQDIGLYTIKVATDPRTLNRVGIYRPSTNI
TTQLELISKWEKKIGRNFKKIHIPEEEIVALTKELPEPENIPVAILHCVFIKGVTM
SYDFKENDVEALTLYPELKFTTIDKLLDIFLHDPPAPASAAF

2.6.2 Amino acid alignment of all ten eugenol synthases

2.7 References

[1] Gang DR, Wang J, Dudareva N, et al., An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, Plant Physiol. 125 (2001) 539-555.

[2] Louie GV, Baiga TJ, Bowman ME, et al., Structure and reaction mechanism of basil eugenol synthase, PLoS ONE 2 (2007) e993.

[3] Reigosa MJ, Pazos-Malvido E, Phytotoxic effects of 21 plant secondary metabolites on *Arabidopsis thaliana* germination and root growth, J. Chem. Ecol. 33 (2007) 1456-1466.

[4] Koeduka T, Fridman E, Gang DR, et al., Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 10128-10133.

[5] Vassao DG, Gang DR, Koeduka T, et al., Chavicol formation in sweet basil (*Ocimum basilicum*): cleavage of an esterified C9 hydroxyl group with NAD(P)H-dependent reduction, Org. Biomol. Chem. 4 (2006) 2733-2744.

[6] Singh P, Ramesha HJ, Agawane SB, et al., Potential dual role of eugenol in inhibiting advanced glycation end products in diabetes: Proteomic and mechanistic Insights, Sci. Rep. 6 (2016) 18798

[7] Huang B, Keeling PL, Hennen-Bierwagen TA, et al., Comparative in vitro analyses of recombinant maize starch synthases SSI, SSIIa, and SSIII reveal direct regulatory interactions and thermo-sensitivity, Arch. Biochem. Biophys. 596 (2016) 63-72

[8] Koeduka T, Louie GV, Orlova I, et al., The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages, Plant J. 54 (2008) 362-374.

[9] Koeduka T, Baiga TJ, Noel JP, et al., Characterization of t-anol/isoeugenol synthase and an O-methyltransferase specific for a C7-C8 propenyl side chain, Plant Physiol. 149 (2009) 384-394.

[10] Min T, Kasahara H, Bedgar DL, et al., Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases, J. Biol. Chem. 278 (2003) 50714-50723.

[11] Singh P, Jayaramaiah RH, Sarate P, et al., Insecticidal potential of defense metabolites from *Ocimum kilimandscharicum* against *Helicoverpa armigera*, PLoS ONE 9(8) (2014) e104377.

[12] Tamura K, Stecher G, Peterson D, et al., MEGA6: Molecular evolutionary genetics analysis Version 6.0, Mol. Biol. Evol. 30 (2013) 2725-2729.

[13] Bradford MM, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248-254.

[14] Araguez I, Osorio S, Hoffmann T, et al., Eugenol production in achenes and receptacles of strawberry fruits is catalyzed by synthases exhibiting distinct kinetics, Plant Physiol. 163 (2013) 946-958.

[15] Akashi T, Koshimizu S, Aoki T, et al., Identification of cDNAs encoding pterocarpan reductase involved in isoflavan phytoalexin biosynthesis in *Lotus japonicus* by EST mining, FEBS Lett. 580 (2006) 5666-5670.

[16] Fujita M, Gang DR, Davin LB, et al., Recombinant pinoresinol-lariciresinol reductases from western red cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions, J. Biol. Chem. 274 (1999) 618-627.

[17] Gang DR, Kasahara H, Xia ZQ, et al., Evolution of plant defense mechanisms: Relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases, J. Biol. Chem. 274 (1999) 7516-7527.

[18] Filling C, Berndt KD, Benach J, et al., Critical residues for structure and catalysis in short-chain dehydrogenases/reductases, J. Biol. Chem. 277 (2002) 25677-25684.

[19] Wang X, He X, Lin J, et al., Crystal structure of isoflavone reductase from alfalfa (*Medicago sativa* L.), J. Mol. Biol. 358 (2006) 1341-1352.

[20] Raina A, Kumar A, Dutta M, Chemical characterization of aroma compounds in essential oil isolated from "Holy Basil" (*Ocimum tenuiflorum* L.) grown in India, Genet. Resou. Crop. Evol. 60 (2013) 1727-1735.

[21] Iijima Y, Davidovich-Rikanati R, Fridman E, et al., The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil, Plant Physiol. 136 (2004) 3724-3736.

[22] Wink M, Introduction: biochemistry, physiology and ecological functions of secondary metabolites. In: M. Wink (Ed.), Biochemistry of plant secondary metabolites, Annu Plant Rev, second ed., Wiley-Blackwell, Hoboken, New Jersey, 2010, pp. 1-19.

[23] Xie Z, Kapteyn J, Gang DR, et al., A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes, Plant J. 54 (2008) 349-361.

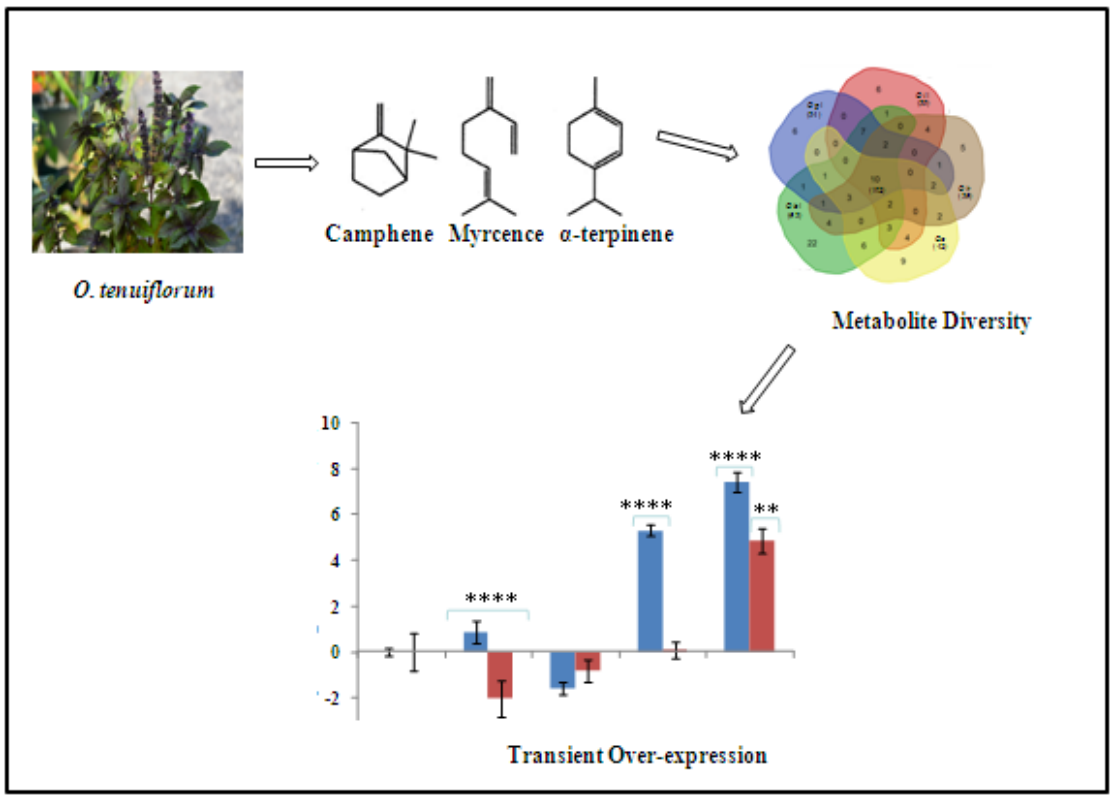
[24] Singh P, Kalunke RM, Giri AP, et al., Towards comprehension of complex chemical evolution and diversification of terpenoids and phenylpropanoid pathways in *Ocimum* species, RSC *Adv*, 5 (2015) 106886-106904.

[25] Zvi MMB, Shklarman E, Masci T, et al., PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers, New Phytol. 195 (2012) 335-345.

[26] Dexter R, Qualley A, Kish CM, et al., Characterization of a petunia acetyl-transferase involved in the biosynthesis of the floral volatile iso-eugenol, Plant J. 49 (2007) 265–275.

Chapter 3

Metabolic Diversity and Characterization of Synthase Involved in Monoterpene Accumulation



Chapter 3

Metabolic Diversity and Characterization of Synthase Involved in Monoterpene Accumulation

Ocimum species produces diverse range of specialized metabolites with immense medicinal properties, however, limited information exists on how these metabolites are accumulated, leading to species diversity. In this chapter, we carried out metabolite profiling of different tissues of five *Ocimum* species and identified major terpenes broadly classifying these species in two distinct chemotypes *viz*. terpene (particularly β-caryophyllene, myrcene, camphor, borneol and selinene) and phenylpropanoid (eugenol, eugenol methyl ether etc.) rich. Transcriptome sequencing of different tissues of three *Ocimum* species resulted in identification of 38 unique transcripts belonging to terpene synthase (TPS) family. Full-length gene cloning, sequence and phylogenetic analysis of three of these transcripts were carried out along with their expression in various tissues of five *Ocimum* species. Terpene metabolite and expression profiling of candidate *TPS* genes in various tissues of *Ocimum* species revealed spatial variances. Putative TPS contig 19414 (*TPS*1) was selected to confirm its role in terpenes biosynthesis. Agrobacterium-mediated transient over-expression assay in the leaves of *O. kilimandscharicum* and subsequent gene expression and metabolite analyses confirmed *TPS*1 as a *cis*-β-terpineol synthase. Overall, present study provided better molecular basis of terpene diversity in *Ocimum* species and could be useful in the enhancement of their various medicinal properties.

3.1 Introduction

Plants have evolved specialized secondary biosynthetic pathways for the synthesis of structurally and functionally complex small molecules, which are essential for the survival. In plants, the pattern of secondary metabolites is complex as it changes in a tissue- and/or organ-specific manner. Secondary metabolites play a key role in the protection of plants against biotic and abiotic stresses. These molecules also play crucial role in, attraction of pollinators and frugivores, survive under given environmental conditions and communications [1]. Terpenes are a large and diverse class of naturally occurring organic compounds present in all forms of living systems. All isoprenoid compounds are constructed from two simple five-carbon building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Biosynthesis of isoprenoids can be divided into four basic steps, which takes place either in cytosol (Mevalonic acid pathway) [2,3] and/or plastids (Methyl erythritol pathway) [4,5,6]: (i) production of the C5 monomers IPP and DMAPP (ii) the head to tail condensation of C_5-units IPP and DMAPP into geranyl diphosphate (GDP), farnesyl diphosphate (FDP) and geranyl geranyl diphosphate (GGDP), (iii) cyclization and/or rearrangement to form the diverse class of terpene carbon backbone and (iv) tailoring and downstream functional group modifications of the terpene skeleton. The compartmentalization of two biosynthetic pathways in plants that leads to the formation of DMAPP/IPP is not absolute, because in several cases one metabolite can exchange between these two pathways [7,8]. Terpene synthases catalyze biosynthesis with high regio- and stereo-chemical precision in a cascade of complex reactions involving highly reactive carbocationic intermediates that undergo a sequence of reactions such as cyclizations, alkylations, rearrangements, deprotonations and hydride shifts [9,10,11,12,13,14,15,16]. Terpenes can be classified into different groups according to the number of five carbon units present in the core structure *viz*. single isoprene unit (C5) are called hemiterpenes, two isoprene units (C10) are called monoterpenes, three isoprene unit (C15) called sesquiterpenes, and so on. Terpene derivatives are also used for the production of numerous pharmaceuticals including taxol (anti-cancer drug) [17], artemisinin (anti-malarial drug) [18].

Monoterpenes consist of two isoprene units with the molecular formula $C_{10}H_{16}$ and are present in secretory tissues such as oil glands of higher plants, insects, fungi and marine organisms. These compounds are widely used in flavor and fragrance

industry, due to their characteristic odour. Monoterpenes can be divided into three subgroups: acyclic (myrcene, geraniol, linalool), monocyclic (limonene, α-terpineol and terpinolene) and bicyclic (α-pinene, sabinene and camphor). Monoterpenes are biosynthesized from geranyl pyrophosphate (GPP) catalyzed by monoterpene synthase. Several monoterpenes possess various pharmacological properties including antibacterial antifungal, antioxidant, and anti cancerous [19]. Generally in plants, monoterpenes are synthesized in plastids through the MEP pathway whereas in other higher organisms and in yeast, they are synthesized through the MVA pathway [20].

The metabolite fraction of *Ocimum* is mainly composed of phenylpropanoid, mono- and sesquiterpenes. We performed terpene profiling in six different tissues of five *Ocimum* species which revealed species specific distribution of mono- and sesquiterpenes in different *Ocimum* species. Transcriptome profiling was carried out for leaf, inflorescence and stem tissue of *O. gratissimum* (*Og*), *O. tenuiflorum* (*Ot*) and *O. kilimandscharicum* (*Ok*) which lead to the identification of 38 putative terpene synthases represented by mono-, sesqui- and diterpenes synthases. Expression analysis of three of these putative synthases (*TPS*1, *TPS*2 and *TPS*3) in different tissues of *Ocimum* species revealed variation in their expression levels in species specific manner. Further, Agrobacterium mediated transient over-expression of *TPS*1 established its role as *cis*-β-terpineol synthase.

3.2 Materials and Methods

3.2.1 Chemicals and plant materials

All chemicals were purchased from Sigma-Aldrich (Sigma Chemical Co., USA), unless stated. *Ocimum* plants belonging to five species *viz*. *Ocimum gratissimum* L. I and II (*Og*I and *Og*II), *O. tenuiflorum* I and II (*Ot*I and *Ot*II), *O. kilimandscharicum* Gürke (*Ok*), *O. americanum* L. (*Oa*) and *O. basilicum* L. I, II, III and IV (*Ob*I, *Ob*II, *Ob*III and *Ob*IV) were grown in a greenhouse at 25-28ºC with ~35-40% humidity and 16 h light and 8 h dark periods. All *Ocimum* plants chemotypes were phenotypically confirmed through taxonomist and samples of these the plants were deposited in the herbarium of Botanical Survey of India, Pune as voucher specimens. All the samples were collected from 3-4 months old plants, immediately frozen in liquid nitrogen and stored at -80ºC until further use. Fresh tissues were used for metabolite extraction and analysis.

3.2.2 Metabolite analysis from different *Ocimum* species

Metabolite analysis was carried out in six different tissues including young leaves (YL; top whorl), mature leaves (ML; third whorl), inflorescence (I), flower (F), stem (S) and root (R) from five *Ocimum* species. Extractions were performed by dichloromethane (DCM) extraction method [21,22]. Briefly, fresh tissues (5 g each) from different plant parts were harvested separately and immediately soaked in 50 mL DCM for 20 h at 28ºC. The combined organic phase was cooled to -20°C for 2 h (for lipid precipitation) and filtered. The contents were dried, weighed, re-dissolved in 2 mL DCM and subjected to GC and GC-MS analysis was carried out for 3 biological replicates in triplicates. GC analyses were carried out on an Agilent 7890A instrument equipped with a hydrogen flame ionization detector and HP-5 capillary column (30 m X 0.32 mm X 0.25 µm, J and W Scientific, USA). Nitrogen was used as the carrier gas at a flow rate of 1 mL/min. The column temperature was raised from 70 °C to 110 °C at 2 °C min^{-1}, then raised to 180 °C at 3 °C min^{-1} and finally at 10 °C min^{-1} raised to 220 °C and held for 2 min. Injector and detector temperatures were 230 °C and 235 °C, respectively. GC-MS was performed on an Agilent 5975C mass selective detector interfaced with an Agilent 7890A gas chromatograph using an HP-5 MS capillary column with helium as the carrier gas and using above mentioned conditions. Compounds were identified by co-injection studies, comparing the retention time and

mass fragmentation pattern with those of reference compounds and also compared acquired mass spectra and retention indices with those of NIST/NBS and the Wiley mass spectral library (software version 2.0, Dec. 2005).

3.2.3 RNA isolation and transcriptome sequencing

Total RNA was extracted from the different tissues using Spectrum^{TM} Plant Total RNA Isolation Kit (Sigma Chemical Co., USA). For transcriptome analysis total RNA from leaf, inflorescence and stem tissues of *Og*, *Ot* and *Ok* was mixed and proceeded with library preparation. Transcriptome library was constructed according to the Illumina TruSeq RNA library protocol outlined in "TruSeq RNA Sample Preparation Guide" (Part # 15008136; Rev. A; Nov 2010) from Genotypic Pvt. Ltd. In brief, mRNA was purified from 1 µg of intact total RNA using oligodT beads (TruSeq RNA Sample Preparation Kit, Illumina). The purified mRNA was fragmented for 2 minute at elevated temperature (94 °C) in the presence of divalent cations and reverse transcribed with Superscript II Reverse transcriptase by priming with random hexamers. Second strand cDNA was synthesized in the presence of DNA polymerase I and RnaseH. The cDNA was cleaned up using Agencourt Ampure XP SPRI beads (Beckman Coulter). Illumina adapters were ligated to the cDNA molecules after end repair and addition of 'A' base and SPRI cleanup was performed after ligation. The library was amplified using 8 cycles of PCR for 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 (Agilent).

3.2.4 De novo transcriptome assembly

Total 10.08 GB paired end raw data was generated with read length of 72 bp and primary QC check of the raw data was performed using the inbuilt tool SeqQC-V2.1. To obtain high quality clean read data for *De novo* assembly, the raw reads were filtered by discarding the reads containing adaptor sequence and poor quality raw reads (Phred score <20). The clean reads were first assembled into contigs using the Velvet_1.1.05 with an optimized hash length of 59. Assembled contigs were given as input for Oasis_0.2.01 to generate transcripts. The redundancy in the output transcripts of Oasis_0.2.01 was removed by using CD-HIT to generate unique unigenes.

3.2.5 Transcriptome annotation

To assign molecular function, biological processes and cellular components of unigenes, functional annotations were performed. ORFs were predicted in all six frames by Virtual Ribosome online program. The longest ORFs were selected for each unigenes and submitted to Pfam-A database to identify protein domain and architecture. Unigenes assigned with Pfam ID were used to perform BLAST2GO search against NCBI Nr database, SwissProt/Uniprot database, Protein Data Bank (PDB) with an E-value$\leq 10^{-5}$. The FASTA format of all the unigenes were submitted to KEGG database to assign KO (KEGG Orthology) number and generate KEGG pathways.

