

TRANSCRIPTOME PROFILING, AND THE CLONING AND  
CHARACTERIZATION OF A MONOTERPENE SYNTHASE FROM THE SEEDS OF  
*CORIANDRUM SATIVUM* L.

by  
MARIANA GALATA

B.Sc., University of British Columbia, 2011

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

in  
THE COLLEGE OF GRADUATE STUDIES

(Biochemistry)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Okanagan)

December 2013

© Mariana Galata, 2013

## ABSTRACT

Plant terpenes are a large and diverse class of naturally-derived compounds, valuable in the medicinal, perfume and culinary industries. The seeds of *Coriandrum sativum* (coriander) produce essential oil (EO) rich in monoterpenes, volatile C<sub>10</sub> terpenes. In this study the coriander seeds were viewed under a scanning electron microscope and stomata as well as vittae structures were observed on the seed surface and in cross-section, respectively. The EO of *C. sativum* seeds was extracted and qualitatively and quantitatively analyzed using gas chromatography-mass spectrometry. The EO terpene content findings agreed with previous literature, with linalool being the most abundant monoterpene in coriander EO.

The transcriptome of coriander seeds, at three developmental stages (early, mid and late) was sequenced via Illumina technology. Analysis of the differential transcript abundance of select terpene biosynthetic genes between these stages revealed that two terpene production pathways are constitutively active in the seeds, with slight upregulation in the mid-developmental stage. All the genes involved with active photosynthesis and fatty acid biosynthesis and metabolism were also identified from the coriander transcript library.

To validate the usability of the transcriptome sequence data, a terpene synthase candidate gene, *C<sub>sy</sub>TRPS*, encoding a 611 amino acid protein was expressed in bacteria and the recombinant protein purified by Ni-NTA affinity chromatography. Enzymatic assays with geranyl diphosphate (GPP), the precursor to monoterpenes, revealed that this 65.86 kDa recombinant protein catalyzed the conversion of GPP to  $\gamma$ -terpinene, with apparent  $V_{max}$ ,  $K_m$ , and  $k_{cat}$  values of  $2.2 \pm 0.2$  pkat/mg,  $66 \pm 13$   $\mu$ M and  $1.476 \times 10^{-4}$  s<sup>-1</sup>, respectively. Knowledge gained from these experiments will facilitate future studies concerning essential and fatty acid oil production in coriander. They will also enable efforts to improve the EO of coriander through metabolic engineering or plant breeding.

## **PREFACE**

All of the scanning electron microscopy (SEM), transcript library analysis, molecular cloning, functional characterization and qPCR work presented in this thesis was performed at the University of British Columbia, Okanagan campus (UBCO) by Mariana Galata, with the exception of the identification, molecular cloning and functional characterization of CsLINS, which was performed by Lukman Sarker. The transcriptome sequencing, including the template preparation, *de novo* assembly and gene expression matrix development was performed as a paid service by staff (in particular Dr. Slava Ilyntskyy) at Plant Biosys, University of Lethbridge, AB, Canada. The SEM work was also a paid service, with hands-on support from David Arkinstall at UBCO and the qPCR was made possible by assistance from Dr. Mark Rheault who provided access to his equipment. The study was designed by Dr. Soheil Mahmoud. A manuscript containing the analysis of the coriander transcript library and molecular cloning and functional characterization of CsLINS and Cs $\gamma$ TRPS was prepared primarily by Mariana Galata with editorial assistance from Dr. Mahmoud and Lukman Sarker. This manuscript is currently under review at the Phytochemistry journal. The scanning electron microscopy, transcript library analysis, molecular cloning and functional characterization work was presented as an oral presentation by Mariana Galata on August 4<sup>th</sup> 2013 at the 52<sup>nd</sup> annual meeting of the phytochemical society of North America (PSNA) hosted by the Oregon State University in Corvallis, OR, USA. Prior to PSNA, parts of this thesis work were presented by Mariana Galata in October 2013 at the Biochemistry and Molecular Biology Seminar Series, UBCO, Kelowna, BC.

## TABLE OF CONTENTS

ABSTRACT .....	ii
PREFACE .....	iii
TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
LIST OF ABBREVIATIONS .....	x
ACKNOWLEDGEMENTS .....	xii
DEDICATION .....	xiii
CHAPTER 1: INTRODUCTION.....	1
1.1 <i>Coriandrum sativum</i> .....	1
1.1.1 Uses of Coriander .....	3
1.2 Terpenoids.....	5
1.2.1 Natural Functions and Industrial Uses of Terpenoids .....	5
1.2.2 Biosynthesis of Terpenoids .....	6
1.3 Terpene Synthases.....	9
1.4 Strategies of Gene Cloning.....	12
1.5 Second Generation Sequencing - Illumina Technology .....	15
1.6 <i>De novo</i> Transcript Assembly .....	18
1.7 Research Objectives .....	20
1.8 Hypothesis and Rationale .....	21
1.9 Significance of Study .....	21
CHAPTER 2: SCANNING ELECTRON MICROSCOPY AND <i>C. SATIVUM</i> ESSENTIAL OIL ANALYSIS .....	23
2.1 Synopsis.....	23
2.2 Materials and Methods .....	23
2.2.1 Plant Material .....	23
2.2.2 Scanning Electron Microscopy (SEM).....	23
2.2.3 Steam Distillation of <i>C. sativum</i> Seeds .....	24
2.3 Results .....	24
2.3.1 SEM Images .....	24

2.3.2 Gas Chromatograms and Mass Spectrometry Fingerprints of <i>C. sativum</i> Seed Volatile Extracts .....	25
2.4 Discussion .....	28
2.4.1 Implications of SEM Images .....	28
2.4.2 Composition of <i>C. sativum</i> Seed Essential Oil.....	29
<b>CHAPTER 3: BIOINFORMATICS ANALYSIS OF ILLUMINA-LED TRANSCRIPTOME SEQUENCING AND <i>DE NOVO</i> TRANSCRIPT ASSEMBLY DATA .....</b>	<b>31</b>
3.1 Synopsis.....	31
3.2 Materials and Methods .....	31
3.2.1 RNA Isolation and Quality Controls .....	31
3.2.2 Details of Illumina Sequencing and <i>De Novo</i> Transcript Assembly .....	31
3.2.3 Bioinformatics .....	32
3.3 Results .....	34
3.3.1 Coriander transcriptome composition .....	34
3.3.2 BLASTx Alignments and Annotations.....	35
3.3.3 Transcript abundances .....	40
3.4 Discussion .....	49
3.4.1 Quality of <i>C. sativum</i> Transcriptome Dataset .....	49
3.4.2 Implications of Gene Expression of Isoprenoid Biosynthetic Genes .....	51
<b>CHAPTER 4: MOLECULAR CLONING AND FUNCTIONAL CHARACTERIZATION OF <math>\gamma</math>-TERPINENE SYNTHASE, AND CLONING AND ATTEMPTED PROTEIN EXPRESSION OF PUTATIVE (S)-LINALOOL SYNTHASE. ....</b>	<b>57</b>
4.1 Synopsis.....	57
4.2 Materials and Methods .....	57
4.2.1 Monoterpene Synthase Candidate Selection .....	57
4.2.2 Cloning of Full Length Monoterpene Synthases .....	58
4.2.3 Recombinant Protein Expression, Crude Enzyme Assay and Purification .....	63
4.2.4. Pure Enzyme Assays and Product Analysis .....	67
4.3 Results .....	68
4.3.1 Monoterpene Synthase Candidates.....	68
4.3.2 Phylogenetic Tree.....	73
4.3.3 Cloning of Full Length Monoterpene Synthases .....	75
4.3.4 Enzyme Kinetics Data for $\gamma$ -Terpinene Synthase.....	82

4.4 Discussion .....	86
4.4.1 Terpene Synthase Sequence Homology and Phylogenetic Analysis.....	86
4.4.2 Truncated vs. full length mTPS3.....	88
4.4.3 Functional Characterization of $\gamma$ -Terpinene Synthase.....	89
CHAPTER 5: CONCLUSIONS.....	93
5.1 Summary of Research.....	93
5.2 Research Novelty .....	94
5.3 Assumptions and Limitations .....	95
5.4 Future Directions.....	96
5.5 Accession Numbers.....	97
REFERENCES.....	99

## LIST OF TABLES

Table 1.1	Scientific classification of the cilantro/coriander plant.....	1
Table 2.1	Coriander seed essential oil (EO) terpene composition.....	28
Table 3.1	Primers for qPCR relative expression analysis of DXS and HMGR genes.....	34
Table 3.2	Summary statistics for alternative k-mer length transcript assembly libraries.....	35
Table 3.3	Number of homologues ( $\geq 90\%$ identity; $X < 1E-40$ ), similar proteins ( $< 90\%$ identity; $X \leq 1E-40$ ), weakly similar proteins ( $< 90\%$ identity; $1E-40 < X \leq 1E-20$ ), and putative proteins ( $< 90\%$ identity; $1E-20 < X \leq 1E-10$ ) for each Oasis transcript library.....	35
Table 3.4	KEGG pathway annotations for <i>C. sativum</i> transcriptome.....	38
Table 3.5	Transcript abundance of genes involved in the DXP and MVA pathways, as well as genes encoding prenyl transferases across three stages of coriander seed development: S1-early, S2-mid and S3-late...	41
Table 3.6	Transcript abundance for genes involved in coriander isoprenoid biosynthesis across three stages of seed development (S1, S2, S3).....	42
Table 3.7	Photosynthetic genes identified in <i>C. sativum</i> seed tissue from RNA-seq data.....	45
Table 3.8	Unigenes in <i>C. sativum</i> seeds involved in fatty acid biosynthesis.....	47
Table 3.9	Percent identities for alignments between coriander transcript library against three known coriander fatty acid nucleotide sequences.....	51
Table 4.1	Coriander essential oil terpene synthase candidate genes.....	69
Table 4.2	Nanodrop spectrophotometer results for plasmids containing mTPS3 with signal peptide extracted from DH10B <i>E. coli</i> cells.....	76
Table 4.3	Nanodrop spectrophotometer results for plasmids containing truncated mTPS3 extracted from DH10B <i>E. coli</i> cells.....	78
Table 4.4	Nanodrop spectrophotometer results for mTPS3_pGEX-4T-1 plasmids extracted from DH10B <i>E. coli</i> cells.....	79
Table 4.5	Nanodrop spectrophotometer results for <i>mTPS2_pET41b(+)</i> plasmids extracted from DH10B <i>E. coli</i> cells.....	80

## LIST OF FIGURES

Figure 1.1	Coriandrum sativum (A) seedling, (B) young foliage, (C) mature foliage, (D) characteristic umbel organization of flowers, (E) late flower blossoms, (F) early flower blossoms and (G) fresh seeds.....	3
Figure 1.2	Mevalonate (MVA) and 1-deoxyxylulose-5-phosphate (DXP) terpenoid prenyl precursor biosynthetic pathways. ....	8
Figure 1.3	The prenyl precursor, geranyl diphosphate, can give rise to a large variety of monoterpene products. These are just a few examples of the monoterpenes found in coriander EO.....	12
Figure 1.4	Illumina sequencing workflow. ....	18
Figure 1.5	Simplistic example of de Bruijn graph layout.....	20
Figure 2.1	Scanning electron microscope images of C. sativum seed tissue. ....	25
Figure 2.2	Gas chromatogram of steam distilled C. sativum seed essential oil.....	27
Figure 3.1	Comparison of C. sativum percentage distribution of gene ontology against two reference databases, M. truncatula and A. thaliana.....	36
Figure 3.2	Breakdown of KEGG annotations for C. sativum seed transcriptome....	37
Figure 3.3	Transcript abundances for key regulatory genes of the 1-deoxy-D-xylulose 5-phosphate (DXP) and mevalonate (HMGR) isoprenoid precursor biosynthetic pathways.....	43
Figure 3.4	Transcript abundances for monoterpene synthases in coriander at three developmental stages.....	44
Figure 3.5	Comparison of relative transcript abundances for DXS1, DXS2 and HMGR genes across three coriander seed developmental stages (S1, S2 and S3) between quantitative RT-PCR and RNA-seq data.....	49
Figure 4.1	Multiple protein sequence alignment of Coriandrum sativum $\gamma$ -terpinene synthase (Cs $\gamma$ TRPS), (S)-linalool synthase (CsLINS), and two other putative C. sativum monoterpene synthases (mTPS2 and mTPS3) against five known plant monoterpene synthases. ....	71
Figure 4.2	Phylogenetic tree showing the evolutionary relationships between plant terpene synthases, including Cs $\gamma$ TRPS and CsLINS.....	74
Figure 4.3	PCR amplification of Cs $\gamma$ TRPS.....	75
Figure 4.4	PCR amplification, gel extraction and diagnostic RE digestion of full length mTPS3.....	76
Figure 4.5	PCR amplification with iProof DNA polymerase, gel extraction and diagnostic RE digestion of truncated mTPS3 with iProof DNA polymerase.....	77
Figure 4.6	Sticky-end PCR amplification and diagnostic PCR amplification of truncated mTPS3.....	79
Figure 4.7	PCR amplification using iProof DNA polymerase and diagnostic RE digestion of mTPS2.....	80
Figure 4.8	Protein expression of mTPS3 at 18, 23, 30 and 37 °C.....	81
Figure 4.9	SDS-PAGE images of protein extracted from Rosetta E. coli cells containing mTPS3 protein of interest.....	82

Figure 4.10	SDS-PAGE of C $\gamma$ TRPS protein extracted from Rosetta (DE3) pLysS bacterial expression cells and purified by Ni-NTA affinity chromatography. ....	83
Figure 4.11	Gas chromatograms (GC) and mass spectrometry (MS) fingerprint for volatile terpene products of C $\gamma$ TRPS.....	84
Figure 4.12	Kinetics data for C $\gamma$ TRPS with GPP substrate.....	85
Figure 4.13	Divalent metal cation cofactor preference and optimization of C $\gamma$ TRPS.....	86

## LIST OF ABBREVIATIONS

MES -	2-(N-Morpholino) ethanesulfonic acid
HMG-CoA -	3-Hydroxy-3-methylglutaryl-CoA
HMGS -	3-Hydroxy-3-methylglutaryl-CoA synthase
MOPSO -	3-Morpholino-2-hydroxypropanesulfonic acid
CDP-ME -	4-Diphosphocytidylyl-methylerythritol
ALP -	Actin Related Protein
ACP -	Acyl Carrier Protein
BLAST -	Basic Local Alignment Search Tool
BSA -	Bovine Serum Albumin
CMK -	CDP-ME Kinase
CDP-MEP -	CDP-ME-2-phosphate
cDNA -	Complementary DNA
CDP -	Copalyl Diphosphate Synthase
Cs $\gamma$ TRPS -	Coriandrum sativum $\gamma$ -Terpinene Synthase
CsLINS -	Coriandrum sativum Linalool Synthase
DMAPP -	Dimethyl Allyl Diphosphate
DTT -	Dithiothreitol
DXR -	DXP Reductoisomerase
DXS -	DXP Synthase
EO -	Essential Oil
EST -	Expressed Sequence Tag
FPP -	Farnesyl Diphosphate
FPPS -	Farnesyl Diphosphate Synthase
GC-MS -	Gas Chromatography-Mass Spectrometry
GO -	Gene Ontology
GPP -	Geranyl Diphosphate
GPPS -	Geranyl Diphosphate Synthase
GGPP -	Geranyl Geranyl Diphosphate
GGPPS -	Geranyl Geranyl Diphosphate Synthase
GSH -	Glutathione
GST -	Glutathione S-transferase
G3P -	Glyceraldehyde-3-phosphate
HDR -	HMBPP Reductase
HDS -	HMBPP Synthase
HMGR -	HMG-CoA Reductase
HBMPP -	Hydroxymethylbutenyl-4-diphosphate
IPP -	Isopentenyl Diphosphate
IPPI -	Isopentenyl Diphosphate Isomerase
IPTG -	Isopropyl- $\beta$ -D-thiogalactopyranoside
KAAS -	KEGG Automatic Annotation Server
KEGG -	Kyoto Encyclopedia of Genes and Genomes
LB -	Lauria Bertani
MDS -	ME-CPP Synthase
MCT -	MEP Cytidylyltransferase

mRNA -	Messenger RNA
MEP -	Methyerythritol-4-phosphate
ME-CPP -	Methyerythritol-2,4-cyclodiphosphate
MK -	Mevalonate Kinase
MDC -	Mevalonate-5-diphosphate Decarboxylase
mTPS -	Monoterpene Synthase
NCBI -	National Center for Biotechnology Information
NIST -	National Institute of Standards and Technology
NPP -	Neryl Diphosphate
NGS -	Next Generation Sequencing
NRT -	No Reverse Transcriptase Control
NPCR -	No Template Control
NI -	Non-induced Control
ORF -	Open Reading Frame
PMSF -	Phenylmethanesulfonylfluoride
PBS -	Phosphate-buffered Saline
PMK -	Phosphomevalonate Kinase
PCR -	Polymerase Chain Reaction
PTV -	Programmable Temperature Vaporizing
qPCR -	Quantitative Polymerase Chain Reaction
RACE -	Rapid Amplification of cDNA Ends
RPKM -	Reads Per Kilobase of Exon Model Per Million Mapped Reads
RT-PCR -	Reverse Transcription Polymerase Chain Reaction
RNA-seq -	RNA Sequencing
SEM -	Scanning Electron Microscopy
sTPS -	Sesquiterpene Synthase
STR -	Short Tandem Repeat
SNP -	Single Nucleotide Polymorphism
SDS-PAGE -	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SOC -	Super Optimal Broth
TPS -	Terpene Synthase
TAIR -	The Arabidopsis Information Resource
TRIS -	Tris(hydroxymethyl)aminomethane

## **ACKNOWLEDGEMENTS**

Firstly, I would like to thank my supervisor, Dr. Soheil Mahmoud, for offering me the opportunity to work in his lab and for sharing his expertise and guidance throughout my project. I also thank my committee members, Drs. Kirsten Wolthers and Mark Rheault. I thank Dr. Wolthers and her graduate students, for their assistance with protein expression and purification trouble shooting as well as the generous donation of certain reagents and bacterial cells. I thank Dr. Rheault and his laboratory technician for their qPCR expertise and access to Dr. Rheault's equipment. A thank you goes to Dr. Slava Ilyntskyy at the University of Lethbridge for tirelessly helping me understand the world of RNA-seq. I also thank Mr. David Arkinstall at UBCO for his technical assistance with the SEM imaging. A special thank you goes to all those involved with keeping the biology and biochemistry departments running smoothly, especially Barb Lucente and Jenny Janok who always go out of their way to help graduate students in this department. Of course, thanks goes to my incredible senior lab-mates, Lukman Sarker and Zerihun Demissie who acted as mentors and expert trouble shooters throughout my project. I also thank my fellow graduate students, many with whom I have become great friends, for all their support and excellent company, while together we rode the roller coaster that is, life as a graduate student.

This project was funded through grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation, and University of British Columbia, awarded to Dr. Soheil Mahmoud.

## **DEDICATION**

For my family, for all their ‘cheer-leading’ when things were ‘up’ and ‘chin-raising’ when things were ‘not so up’. I couldn’t have done it without you.

## CHAPTER 1: INTRODUCTION

### 1.1 *Coriandrum sativum*

*Coriandrum sativum* is a hardy annual plant belonging to the *Apiacea* family, which is natively found growing in Mediterranean Europe and Western Asia. See Table 1.1 for coriander's scientific classification. It is now being cultivated in various temperate countries (Reuter, 2008). The term cilantro refers to the immature *C. sativum* plant, or simply the leaves. This name was derived from the Spanish common name for the plant, and since "cilantro" is widely used in Spanish speaking regions, cilantro has become the commonly accepted name for the immature leafy portions of this plant. Upon reaching maturity, the dry fruits, or seeds, are what have become known as coriander fruits. The spicy aroma of coriander has been described as warm and nutty with a hint of citrus but some describe the flavor to be soapy or bitter (Kubo, 2004 and Msaada, 2009b).

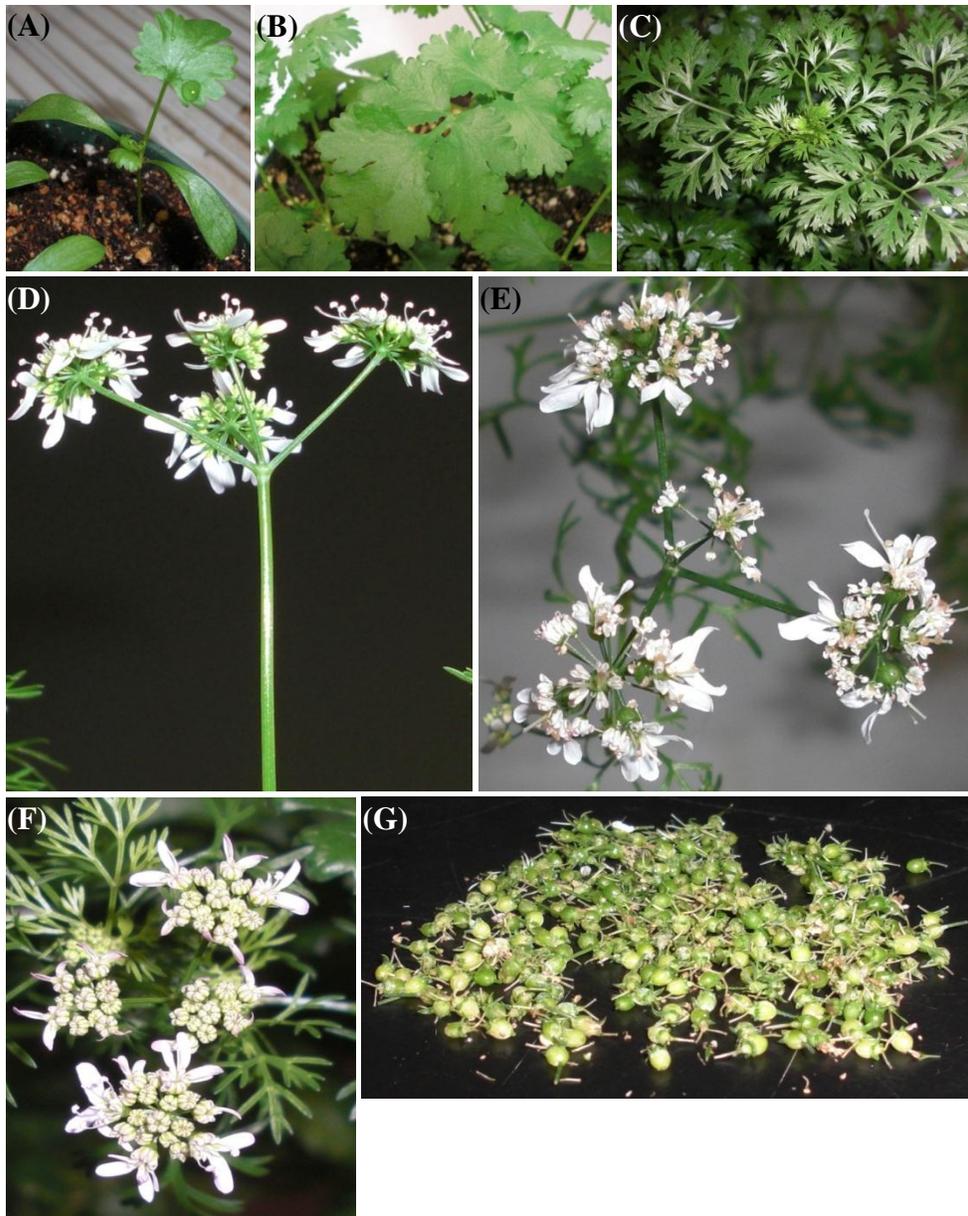
**Table 1.1** Scientific classification of the cilantro/coriander plant.

Kingdom	Plantae
Division	Angiospermae
Class	Dicotyledonae
Series	Calyciflorae
Order	Apiales
Family	Apiaceae
Genus	<i>Coriandrum</i>
Species	<i>sativum</i>

*C. sativum* grows to ca. 25-60 cm or 6-24 inches in height. Its umbel flowers are pale pink to white, Figure 1.1 D-F, and emit a potent fragrance, characteristic of cilantro and coriander. This plant's natural flowering season is from June to July, at which point it yields the round fruits known as coriander. There are two varieties of *C. sativum*: *vulgare* and *microcarpum*. The former has larger fruits with essential oil (EO) yields of 0.1-0.35% (v/w) while the latter has smaller fruits with EO yields of 0.8-1.8% (v/w) (Burdock, 2009). The coriander seed, is a globular dry schizocarp, about 3-5 mm in size, with two mericarps (Yeung, 2011), Figure 1.1 G. One to two true seeds can be found within each schizocarp. This seed is composed of three tissues: outer seed coat,

endosperm, and the embryo. The embryo itself is very small in comparison to the endosperm, which provides nutrients to the developing embryo at the time of germination.