3.2.6 Sequence and phylogenetic analysis

The sequence analyses were carried out on Bioedit software (Ibis Biosciences, USA); alignment was done using the CLUSTALW2 program (www.ebi.ac.uk/Tools/msa/clustalw2/). Nucleotide sequences were translated using the ExPASy translate tool (http://web.expasy.org/translate/) and BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html) was used for marking identical and similar amino acid residues.To determine the evolutionary origin of *TPS*1, *TPS*2 and *TPS*3, phylogenetic tree was constructed. Multiple sequence alignment was performed using ClustalX 2.1 software (http://www.clustal.org/clustal2/) and the phylogenetic tree was reconstructed with MEGA6 [23] using the neighbour-joining method with 1000 bootstrap iterations.

3.2.7 Semi-quantitative and quantitative RT-PCR

The DNase-treated RNA (4 µg) was used for cDNA preparation in a 20 µL reaction using SuperScript^{TM} III reverse-transcriptase system (Thermo Scientific). For semi-quantitative PCR, 2 µg DNase-treated RNA, each of young and mature leaves were mixed together and cDNA was synthesized as described. Semi-quantitative RT-PCR (sqRT-PCR) was performed for all three *TPS*s in 10 µL reaction consisting of 5 µL of 2X Jumpstart readymix, 1 µL each of 10 µM gene specific forward and reverse primers, 1 µL of cDNA optimized with endogenous control (18S rRNA) and nuclease free water was added to volume of 10 µL. PCR was carried out for 30 cycles and the amplified products were run on 1% agarose gel. Quantitative RT-PCR (qRT-PCR) was performed by using SYBR Green chemistry. A typical reaction consisted of 5 µL

of SYBR Green master-mix, 0.5 µL each of forward and reverse primer (10 µM) and 1 µL of diluted cDNA (1:2) with nuclease-free water added to make up a volume of 10 µL. For qRT-PCR reactions, elongation factor (EF-1α) was used as an endogenous control and the reactions were carried out in triplicates for 3 biological replicates in the 7500 Fast Real Time PCR System (Thermo Scientific). Annealing temperature was kept at 58°C and cycling conditions were kept as per the manufacture's instruction. Primer sequences are as given in (Table 1).

Table 1: List of primer sequences

Primer Code	**Primer sequence**
TSP1 NdeIF	5'-AATACATATGATGGCAATCTGCAATTTCCCATCAACTCC-3'
TPS1 BamHIR	5'-AATAGGATCCCTAATGGAACAATTTAGTAGTGTTGACGAG-3'
RT-TPS1-F	5'-CTATGCTCATCGACAGTC-3'
RT-TPS1-R	5'-TCTGTTGTGCCATATAGATC-3'
RT-TPS2-F	5'-ATCACTCAACAATCATCACT-3'
RT-TPS2-R	5'-AAGTCTTCAATCAACTCCAA-3'
RT- TPS3-F	5'-TTCAACAAGAACTACAAGAGA-3'
RT- TPS3-R	5'-GCATTATCCTCTCATATCTGT-3'
EF-1α-F	5'-TGGTGTCATCAAGCCTGGTATGGT-3'
EF-1α-R	5'-ACTCATGGTGCATCTCAACGGACT-3'
18S rRNA F	5'-TCGAAACCTGCAAAGCAGACC-3'
18S rRNA R	5'-GATTCTGCAATTCACACCAAGTATCG-3'

3.2.8 Cloning and in-vivo expression analysis of *TPS*1

Putative *Ocimum* TPS encoding genes were identified using Blastx by comparing known TPS genes with de novo assembled RNA-Seq *Ocimum* transcriptome sequences. Coding sequence of *TPS*1 was obtained from *Ot* young leaf cDNA using full-length ORF primers (Table 1). *TPS*1 full-length ORF was cloned into the pRI 101-AN vector for over-expression in plants, using *Nde*I and *BamH*I restriction sites in forward and reverse primer (Table 1), respectively. After verifying the construct by sequencing, it was transformed into the *Agrobacterium tumefaciens* (GV3101) chemically competent cells and plated on Luria-agar plate containing 50µg/mL kanamycin and 25 µg/mL rifampcin. *A. tumefaciens* (GV3101) cells carrying over-expression construct were grown in 2mL LB media with the above mentioned antibiotics at 28 °C on rotary shaker set at 180 rpm for 2 days. This starter culture was used to inoculate 10 mL LB with the antibiotics and incubated over-night at 28 °C and 180 rpm. After incubation, cells were pelleted down by centrifugation at 10,000 rpm for 10 mins at 4 °C. The cells were washed with half strength MS media, pH 5.4 and pelleted again. This cell pellet was re-suspended in half strength MS media, pH 5.4 to bring it to an $O.D._{600nm}$ of 0.3-0.5. This suspension was used for syringe assisted infection of the adaxial surface of the leaves. Three plants each were taken for over-expression, empty vector and uninfected control. *A. tumefaciens* (GV3101) cells harboring empty pRI 101-AN vector was used as control. The plants were kept at 22 °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. The samples for volatile and real time analysis were collected after 4 and 8 days, and were processed as described in the earlier sections.

3.2.9 Statistical Analysis

GC-MS analyses of volatiles and gene expression analysis were performed with four biological and three technical replicates each. Data sets were represented as mean ± standard deviation. Significant differences between control and treated plants for metabolites and gene expression were determined using two-way ANOVA followed by Bonferroni's multiple comparisons test.

3.3 Results

3.3.1 Terpene diversity in different tissues across five *Ocimum* species

The terpene analysis indicated both quantitative (Figure 1) and qualitative (Figure 2, Table 2) differences in distribution of these compounds across different *Ocimum* species. We also analyzed the abundances of these metabolites in different tissues individually (Figure 3 and Figure 4 a-e). Ten metabolites were found to be common among different species, and were equally distributed between mono- and sesquiterpenes. These ten common terpenes among *Ocimum* species represents as specific chemical markers to identify these plants. In young leaves, linalool, ocimene, camphor and eucalyptol were the major monoterpenes in *Og*, *Ok* and *Oa*. Monoterpenes, ocimene and myrcene, were found to be present in all the species studied. Young leaves were observed to be rich sesquiterpenes in *Og* and *Ot*. Two species, *Ok* and *Oa* have comparable distribution of both mono- and sesquiterpenes. Among sesquiterpenes, β-caryophyllene was detected in all the *Ocimum* species. A similar trend of metabolite distribution was observed in other tissues (Figure 4a-e), where *Og* and *Ot* were found to be rich in sesquiterpenes, whereas in four different varieties *of Oba* I-IV, monoterpenes dominate the metabolite profile. This species-specific metabolite distribution was also evident from the cluster analysis, where species rich in either type of metabolite, incline to always cluster together. In all the different tissues analyzed, monoterpene rich *Oba* species and sesquiterpene rich species, *Og* and *Ot*, cluster together.

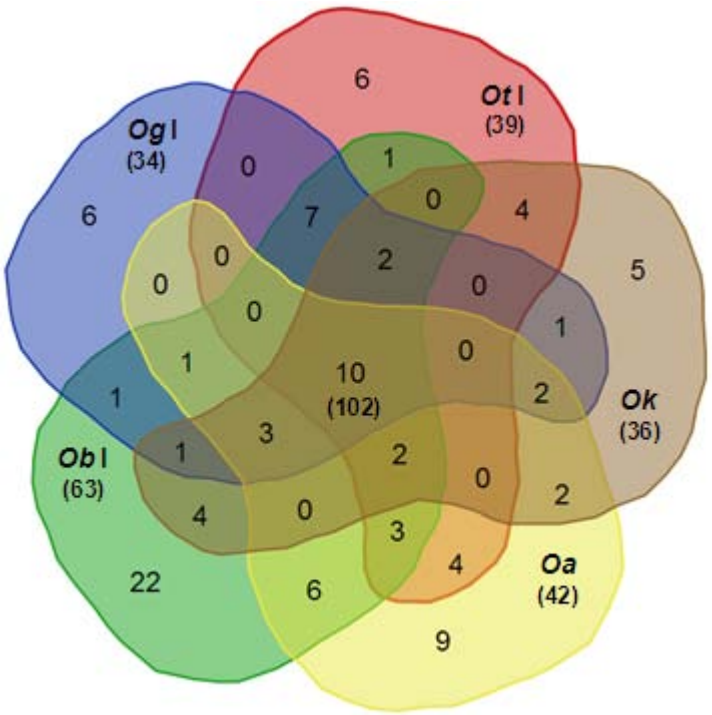


Figure 1: Venn diagram showing the unique and common metabolites from young leaves among five different species. Number in parentheses indicates the number of total metabolites identified in a particular species; the other numbers indicate metabolites that are unique and common between the species.

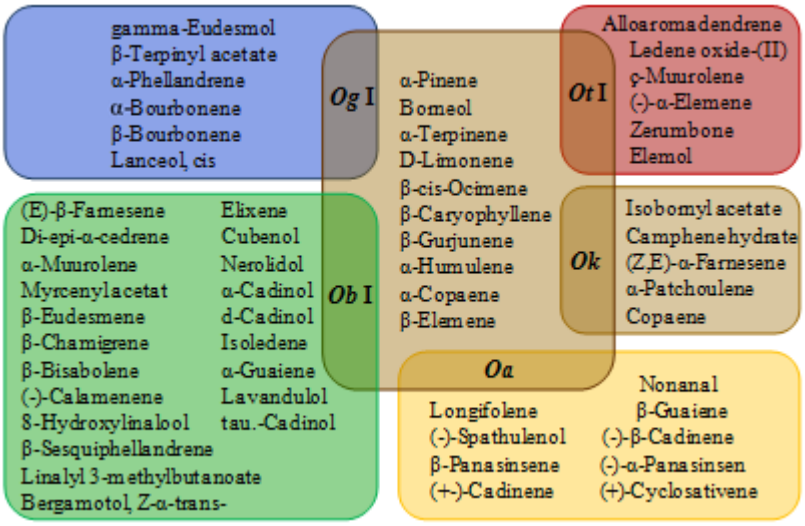


Figure 2: Venn diagram showing the qualitative distribution of metabolites among different species of *Ocimum*.

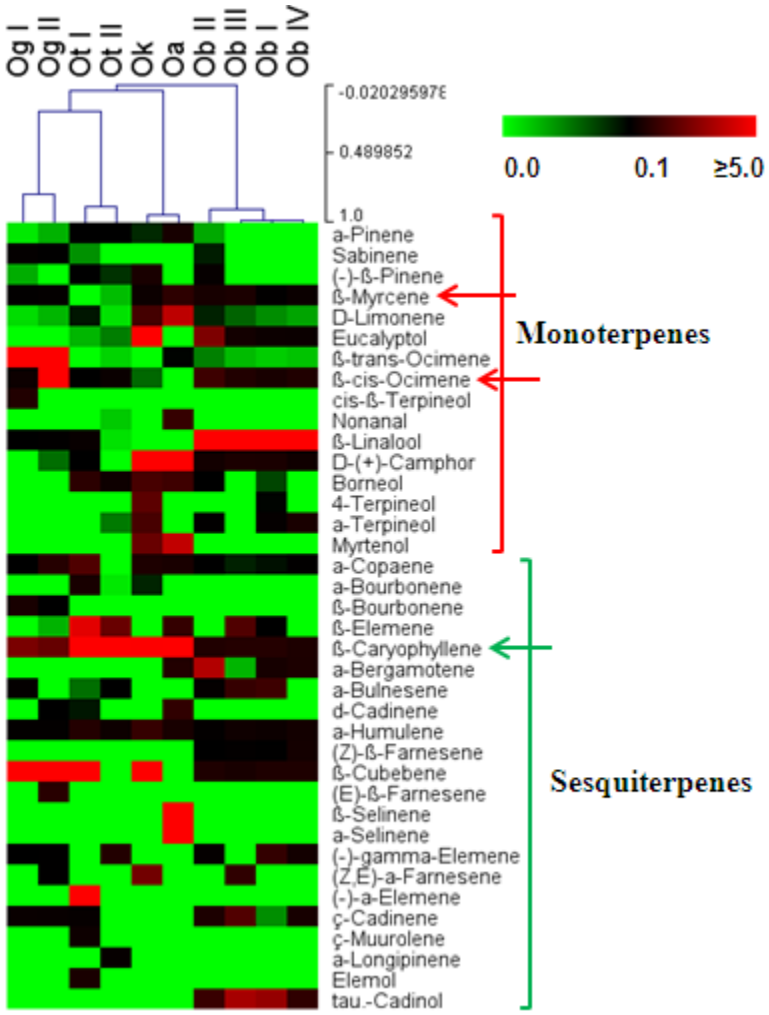
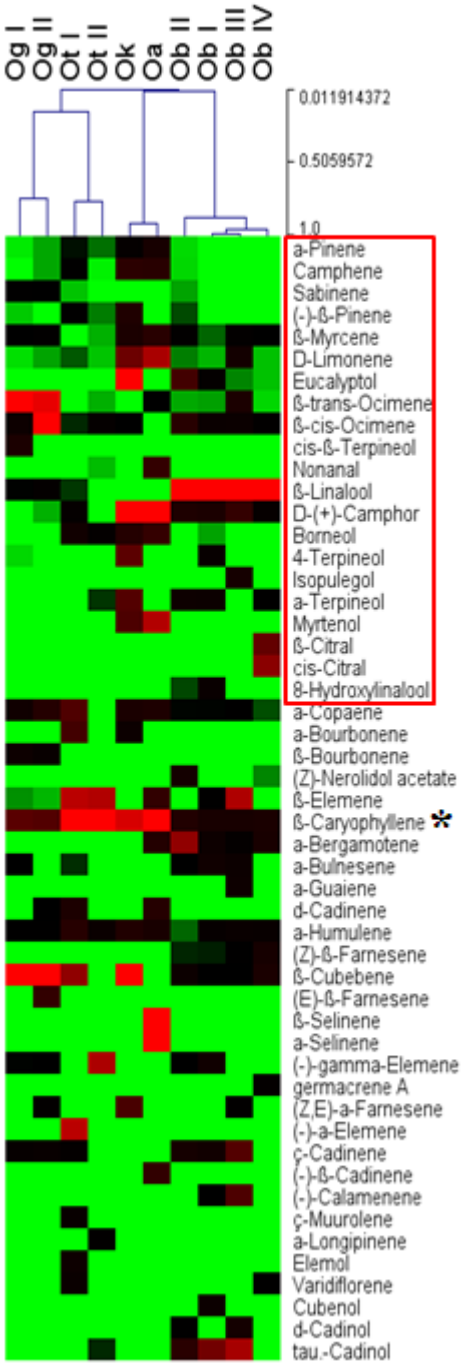


Figure 3: Heat map showing the distribution of top 10 metabolites from different *Ocimum* species in young leaves. Arrows indicate the metabolites that were present in all the species.



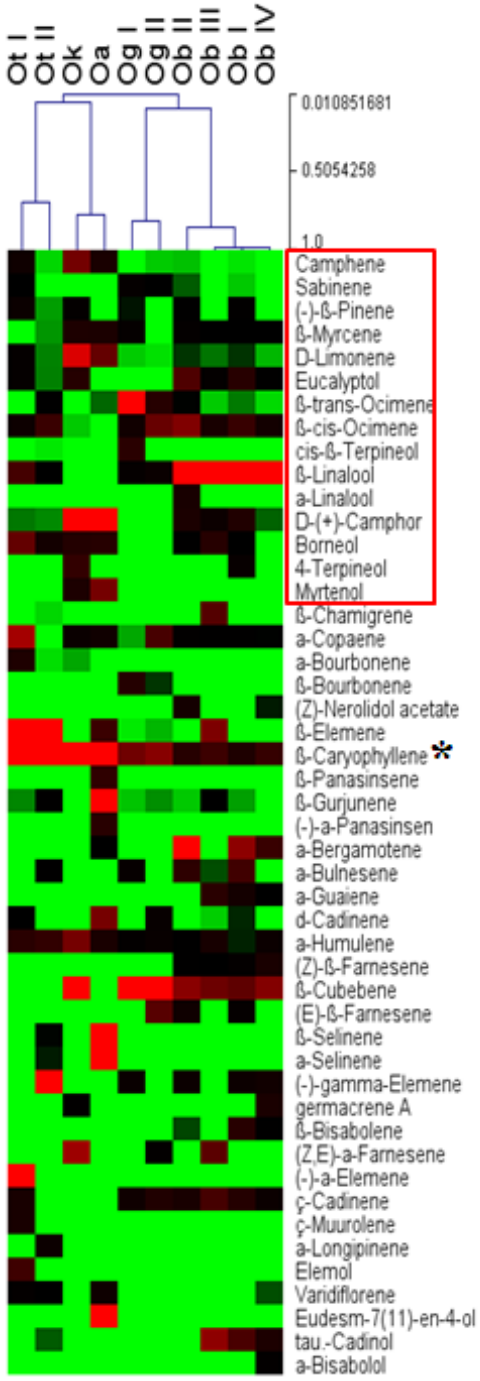


Figure 4a Mature Leaves

Figure 4b Inflorescence

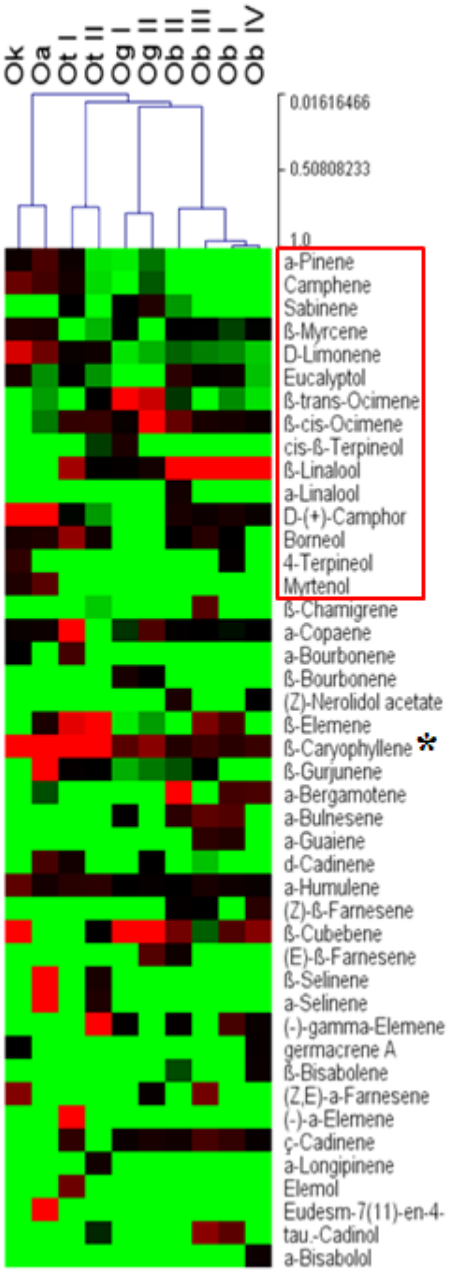


Figure 4c Flower

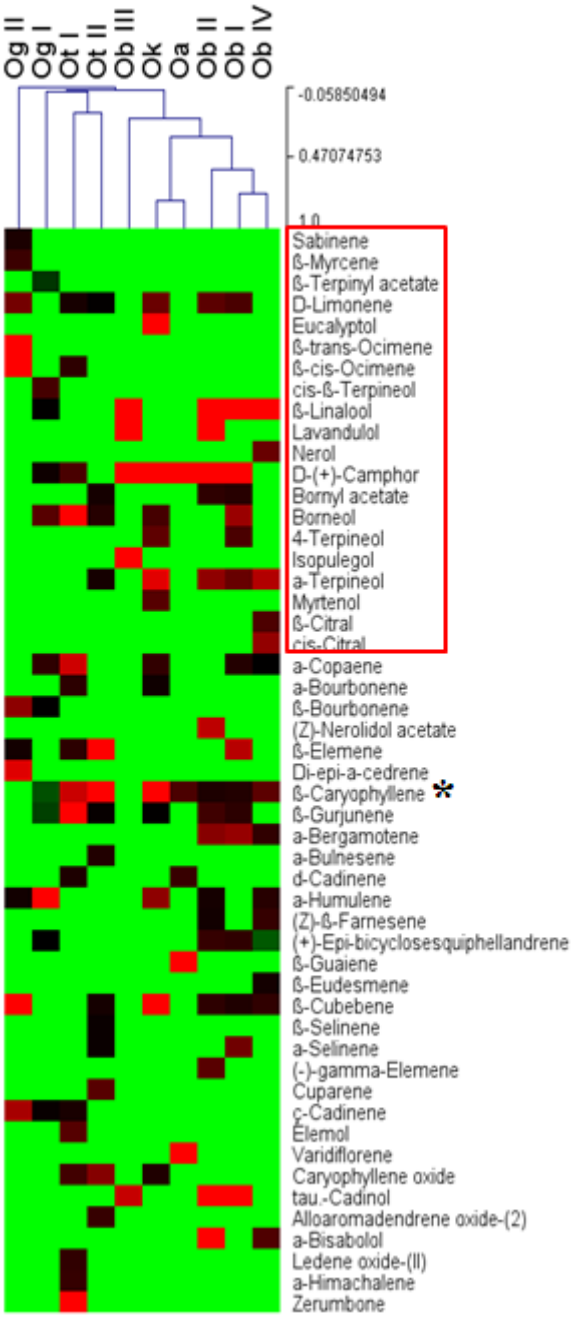


Figure 4d Stem

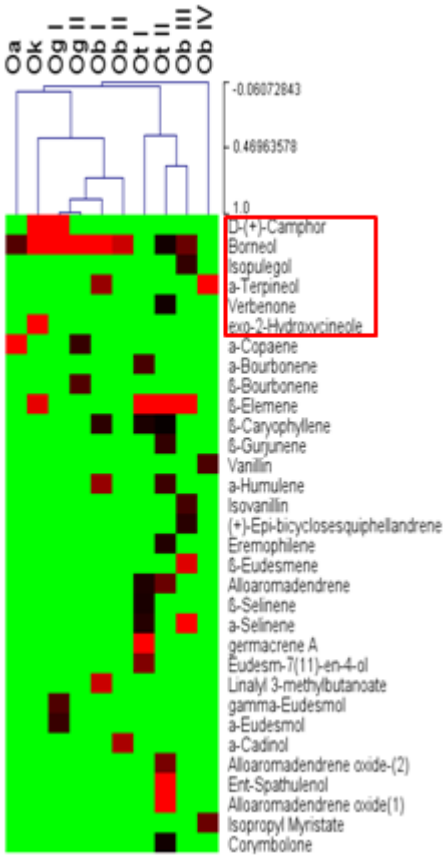


Figure 4e

Figure 4a-f: Heat map showing relative percentage of terpenoids from 10 *Ocimum* plants (a) Mature leaves (b) Inflorescence (c) Flower (d) Stem and (e) Root. Different species and varieties are hierarchically clustered based on their abundance diversity in mono- and sesquiterpenes. Monoterpenes are marked in red box. Asterisk represents the common metabolite present in all the species.

3.3.2 Transcriptome sequencing identifies 38 putative TPS genes in selected *Ocimum* species

1 µg of pooled total RNA from three *Ocimum* species was used for transcriptome sequencing (Fig. 5). A total of 40848829 (40.85 million) paired end reads, each of 72 bp, were generated by Illumina GA II platform. In this, 38.25 millions (93.62%) high quality reads were obtained with >20 phred score, reads of low quality were trimmed before further analysis. Velvet assembly generated a total of 220575 contigs with hash length of 55. These were used as input for Oases assembly to generate 253601 transcripts with average length of 255.9 bp. These transcripts were clustered to remove the redundancy using CD-HIT and generated 202081 clustered transcripts of varied lengths (Fig. 6). Blast annotation of these transcripts was carried over with proteins of eudicotyledons family and the unannotated transcripts were further annotated with ESTs of *Lamiids* family. CAP3 analysis on these transcripts, gave 31034 contigs and 25855 singlets. Pathway annotation was carried out by KAAS (KEGG Automatic Annotation Server). Out of the 56889 transcripts, only 4294 were assigned 1888 unique KO numbers. A web based server Virtual Ribosome identified 20302 transcripts to have an open reading frame (ORF) more than 100 amino acids (aa), whereas 88 were without any ORF. Pfam analysis was carried out on the peptide sequence of the transcripts>100aa, where Pfam ID was assigned to 15737 transcripts. In particular, we identified 38 terpene synthases, of which, 12 had similarities with monoterpene synthases, 13 were related to sesquiterpene synthases, 3 and 7 represented di- and triterpene synthases, respectively, and 3 putative prenyl transferases. In an attempt to understand the terpene accumulation in *Ocimum* species, three terpene synthase contigs were selected and designated as *TPS*1 (Contig 19414), *TPS*2 (Contig 13517) and *TPS*3 (Contig 9641).

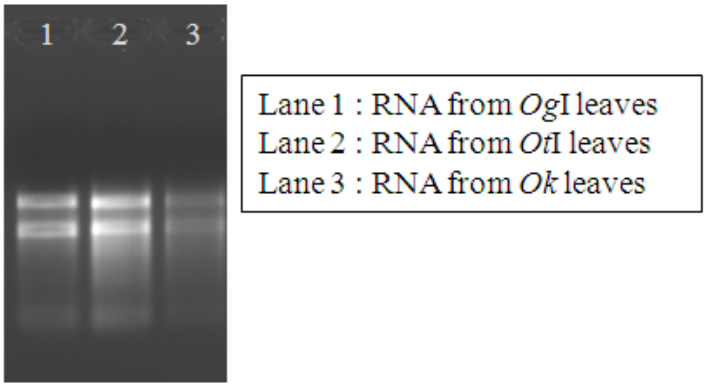


Figure 5: Total RNA from leaf tissue of three *Ocimum* species.

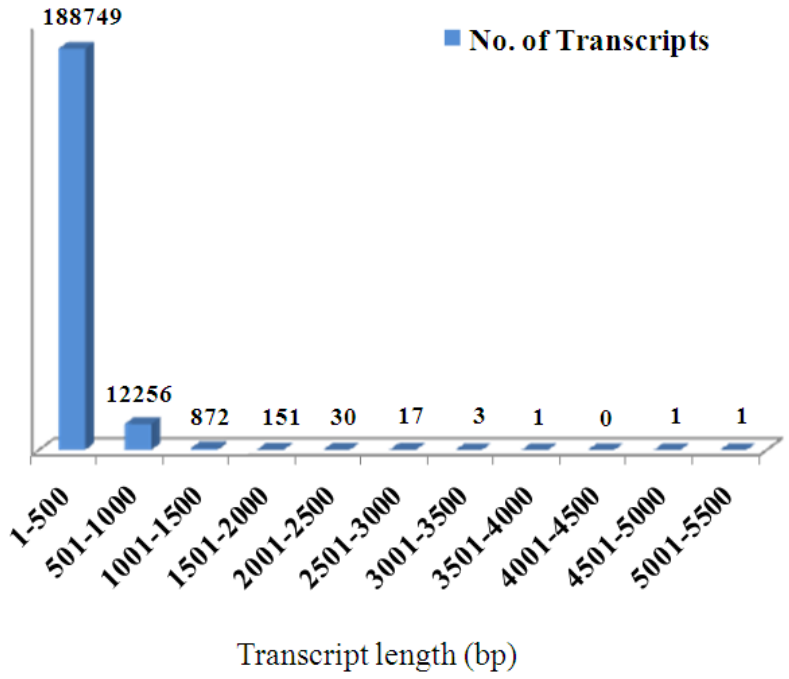


Figure 6: Bar graph indicating transcript length distribution.

3.3.3 Three *TPS* shares sequence similarity with other monoterpene synthases

The three TPS were cloned and sequenced (Figure 7). Upon sequence analysis of these three contigs, presence of conserved domains and other characteristics of TPS gene family was evident (Figure 8), including plastidial targeting sequence, which are characteristics of monoterpene synthases. *TPS*1 consists of 579 aa (63.7 kDa), *TPS*2 of 470 aa (51.7 kDa) and *TPS*3 of 580 aa (63.8 kDa). Upon phylogenetic analysis, *TPS*2 and 3 grouped together with other monoterpene synthases belonging to TPS-b family of terpene synthases. *TPS*1 grouped together with linalool/nerolidol synthase from *S. lycopersicum* and nerol synthase from *G. max*, both of which belonged to TPS-g family of terpene synthases (Figure 9). This family includes mono- and sesquiterpene synthase from angiosperms.

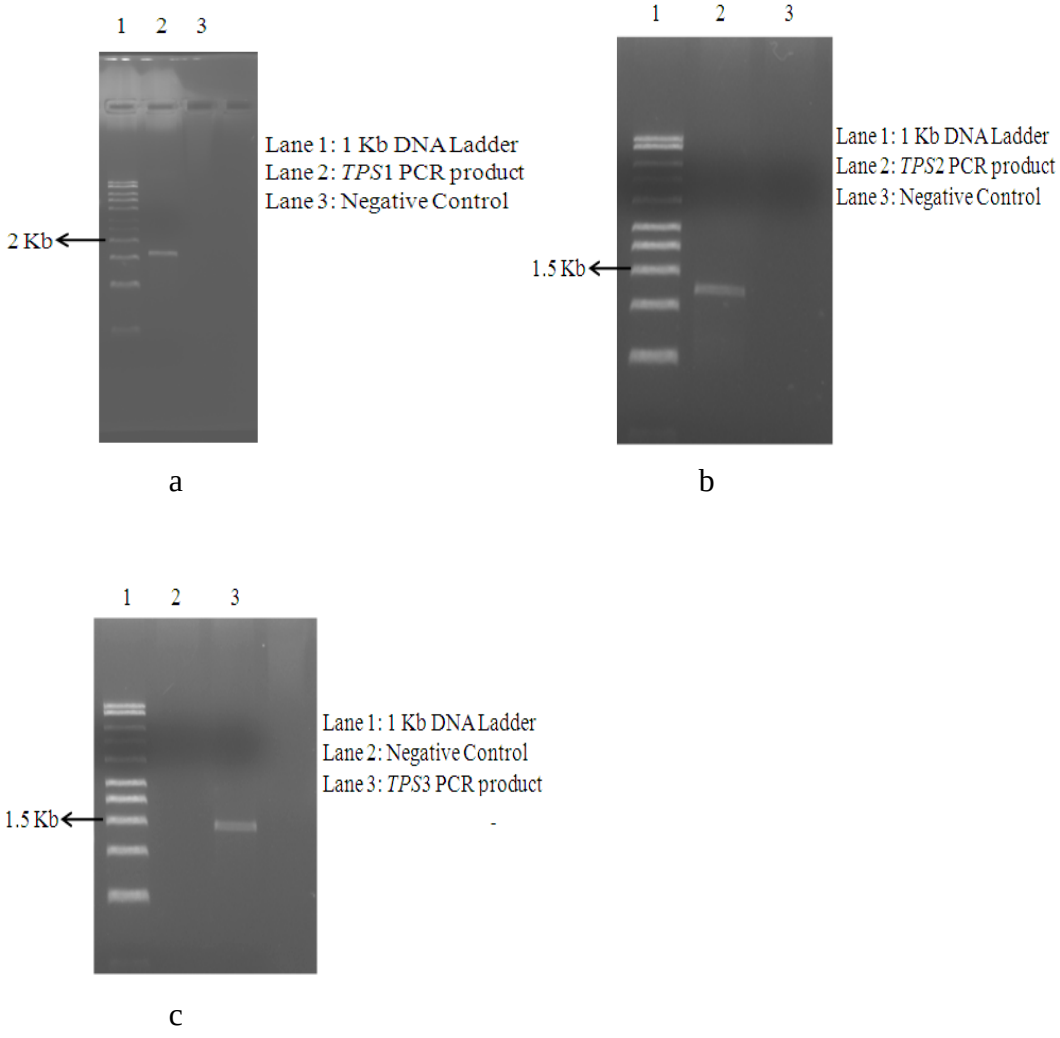


Figure 7: Full length PCR of a) *TPS*1, b) *TPS*2 and c) *TPS*3

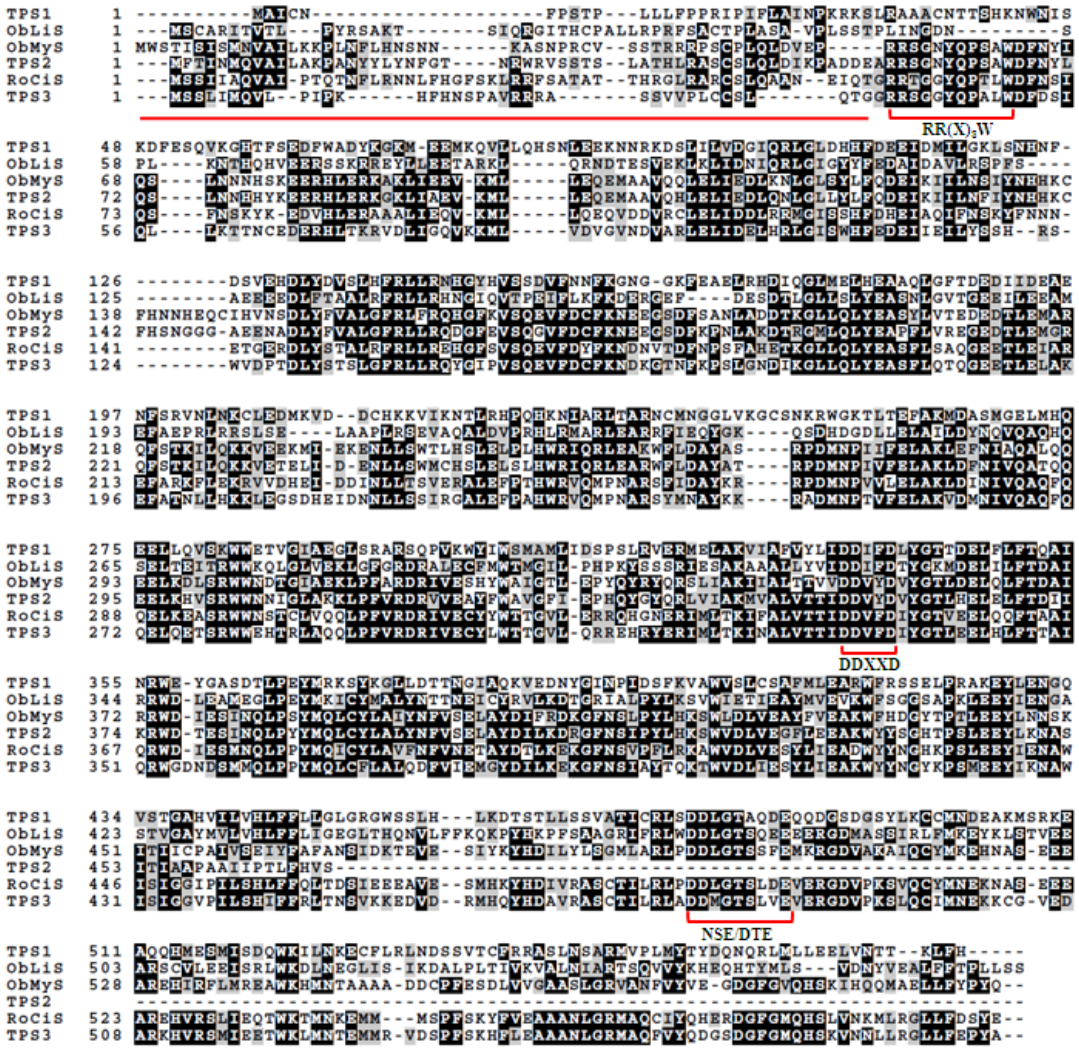


Figure 8: Sequence alignment of TPSs with other reported monoterpene synthases. Conserved regions are marked in red. *Ob*FS - *Ocimum basilicum* Fenchol synthase, *Ob*MyS - *Ocimum basilicum* Myrcene Synthase. *TPS*1, *TPS*2 and *TPS*3 from *Ocimum* species. *Ro*CiS - *Rosmarinus officinalis* Cineole Synthase. Conserved regions are marked in brackets and signal sequence indicated by line.