The dried *C. sativum* seed EO is a blend of monoterpenes. Linalool is the most abundant terpene found in coriander EO, making up approximately 72% of the total terpene content. The oil composition changes, depending on the maturity of the seed. As it matures, it develops from a small green, to a large brown fruit. If its development were broken into four stages, at the first stage, EO composition includes 36% linalool, 35% geranyl acetate, and assorted mono- and sesquiterpenes, each in trace amounts. The second stage is characterized by 40% linalool, 8% geranyl acetate, 4% camphor, 3% menthol, and other mono- and sesquiterpenes in trace amounts. At the third stage the oil composition consists of 45% linalool, 22% monoterpene esters (ex. geranyl acetate), mono- and sesquiterpene hydrocarbons (e.g., limonene) and ketones (e.g., camphor) were in reduced amounts. In the fourth stage, the EO terpene content is dominated by 78% monoterpene alcohols, with 72% of this being linalool, 5% monoterpene hydrocarbons, 2% monoterpene esters, and only 1% monoterpene ketones. These differences in the EO content of *C. sativum* seeds suggests there are modifications to the specialized metabolism as the seeds mature (Msaada, 2009b). These terpenes have been found to be of great importance as they are multi-industrially useful, and they can be obtained from their plant sources by steam distillation.



**Figure 1.1** *Coriandrum sativum* (A) seedling, (B) young foliage, (C) mature foliage, (D) characteristic umbel organization of flowers, (E) late flower blossoms, (F) early flower blossoms and (G) fresh seeds.

### 1.1.1 Uses of Coriander

Coriander has a long history of use which goes all the way back to ca. 400 BC where it was used as a Greek traditional medicine by Hippocrates. Greeks and Romans also used

it to flavor their wine. This herb was dubbed the “spice of happiness” by the Egyptians which considered it to be an aphrodisiac. They also used it as a cooking ingredient and to treat digestive upsets. When it was finally introduced to Great Britain in the 13<sup>th</sup> century, it was employed in midwifery to accelerate the child birthing process, and thus the seeds of *C. sativum* have been widely used in a variety of manners for over 7000 years (Burdock, 2009).

Today, the EO distilled from coriander can be used in certain food condiments and liqueurs (Burdock, 2009). These seeds are a rich source of lipids, 28.4% of the total seed weight, which may be of great importance in the food industry (Yeung, 2011). Alternatively, *C. sativum* seeds can be ground up and used to spice Indian cuisine, for example, Indian curry. The fresh cilantro leaves are also highly valued in both Indian and Mexican cuisine, where they are the key ingredient in dishes such as Mexican salsa and guacamole. In parts of Asia, e.g., Thailand, the roots are pounded into a paste and used to flavor curries and soups.

There have been studies which found that coriander EO has antimicrobial properties, and is highly effective against many gram positive and gram negative bacteria, for example, *Staphylococcus aureus* (Delaquis, 2002; Singh, 2002; Lo Cantore, 2004; Kubo, 2004 ). Coriander EO was also shown to possess anti-bacterial properties against *Salmonella choleraesuis*, the pathogen responsible for one of the most frequently occurring food borne illnesses, salmonellosis, thus suggesting that addition of coriander EO in certain foods can act as a food spoilage preventative (Kubo, 2004). One study found small quantities of coriander EO to exhibit moderate anti-inflammatory activity when used on UV-B induced cases of erythema (Reuter, 2008). This seed EO was also found to improve blood glucose control *in vitro* and thus it held promise for use as an antihyperglycemic (Gallagher, 2003). Additionally, coriander EO has been reported to possess antioxidant (Wangensteen, 2004), anticarcinogenic and antimutagenic properties (Chithra, 2000).

The pleasant fragrance of coriander EO, combined with its antimicrobial activity, make it a desirable candidate for use in antiseptics, skin care products, and various cosmetics. The pharmaceutical industry uses this oil as an additive in certain oral

medicines to improve the foul taste, giving them more pleasant flavours. Coriander EO has also been used in bactericides against plant pathogenic bacteria in the agricultural industry (Lo Cantore, 2004).

## **1.2 Terpenoids**

Terpenoids or isoprenoids encompass over 40,000 structurally and functionally diverse natural products. Although present in all living organisms, they are most abundant in plants, where they perform essential functions in growth and development (e.g., as plant growth regulators) and crucial ecological roles (e.g., in defense and pollinator attraction). Plant terpenoids are natural metabolic products, made up of five carbon compounds (isoprene units) which can be biosynthesized ubiquitously in all plant tissues including the: leaves, flowers, buds, stems, roots, and seeds (Bakkali, 2008). There are seven major classes of terpenoids, categorized according to the number of isoprene units that make up their backbone structure (Yazaki, 2006; Buchanan, 2002). These include the mono- ( $C_{10}$ ), sesqui- ( $C_{15}$ ), di- ( $C_{20}$ ), sester- ( $C_{25}$ ), tri- ( $C_{30}$ ), tetra- ( $C_{40}$ ), and polyterpenes ( $C_n$ ).

### **1.2.1 Natural Functions and Industrial Uses of Terpenoids**

Terpenoids were traditionally considered secondary metabolites. However, they have actually been found to play a huge variety of important roles in plants, such as in pollinator attraction, by way of pigmentation (e.g., lycopene) (Fraser, 2004) or aromas, thus, these compounds are now termed “specialized metabolites”. The monoterpenes, linalool and camphene, serve as semiochemicals in pollinator attraction (Mahmoud, 2002). Some terpenoids with antimicrobial properties play defensive roles in plants (Pichersky, 1995), such as the sesquiterpene,  $\alpha$ -cedrene (Barrero, 2005), and the diterpene, horminone (Ulubelen, 2003). There are also indirect methods of defense, manifested in the form of tritrophic level interactions. For example, *Phaseolus lunatus* (lima bean) defends itself from herbivorous spider mites by releasing a cocktail of mono- and sesquiterpenes which serve as chemical attractants to predaceous mites (Bruin, 1992 and Mumm, 2008). Terpenoids can also give rise to hormones such as, gibberellins,

abscisic acid, and phytosterols, which are actively involved in regulation of plant growth and development (McGarvey, 1995). Moreover, terpenoids are found as prenyl side chains, essential for biological activity of some proteins, polysaccharides, and other important cellular metabolites (Rodriguez-Concepcion, 2010; Fraser, 2004).

In addition to plants' natural uses for terpenoids, there are many industrial uses for these natural compounds. Some of these include fragrances in hygiene products, cosmetics, perfumes, air fresheners, and cleaning products (Croteau, 2005; Pichersky, 1995). Terpenoids are also widely used in the culinary industry as flavoring ingredients in various foods and beverages (Pichersky, 1995); for example, the monoterpenes pinene and linalool are responsible for giving food a citrus flavor (Msaada, 2009b). Several bioactive terpenoids are of great medicinal value as they have been shown to exhibit certain medicinal properties such as the diterpene, taxol, from the yew tree which is used as a chemotherapeutic agent (Koshroushahi, 2006; Wu, 2003) or the sesquiterpene from the *Artemisia* genus which is a proven anti-malarial compound (Brown, 2010; Nafis, 2011).

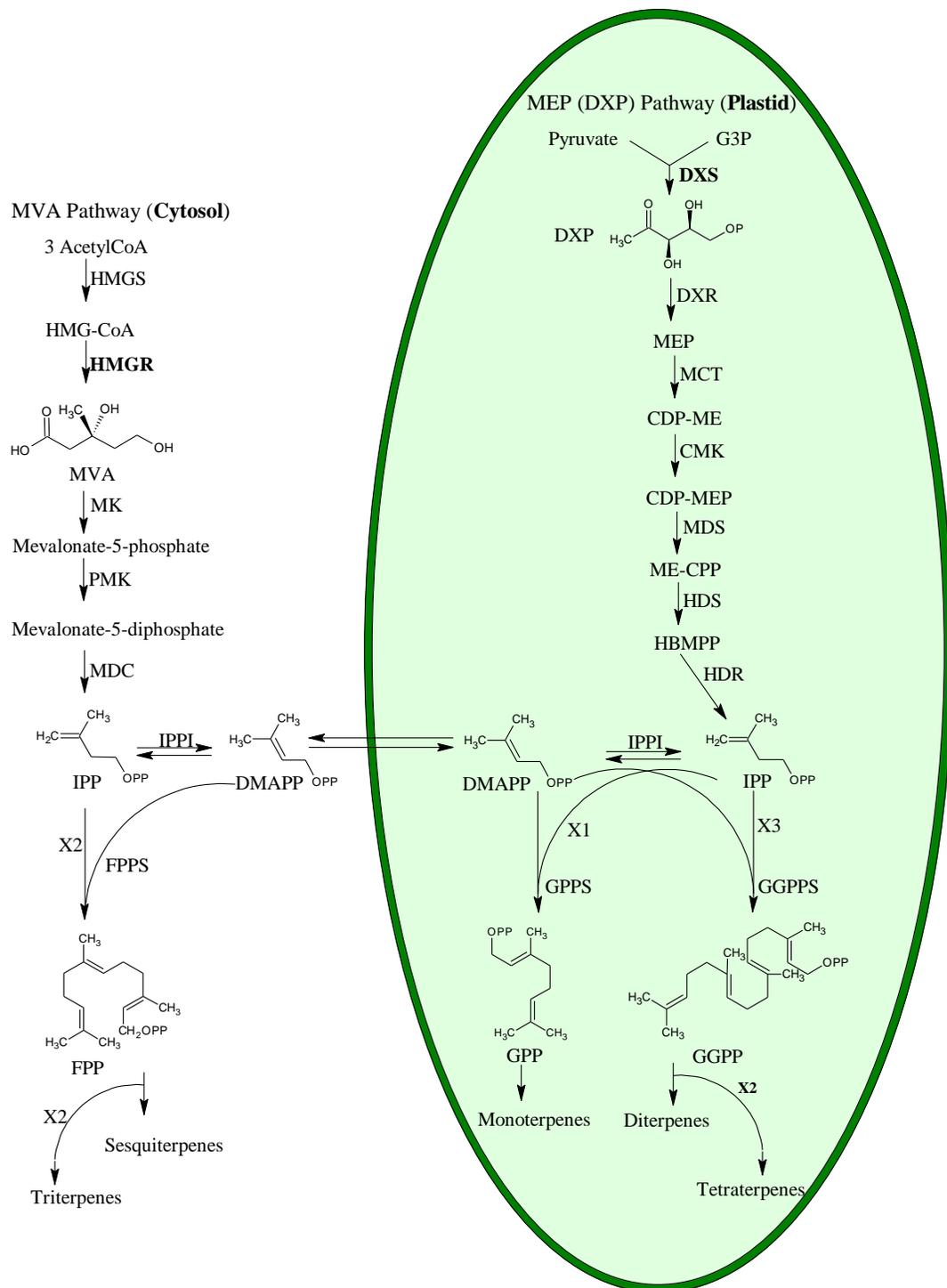
### **1.2.2 Biosynthesis of Terpenoids**

Certain terpenoids (e.g., essential oil constituents) are cytotoxic to plants, especially in concentrated amounts. Therefore, those plants which accumulate large quantities of such terpenoids as essential oils or defensive resins use specialized structures such as glandular trichomes or resin ducts, for storage of these compounds. In this way, the toxic terpenoids are sequestered in certain locations of a plant (Bakkali, 2008).

Biosynthesis of isoprenoids occurs in four general stages. First is the synthesis of isopentenyl diphosphate (IPP) and its isomer dimethyl allyl diphosphate (DMAPP) via the mevalonate (MVA) and 1-deoxyxylulose-5-phosphate (DXP), aka. methylerythritol-4-phosphate (MEP) pathways, which are located in separate compartments within a plant cell (Adorjan, 2010; Rodriguez-Concepcion, 2010), Figure 1.2. The DXP pathway is located in the plastid and through this pathway, mono-, di-, and tetraterpenes are produced (Hasunuma, 2008; Rodriguez-Concepcion, 2010). The MVA pathway is

located in the cytosol, where it is responsible for the production of sesqui-, tri-, and polyterpenes (Ganjewala, 2009; Rodriguez-Concepcion, 2010).

The second phase of terpenoid biosynthesis involves the formation of prenyl diphosphates, the linear precursors to various isoprenoids (Mahmoud, 2002; Rodriguez-Concepcion, 2010). During this stage, one molecule of each of DMAPP and IPP are condensed to produce geranyl diphosphate (GPP) (McGarvey, 1995; Buchanan, 2002), the linear precursor to monoterpenes. Two IPPs and one DMAPP are condensed to yield farnesyl diphosphate (FPP), the precursor for sesqui- and triterpenes. Three IPPs and one DMAPP will condense to form geranyl geranyl diphosphate (GGPP), the precursor for di- and tetraterpenes (Mahmoud, 2002), Figure 1.2. The third step involves the synthesis of parent skeletons for each terpenoid class from the respective prenyl diphosphates (GPP, FPP and GGPP), via terpene synthases (Nagegowda, 2010; Buchanan, 2002). Finally, there is modification of the terpene parental skeletons, which include electrophilic addition of side groups or rearrangement of the molecule to give the final terpene products (McGarvey, 1995; Buchanan, 2002).



**Figure 1.2** Mevalonate (MVA) and 1-deoxyxylulose-5-phosphate (DXP) terpenoid prenyl precursor biosynthetic pathways. X1 (one of designated molecule), X2 (two of designated molecule), X3 (three of designated molecule).

### 1.3 Terpene Synthases

Several monoterpene synthases have been purified and functionally characterized from angiosperms, gymnosperms, and bryophytes. All have very similar properties, except that angiosperm monoterpene synthases require a divalent metal cation for activity, while those of gymnosperms require monovalent metal cations, as well as a higher optimal pH, for activity. Terpene synthases (TPS) are soluble enzymes, localized to either the plastid (mono- and diterpene synthases), or the cytosol (sesquiterpene synthases) (Sugiura, 2011) as mentioned before. The plastidially targeted terpene synthases have N-terminal transit, or signal peptides, which transport nuclear encoded terpene synthase immature proteins to the plastids, e.g., leucoplasts, for processing into mature proteins. These transit peptide sequences are characteristically rich in serine and threonine and lack acidic amino acids. During expression of these cloned terpene synthases into active protein by plasmid expression vectors in *E. coli*, this transit peptide can promote the formation of inclusion bodies, thus compacting the terpene synthase's shape and rendering it inactive (Bohlmann, 1998). By truncating these transit peptides from the terpene synthase sequence, high yield soluble protein can be obtained.

Most plant terpene synthases have a characteristic ability to yield multiple products from a single substrate with high regio- and stereospecificity (Tholl, 2006; Degenhardt, 2009). For example, the (-)-pinene synthases from Douglas and Grand Fir can produce both (-)- $\alpha$ -pinene and (-)- $\beta$ -pinene from GPP (Sugiura, 2011). A suggested reason for this is that some of the minor products formed may have been derived from reaction intermediates which underwent early termination during their metabolic pathway (Degenhardt, 2009); also some carbocationic intermediates generated during the reaction may have many possible metabolic fates (Chen, 2003). Another suggestion is that of multi-product formation as being an evolutionary adaptation of TPSs to produce the greatest number of products using the least amount of genetic and enzymatic machinery as possible. Active site conformation flexibility of the TPSs has also been proposed as a factor contributing to multi-product formation from a single TPS (Degenhardt, 2009). Examples of multi-product formation are seen in  $\gamma$ -humulene synthase of *Abies grandis*,

which produces 52 different sesquiterpenes from the FPP substrate (Steele, 1998) and sabinene synthase from Sage which produces five monoterpene products from the GPP substrate (Wise, 1998).

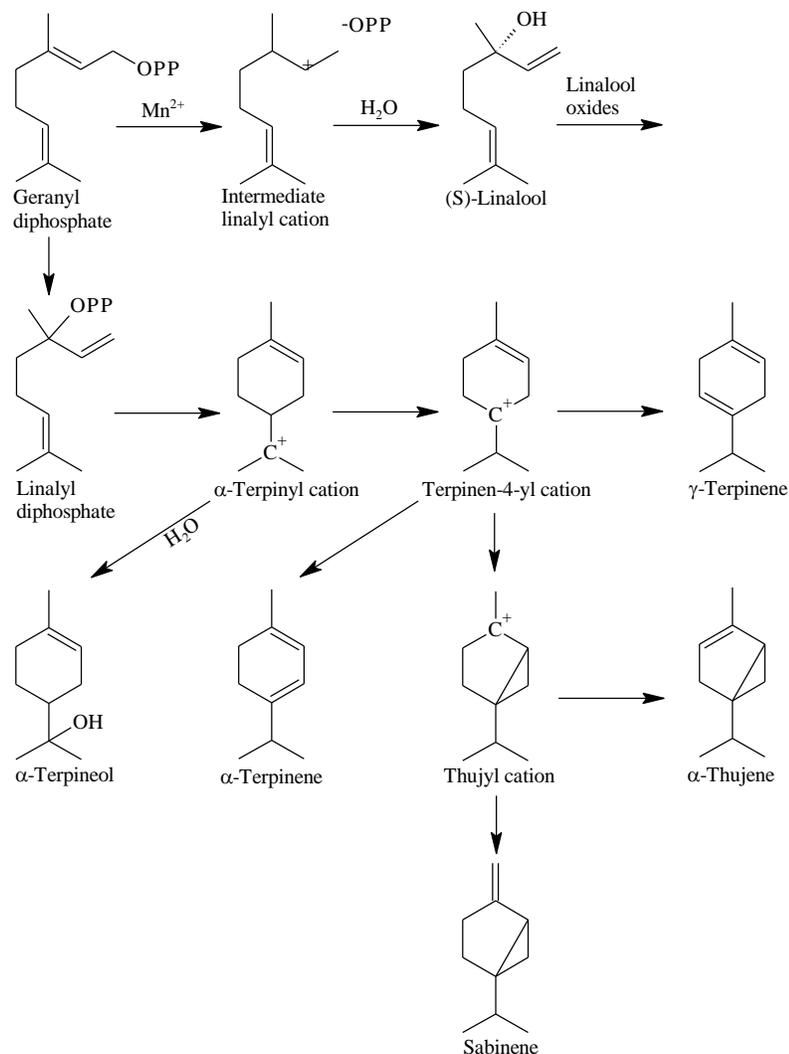
Plant TPS diversity seen today has evolved over time as a byproduct of breeding various plant cultivars, due to variation of alleles and subtle differences in the primary structure of closely related TPS genes. There are situations where alleles which code for non-functional TPS genes are transcribed, resulting in non-functional proteins, or pseudo proteins. In basil, the mixture of mono- and sesquiterpenes in the EO are the products of variations in the expression of several functional TPS genes. Variations in expression can arise in the form of any post-transcriptional regulation such as inhibitory regulatory proteins (blocking RNA expression) and alternative RNA splicing. This can result in finding additional terpene products *in vivo* in basil which would not otherwise be found as products of the bacterial expression systems generally used to clone TPS genes (Iijima, 2004).

Enzymes in the terpene biosynthetic pathway have been classified into seven major classes or subfamilies, TPS-a to TPS-g (Bohlmann, 1998; Dudareva, 2003). All the members within these classes share at least 40% similarity (Bohlmann, 1998). The most extensively studied TPS families include, TPSa, TPSb, TPSd, and the distantly related TPSf, which contains linalool synthase. TPSa includes the angiosperm sesquiterpene synthases, TPSb includes the angiosperm monoterpene synthases (mTPS), TPSc includes copalyl diphosphate synthases ((-)-CDP) (diterpenoid biosynthesis), and TPSd includes the gymnosperm mTPSs. TPSa, TPSb, and TPSd share the C-terminal active site domain and the same N-terminal domain. Mono-TPSs are generally around 600-650 amino acids in length, approximately 50-70 amino acids longer than sesquiterpene synthases (sTPS). This length difference is mostly due to the N-terminal transit peptides necessary for plastidial targeting of mTPSs.

All TPSs contain the aspartate rich DDXXD motif necessary for coordination of the interaction between divalent metal ions and the diphosphate moiety of the substrate. Another motif which is at least partially conserved between TPSs is the NSE/DTE motif with a sequence of (L,V)(V,L,A)-(N,D)D(L,I,V)x(S,T)xxxE, which also helps to

coordinate with the DDXXD motif during binding of the diphosphate substrate (Degenhardt, 2009). Another relatively well conserved motif is the tandem arginines motif, RRX<sub>8</sub>W, approximately 60 residues upstream from the N-terminus of a TPS, involved in isomerization of GPP to a cyclization intermediate, only found in terpene synthases which produce cyclic products. The absence of the RRX<sub>8</sub>W motif from an mTPS results in the production of acyclic terpenes (Williams, 1998).

Hemi-, mono-, sesqui-, and diterpene synthases are evolutionarily related to each other and structurally distinct from tri- and tetraterpene synthases (Tholl, 2006). Within the mTPSs, all share a common carbocation mechanism of action, which begins with a metal ion-dependent ionization of the substrate, yielding a cationic intermediate. This intermediate then undergoes a series of rearrangements including cyclizations or hydride shifts until there is proton loss or a nucleophile is added (Degenhardt, 2009) (Figure 1.3).



**Figure 1.3** The prenyl precursor, geranyl diphosphate, can give rise to a large variety of monoterpene products. These are just a few examples of the monoterpenes found in coriander EO.

#### 1.4 Strategies of Gene Cloning

Cloning a gene is the process by which a clone of a particular gene is made. To date, there are different approaches to gene cloning. Firstly, the nucleic acid sequence information for a target gene must be obtained and then that sequence is inserted into a host organism which will produce many copies of the gene of interest. *Escherichia coli* bacteria is generally the host organism chosen for the production of recombinant DNA as

this prokaryotic model organism is very widely studied, is quite easily and inexpensively grown and has harmless strains available for molecular work (Glick, 2010).

Applications of gene cloning include gene expression analysis and studies of gene functionality, for example, via site directed mutagenesis (Alba, 2004; Takeda, 1998). Gene cloning can also lead to the production of recombinant proteins in bacterial expression systems. An example of this would be in gene therapy where cells which lack a particular functional gene resulting in a genetic disease, can be cured by providing those cells with that missing gene (Pfeifer, 2001). Additionally, transgenic organisms can be produced from cloned genes. For example, in the agriculture and pharmaceutical industries, numerous improved plants (e.g., herbicide or pest resistant crops as well as plants which produce enriched desirable products) are the result of gene cloning leading to the insertion of genetically modified material into those plants (Rigano, 2013; Zaidi, 2012).

Approaches to gene cloning include the use of degenerate primers (Linhart, 2005), construction of genomic or cDNA libraries (Alba, 2004), and transcriptome sequencing (Bentley, 2006). Degenerate primers have some nucleic acid positions which can code for a number of different nucleotides. The degeneracy of a primer is defined by the number of nucleotide sequence combinations that primer contains as a result of the multiple base possibilities at various nucleotide positions. Various nucleic acid triplets or codons, encode certain amino acids. For example, three different codons encode isoleucine. By using degenerate primers, related yet distinct nucleotide sequences can be amplified in addition to nucleic acid targets for which only their protein sequences are known (Kwok, 1994; Linhart, 2005). In order to design appropriate degenerate primers, the protein sequences of genes related to the target gene are aligned and searched for conserved regions. Degenerate primers are then designed at these conserved regions (Shen, 2003). A target sequence is first amplified via the degenerate primers and cloned into a cloning or sequencing vector. The plasmids are then sequenced to determine the actual nucleic acid sequence of the amplified target, generally the middle portion of a gene sequence. Using this known portion of sequence, a new specific primer set can be designed to determine

the ends of the gene sequence using rapid amplification of cDNA ends (RACE) (Frohman, 1988).

Nucleic acid libraries are methods for cataloging either the entire genomic or transcriptomic content of an organism. This approach first clones as many genes as possible, and then from that, finds the genes of interest. This is done by transforming single genes into individual cells in a population of host bacteria. These methods enable the study of individual genes or trends in genes, for example. There are two types of nucleic acid libraries: genomic and cDNA. The former represents an organism's entire genome while the latter represents an organism's entire transcriptome, or only those genomic portions which make up coding regions of DNA. If studying the regulatory or intron regions of a gene, a genomic library would be the most useful of the two because this type of sequence information is absent from the fully processed mRNA's used to make the cDNA library. Genomic libraries are also useful when looking to clone a gene where the cell type in which it is expressed is unknown. However, if the coding region of a gene is preferred, then a cDNA library would be most useful. These cDNA libraries are also useful in studying the specialized functions of particular cell types or studying changes in gene expression patterns over the course of development or in response to certain environmental signals (Campbell, 2005).