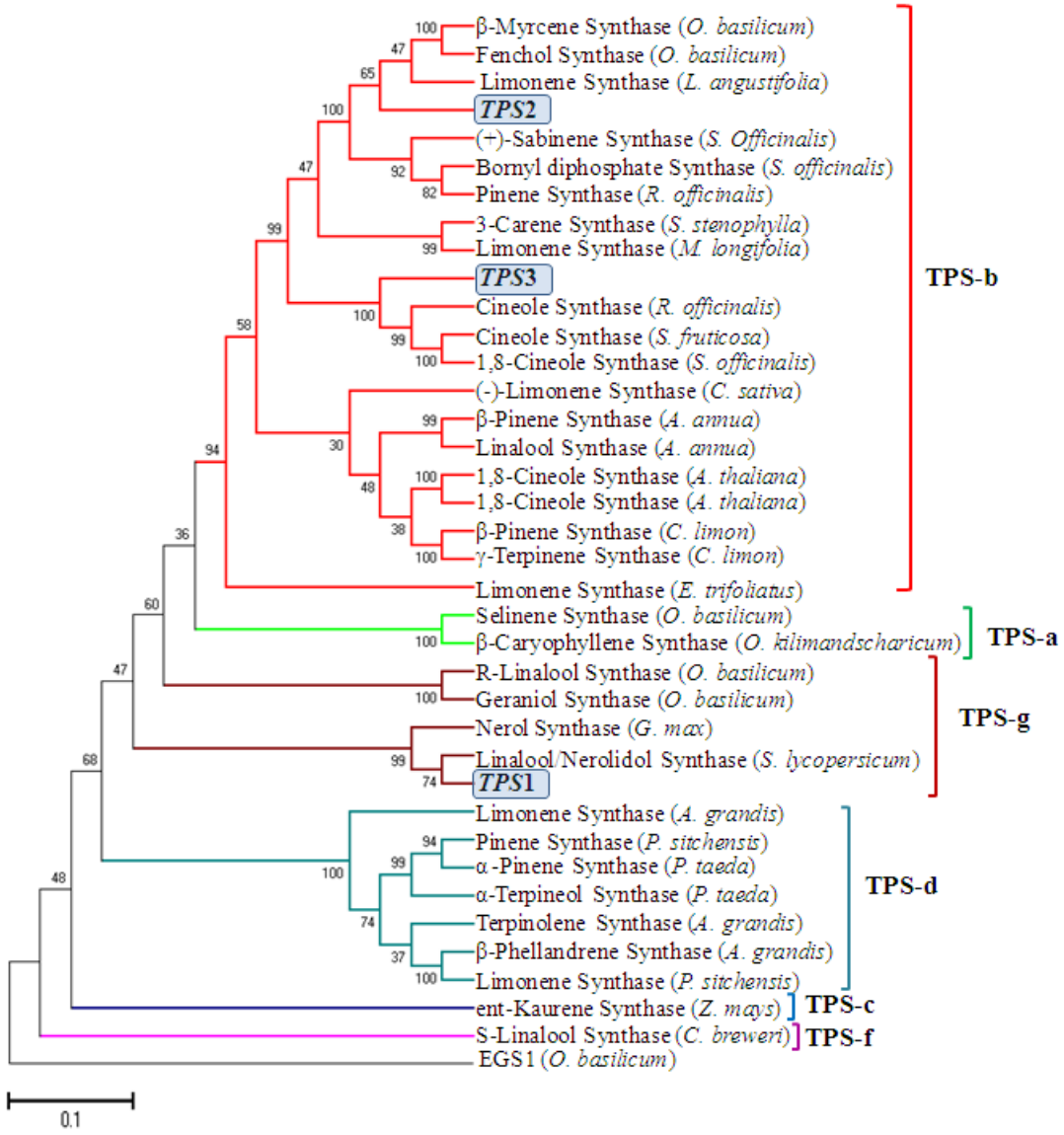


Figure 9: Phylogenetic analysis of *TPS*1, *TPS*2 and *TPS*3 (marked in blue box) with other reported TPS sequences from different classes (a- f). Neighbor joining tree was constructed using MEGA 6 with 1000 bootstrap iterations

3.3.4 Tissue specific expression analysis of three *TPS*s among different *Ocimum* species

Transcript abundance of these three *TPS*s was analyzed in leaves of candidate five *Ocimum* species, and significant tissue and species specific variations were observed. Across different tissues, in general, inflorescence represented highest expression levels of these contigs, with stem having the least. *TPS*1 transcript abundance was higher leaf tissue of in *Ot*, *Oba*II and *Oa*, while *TPS*2 expression was higher in all the

five species analyzed (Figure 10). In qRT-PCR (Figure 11), except for *Oba* species, *TPS*2 had very high expression levels in all tissues of the *Ocimum* species that were analyzed. In case of *Ot*I and II, *TPS*1 and 2 had similar expression profiles. Apart from *Ot*, *TPS*1 was also found to be the most abundant contig in *Ob*II, whereas in *Og* and *Ok* it had poor expression profiles. *TPS*3, on the other hand, was observed to be abundant only in *Ok*, in both semi-quantitative and qRT-PCR.

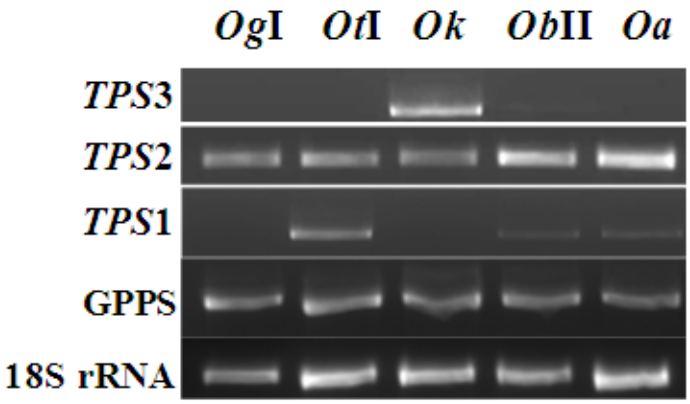


Figure 10: Expression analysis of GPPS, *TPS*1, *TPS*2 and *TPS*3 by sqRT-PCR using 18S rRNA as reference gene in young leave tissue of five *Ocimum* species.

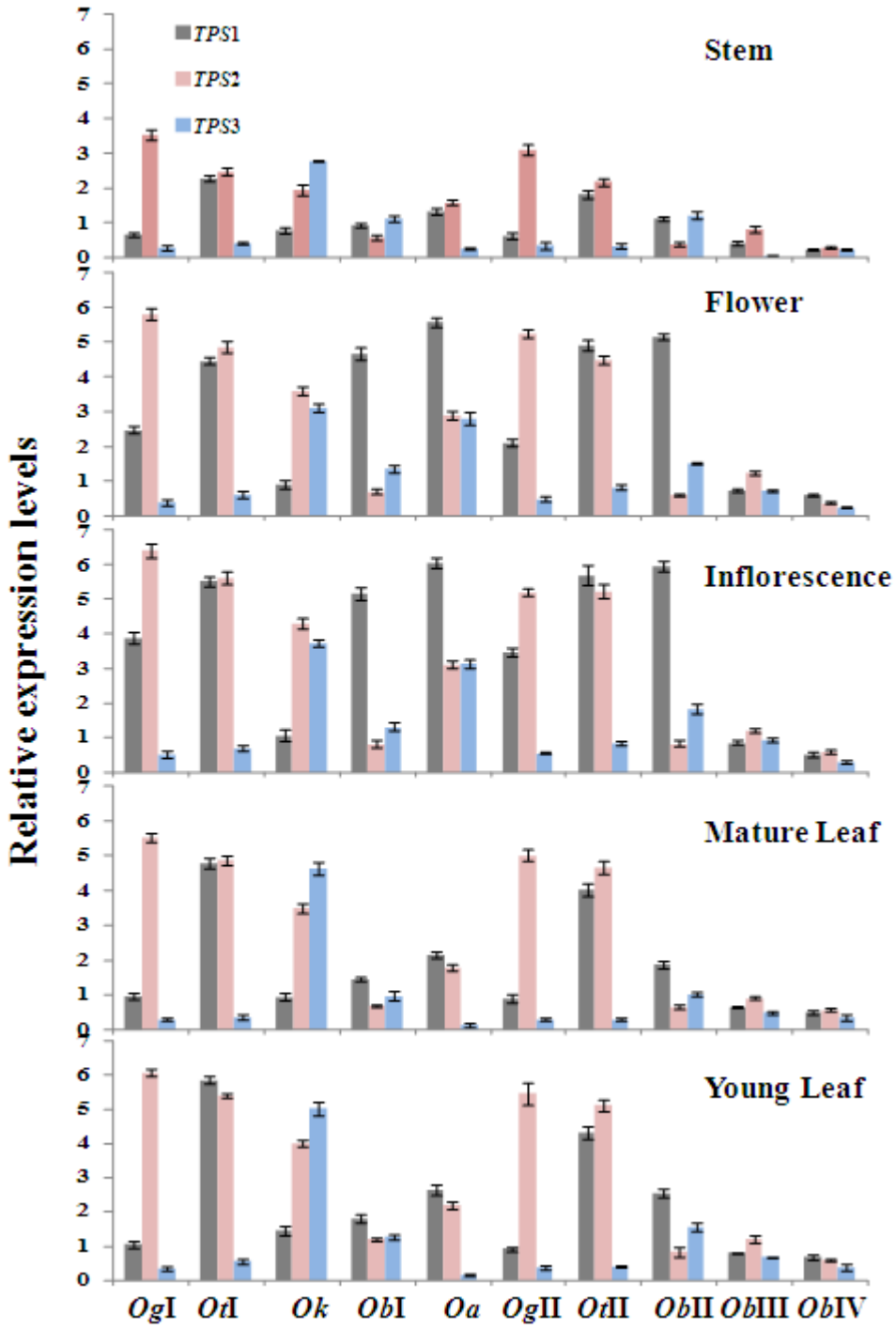


Figure 11: qRT-PCR analysis of *TPS*1, *TPS*2 and *TPS*3 with Actin as endogenous reference in different tissues of five species of *Ocimum*. *Og- Ocimum gratissimum, Ot- Ocimum tenuiflorum, Ok- Ocimum kilimandscharicum, Ob- Ocimum basilicum, Oa- Ocimum americanum.*

3.3.5 Transient over- expression of *TPS*1 indicates *cis*-β-terpineol synthase activity

In order to analyze gene function of *TPS*1, transient over-expression analysis was carried on *Ok* plants. Accumulation of metabolites and transcript upon over-expression of *TPS*1 was estimated by comparing them with that of control and empty vector infected plants (Figure 12 and 13). In qRT-PCR analysis, two-fold increase in *TPS*1 was noted, in both local and systemic tissues, which was reflected in 7% increase in accumulation of cis-β-terpineol in 8DAI (Figure 14). There was no significant difference in accumulation of any other metabolites upon over-expression of *TPS*1 in *Ok* leaves. As compared to that of metabolite analysis, transcript abundance decreased at 8 DAI, an observation which was consistent with the reports suggesting instability of non-integrated T-DNA.

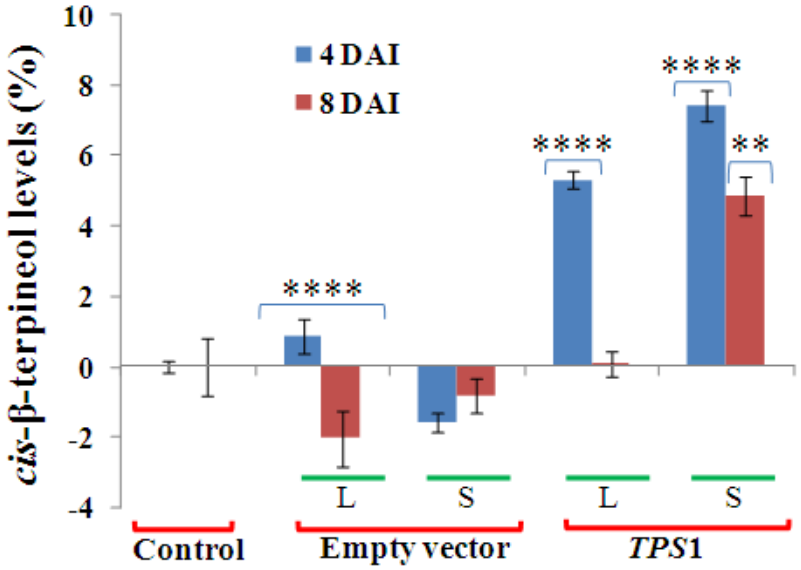


Figure 12: cis-β-terpineol levels upon transient over-expression of *TPS*1 in *Ok*. Experiment was performed with 4 biological and 3 technical replicates (L-local tissue, S- systemic tissue). Metabolite level was checked in leaf tissues collected after 4 and 8 days after infection (DAI). Data expressed in terms of relative area % and bars represents the SD values. Two way ANOVA followed by Bonferroni's multiple comparisons suggested significant difference between the data at $p<0.001$ (indicated as '***'), $p<0.01$ (indicated as '**').

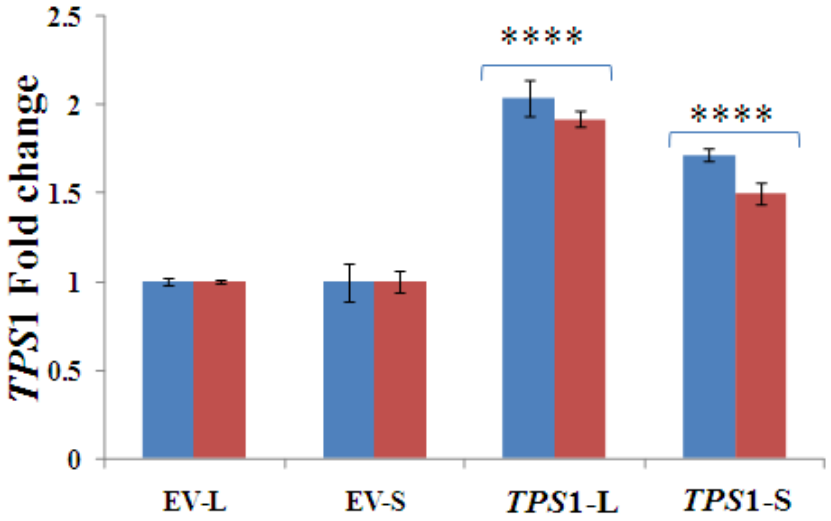


Figure 13: *TPS*1 expression levels determined by qRT-PCR taking actin as endogenous control. The values are expressed in comparison to the transcript levels in empty vector. Two way ANOVA followed by Bonferroni's multiple comparisons suggested significant difference between the data at $p<0.001$ (indicated as '***'), $p<0.01$ (indicated as '**').

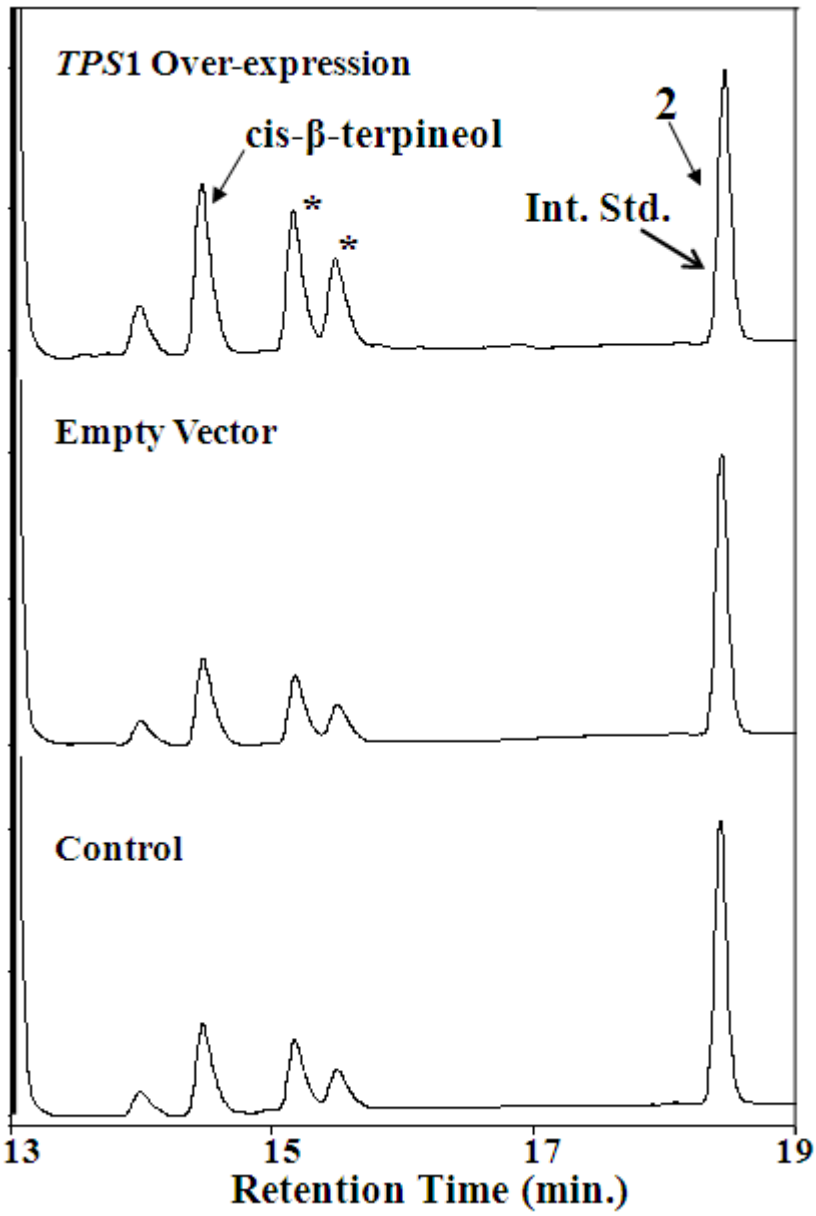


Figure 14: GC-chromatograms of metabolite analysis of control, empty vector infected and *TPS*1 infected plants. 1- cis-β-terpineol and 2- Linalyl acetate (internal standard). Asterisk refers to the hydrocarbon dodecane observed in the GC-MS runs.

3.4 Discussion

Ocimum, a member of the *Lamiaceae*, is known to produce various secondary metabolites, in different proportions across different species. The predominant among these compounds are terpenes and phenylpropanoids, with β-caryophyllene, limonene, myrcene and camphor as the major terpenes present. In this chapter, we have analyzed the terpene diversity present in six different tissues of five *Ocimum* species. It was observed that there exists species specific distribution of terpenes, with *Og* and *Ok* species rich in monoterpenes and the *Oba* species in sesquiterpenes. Moreover, some of the terpenes like ocimene, myrcene and β-caryophyllene were found to be present in all the tissues that were analyzed. Transcriptome profiling was carried out on leaf, inflorescence and stem tissue from *Og*, *Ot* and *Ok* species. Transcripts were annotated for terpene biosynthesis pathway and 38 putative terpene synthases were identified. Out of these, three were selected (*TPS*1, 2 and 3) for further analysis and gene function identification. Upon sequence analysis of these *TPS*s with other reported sequences from NCBI database, presence of different conserved domains was observed. On the basis of amino acid sequence and protein function, terpene syntheses have been organized into sub-families, which includes, three angiosperm specific family (TPS-a, TPS-b and TPS-g), a gymnosperm specific subfamily (TPS-d), TPS-c is most conserved among the land plants and TPS-e and f is conserved among the vascular plants [24,25,26]. All three *TPS*s showed the presence of conserved features of plant terpene synthases like aspartate-rich region (DDxxD) which is involved in catalysis [27], xDx_6E' motif for metal cofactor binding [28]. A RR(X8) W motif is present in the N-terminal region of TPS2 and 3, which participates in the ionization of the substrate and is characteristic of TPS-a and b subfamilies [29,30]. Phylogenetic analysis indicated *TPS*1 to belong to TPS-g and *TPS*2 and 3 to TPS-b sub-family, both these family include monoterpenes from different plant species. Transcript analysis of these three *TPS*s was also performed using semi-quantitative and quantitative real time PCR, and tissue and species specific expression pattern of *TPS* expression was observed. *TPS*2 was found to have higher expression levels in *Og* and *Ot* species, and *TPS*1 in *Ot* species.*TPS*3 expression levels were higher in *Ok* species, and apart from this, its expression level was comparable to other two *TPS*s, only in inflorescence of *Ot* species. Transient expression using *Agrobacterium* is frequently used as it allows rapid analysis of gene function. Several reports are available where

transient expression mediated by *Agrobacterium* was used for studying gene expression [31], gene silencing [32] and protein-protein interaction [33]. Gene function analysis of *TPS*1 was done by transient over-expression using Agrobacterium and the study indicated it to be *cis*-β-terpineol synthase.

3.5 Conclusion

Terpene profiling of five different *Ocimum* species indicated significant variations across different tissues and species. Transcriptome analysis led to detection of 38 unique *TPS* transcripts. Sequence and phylogenetic analysis of three full-length transcripts identified them as monoterpene syntheses. Transcript abundance in five tissues across *Ocimum* species revealed higher abundance of *TPS*2 in all the species, whereas *TPS*1 and *TPS*2 showed species-specific abundances. Agro-infiltration based transient over-expression assay of *TPS*1 indicated it to be *cis*-β-terpineol synthase. Overall, this study provided detailed terpene profiling across different *Ocimum* species and established presence of *cis*-β-terpineol synthase (*TPS*1).

3.6 Appendix: Amino acid sequence of *TPS*1, *TPS*2 and *TPS*3

>*TPS*1

MAICNFPSTPLLLFPPRIPIFLAINPKRKSLRAAACNTTSHKNWNISKDFESQVK
GHTFSEDFWADYKGKMEEMKQVLLQHSNLEEKNNRKDSLILVDGIQRLGLD
HHFDEEIDMILGKLSNHNFDSVEHDLYDVSLHFRLLRNHGYHVSSDVFNNFK
GNGGKFEAELRHDIQGLMELHEAAQLGFTDEDIIDEAENFSRVNLNKCLEDM
KVDDCHKKVIKNTLRHPQHKNIARLTARNCMNGGLVKGCSNKRWGKTLTEF
AKMDASMGELMHQEELLQVSKWWETVGIAEGLSRARSQPVKWYIWSMAM
LIDSPSLRVERMELAKVIAFVYLIDDIFDLYGTTDELFLFTQAINRWEYGASDT
LPEYMRKSYKGLLDTTNGIAQKVEDNYGINPIDSFKVAWVSLCSAFMLEARW
FRSSELPRAKEYLENGQVSTGAHVILVHLFFLLGLGRGWSSLHLKDTSTLLSS
VATICRLSDDLGTAQDEQQDGSDGSYLKCCMNDEAKMSRKEAQQHMESMIS
DQWKILNKECFLRLNDSSVTCFRRASLNSARMVPLMYTYDQNQRLMLLEEL
VNTTKLFH

>*TPS*2

MFTINMQVAILAKPANYYLYNFGTNRWRVSSTSLATHLRASCSLQLDIKPAD
DEARRSNYQPSAWDFNYLQSLNNHHYKEERHLERKGKLIAEVKMLLEQEMA
AVQHLELIEDLQNLGLLYLFQDEIKIILNFIYNHHKCFHSNGGGAEENADLYFV
ALGFRLLRQDGFEVSQGVFDCFKNEEGSDFKPNLAKDTRGMLQLYEAPFLVR
EGEDTLEMGRQFSTKILQKKVETELIDENLLSWMCHSLELSLHWRIQRLEAR
WFLDAYATRPDMNPIVFELAKLDFNIVQATQQEELKHVSRWWNNIGLAKKL
PFVRDRVVEAYFWAVGFIEPHQYGYQRLVIAKMVALVTTIDDVYDVYGTLH
ELELFTDIIKRWDTESINQLPYYMQLCYLALYNFVSELAYDILKDRGFNSIPYL
HKSWVDLVEGFLEEAKWYYSGHTPSLEEYLKNASITIAAPAAIIPTLFHVS

>*TPS*3

MSSLIMQVLPIPKHFHNSPAVRRRASSVVPLCCSLQTGGRRSGGYQPALWDF
DSIQLLKTTNCEDERHLTKRVDLIGQVKKMLVDVGVNDVARLELIDELHRLG
ISWHFEDEIIEILYSSHRSWVDPTDLYSTSLGFRLLRQYGIPVSQEVFDCFKND
KGTNFKPSLGNDIKGLLQLYEASFLQTQGEETLELAKEFATNLLHKKLEGSDH
EIDNNLLSSIRGALEFPAHWRVQMPNARSYMNAYKKRADMNPTVFELAKVD
MNIVQAQFQQELQETSRWWEHTRLAQQLPFVRDRIVECYLWTTGVLQRREH
RYERIMLTKINALVTTIDDVFDIYGTLEELHLFTTAIQRWGDNDSMMQLPPYM
QLCFLALQDFVIEMGYDILKEKGFNSIAYTQKTWVDLIESYLIEAKWYYNGY

KPSMEEYIKNAWISIGGVPILSHIFFRLTNSVKKEDVDRMHQYHDAVRASCTI
LRLADDMGTSLVEVERGDVPKSLQCIMNEKKCGVEDARKHVRSMIEETWKL
MNTEMMRVDSPFSKHFLEAAANLGRMAQFVYQDGSDGFGMQHSKVNNLLR
GLLFEPYA

3.7 References

[1] Dixon RA, Plant natural products: the molecular genetic basis of biosynthetic pathway, Curr. Opin. Biotech. 10 (1999) 192-197.

[2] McGarvey DJ and Croteau R, Terpenoid metabolism, Plant Cell 7 (1995) 1015-26.

[3] Bach TJ, Some new aspects of isoprenoid biosynthesis in plants—A review, Lipids. 30 (1995) 191-202.

[4] Rohmer M, Knani M, Simonin P, et. al., Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate, Biochem. J 295 (1993) 517-524.

[5] Sprenger GA, Schorken U, Wiegert T, et. al., Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol, Proc. Natl. Acad. Sci. USA 94 (1997) 12857-12862.

[6] Eisenreich W, Schwarz M, Cartayrade A, et. al., The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms, Chem. Biol. 5 (1998) R221-R233.

[7] Burke C and Croteau R, Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate, J. Biol Chem 277 (2002) 3141-3149.

[8] Thiel R and Adam KP, An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants, Phytochem. 59 (2002) 269-274.

[9] Cane DE, Isoprenoid biosynthesis: Stereochemistry of the cyclization of allylic pyrophosphates, Accounts Chem. Res. 18 (1985) 220-226.

[10] Croteau R, Biosynthesis and catabolism of monoterpenoids, Chem. Rev. 87 (1987) 929-954.

[11] Cane DE, Enzymic formation of sesquiterpenes, Chem. Rev. 90 (1990) 1089-1103.

[12] Lesburg CA, Caruthers JM, Paschall CM, et. al., Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism, Curr. Opin. Struc. Biol. 8 (1998) 695-703.

[13] Davis EM and Croteau R, Cyclization Enzymes in the biosynthesis of monoterpenes, sesquiterpenes and diterpenes, Springer, Vol. 209 (2000) p 53-95.

[14] Poulter CD, Farnesyl diphosphate synthase: A paradigm for understanding structure and function relationships in E-polyprenyl diphosphate synthases, Phytochem. Rev. 5 (2006)17-26.

[15] Thulasiram HV and Poulter CD, Farnesyl Diphosphate Synthase: The art of compromise between substrate selectivity and stereo-selectivity, J. Am. Chem. Soc. 128 (2006)15819-15823.

[16] Tantillo DJ, Biosynthesis via carbocations: Theoretical studies on terpene formation, Nat. Prod. Rep. 28 (2011)1035-1053.

[17] Jennewein S and Croteau R, Taxol: biosynthesis, molecular genetics, and biotechnological applications. Appl. Microbiol. Biot. 57 (2001)13-19.

[18] Rodriguez-Concepion M, The MEP Pathway: A new target for the development of herbicides, antibiotics and anti-malarial drugs, Curr. Pharm. Design. 10 (2004) 2391-2400.

[19] Garcia R, Alves ESS, Santos MP, et. al., Antimicrobial activity and potential use of monoterpenes as tropical fruits preservatives, Braz. J. Microbiol. 39 (2008) 163-168.

[20] Hampel D, Mosandl A, and Wust M, et. al., Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways, Phytochem. 66 (2005) 305-311.

[21] Ramesha HJ, Anand A, Beedkar SD, et. al., Functional characterization and transient expression manipulation of a new sesquiterpene synthase involved in β-caryophyllene accumulation in *Ocimum*, Biochem. Biophys. Res. Commun. 473 (2016) 265-271.

[22] Anand A, Ramesha HJ, Beedkar SD, et al., Comparative functional characterization of eugenol synthase from four different *Ocimum* species: Implications on eugenol accumulation, BBA-Proteins Proteom. 1864 (2016) 1539-1547.

[23] Tamura K, Stecher G, Peterson D, et. al., MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0., Mol. Biol. Evol. 30 (2013) 2725–2729.

[24] Bohlmann J, Crock J, Jetter R, et al., Terpenoid based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible

(*E*)- α bisabolene synthase from grand fir (*Abies grandis*), Proc. Natl. Acad. Sci. USA 95 (1998) 6756–6761.

[25] Chen F, Tholl D, Bohlmann J, et. al., The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom, Plant J. 66 (2011) 212-229.

[26] Benabdelkader T, Guitton Y, Pasquier B, et al., Functional characterization of terpene synthases and chemotypic variation in three lavender species of section stoechas, Plant Physiol. 153 (2015) 43–57.

[27] Rynkiewicz MJ, Cane DE and Christianson DW, Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade, Proc. Natl. Acad. Sci. USA 98 (2001) 13543–13548.

[28] Degenhardt J, Koellner TG and Gershenzon J, Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, Phytochem. 70 (2009) 1621–1637.

[29] Williams DC, McGarvey DJ, Katahira EJ, et. al., Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair, Biochem. 37 (1998) 12213−12220.

[30] Martin DM and Bohlmann J, Identification of *Vitis vinifera* (−)-α-terpineol synthase by *in silico* screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase, Phytochem. 65 (2004) 1223−1229.

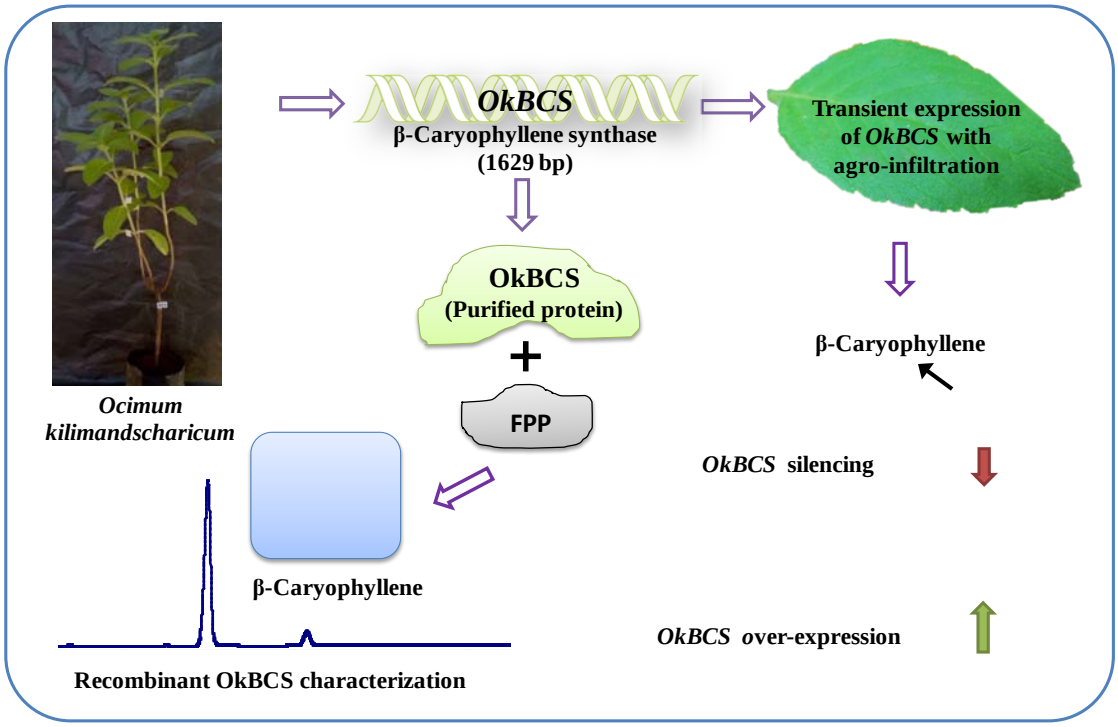
[31] Kapila J, Rycke RD and Montagu VM, An *Agrobacterium*-mediated transient gene expression system for intact leaves, Plant Sci. 122 (1997) 101–108.

[32] Schöb H, Kunz C and Meins F, Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing, Mol. Gen. Genet. 256 (1997) 581–585.

[33] Ferrando A, Koncz-Kálmán Z, Farràs R, et. al., Detection of in vivo protein interactions between Snf1-related kinase subunits with intron-tagged epitope-labelling in plants cells, Nucleic Acids Res. 29 (2001) 3685–3693.

Chapter 4

Functional Characterization of β-Caryophyllene Synthase – a Sesquiterpene Cyclase



Chapter 4

Functional Characterization of β-Caryophyllene Synthase – a Sesquiterpene Cyclase

The genus *Ocimum* has a unique blend of diverse secondary metabolites, with major proportion of terpenoids including mono- and sesqui-terpenes. Although, β-Caryophyllene, a bicyclic sesqui-terpene, is one of the major terpene found in *Ocimum* species and known to possess several biological activities, not much is known about its biosynthesis in *Ocimum*. Here, we describe isolation and characterization of β-caryophyllene synthase gene from *Ocimum kilimandscharicum* Gürke (*OkBCS*- GenBank accession no. KP226502). The open reading frame of 1,629 bp encoded a protein of 542 amino acids with molecular mass of 63.6kDa and pI value of 5.66. The deduced amino acid sequence revealed 50-70% similarity with known sesqui-terpene synthases from angiosperms. Recombinant OkBCS converted farnesyl diphosphate to β-caryophyllene as a major product (94%) and 6% α-humulene. Expression variation of *OkBCS* well corroborated with β-caryophyllene levels in different tissues from five *Ocimum* species. *OkBCS* transcript revealed higher expression in leaves and flowers. Further, agro-infiltration based transient expression manipulation with *OkBCS* over-expression and silencing confirmed its role in β-caryophyllene biosynthesis. These findings may potentially be further utilized to improve plant defense against insect pests.

4.1 Introduction

Terpenes are the largest group of natural products, present ubiquitously with immense diversity in their structure and function [1]. They perform versatile functions including communication and defense in plants [2]. Several terpenes that offer distinct fragrance to flower are crucial in attracting many insects and other pollinators [3]. Plants use different defense strategies for their survival and one of them is to recruit natural enemies of herbivores using induced volatiles. Among herbivore-induced volatiles, terpenoids that comprise mono- and sesqui-terpenes, are the most commonly utilized compounds as tactical arsenal [4]. Like other members of the *Lamiaceae* family, *Ocimum* plant synthesizes and accumulates these volatile compounds in the secretory capitate and peltate trichomes which are located on the surface of aerial parts of plants [5]. Mono- and sesqui-terpenes are the main constituents of volatile components, which impart them unlimited medicinal properties [6] as well as a characteristic flavor and taste of *Ocimum* [7]. Several species from the genus *Ocimum* are known to possess insecticidal and other important bioactive properties [8,9].

β-caryophyllene, a bicyclic sesqui-terpene, is widely distributed in plant kingdom including *Ocimum* species [10]. Like most terpenes, it contributes unique aroma in essential oils from numerous plants. β-caryophyllene might be effective for treating cancer, anxiety and depression [11,12] and FDA (Food and Drug Administration, USA) has approved it as food additive. Additionally due to insecticidal nature, it could be potentially useful in plant defense [9,13]. Though terpenoid biosynthesis from genus *Ocimum* has received considerable attention, very little is known on β-caryophyllene biosynthesis in *Ocimum.* Molecular characterization of β-caryophyllene synthase from *Ocimum kilimandscharicum* Gürke (*OkBCS*) was carried out. Correlation of *OkBCS* transcript levels with metabolic variations in different tissues from five *Ocimum* species was performed. Further, transient manipulation of *OkBCS* expression in terms of silencing and over expression validated its role in β-caryophyllene biosynthesis and accumulation.