Complementary DNA libraries are extremely valuable for the discovery of new genes and assignment of gene function. This gene discovery works through expressed sequence tag sequencing. Expressed sequence tags (EST) are a reliable and effective source of gene expression data. ESTs are generated by sequencing either the 5' and or 3' ends of randomly isolated gene transcripts which have been reverse transcribed into cDNA (Alba, 2004). An EST represents only a portion of the coding sequence (200-900 nucleotides); it is generated by partial sequencing of large numbers of individual clones from a cDNA library. These ESTs serve as tags which represent expressed or transcribed regions of genes; they are not supposed to serve as a whole gene sequence but are of sufficient quality that they will have high homology to a portion of a known gene sequence in a database. Thus, putative protein coding regions can be identified within a cDNA library. This is a quick and relatively inexpensive method for discovering new genes. ESTs can

also be used as genetic markers in mapping projects or in the design of PCR primers in situations where ESTs can be generated from both ends of a cDNA clone (Bouck, 2007). However, because EST's represent only those genes which are actively transcribed, it is not possible to use this method of sequencing alone to obtain the whole organism's genetic content (Alba, 2004).

The construction of a cDNA library begins with the extraction of mRNA (or total RNA) from the tissue of interest. Individual mRNA molecules are reverse transcribed to cDNA, cloned into an appropriate vector, and then transformed into *E. coli* hosts. A large and random assortment of colonies is selected for partial sequencing of the cloned genes and thus, a cDNA library is created (Campbell, 2005). Disadvantages of cDNA libraries include their poor sensitivity to low abundance transcripts and the large amount of resources and time required for the cloning and sequencing involved in this method (Alba, 2004).

### **1.5 Second Generation Sequencing - Illumina Technology**

Second - or next - generation sequencing (NGS) technology has provided the ability to perform millions of sequencing reactions at extremely high rates in parallel (Bentley, 2006). As NGS has gained popularity over the years it has become increasingly cost-effective, especially in the case of whole transcriptome or genome sequencing (Wall, 2009). Sequencing a single human genome via the Sanger method with capillary electrophoresis would take approximately 30 instruments for one year and cost around one million dollars. NGS allows 24-48 human genomes to be sequenced for about one hundred thousand dollars (Bentley, 2006).

There are different NGS technologies available such as, sequencing by synthesis and sequencing by ligation (Niedringhaus, 2011), and each technology is slightly modified for the various sequencing platforms which have been developed. Illumina and Pyrosequencing are two platforms which use sequencing by synthesis technology in which fluorescently labeled nucleotides are added to fragmented DNA templates. As each nucleotide is incorporated a fluorescent signal is given off and used to identify which

nucleotide was added, thus the DNA sequence is obtained as the complementary DNA strand is synthesized.

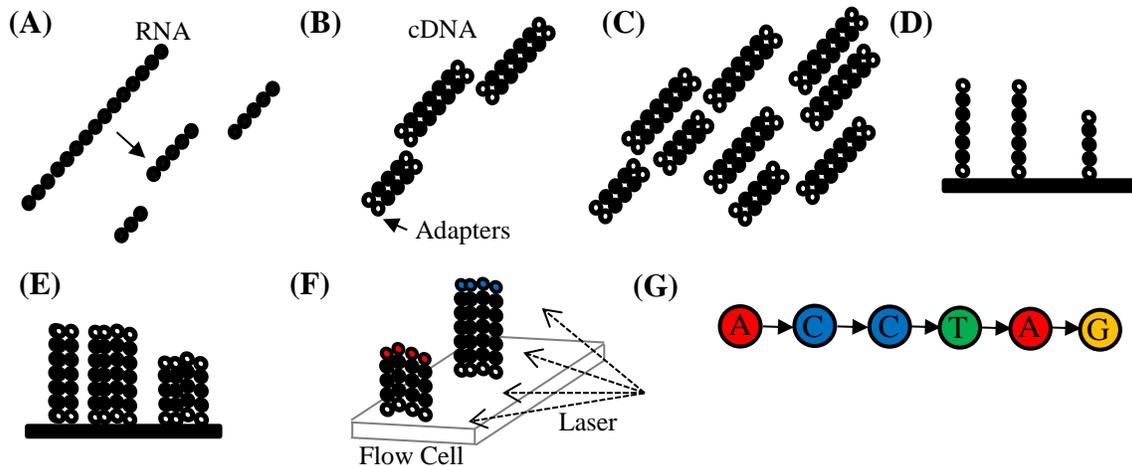
In this project, Illumina sequencing by synthesis technology was used. The Illumina Genome Analyzer IIX is capable of producing up to 640 million short paired-end reads per flow cell. The workflow of this sequencer involves three components: library preparation, cluster generation, and sequencing. Beginning with library preparation, the RNA templates are first assessed for integrity and poly A tail selected. The templates are then chemically fragmented and reverse transcribed to produce cDNA. Blunt ends of the cDNA are extended with polyA overhangs to allow for adapter ligation. The DNA templates are then PCR enriched and quantified (Illumina Inc, 2010). In the following stage, cluster generation, the flow cell surface is covered with a lawn of oligos to which the double stranded DNA templates are bound via the oligo adapters. Each library template is clonally amplified, resulting in hundreds of millions of unique clusters. Upon completion of this amplification the reverse strands are cleaved and washed away, leaving behind single stranded DNA templates (Illumina Inc, 2010).

To begin the sequencing, template ends are blocked and hybridized with the sequencing primers. The hundreds of millions of clusters are all sequenced simultaneously in parallel. Fluorescently labeled, reversibly terminated nucleotides are added to the flow cell. After each nucleotide incorporation cycle (round) the clusters are excited by a laser, leading to emission of a colour which is used to identify the newly incorporated base. At this point the fluorescent label and blocking group are cleaved off, allowing for addition of the next nucleotide. This sequencing cycle is repeated to determine the complete sequence of the bound template, one base at a time. The main disadvantage with new high throughput sequencing technologies such as Illumina is the challenge of finding assembly software that can deal with the massive amounts of short read data which is produced (Liu, 2011). A summary of Illumina sequencing can be seen in Figure 1.4.

Unlike Illumina sequencing, Pyrosequencing does not reversibly terminate DNA synthesis after incorporation of each new nucleotide. Because of this, Illumina sequencing can more accurately read through homopolymers or any highly repetitive

regions (Ahmadian, 2006). However, Pyrosequencing does yield longer reads than Illumina sequencing (230-400 vs. 75-100), which reduces the sequence assembly burden (Niedringhaus, 2011).

Sequencing by ligation, unlike sequencing by synthesis, does not use DNA polymerase to create a second DNA strand and thus determine the nucleic acid sequence by synthesis, but rather uses DNA ligase to determine the DNA sequence. DNA ligase is sensitive to DNA structure and will only ligate to nucleic acid strands which have true base-pair complementarity. A single stranded target DNA molecule is flanked by known sequence primers and very short fluorescently labeled oligonucleotides which represent all the possible variations of complementary bases are added. Once the complementary oligonucleotide binds, a fluorescent signal is released and the nucleotide incorporated is determined from that signal. The next round of labeled oligonucleotides are then added. By using these short oligonucleotides (~9 nucleotides long), only every Nth nucleotide on the DNA strand is sequenced. Thus, the same DNA fragment is sequenced more than once with oligonucleotides of different lengths to cover any gaps not sequenced during the preceding round(s). Since the same DNA fragment is sequenced multiple times, this improves the sequence accuracy however; this technology disallows the production of reads longer than 25-35 base pairs, making the read assembly process more difficult than with sequencing by synthesis (Niedringhaus, 2011; Myllykangas, 2012).



**Figure 1.4** Illumina sequencing workflow. (A) Chemical fragmentation of RNA template, (B) cDNA synthesis and addition of adapters, (C) cDNA enrichment via PCR amplification, (D) single stranded cDNA templates attached to flow cell surface and (E) clonal PCR amplification. (F) Sequencing by synthesis occurs via the addition of sequencing reagents including four fluorescently labeled reversibly terminated nucleotides. (G) The fluorescence emitted by each cluster upon laser excitation is captured after addition of each nucleotide.

## 1.6 De novo Transcript Assembly

Perhaps the greatest limitation of NGS technologies is the short length of raw reads produced. It is very difficult to assemble such short reads into accurate gene/transcript sequences due to issues including the great impact which base calling errors have in short reads and the difficulty of assembling long repetitive regions which span a length greater than that of a read (Paszkiwics, 2010). Today, various software have been developed to counter this problem; there are upwards of 10 publicly available programs including, ABySS, SOAP denovo and Velvet. Re-sequencing, in which reference sequence information was previously available, made short read assembly of re-sequenced organisms possible via alignment methods during the sequence validation process. However, more complex assembly algorithms became necessary when *de novo* sequencing arose, a situation where no reference sequence information is available (Bentley, 2006). In this project the Oasis transcript assembler based on the Velvet program was used to assemble the raw Illumina reads into transcript sequences.

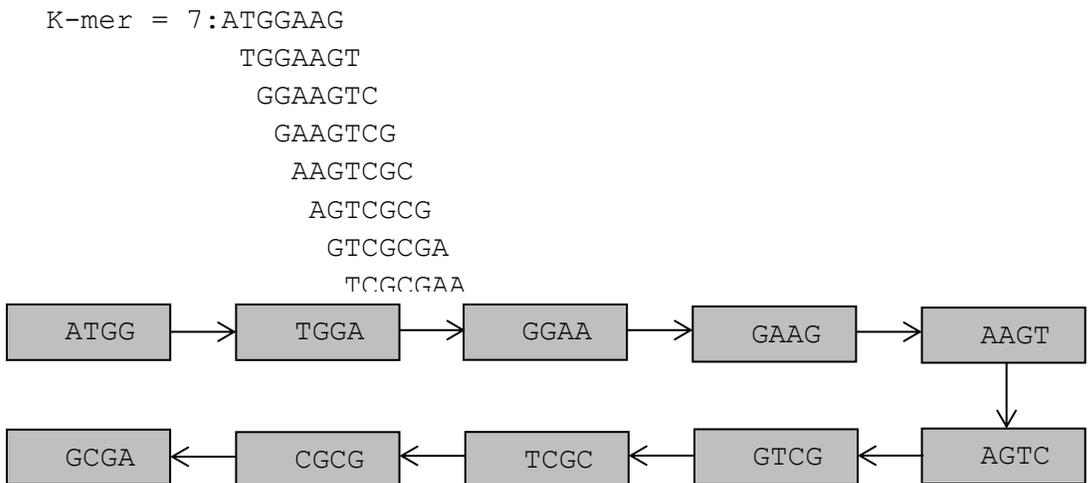
The Velvet algorithms use de Bruijn graphs for genomic and transcriptomic de novo sequence assembly. Reads are represented as mapped pathways through deBruijn graphs, which the velvet algorithms alter and reshape to determine repeats and remove read errors (Zerbino, 2008). The basic layout of de Bruijn graphs is shown in Figure 1.5, where a series of overlapping k-mers right beside each other, overlap by k-1 mers. From this arrangement, node sequences can be determined and used to construct a de Bruijn graph.

Each node has a twin directly below, which ensures that the overlaps between reads from opposite strands are considered. The twin nodes represent the reverse series of reverse complementary k-mers (Zerbino, 2008). These nodes are connected with directional arrows, or arcs. Each arc, or directional arrow in the graph is only visited once, like having designed the most efficient, least redundant walking route through a city (Compeau, 2011). From this point, the actual transcript sequences can be obtained by following the path through the de Bruijn graph (Zerbino, 2008).

Advantages to using k-mer overlaps rather than the whole reads, is that high redundancy of short read data sets is reduced. Mid-assembly errors are also more easily avoided. Smaller k-mers increase connectivity, or provide higher sensitivity of the deBruijn graphs by increasing the chances of having overlapping regions between two reads. As a result, a larger number of contigs of shorter average length are obtained.

Oasis takes the contigs generated by Velvet and assembles them into transcripts, organized into many loci. In this situation they are not genetic loci but rather “locus” here refers to a set of highly similar transcripts. There are one or more transcripts per locus. The Oases program was specifically developed for transcriptome assembly from short reads (Garg, 2011).

Sequence:  
ATGGAAGTCGCGAATC



**Figure 1.5** Simplistic example of de Bruijn graph layout.

### 1.7 Research Objectives

*Objective 1:* To produce a *de novo* transcriptome assembly for the non-model, *C. sativum* plant seeds.

*Objective 2:* To determine the relative transcript abundance at three seed developmental stages for the key genes (*DXS* and *HMGR*) involved in each of the isoprenoid biosynthetic pathways, 1-deoxyxylulose-5-phosphate and mevalonate, respectively.

*Objective 3:* To molecularly clone and functionally characterize monoterpene synthases which are responsible for the production of the majority of the monoterpenes in coriander essential oil.

*Objective 4:* To visually identify the secretory duct/canal structures known as vittae where coriander seed essential oils are stored.