4.2 Materials and Methods

4.2.1 Plant material

Authenticated *O. kilimandscharicum* Gürke (*Ok*) were grown in a greenhouse at 25-28ºC with ~35-40% humidity and 16 h light and 8 h dark periods. Plant samples of all the species were deposited in the herbarium of Botanical Survey of India, Pune as voucher specimens. All the samples were collected from 2-3 months old plants, immediately frozen in liquid nitrogen and stored at -80 ºC until further use. Fresh tissues were used for metabolite extraction and analysis.

4.2.2 Reagents

All chemicals were purchased from Sigma-Aldrich (Sigma Chemical Co., USA), unless stated. The TA cloning kit with pCR 2.1 vector, Zero Blunt Vector, pET102 directional cloning kit, PCR purification kit, gel extraction kit and Accuprime proof reading polymerase were purchased from Thermo Scientific (USA).

4.2.3 Isolation of β-caryophyllene synthase *(BCS)* from *Ocimum kilimandscharicum*

Putative *Ocimum* terpene synthase (TPS) encoding genes were identified using Blastx by comparing known *TPS* with de novo assembled *Ocimum* transcriptome. Total RNA was isolated from leaves using Spectrum Plant Total RNA Kit (Sigma Chemical Co.) and cDNA was synthesized using SuperScript™III reverse transcriptase kit (Thermo Scientific). Coding sequence of *OkBCS* was obtained from this cDNA using full-length open reading frame (ORF) primers (Table 1). Full-length ORF of cDNA was cloned in Zero blunt vector (Thermo Scientific) and sequenced.

4.2.4 Sequence analysis

Sequence analyses were carried out with BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Nucleotide sequences were translated using ExPASy (http://web.expasy.org/translate/) and BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html) was used for marking identical and similar amino acids. Multiple sequence alignment was performed using ClustalX 2.1 (http://www.clustal.org/clustal2/) and neighbour-joining (NJ) tree of 33 TPSs was reconstructed with MEGA 6 [14] with 1000 iterations.

4.2.5 Heterologous protein expression and purification

For directional cloning into pET102 expression vector (Thermo Scientific), 5'-CACC-3' was added to the 5′ end of the forward primer and stop codon was removed in reverse primer to allow for C-terminal His-tag expression. The complete *OkBCS* ORF was amplified with AccuPrime™ Pfx DNA Polymerase (Thermo Scientific) and cloned into the pET102 vector producing pET102-*OkBCS*, in which the *OkBCS* coding sequence was fused with a 6X His-tag-coding extension at the C-terminus and transformed into *E. coli* TOP10 chemically competent cells. The construct was verified by DNA sequencing. Plasmid of the pET102 harboring full-length transcripts of BCS was introduced into Rosetta 2 (DE3) chemically competent cells. A single colony was inoculated in 5 mL of Luria Bertani broth (LB) and incubated on a rotary shaker (200 rpm) at 37ºC for 12 h. This culture was used for inoculating 50 mL "terrific broth" (TB) in a 250 mL Erlenmeyer flask and incubated on a rotary shaker (200 rpm) at 37ºC till optical density reached 0.6-0.8. The culture was induced with 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated on a rotary shaker for 12 h at 18ºC. The cell pellet (10 grams) was re-suspended in lysis buffer (50 mM MOPSO, pH 7.4, 300 mM NaCl, 1mg/mL lysozyme, 1mM PMSF and 0.5% CHAPS; 5mL/gram of pellet) and incubated on ice for 30 minutes, followed by sonication using a probe sonicator for 10 cycles of 30 sec burst and 1 minute cooling with amplitude set at 85%. Lysed suspension was centrifuged at 10, 000 x g for 15 minutes. The supernatant was mixed with 2 mL of washed Ni-NTA slurry and kept on a rocker for 1 hour at 4 °C. The resin was washed with wash buffer (50 mM MOPSO, pH 7.4, 300 mM NaCl and 20 mM imidazole) and bound protein was eluted using elution buffer (wash buffer supplemented with 250 mM imidazole), fractions run on 12% poly acryl amide-SDS gel. The fractions containing pure protein were pooled together and desalted in buffer (50 mM MOPSO, pH 7.4) using Sephadex G-25 (Hi-prep 26/10 desalting column). The protein concentration of each fraction was estimated by Bradford's method [15] and run on 12% poly acryl amide-SDS gel to check for purity and homogeneity.

4.2.6 Biochemical characterization of recombinant OkBCS

Enzyme assays were performed in 10 mL glass tubes, using 10 µg of purified recombinant protein, 50 µM substrate (GPP or FPP), 10 mM $MgCl_2$ and 1 mM DTT in assay buffer to a total volume of 250 µL. The reactions were incubated for 30 min at 30°C. Later, assay mixture was extracted twice with 2 mL of n-Hexane. The combined organic phase was dried over anhydrous sodium sulphate and concentrated to 50 µL by flushing it with nitrogen. Enzymatic products were analyzed by GC-MS for identification of volatiles. Rosetta cells transformed with empty vector and assays without substrates were used as controls. Steady state kinetics of recombinant OkBCS (rOkBCS) was performed in assay buffer containing 10 µM purified protein and substrate concentrations varying from 10-600 µM. The product amount and ratio were calculated using GC and used for the determination of K_m and k_{cat} values using GraphPad Prism ver. 6 (http://www.graphpad.com/scientific-software/prism/).

4.2.7 Volatile extraction and GC-MS analysis

The volatiles β-caryophyllene and α-humulene were measured in six different tissues including young leaves, mature leaves, inflorescence, flower, stem and root from five *Ocimum* species. Fresh tissues (5 gm each) from 2-month-old plants were harvested separately in triplicates and immediately used for volatile extraction using solvent extraction method. These tissues after harvesting were immediately soaked in 50 mL DCM for 20 h at 28 ºC. The combined organic phase was cooled to -20 °C for 2 h (for lipid precipitation) and filtered. The contents were dried, weighed, re-dissolved in 2 mL DCM and subjected to GC and GC-MS analysis for 3 biological replicates. GC analyses were carried out on an Agilent 7890A instrument equipped with a hydrogen flame ionization detector and HP-5 capillary column (30 m X 0.32 mm X 0.25 µm, J and W Scientific, USA). Nitrogen was used as the carrier gas at a flow rate of 1 mL/min. The column temperature was raised from 70°C to 110°C at 2°C min^{-1}, then raised to 180°C at 3°C min^{-1} and finally at 10°C min^{-1} raised to 220°C and held for 2 min. Injector and detector temperatures were 230°C and 235°C, respectively. GC-MS was performed on an Agilent 5975C mass selective detector interfaced with an Agilent 7890A gas chromatograph using an HP-5 MS capillary column with helium as the carrier gas and using above mentioned conditions. Compounds were identified by co-injection studies, comparing the retention time and mass fragmentation pattern

with those of reference compounds and also compared acquired mass spectra and retention indices with those of NIST/NBS and the Wiley mass spectral library (software version 2.0, Dec. 2005).

4.2.8 *OkBCS* expression analysis

For quantitative RT-PCR (qRT-PCR), 1µg of total RNA was used for cDNA synthesis using High capacity cDNA reverses transcription kit (Applied Biosystems, USA). qRT-PCR was performed on an Applied Biosystems 7500 HT Fast Real-Time PCR. A typical 10 µL reaction consisted of 5 µL of SYBR green master-mix, 0.2 µL of 10µM forward and reverse gene specific primers (Table 1) and 1 µL of diluted cDNA (1:2) with nuclease-free water added to make up a volume of 10 µL. A relative quantification of gene expression was performed using elongation factor-1α (EF-1α) as a reference gene. The difference in relative expression levels were calculated from the $2^{-\Delta\Delta Ct}$ value after normalization of *OkBCS* data to EF-1α. All analyses were performed in triplicates using five biological replicates for each sample.

Table 1: List of primers for functional characterization of OkBCS

Primer Name	5'-Sequence-3'
OkBCS-ORF-F	ATGGCTGCTTCCATCTCGAATGAC
OkBCS-ORF-R	TCAAATTATGATAGGGTGAACGAG
*OkBCS-Sal*I-F	AATAGTCGACATGGCTGCTTCCATCTCGAATGACAATG
*OkBCS-Eco*RI-R	AATAGAATTCTCAAATTATGATAGGGTGAACGAGGATGGA
OkBCS-Sense-F	AATAGTCGAC ATGGCTGCTTCCATCTCGAATGAC
OkBCS-Sense-R	AATAGGTACCAGACTCCCACGATCCAAAAGT
OkBCS-Antisense-F	AATAGGATCCAGACTCCCACGATCCAAAAGT
OkBCS-Antisense-R	AATAGAGCTC ATGGCTGCTTCCATCTCGAATGAC
OkBCS-qRT-F	CTTACAAGGTGGTGGAAG
OkBCS-qRT-R	TTGGTCAGTAATCTTCGTG
EF-1α -qRT-F	TGGTGTCATCAAGCCTGGTATGGT
EF-1α -qRT-R	ACTCATGGTGCATCTCAACGGACT
Intron-*Kpn*I-F	AATAGGTACCGCAGAAAATATACGAG
Intron-*BamH*I-R	AATAGGATCCATGTTGGTCAATAGCA

4.2.9 Agrobacterium mediated transient over-expression and silencing of *OkBCS*

Full-length *OkBCS* was cloned into the pRI 101-AN vector (Fig. 1) for over-expression in plants, using *Sal*I and *Eco*RI restriction sites in forward and reverse primers, respectively (Table 1). An 826 bp region from open reading frame was used as sense fragment and its complementary sequence as anti-sense fragment. A 500 bp fragment from wheat starch branching enzyme was used as intron. Silencing construct was prepared by cloning sense, intron and anti-sense fragments in tandem (Fig. 2) in pRI 101-AN vector. Sense fragment was cloned using *Sal*I and *Kpn*I restriction sites, transformed in chemically competent TOP10 *E. coli* cells and positive clones were verified by restriction digestion and sequencing. Similarly, intron was cloned using *Kpn*I and *Bam*HI restriction sites and anti-sense fragment was cloned using *Bam*HI and *Sac*I restriction sites. After sequence verification, it was transformed into *Agrobacterium tumefaciens* (GV3101) chemically competent cells on Luria-agar plate containing 50µg/mL kanamycin and 25µg/mL rifampcin. *A. tumefaciens* cells carrying over-expression or silencing construct were grown in 2mL LB media containing these antibiotics at 28 °C on rotary shaker set at 180 rpm for 2 days. This

starter culture was used to inoculate 10 mL LB with the antibiotics and incubated over-night at 28°C and 180 rpm. Later, cells were pelleted down by centrifugation at 10,000 rpm for 10 mins at 4°C. Cells were washed with half strength MS media, pH 5.4 and pelleted again. This cell pellet was re-suspended in half strength MS media, pH 5.4 to bring it to an $O.D._{600}$ of 0.3-0.5. This suspension was used for syringe-assisted infection on the adaxial surface of the leaves. Five *Ok* plants, each were taken for *OkBCS* over-expression and silencing. *A. tumefaciens* cells harboring empty pRI 101-AN vector along with plants without any treatment was used as control. All the plants were kept at 24°C, 16 h light and 8 h dark for acclimatization (4-6 days) before infection and maintained at the same conditions throughout the experiment. Samples for metabolic and real-time analyses were collected at 4 and 8 days after infection (DAI).

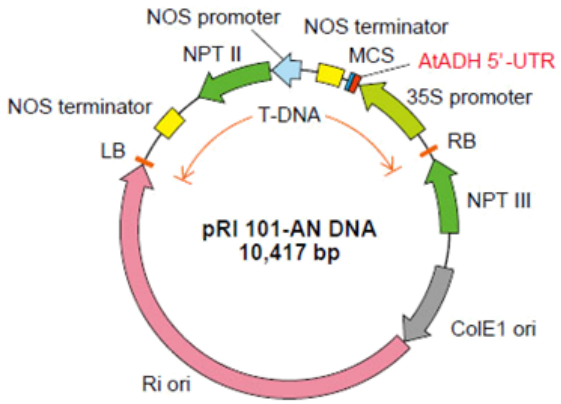


Figure 1: pRI-101-AN vector map

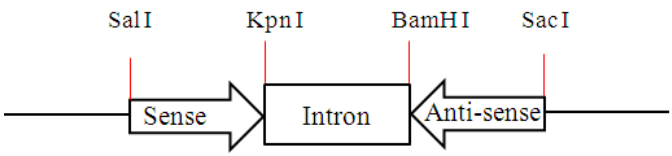


Figure 2: General scheme for the preparation of silencing constructs

4.2.10 Statistical analysis

GC-MS analyses of volatiles and gene expression analysis were performed with five biological and three technical replicates each. Data sets were represented as mean ± standard deviation. Significant differences between control and treated plants for metabolites and gene expression were determined using two-way ANOVA followed by Bonferroni's multiple comparisons test.

4.3 Results

4.3.1 *OkBCS* shares similarities and ancestry with other sesquiterpene synthases

A transcript (Contig44) from *Ocimum* transcriptome exhibited high sequence similarities with known sesqui-terpene synthase genes. This ORF was cloned from young leaf cDNA of *Ok* plant and designated as *OkBCS* (GenBank ID: KP226502) (Fig. 3). It was 1,629 bp long and encoded a protein of 542 amino acids with 63.63 kDa mass and pI of 5.66. Sequence comparison of *OkBCS* with other terpene synthases indicated high sequence similarity (Fig. 4). This alignment revealed the presence of several highly conserved regions of sesqui-terpene synthases (Fig. 4), which included the conserved aspartate-rich region (DDxxD), 'xDx$_6$E' and 'RR(X)$_8$W' motif, present in the N-terminal region of *OkBCS* and downstream of the N-terminal transit peptide. Another conserved region, 'RxR' motif was also found in *OkBCS*. NJ tree placed *OkBCS* closely to *LaBCS*, along with cluster of other *BCS*s and sesqui-terpenes from *Ocimum* (Fig. 5).

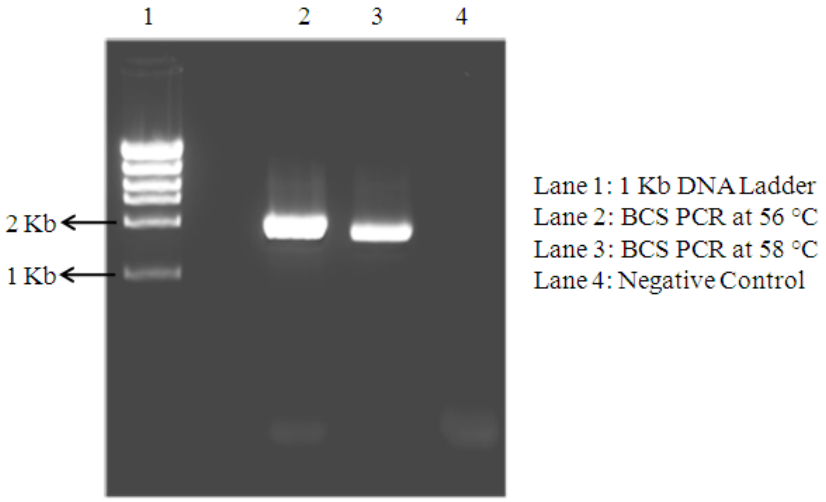


Figure 3: Full length PCR *OkBCS* with *Ok* young leaf cDNA.

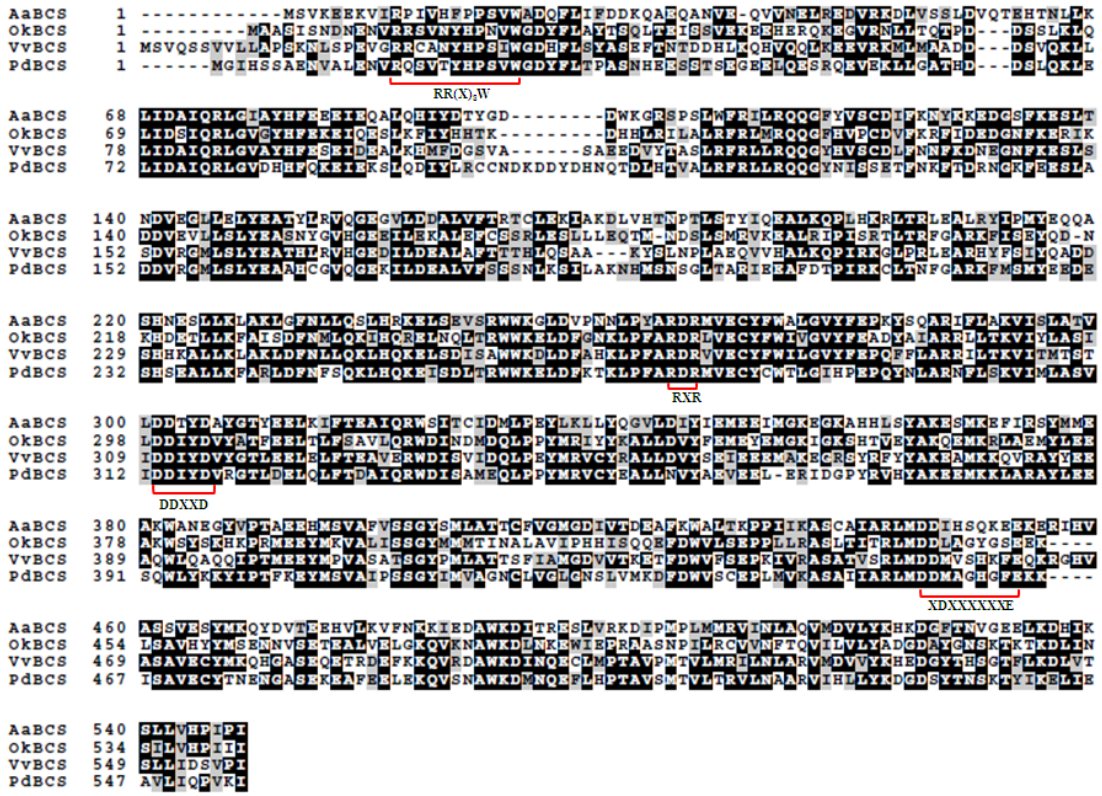


Figure 4: Amino acid sequence alignment of BCS from *O. kilimandscharicum* (OkBCS, AKA94109) with reported sequences from *Phyla dulcis* (AFR23370), *Mikania micrantha* (ACN67535), *Artemisia annua* (AAL79181) and *Vitis vinifera* (NP_001268204). Alignment was done by CLUSTALW and shaded using BOXSHADE 3.21. Identical residues are shaded black and similar residues are marked gray. Different conserved domains are indicated in red.

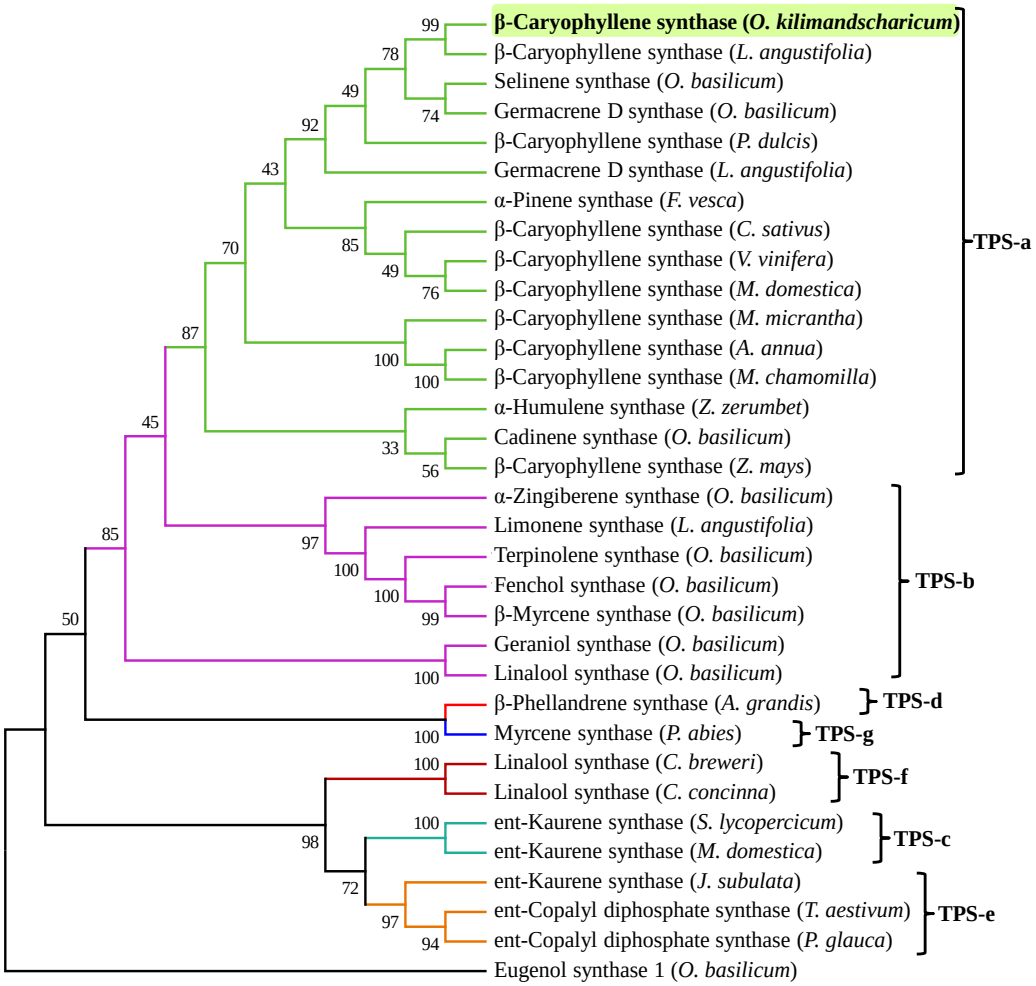


Figure 5: Neighbour-joining tree showing phylogeny of OkBCS with other TPSs using MEGA6.

4.3.2 Heterologous expression and functional characterization of OkBCS

Recombinant OkBCS was purified to homogeneity from the cell-free homogenates by Ni^{2+}- affinity chromatography (Fig. 6).For OkBCS enzyme assay, purified protein was incubated with farnesyl diphosphate (FPP) and geranyl pyrophosphate (GPP) as substrates [16]. The cyclization mechanism of BCS has been described previously (Fig. 7) [17]. GC and GC-MS analysis of the assay extract with FPP indicated the presence of one major (94%) and minor (6%) metabolites (Fig. 8). The compounds were identified as β-caryophyllene and α-humulene by comparing the EI mass fragmentation pattern, retention time and GC co-injection studies with authentic standards (Fig. 3). The K_m value of OkBCS with FPP was125.5µM and required Mg^{2+} as co-factor. Assays with empty vector as control or without FPP as a substrate did not yield any product. OkBCS was not active with GPP as substrate, however, incubation with GPP showed the traces of geraniol (the hydrolyzed product of GPP).

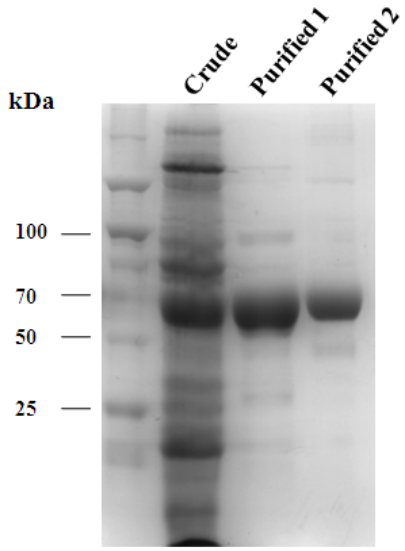


Figure 6: His-tag purified (purified 1) and desalted (purified 2) recombinant OkBCS on 12% SDS-PAGE gel.

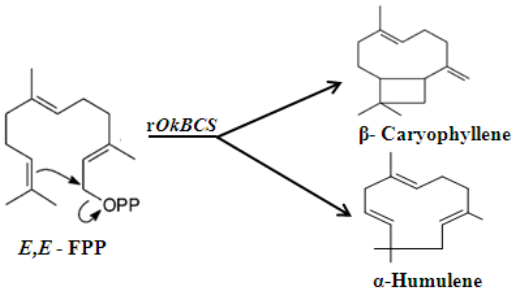


Figure 7: General scheme of conversion of *E,E*-FPP to β-caryophyllene and α-humulene

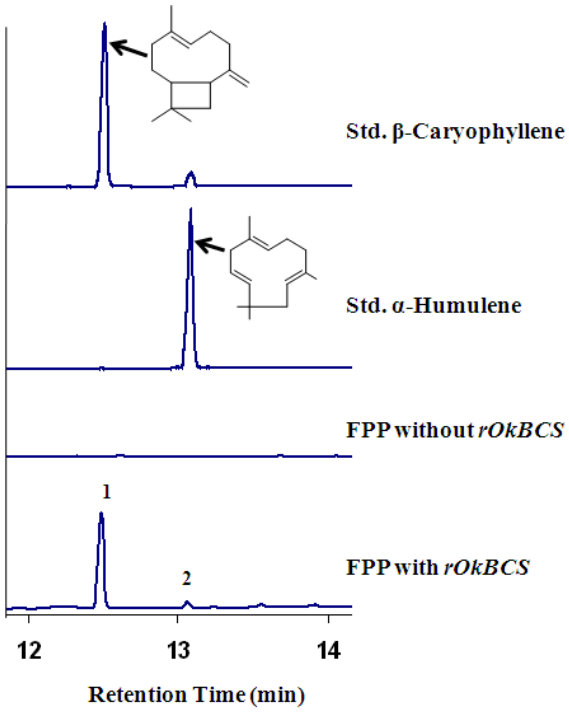


Figure 8: GC separation of authentic standards of (i) β-caryophyllene, (ii) α-humulene (iii) FPP without rOkBCS and (iv) *in vitro* assay products of FPP with rOkBCS; 1- β-caryophyllene and 2-α-humulene.

4.3.3 Volatile accumulation in different organs among *Ocimum* species

To survey natural variation of β-caryophyllene, we analyzed volatile profiles of six different tissues from five different *Ocimum* species. In all the species of *Ocimum*, both β-caryophyllene and α-humulene were present in a ratio of about15:1. In *Ot*, *Ok* and *Oa*, β-caryophyllene was the major sesqui-terpene ranging from 6-18% relative abundance (Fig. 9). Highest levels were found in *Ot* II inflorescence (18%) and in floral and leaf tissues (16%) of *Ot*. All the tissues of *Ok* and *Oa* had around 6% β-caryophyllene except in root, while 2% in *Og*I and II. Trace level (~1%) of β-caryophyllene was found in all the tissues of *Ob*I to IV, and in stem and root of *Ob*III. Abundance of α-humulene was relatively low to the levels of β-caryophyllene in all the analyzed *Ocimum* species across the tissues. Results of two way ANOVA for β-caryophyllene levels showed a statistically significant variation of 58.44% at $P<0.0001$ in species and 14.8% at $P=0.001$ between tissues types.

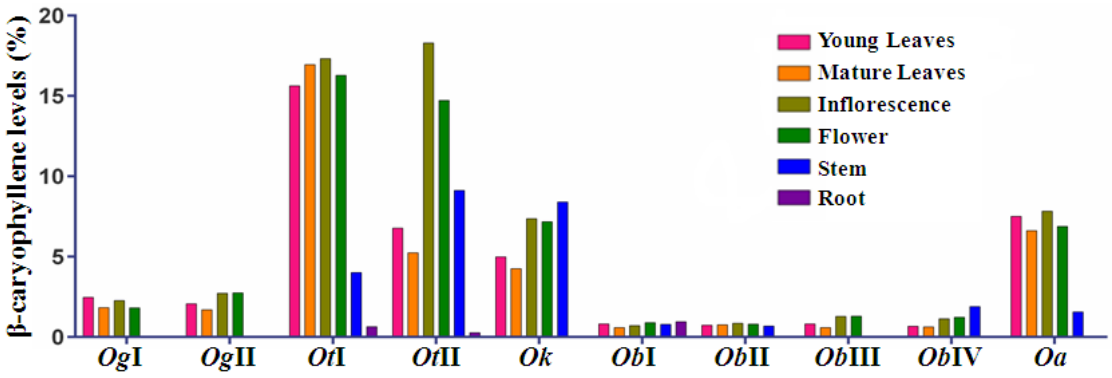


Figure 9: β-caryophyllene levels in six tissues of five different *Ocimum* species in relative area %. *Og- Ocimum gratissimum, Ot- Ocimum tenuiflorum, Ok- Ocimum kilimandscharicum, Ob- Ocimum basilicum, Oa- Ocimum americanum.*

4.3.4 Species and tissue specific mRNA accumulation patterns of *OkBCS*

Transcript abundance of *OkBCS* in different tissues across five *Ocimum* species indicated a positive correlation between gene expression and metabolite levels in all the species, except *Ot* (Fig. 10). *OkBCS* showed higher expression in young leaves and inflorescence, while lower expression in mature leaves, flowers and stem. Further across five *Ocimum* species, *OkBCS* showed the highest expression in *Ok* followed by *Oa, Ot* and *Og* while *Ob* had trace levels. Results of two-way ANOVA showed significant interaction between different tissues and *Ocimum* species (total variance of 5.17%, $P<0.0001$).

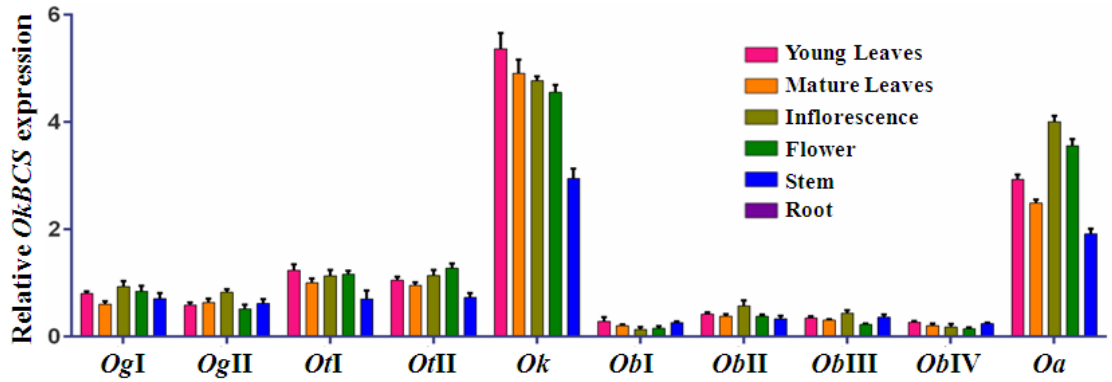


Figure 10: *OkBCS* expression analyses in different tissues (as shown in different colors) of five *Ocimum* species. Absolute quantification was done using EF-1α as the endogenous control.

4.3.5 Role of *OkBCS* in β-caryophyllene biosynthesis

To demonstrate *OkBCS* expression patterns and their effect on metabolite levels, transient over-expression and silencing assays were established in *Ok* plants by syringe assisted agro-infiltration method. Changes in metabolites and transcripts were calculated by comparing with values from non-infected control plants at respective stages (Fig. 11 and 12). Plants infected with empty vector constructs were also used to nullify any effect of tissue injury as wounding might have led to increase in β-caryophyllene levels [18]. In qRT-PCR analysis, silencing effect was more prominent in systemic leaves with 95% reduction in *OkBCS* transcripts against 51% decrease in local leaves at 4 DAI. However, this silencing effect was less at 8 DAI as systemic tissues showed only 45% reduction in transcripts of *OkBCS* while 31% decrease in local tissues. Plants infected with empty vector constructs had slight increase in transcript levels in local tissues at both stages that could be attributed to wounding (Fig. 11). Overall, two-way ANOVA followed by Bonferroni's multiple comparisons test suggested significant difference in *OkBCS* transcripts between control and *OkBCS* silenced plants ($P< 0.01$) (Fig. 11). Further, this silencing effect in transcript abundance was also evident from changes in metabolite levels (Fig. 12). β-caryophyllene level decreased by 20% in silenced plants compared to controls at 4 DAI (Fig. 12 and 13) with α-humulene following the similar trend. However apart from these two metabolites, no other terpenes displayed any significant change in their levels at both stages.

In *OkBCS* over-expression studies, transcript level was 400% higher than controls on 4 DAI while this was decreased to 250% at 8 DAI in local leaves. The same trend was observed in case of systemic tissues with 160% increase in transcript level at 4 DAI, which reduced to 27% at 8 DAI (Fig 11). However, metabolite accumulation was not consistent with increase in the *OkBCS* transcript at both the stages. At 4 DAI, β-caryophyllene increased to 22% and 32% in local and systemic tissues, respectively, compared to control plants while it rose to 52% and 56% in local and systemic leaves, respectively at 8 DAI.

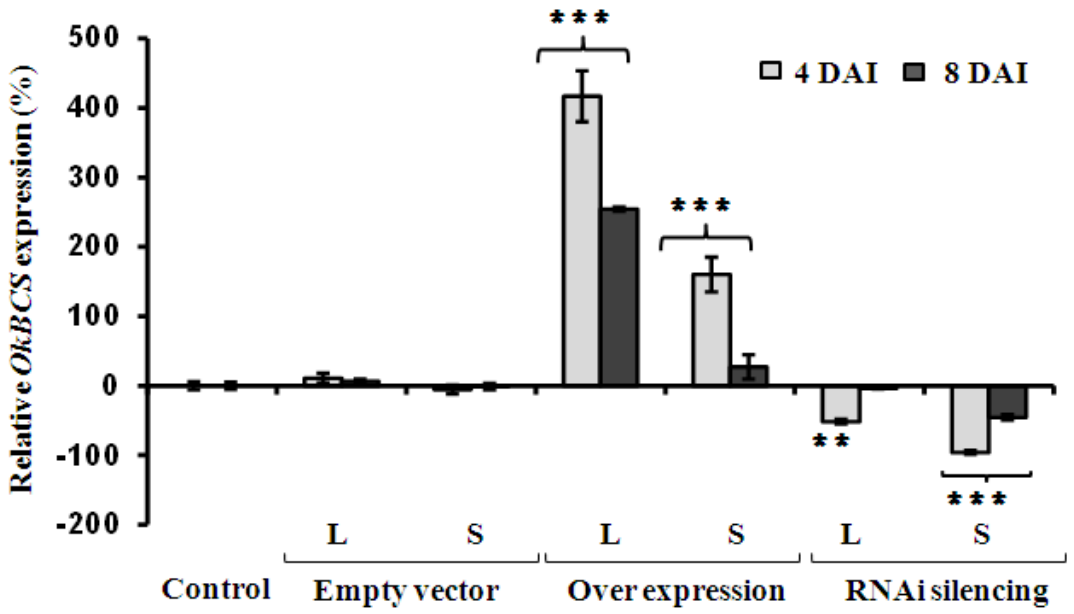


Figure 11: *OkBCS* expression levels determined by qRT-PCR upon *OkBCS* over-expression and RNAi silencing (L-local; S-systemic). Values represent relative change (%) in comparison to β-caryophyllene levels in control. Two way ANOVA followed by Bonferroni's multiple comparisons suggested significant difference between the data at $p<0.001$ (indicated as '***'), $p<0.01$ (indicated as '**').

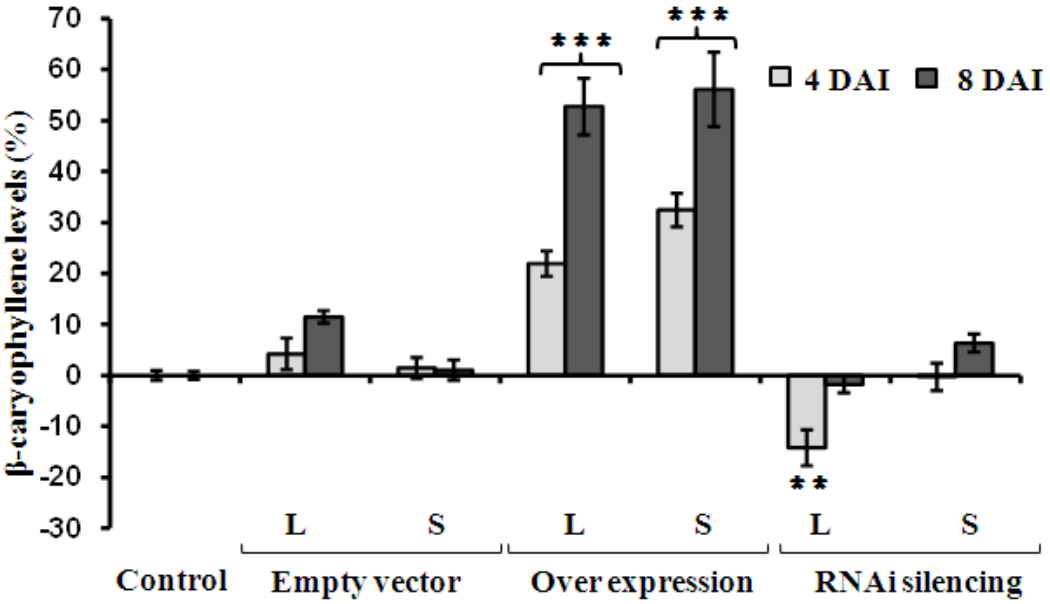


Figure 12: β-caryophyllene levels upon *OkBCS* over-expression and RNAi silencing (L-local; S-systemic). Values represent relative change (%) in comparison to β-caryophyllene levels in control. Two way ANOVA followed by Bonferroni's multiple

comparisons suggested significant difference between the data at $p<0.001$ (indicated as '***'), $p<0.01$ (indicated as '**').