## 1.8 Hypothesis and Rationale

*Hypothesis 1:* The key regulatory genes, DXS and HMGR, involved in the DXP and MVA isoprenoid biosynthetic pathways, respectively, as well as monoterpene synthase genes, exhibit differential transcript abundance between three different stages of *C. sativum* seed development.

*Rationale 1:* Msaada et al., in 2009b showed that the monoterpene (and a few sesquiterpene) constituents of coriander essential oil change over the course of the seed's development. The question here was: May this change in EO monoterpene composition be in part due to differential transcription of select isoprenoid biosynthetic genes, DXS and HMGR and monoterpene synthase genes?

*Hypothesis 2:* A few monoterpene synthase genes are responsible for the production of all monoterpenoid compounds in coriander EO.

*Rationale 2:* Monoterpenes are produced by monoterpene synthase enzymes in all plants. Many monoterpene synthases known today yield various monoterpene products from a single type of substrate, geranyl diphosphate (e.g., SamonoTPS1 from Sandalwood (Jones, 2008)). There is a high abundance of the monoterpene, linalool, in coriander EO, whereas all other monoterpenes present are found in much smaller amounts. The questions here were: Are monoterpene synthases expressed in coriander seeds, and do some of them produce multiple monoterpene products from geranyl diphosphate?

## 1.9 Significance of Study

*C. sativum* is both an important culinary herb and EO crop. Coriander EO has been shown to exhibit medicinal activity which is likely due to its monoterpene content, for example as an anti-hyperlipidemic and an anxiolytic (Dhanapakiam, 2008; Mahendra, 2011). Further, coriander contains large quantities of the monounsaturated fatty acid, petroselinic acid, which is useful in the production of detergents and nylon polymers

(Msaada, 2009a). Although coriander is clearly an important crop, genomic resources for this plant have not been developed. This project is the first instance where an NGS-transcript library for coriander was developed and where the cloning of a terpene synthase gene from this plant was accomplished.

In addition to facilitating gene discovery, the *de novo* transcriptome assembly of this non-model plant contributes to the advancement of genetics and plant breeding research for this underutilized crop. For example, the genetic composition and gene functionality information provided by this research can lead to identification of molecular markers such as single tandem repeats and single nucleotide polymorphisms (Li et al., 2012; Li et al., 2013). Future development of molecular markers will allow this specialty oil crop plant to be industrially improved via marker-assisted selective breeding. Plant seeds are excellent storage vessels, as they maintain oil integrity and have minimal storage requirements (Misharina, 2001). This makes seeds great targets for research concerning production of industrially valuable natural products, including terpenoids.

## CHAPTER 2: SCANNING ELECTRON MICROSCOPY AND *C. SATIVUM* ESSENTIAL OIL ANALYSIS

### 2.1 Synopsis

*Coriandrum sativum* seeds, whole seed and cross-section, were viewed under a scanning electron microscope. Distinct stomata structures were found on the whole seed surface. Four vittae (secretory canals) were clearly visible in the seed cross-section. The volatile components from *C. sativum* seeds were extracted by steam distillation-solvent extraction and the monoterpene content was analyzed via gas chromatography-mass spectrometry.

Seventeen monoterpenes were identified with linalool being the most abundant, making up 78.96% of the total EO terpene content. These results were in agreement with previous literature findings.

### 2.2 Materials and Methods

#### 2.2.1 Plant Material

*C. sativum* seeds were germinated and plants grown in a growth chamber at 25 °C and 150  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light intensity, with a 16:8 photoperiod, until plants were bearing seeds of varying maturity levels. Seeds of three distinct size (maturity) stages, small (1-2.5 mm diameter) – sample 1, medium (2.5-4 mm diameter) – sample 2, and large (>4 mm diameter) – sample 3, were harvested and immediately frozen in liquid nitrogen and stored at -80°C.

#### 2.2.2 Scanning Electron Microscopy (SEM)

Large coriander seeds were submersed in liquid nitrogen before being cross-sectioned with a razor blade for SEM. Both whole and cross-sectioned seeds were coated with 7.5 nm of a palladium-platinum alloy using a Cressington Sputter Coater 208HR and Thickness Controller MTM20 (Cressington Scientific Instruments Ltd., Watford, England). SEM imaging was performed on a Tescan Mira3 XMU Field Emission Scanning Electron Microscope (TESCAN, Brno, Czech Republic).

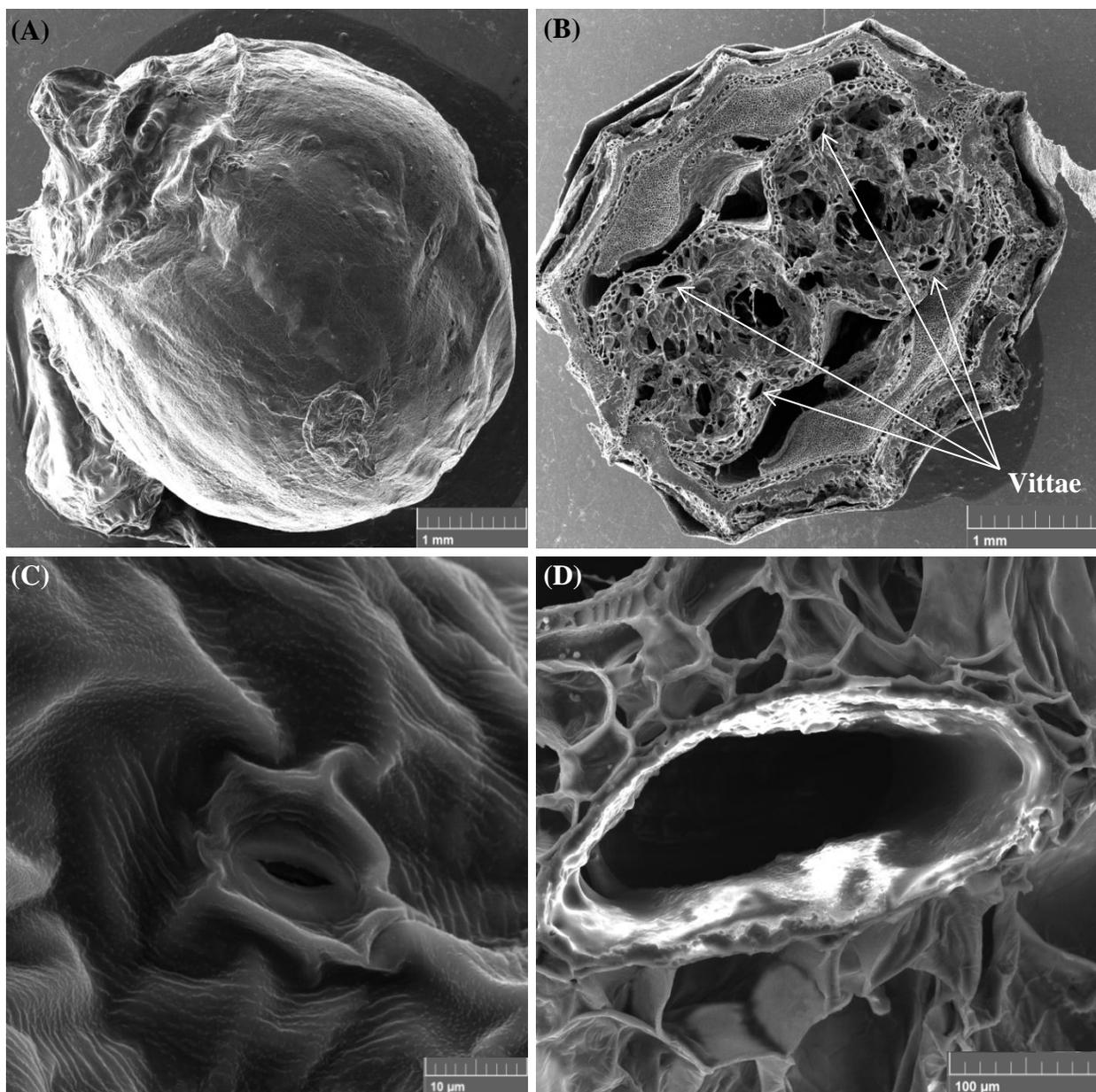
### **2.2.3 Steam Distillation of *C. sativum* Seeds**

Analysis of *C. sativum* EO was performed using 0.80 g of a blend of 1-4.5 mm diameter *C. sativum* seeds. The seed EO was distilled for 45min in 10 ml of pentane containing 10 µl of 10 mg/ml menthol as internal standard using a Clevenger-type apparatus. The organic layer was collected at room temperature and atmospheric pressure. The oil was analyzed by gas chromatography-mass spectrometry (GC-MS) using a Varian GC 3800 Gas Chromatographer coupled with a Saturn 2200 Ion Trap mass detector as previously described (Demissie et al., 2011). The GC was equipped with a 30 m × 0.25 mm capillary column coated with a 0.25 µm film of acid-modified polyethylene glycol (ECTM 1000, Altech, Deerfield, IL, USA); as well as a CO<sub>2</sub> cooled 1079 Programmable Temperature Vaporizing (PTV) injector (Varian Inc., USA). Two microliter concentrated samples were injected at 40 °C. The oven temperature was initially maintained at 40 °C for 3 minutes, followed by a two-step temperature increase, first to 130 °C (ramp rate of 10 °C per minute) and then to 230 °C (ramp rate of 50 °C per minute), and held at 230 °C for 8 minutes. The helium carrier gas flow rate was 1 ml per minute. Essential oil volatile products were identified by comparison of their mass spectra to those of authentic standards (when available), or to those in the National Institute of Standards and Technology (NIST) library database (NIST MS Search v.2.0). Two separate 0.80 g batches of *C. sativum* seeds were steam distilled simultaneously and each extract was injected twice through the GC-MS.

## **2.3 Results**

### **2.3.1 SEM Images**

High resolution images were obtained of coriander whole seed and cross-section (Figure 2.1 A and B). Four vittae, or secretory canals, are clearly visible in Figure 2.1 B and a close-up in Figure 2.1 D. Typical stomata, with guard cells, are present on the surface of the coriander seed (Figure 2.1 C).

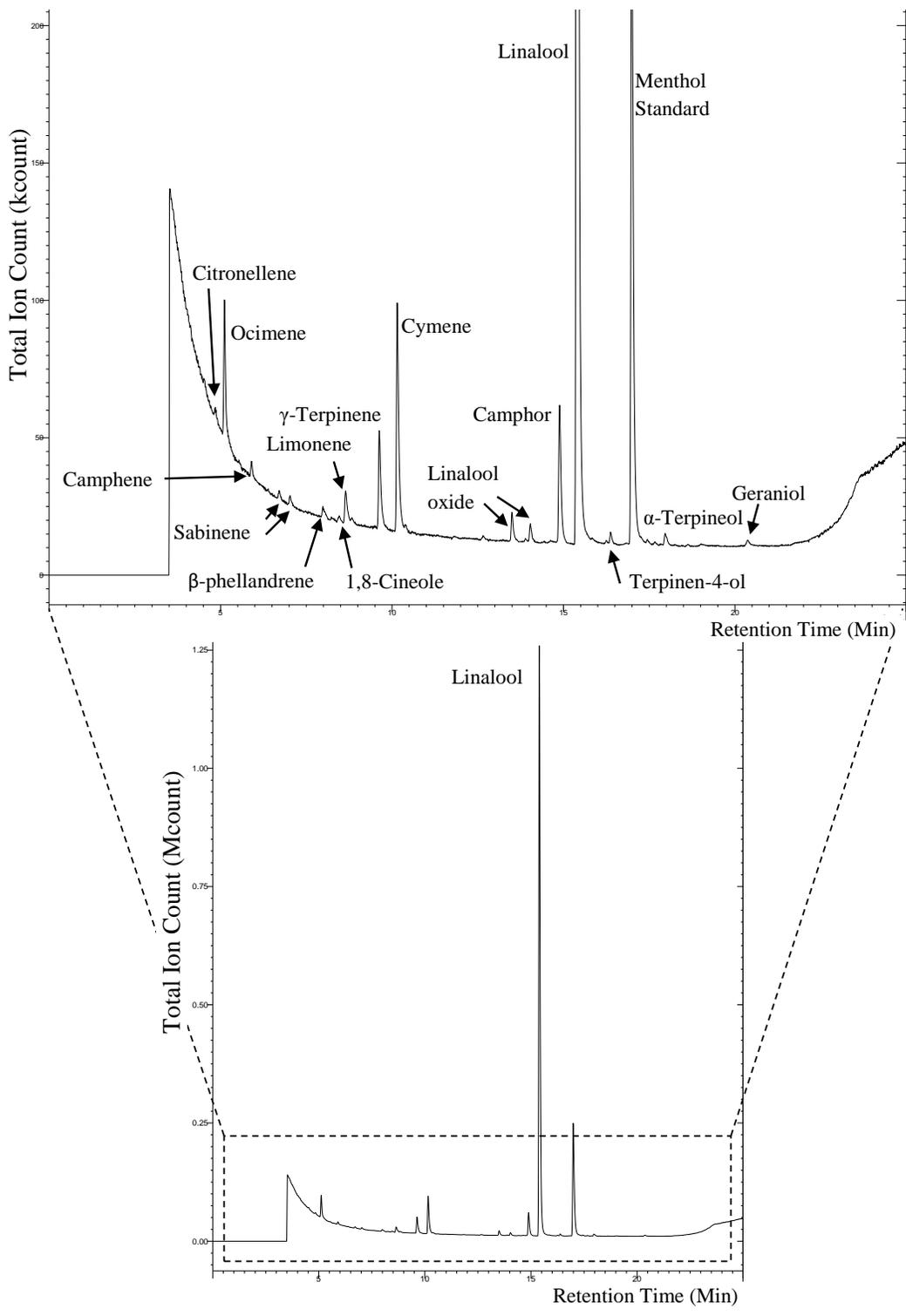


**Figure 2.1** Scanning electron microscope images of *C. sativum* seed tissue. (A) Whole seed, (B) cross-section with four vittae visible, (C) stomata on seed surface (D) close-up of a vitta.

### ***2.3.2 Gas Chromatograms and Mass Spectrometry Fingerprints of C. sativum Seed Volatile Extracts***

The total terpene content of coriander seed volatile extracts can be seen in the gas chromatogram in Figure 2.2. Double peaks observed for sabinene and linalool oxide is due to database assigning each of these peaks' identity according to a probability based

on similarity of experimental peaks' retention time and mass spectrometry fingerprints with those in the database. To confirm the true identity of each of these 'double peaks', authentic standards of linalool oxide and sabinene should be run. Authentic standards for monoterpene compounds making up less than 2% of total terpene content were not run as these products were very minor EO constituents. A total of 17 different monoterpene products were detected in the pentane fractions of the coriander seed distillate. Peak areas were determined by peak integrations and used to calculate the percent of total EO composition for each terpene detected. These values are given in Table 2.1.



**Figure 2.2** Gas chromatogram of steam distilled *C. sativum* seed essential oil.

**Table 2.1** Coriander seed essential oil (EO) terpene composition. Error bounds are in percent standard deviation.

Terpene	% of Total EO
Linalool	78.96 ± 2.86
Cymene	6.38 ± 1.28
Ocimene	4.46 ± 3.39
Camphor	3.62 ± 0.38
γ-Terpinene	2.66 ± 1.34
Limonene	1.13 ± 0.04
Linalool oxide	1.10 ± 0.25
Geraniol	0.67 ± 0.54
β-phellandrene	0.50 ± 0.14
Sabinene	0.48 ± 0.41
Camphene	0.32 ± 0.05
Terpinen-4-ol	0.25 ± 0.09
Borneol	0.18 ± 0.14
α-Terpineol	0.18 ± 0.15
Terpinolene	0.13 ± 0.05
1,8-Cineole	tr
Citronellene	tr

\* “tr” indicates trace amounts detected (<0.1%)

## 2.4 Discussion

### 2.4.1 Implications of SEM Images

The vittae visible in Figure 2.1 B and Figure 2.1 D are the storage sites of coriander seed EOs (Parthasarthy, 2008; Purseglove, 1981) unlike some angiosperms which store EOs in glandular trichomes (Turner, 2004) and gymnosperms which store EOs in resin ducts (Wu, 1997). The presence of stomata on seeds is uncommon, although it has been previously reported (Jernstedt, 1979; Rugenstein, 1981). To date, the exact function of stomata on plant seeds has not been elucidated with certainty. Proposed functions include the facilitation of gas exchange in photosynthesizing seeds, during embryo development as well as playing a role during imbibition (Jernstedt, 1979; Werker, 1997; Paiva, 2006). Green plant tissues, including green plant seeds, have entered the greening process and are actively photosynthesizing (Tschiersch, 2011). According to the RNA-seq data coriander seeds strongly express all photosynthetic genes (Table 3.5). The strong expression of photosynthetic genes coupled with the observation that developing coriander seeds are green, indicates that these tissues are actively photosynthesizing

during their development. Once coriander seeds have matured and fallen from the parent plant they become desiccated and their seed coat browns and hardens. Photosynthesis is likely inactive at this point yet the stomata may still function in imbibition, allowing water to enter the seed to initiate germination.

#### ***2.4.2 Composition of C. sativum Seed Essential Oil***

*C. sativum* seed EO composition has been extensively studied; the EO analysis performed in this thesis confirmed percent terpene composition results previously described (Bhuiyan, 2009; Misharina, 2001; Msaada, 2009b; Potter, 1996; Sriti, 2009). The gas chromatogram depicted in Figure 2.2 indicates that the major component of coriander EO is linalool, followed in abundance by cymene, ocimene, camphor and  $\gamma$ -terpinene. Those volatile terpene products present in amounts less than 2% of total EO terpenes include limonene, linalool oxide, geraniol,  $\beta$ -phellandrene, sabinene, camphene, terpinene-4-ol, borneol,  $\alpha$ -terpineol, terpinolene, 1,8-cineole and citronellene (Figure 2.2 and Table 2.1).

Msaada (2009b) identified 36 monoterpenes and four sesquiterpenes in coriander EO, Misharina (2001) identified 23 monoterpenes and no sesquiterpenes, Bhuiyan (2009) identified 37 monoterpenes and eight sesquiterpenes, Sriti (2009) identified 33 monoterpenes and three sesquiterpenes. In this EO analysis 17 monoterpenes and no sesquiterpenes were identified. The differences observed in the EO analyses can be attributed to the varying regional and seasonal conditions in which these *C. sativum* plants were grown as it has been previously shown that these variables affect EO composition (Ravi, 2006; Jerkovic, 2001). Additionally, the seeds used in this project were harvested as fresh seeds and then immediately stored at -80 °C, thus frozen seeds were used for the steam distillation experiment. Perhaps the act of freezing and thawing the seeds had an effect on the volatile terpene content of these seeds, reducing those found in already trace amounts to below the detection limit. According to the study by Misharina in 2001, *C. sativum* seeds maintained their EO integrity when stored in a dry dark place for up to a year. Therefore, drying the seeds after harvest, rather than freezing, may have been a better storage strategy. Biological variability between individuals of the

same species, even those grown in similar environments is also a contributing factor to EO composition differences. A small genetic difference can result in a large phenotypic difference, resulting in individuals of the same species with significantly different chemotypes (Bhuiyan, 2009).

In conclusion, upon analysis via scanning electron microscopy, evidence of stomata and vittae were found on the surface and in the cross-section of *C. sativum* seeds, respectively. These findings indicate that coriander actively photosynthesizes at some point of its development; additionally, the visualization of vittae on these coriander individuals confirms previous reports of these EO storage structures. GC-MS analysis of the coriander EO which was extracted by steam distillation allowed the identification of 17 monoterpenes, with linalool being the most abundant, 79% of the total EO terpene content. These findings also parallel those previously reported.

## **CHAPTER 3: BIOINFORMATICS ANALYSIS OF ILLUMINA-LED TRANSCRIPTOME SEQUENCING AND *DE NOVO* TRANSCRIPT ASSEMBLY DATA.**

### **3.1 Synopsis**

Illumina technology was used to produce 33,330,312 raw single-end reads which were assembled into a transcript library consisting of 65,306 transcript sequences. The transcripts were aligned against NCBI, Uniprot and TAIR databases using the BLASTx algorithm, and were also KEGG and GO annotated. These alignment and annotation methods allowed the suggested identification of genes involved with isoprenoid biosynthesis, photosynthesis and fatty acid biosynthesis and metabolism.

Analysis of the normalized transcript read counts showed that some isoprenoid biosynthetic transcripts were differentially abundant between three coriander seed developmental stages. Selected isoprenoid biosynthetic genes, *DXS1*, *DXS2* and *HMGR*, were relatively quantified by real time PCR and found there were significant differential transcript abundances for *DXS1* and *DXS2*. Also, biological variability was observed within the *Coriandrum sativum* species.

### **3.2 Materials and Methods**

#### ***3.2.1 RNA Isolation and Quality Controls***

Total RNA was extracted from the three developmental samples (S1, S2 and S3) of *C. sativum* seeds. The RNA was extracted using an Omega E.Z.N.A Plant RNA Kit, including DNase 1 digestion following exactly the protocol recommended by the manufacturer (OMEGA bio-tek, GA, USA). The RNA was quantified and its purity determined using a UV-Vis spectrophotometer (Nanodrop ND-100, Wilmington, DE, USA). RNA integrity was assessed on a 1.5% agarose gel.

#### ***3.2.2 Details of Illumina Sequencing and De Novo Transcript Assembly***

Complementary DNA was generated for each seed RNA sample and the transcriptome of the three multiplexed samples were single-end sequenced on a single flow cell lane

using an Illumina GAIIx sequencer (Illumina Inc., San Diego, CA, USA). Raw reads obtained from the Illumina genome analyzer, were searched for adapters and quality trimmed using Btrim software (PMID:21651976). The individual reads were then each broken down into three different k-mer length fragments: 21, 27, and 31, so that they could be assembled into contigs (three different contig libraries) via the de Bruijn graph algorithms in the publicly available Velvet software (v.1.0; Zerbino, 2008). The resulting contig libraries became the input for the also publicly available Oases transcript assembler with default settings (initial release; Schulz, 2012). The sequencing and library construction were performed by service providers at the Plant Biotechnology Laboratory, University of Lethbridge (Lethbridge, AB, Canada).

### **3.2.3 Bioinformatics**

#### *3.2.3.1 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Annotations*

Using the BLASTx algorithm, *C. sativum* transcripts were aligned against the whole non-redundant database with a cut-off e-value less than 1E-10. All unique transcript sequences were aligned against sequences in The Arabidopsis Information Resource (TAIR v.2.2.8) and the Universal Protein Resource Knowledgebase (UniProtKB) protein databases via the BLASTx algorithm. All sequences with blast hits were annotated with Gene Ontology (GO) terms by service providers at the National Research Council Plant Biotechnology Institute (NRC-PBI, Saskatoon, SK, Canada). Sequences were also assigned Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations via the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007).

#### *3.2.3.2 Isoprenoid Biosynthetic Gene Expression*

Differential transcript expression was performed via RNA-Seq by Expected Maximization (RSEM) algorithm (Li, 2011) and expression comparisons were based on log<sub>2</sub> ratios of samples 2 and 3 to sample 1. Raw transcript expression data was obtained by service providers at the Plant Biotechnology Laboratory (University of Lethbridge) and raw read counts were received as a tab delimited file containing gene ID's,

expression level and  $\log_2$  ratios. Using GO and KEGG annotations, all the genes involved with isoprenoid biosynthesis (DXP and MVA pathway as well as terpene synthases/cyclases) were identified and their differential expression patterns analyzed. The GO and KEGG annotations were also used to identify all photosynthetic genes as well as fatty acid biosynthetic genes.

As a secondary evaluation of transcript abundances found in the RNA-seq data, relative transcript abundances of *DXS* and *HMGR* genes were also assessed by quantitative PCR using the CFX96 real-time PCR detection system with SsoFast EvaGreen Supermix (Bio-Rad, Mississauga, ON, Canada) and 500 nM primer (ea.) concentrations in a 20  $\mu$ l reaction volume. Primers specific for *DXS1*, *DXS2*, *HMGR*, and a reference gene were designed via GenScript Real-time PCR (TaqMan) Primer Design (<https://www.genscript.com/ssl-bin/app/primer>) (Table 3.1). RNA extracted as described above was treated with additional RTS-DNase Kit (MO BIO Laboratories Inc., Carlsbad, CA, USA) as per manufacturer's instructions to remove any remaining traces of genomic DNA contamination. One microgram of RNA (extracted as per above) was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Mississauga, Ontario, Canada) according to the manufacturer's protocol.

A thermal gradient of 48 to 63 °C was run to determine the optimal temperature for all primer sets. A calibration curve was prepared for each primer set except primers for *DXS2* (due to shortage of reagents). Primer efficiencies for each of *DXS1*, *HMGR* and actin were found to be 109.3, 109.7 and 93.5%, respectively. Data was corrected for these primer efficiencies, in the case of *DXS2* an auto-efficiency of 100% was assumed.