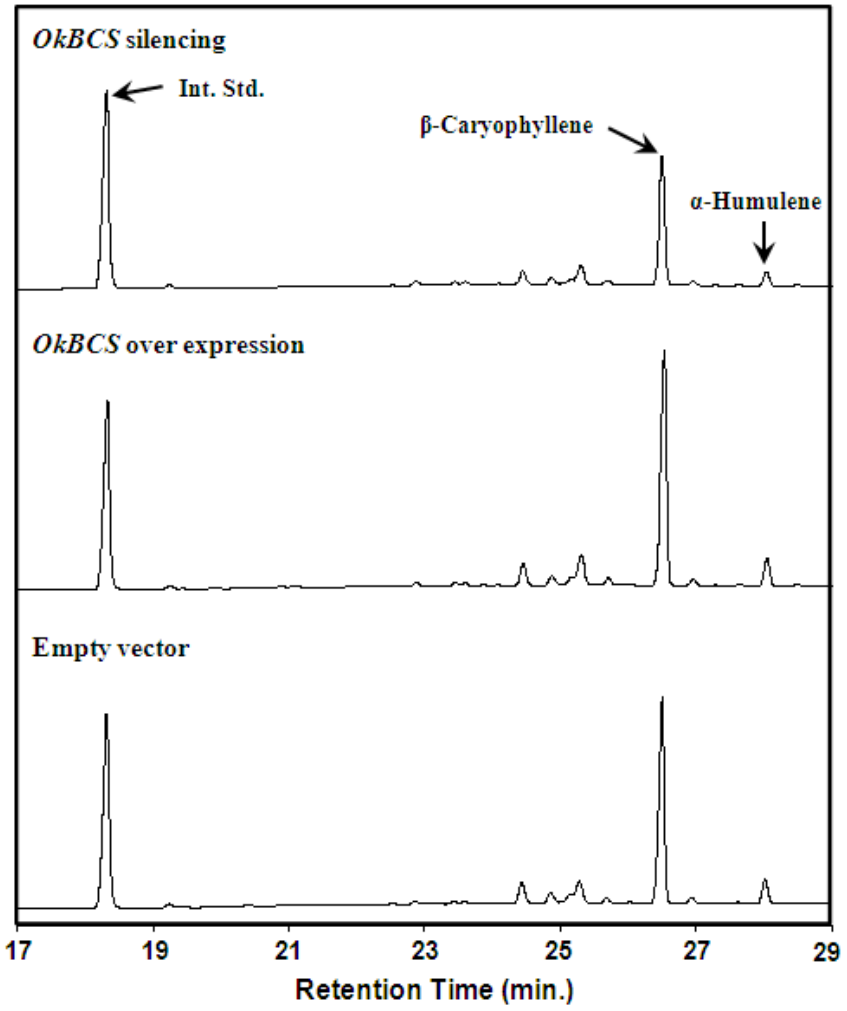


Figure 13: GC chromatograms; empty vector, *OkBCS* over-expressed and *OkBCS* silenced plants in local leaves at 8 DAI. Internal standard - Linalyl acetate

4.4 Discussion

Ocimum is a member of the *Lamiaceae* family and is known to biosynthesize and accumulate different proportion of various secondary metabolites in different plant parts. The predominant among these metabolites are mono- and sesqui-terpenes. Sesquiterpenes are the most diverse of the terpenes, synthesized by the head to tail fusion of condensation of IPP and GPP or IPP and DMAPP. Here, we have performed the volatile analysis of six different tissues of five *Ocimum* species to analyze the accumulation pattern of β-caryophyllene, a major sesquiterpene found in several *Ocimum* species and other medicinal plants. Usually, β-caryophyllene is found in nature as a mixture with isomer α–caryophyllene, also referred to as α-humulene [19]. It was observed in our analysis that the abundance of β-caryophyllene over α-humelene, in all the species that were studied, remains almost constant, to a ratio of 15:1. Also, across the species variations in accumulation pattern was also observed, with *Ot* species being the richest in β-caryophyllene contents, while *Ob* had the least accumulation. Upon sequence comparison of *OkBCS* with other reported terpene synthases including *BCS* from *Lavandula angustifolia* Mill. L. (70%) (AGL98419; [20]) and (-)- germacrene D synthase from *Ocimum basilicum* (58%) (Q5SBP6; [21]), along with *BCS*s from *Lippia dulcis* Trevir. (J7LJN5; [22]), *Matricaria recutita* L. (I6RAQ6; [23]), *Artemisia annua* L.(Q8SA63; [24]) and *Vitis vinifera* L. (E5GAF4; [25]), several similarities were identified. Domains crucial for the activity of these proteins, like 'DDxxD' which is crucial for the substrate binding [26] and 'xDx_6E' motif for metal cofactor binding [27], were found to be present. No chloroplast-targeting sequence was identified in *OkBCS*, which indicated its function as a cytosolic sesqui-terpene synthase. 'RR$(X)_8$W' that is assumed to participate in the ionization of the substrate [28] and is characteristic of majority of the members of subfamilies TPS-a and TPS-b [29]. 'RxR' which was located 35 amino acids upstream of the DDxxD motif and known to form complex of diphosphate group after substrate ionization [30]. Phylogenetic analysis of *OkBCS* placed it in a group with other terpene synthase of TPS-a subfamily. Thus, sequence analysis and phylogenetic tree suggested that *OkBCS* could have evolved from a common ancestor more closely related to TPS-a subfamily. Upon kinetic characterization of OkBCS, the parameters were observed to be higher as compared to other sesqui-terpene synthases, *viz*., from *Zingiber zerumbet* (L.) Sm. (α-humulene synthase, K_m=32 μM) and *Laurus nobilis* L.

(LnTPS3, K_m=43.4 μM) which had about 3-4 times lower K_m values [31,32]. Additionally, similar $K_m$values were observed from other sesqui-terpene synthases [33-35]. The V_{max} 2.55 $\mu M \cdot s^{-1}$ and k_{cat} 4.06 s^{-1} were also high compared to many other reported sesqui-terpene synthases and were in the typical range of enzymes of the secondary metabolism [34]. The relatively high K_m value suggested a low binding affinity of OkBCS to FPP. OkBCS resembled in principle reaction parameters and general properties of BCS from *Artemisia annua* [24] and *Lavandula angustifolia* [20]; and belonged to typical class of sesqui-terpene synthases from angiosperms [36]. Expression analysis of *OkBCS* indicated a positive correlation between transcript abundance and metabolite accumulation, except for *Ot* and *Ok* species (Fig. 8 and 9). In *Ot*, it was observed that in spite of low *BCS* levels, expression of β-caryophyllene was higher in almost all the tissues of *Ot*. This could be due to contribution from other TPS in β-caryophyllene production as byproduct [31,37] which lead to overall higher β-caryophyllene levels. A reverse trend to that of *Ot* was observed in case of *Ok*, which could be attributed to differential gene expression of pathway enzymes and transcription factors [38,39]. Transient expression using *Agrobacterium* is frequently used as it allows rapid analysis of gene function. Several reports are available where transient expression mediated by *Agrobacterium* was used for studying gene expression [40], gene silencing [41] and protein-protein interaction [42]. It was observed that the over-expression and silencing was specific for β-caryophyllene and α-humulene, as there was no effect on accumulation of any other metabolite. Also, the results indicated that the either of the effect tend to wither away with from day 4 to day 8, which was consistent with the available reports indicating that the effect withered away upon time during transient expression, owing to the instability of non-integrated T-DNA copies [43].

4.5 Conclusion

β-caryophyllene synthase was functionally characterized from *Ocimum kilimandscharicum* for the first time. OkBCS carried out the cyclization of FPP into β-caryophyllene as a major product along with α-humelene as a minor metabolite. *OkBCS* expression variation correlated with metabolite levels among five *Ocimum* species. Agro-infiltration based transient expression assay was established in *Ocimum* that clearly demonstrated the role of OkBCS in β-Caryophyllene accumulation. These findings may pave the way to improve disease resistance in crop plants and enriched medicinal properties.

4.6 Appendix

4.1 OkBCS amino acid sequence

>OkBCS

MAASISNDNENVRRSVNYHPNVWGDYFLAYTSQLTEISSVEKEEHERQKEGV
RNLLTQTPDDSSLKLQLIDSIQRLGVGYHFEKEIQESLKFIYHHTKDHHLRILA
LRFRLMRQQGFHVPCDVFKRFIDEDGNFKERIKDDVEVLLSLYEASNYGVHG
EEILEKALEFCSSRLESLLLEQTMNDSLSMRVKEALRIPISRTLTRFGARKFISE
YQDNKHDETLLKFAISDFNMLQKIHQRELNQLTRWWKELDFGNKLPFARDR
LVECYFWIVGVYFEADYAIARRLLTKVIYLASILDDIYDVYATFEELTLFSAVL
QRWDINDMDQLPPYMRIYYKALLDVYFEMEYEMGKIGKSHTVEYAKQEMK
RLAEMYLEEAKWSYSKHKPRMEEYMKVALISSGYMMMTINALAVIPHHISQ
QEFDWVLSEPPLLRASLTITRLMDDLAGYGSEEKLSAVHYYMSENNVSETEA
LVELGKQVKNAWKDLNKEWIEPRAASNPILRCVVNFTQVILVLYADGDAYG
NSKTKTKDLINSILVHPIII

4.7 References

[1] Rock CD, Zeevaart JAD, The ABA mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 7496–7499.

[2] Das A, Lee SH, Hyun TK, et al., Plant volatiles as method of communication, Plant Biotechnol. Rep. 7 (2013) 9–26.

[3] Pichersky E, Gershenzon J, The formation and function of plant volatiles: perfumes for pollinator attraction and defense, Curr. Opin. Plant Biol. 5 (2002) 237–243.

[4] Kessler A, Baldwin IT, Plant responses to insect herbivory: The emerging molecular analysis, Annu. Rev. Plant Biol. 53 (2002) 299–328.

[5] Gang DR, Wang J, Dudareva N, et al., An investigation of the storage and biosynthesis of phenylpropenes in sweet basil, Plant Physiol. 125 (2001) 539–555.

[6] Hakkim FL, Arivazhagan G, Boopathy R, Antioxidant property of selected *Ocimum* species and their secondary metabolite content, J. Med. Plants Res. 2 (2008) 250–257.

[7] Prakash P, Gupta N, Therapeutic uses of *Ocimum sanctum* Linn (Tulsi) with a note on eugenol and its pharmacological actions: a short review, Indian J. Physiol. Pharmacol. 49 (2005) 125–131.

[8] Ogendo JO, Kostyukovsky M, Ravid U, et al., Bioactivity of *Ocimum gratissimum* L. oil and two of its constituents against five insect pests attacking stored food products, J. Stored Prod. Res. 44 (2008) 328–334.

[9] Singh P, Ramesha HJ, Sarate P, et al., Insecticidal potential of defense metabolites from *Ocimum kilimandscharicum* against *Helicoverpa armigera,* PLoS One. 9 (2014) e104377

[10] Bayala B, Bassole IHN, Gnoula C, et al., Chemical composition, antioxidant, anti-inflammatory and anti-proliferative activities of essential oils of plants from Burkina Faso, PLoS One. 9 (2014) e92122.

[11] Bahi A, Mansouri S Al, Memari E Al, et al., β-Caryophyllene, a CB_2 receptor agonist produces multiple behavioral changes relevant to anxiety and depression in mice, Physiol. Behav. 135 (2014) 119–124.

[12] Dahham SS, Tabana YM, Iqbal MA, et al., The anticancer, antioxidant and antimicrobial properties of the sesquiterpene β-Caryophyllene from the essential oil of *Aquilaria crassna*, Molecules. 20 (2015) 11808–11829.

[13] Huang M, Sanchez-Moreiras AM, Abel C, et al., The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)-β-caryophyllene, is a defense against a bacterial pathogen, New Phytol. 193 (2012) 997–1008.

[14] Srivastava PL, Daramwar PP, Ramakrishna K, et al., Functional characterization of novel sesquiterpene synthases from Indian sandalwood, *Santalum album*, Sci. Rep. 5 (2015) 10095.

[15] Bradford MM, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248-254.

[16] Starks CM, Back KW, Chappell J, et al., Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase, Science 277 (1997) 1815–1820.

[17] Thulasiram HV, Phan RM, Rivera SB, et al., Synthesis of deuterium-labeled derivatives of dimethylallyl diphosphate, J. Org. Chem. 71 (2006) 1739-1741.

[18] Ferrando A, Koncz-Kálmán Z, Farràs R, et al., Detection of in vivo protein interactions between Snf1-related kinase subunits with intron-tagged epitope-labelling in plants cells., Nucleic Acids Res. 29 (2001) 3685–3693.

[19] Cane DE, Sesquiterpene biosynthesis: Cyclization mechanisms, in: Compr. Nat. Prod. Chem., 1999: pp. 155–200.

[20] Tamura K, Stecher G, Peterson D, et al., MEGA6: Molecular evolutionary genetics analysis Version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.

[21] Jullien F, Moja S, Bony A, et al., Isolation and functional characterization of a τ-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*, Plant Mol. Biol. 84 (2014) 227–241.

[22] Iijima Y, Davidovich-Rikanati R, Fridman E, et al., The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil, Plant Physiol. 136 (2004) 3724–3736.

[23] Attia M, Kim SU, Ro DK, Molecular cloning and characterization of (+)-epi-α-bisabolol synthase, catalyzing the first step in the biosynthesis of the natural sweetener, hernandulcin, in *Lippia dulcis*, Arch. Biochem. Biophys. 527 (2012) 37–44.

[24] Irmisch S, Krause ST, Kunert G, et al., The organ-specific expression of terpene synthase genes contributes to the terpene hydrocarbon composition of chamomile essential oils, BMC Plant Biol. 12 (2012) 84.

[25] Olofsson L, Engstrom A, Lundgren A, et al., Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L, BMC Plant Biol. 11 (2011) 45.

[26] Martin DM, Aubourg S, Schouwey MB, et al., Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays, BMC Plant Biol. 10 (2010) 226.

[27] Rynkiewicz MJ, Cane DE, Christianson DW, Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13543–13548.

[28] Degenhardt J, Koellner TG, Gershenzon J, Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, Phytochemistry. 70 (2009) 1621–1637.

[29] Williams DC, McGarvey DJ, Katahira EJ, et al., Truncation of limonene synthase preprotein provides a fully active “Pseudomature” form of this

monoterpene cyclase and reveals the function of the amino-terminal arginine pair, Biochemistry. 37 (1998) 12213–12220.

[30] Bohlmann J, Meyer-Gauen G, Croteau R, Plant terpenoid synthases: molecular biology and phylogenetic analysis, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 4126–4133.

[31] Dehal SS, Croteau R, Partial purification and characterization of two sesquiterpene cyclases from sage (*Salvia officinalis*) which catalyze the respective conversion of farnesyl pyrophosphate to humulene and caryophyllene, Arch. Biochem. Biophys. 261 (1988) 346–356.

[32] Alemdar S, Hartwig S, Frister T, et al., Heterologous expression, purification, and biochemical characterization of α-Humulene synthase from *Zingiber zerumbet* Smith, Appl. Biochem. Biotechnol. 178 (2015) 474-89.

[33] Mossaab Y, Tholl D, Cormier G, et al., Identification and characterization of terpene synthases potentially involved in the formation of volatile terpenes in carrot (*Daucus carota* L.) roots, J. Agric. Food Chem. 63 (2015) 4870–4878.

[34] Tholl D, Chen F, Petri J, et al., Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers, Plant J. 42 (2005) 757–771.

[35] Bar-Even A, Noor E, Savir Y, et al., The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters, Biochemistry. 50 (2011) 4402–4410.

[36] Nieuwenhuizen NJ, Wang MY, Matich AJ, et al., Two terpene synthases are responsible for the major sesquiterpenes emitted from the flowers of kiwifruit (*Actinidia deliciosa*), J. Exp. Bot. 60 (2009) 3203–3219.

[37] Budavari S, O'Neil MJ, Smith A, et al., The Merck index: an encyclopedia of chemicals, drugs, and biologicals, 12th ed., Whitehouse Station, N.J; Merck & Co Inc, 87, 1996.

[38] Singh P, Kalunke RM, Giri AP, Towards comprehension of complex chemical evolution and diversification of terpenoids and phenylpropanoid pathways in *Ocimum* species, RSC *Adv*, 5 (2015) 106886-106904.

[39] Zvi MMB, Shklarman E, Masci T, et al., PAP1 transcription factor enhances production of phenylpropanoid and terpenoid scent compounds in rose flowers, New Phytol. 195 (2012) 335-345.

[40] Seemann M, Zhai G, de Kraker JW, et al., Pentalenene synthase; Analysis of active site residues by site-directed mutagenesis, J. Am. Chem. Soc. 124 (2002) 7681–7689.

[41] Kapila J, Rycke RD, Van MM, et al., An *Agrobacterium*-mediated transient gene expression system for intact leaves, Plant Sci. 122 (1997) 101–108.

[42] Schöb H, Kunz C, Meins F, Silencing of transgenes introduced into leaves by agroinfiltration: a simple, rapid method for investigating sequence requirements for gene silencing, Mol. Gen. Genet. 256 (1997) 581–585.

[43] Cai Y, Jia JW, Crock J, et al., A cDNA clone for beta-caryophyllene synthase from *Artemisia annua*, Phytochemistry. 61 (2002) 523–529.