The following program was used for qPCR: 95 °C for 30 seconds followed by 40 cycles of 95 °C for 10 seconds, 60.5 °C for 30 seconds and a plate read. For the qPCR analysis, three biological replicates and two technical replicates were prepared. Relative normalized expression values for *DXS1*, *DXS2* and *HMGR* were calculated using CFX96 Data Manager Software (Bio-Rad). Reference gene, actin-related protein 5 was checked for stability between the three coriander developmental stages by analysis of RPKM-normalized read counts from the gene expression matrix.

**Table 3.1** Primers for qPCR relative expression analysis of DXS and HMGR genes.

Primer Name	Target Gene	Amplicon Size (bp)	Primer Sequence
DXS1_p15_F	DXS1	92	5' GCGGTTCAAAGTTGTTTGG 3'
DXS1_p15_R	DXS1		5' TCCAGTGGTTTACAGAAGCG 3'
DXS2_p8_F	DXS2	106	5' TCCATTCTCAAGGTCTTCC 3'
DXS2_p8_R	DXS		5' TCATCTTCCGTTCTTCACCA 3'
HMGR_F	HMGR	145	5' TTATAGCCACGGGTCAGGA 3'
HMGR_R	HMGR		5' TGTGTTCCACCTCCAAGTGT 3'
Actin_p6_F	ALP5*	105	5' GACGAGGATGAGGCAGAGTT 3'
Actin_p6_R	ALP5		5' TGGAGCATCAGAAACAGAGG 3'

\*ALP5 (Actin-related protein 5)

### 3.3 Results

#### 3.3.1 *Coriander transcriptome composition*

Transcriptome sequencing yielded a total of 33,330,312 raw reads of 65bp length each, 10,638,013 from Sample 1 (S1; small seeds), 12,513,426 from Sample 2 (S2; medium seeds) and 10,178,873 from Sample 3 (S3; large seeds). *De novo* assembly was performed with three k-mer length libraries, 21, 27 and 31 which have the following parameters, number of transcript, longest transcript, median transcript length and number of transcripts aligned with e-value less than 1E-10, shown in Table 3.1. All unique transcript sequences were aligned against sequences in The Arabidopsis Information Resource TAIR (v.2.2.8) and UniProtKB databases, resulting in 55,689 sequences with blast hits.

The number of homologues having greater than 90% identity to protein in public databases was very similar between transcript libraries. However, the 21 k-mer library did have a higher number of homologues with less than 90% identity to proteins in public databases (Table 3.3). The 21-mer transcript library was chosen as the 'best' quality, see discussion for further details; thus, this library was used for all bioinformatics analysis (e.g., BLASTx alignments and gene annotations).

**Table 3.2** Summary statistics for alternative k-mer length transcript assembly libraries.

k-mer length (nucleotides)	Number of transcripts	Median transcript length (bp)	Number of transcripts aligned with e-value < 1E-10
21	65306	519	40492
27	61767	409	36471
31	62572	328	35221

**Table 3.3** Number of homologues ( $\geq 90\%$  identity;  $X < 1E-40$ ), similar proteins ( $< 90\%$  identity;  $X \leq 1E-40$ ), weakly similar proteins ( $< 90\%$  identity;  $1E-40 < X \leq 1E-20$ ), and putative proteins ( $< 90\%$  identity;  $1E-20 < X \leq 1E-10$ ) for each Oases transcript library.

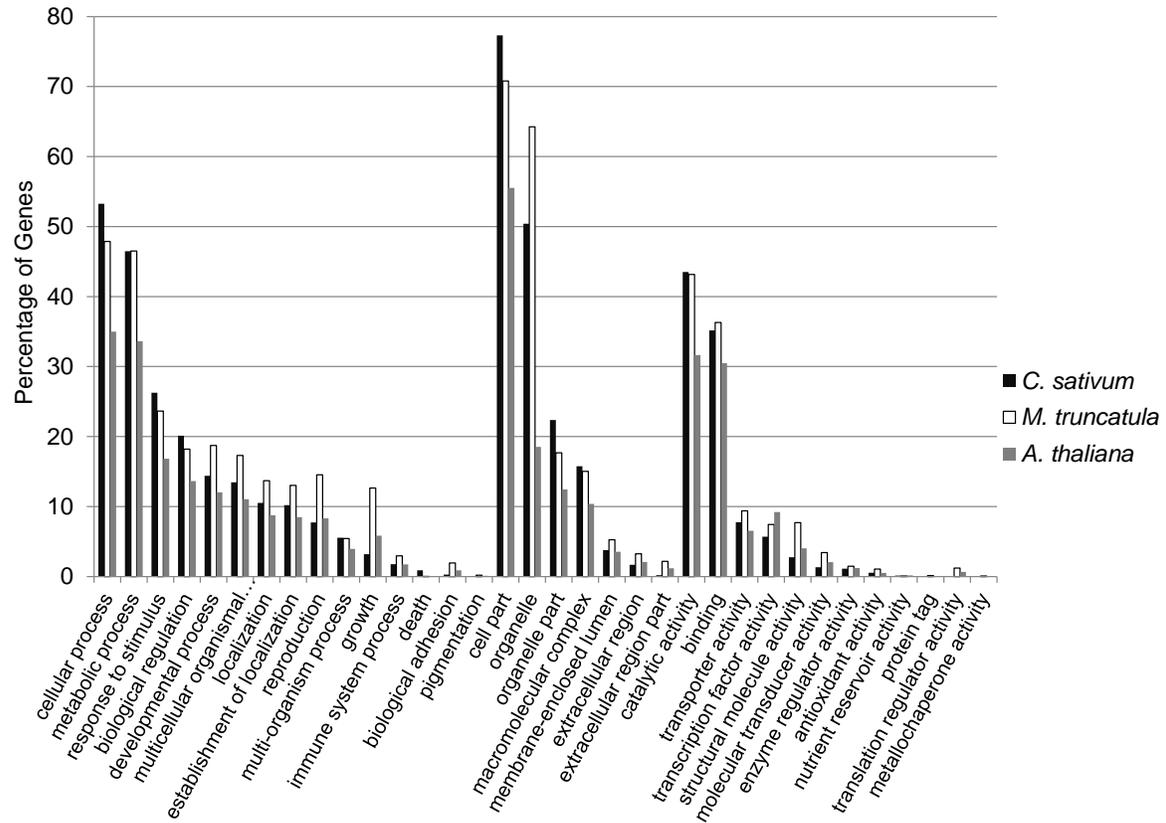
	K-mer 21		K-mer 27		K-mer 31	
Number of Alignments	Number of transcripts	Percent total	Number of transcripts	Percent total	Number of transcripts	Percent total
Total	65306	100	61761	100	62572	100
Homologue	1262	1.9	1264	2	1150	1.8
Similar	25472	39	21546	34.9	18966	30.3
Weak similar	8059	12.3	7559	12.2	8085	12.3
Putative	5699	8.73	6102	9.9	7020	11.2

### 3.3.2 BLASTx Alignments and Annotations

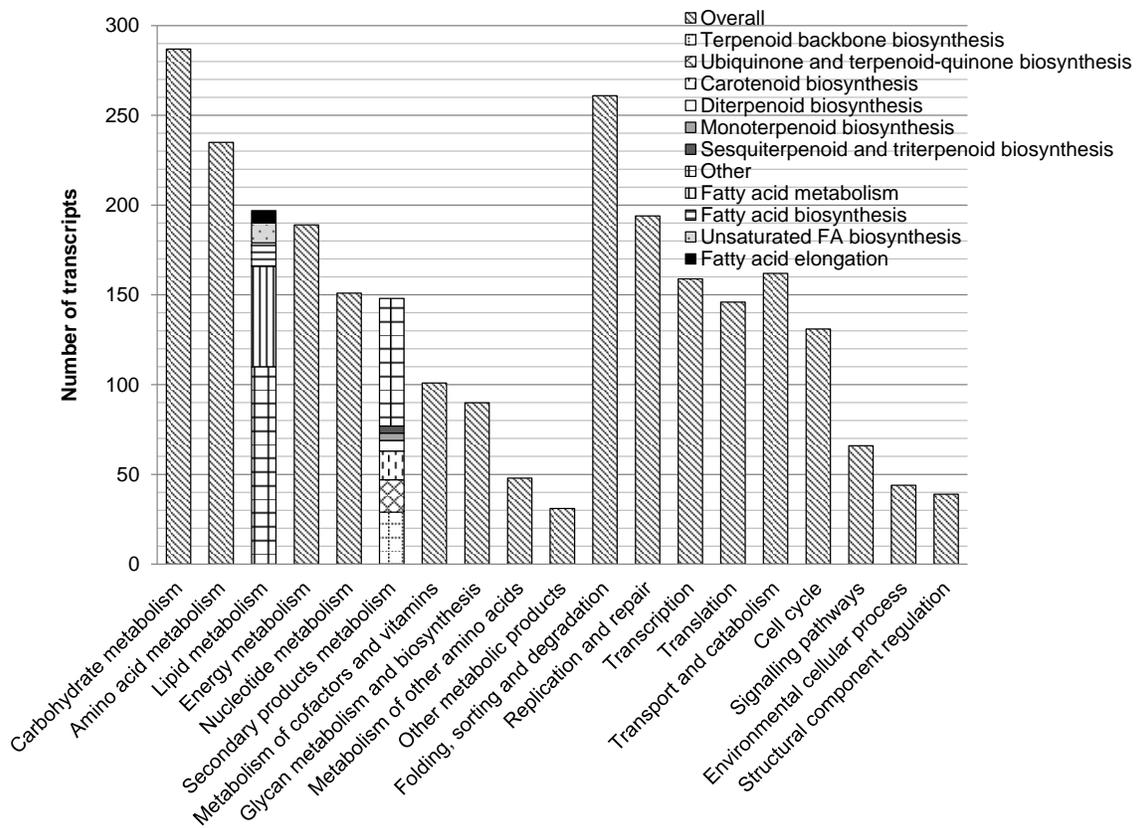
Of the 65,306 unique sequences from the transcript library chosen as the ‘best’ (21-mer), 35,928 were assigned at least one gene ontology (GO) term. Among these, 26,882 (41.16%), 19,025 (29.13%) and 32,405 (49.62%) sequences were assigned at least one GO term in the biological processes, cellular component and molecular function categories, respectively. Distribution of the most abundant GO terms for biological processes, molecular functions, and cellular components is summarized in Figure 3.1. In Figure 3.1 the GO percent distributions for coriander are also compared against two known reference plants, *Medicago truncatula* and *Arabidopsis thaliana*.

KEGG annotations categorized 1508, 760 and 442 transcripts under metabolism, genetic information processing and cellular processes, respectively. Within those transcripts grouped under metabolism, 287 corresponded to carbohydrate metabolism, 235 to amino acid metabolism, 197 to lipid metabolism (includes all genes involved in fatty acid biosynthesis/metabolism), 189 to energy metabolism, 179 to secondary (specialized) products metabolism (includes all genes involved in terpenoid, phenylpropanoid, flavonoid, alkaloid and polyketide metabolism), 151 to nucleotide

metabolism, 101 to cofactors/vitamins metabolism and 169 to other processes (e.g., other amino acids and glycan metabolism) (Figure 3.2 and Table 3.4).



**Figure 3.1** Comparison of *C. sativum* percentage distribution of gene ontology against two reference databases, *M. truncatula* and *A. thaliana*.



**Figure 3.2** Breakdown of KEGG annotations for *C. sativum* seed transcriptome.

**Table 3.4** KEGG pathway annotations for *C. sativum* transcriptome; number of transcripts for each KEGG annotation shown.

<b>Metabolism</b>		<b>Genetic Information processing</b>	
<b><i>Carbohydrate metabolism</i></b> <b>287</b>		<b><i>Folding, sorting and degradation</i></b> <b>261</b>	
Amino sugar and nucleotide sugar metabolism	37	Protein processing in endoplasmic reticulum	74
Starch and sucrose metabolism	30	Ubiquitin mediated proteolysis	54
Glycolysis / Gluconeogenesis	29	RNA degradation	47
Pyruvate metabolism	28	Proteasome	34
Glyoxylate and dicarboxylate metabolism	23	Protein export	26
Citrate cycle (TCA cycle)	19	SNARE interactions in vesicular transport	17
Inositol phosphate metabolism	18	Sulfur relay system	9
Fructose and mannose metabolism	17		
Propanoate metabolism	15	<b><i>Replication and repair</i></b> <b>194</b>	
Pentose phosphate pathway	15	mRNA surveillance pathway	47
Galactose metabolism	14	Nucleotide excision repair	37
Ascorbate and aldarate metabolism	14	DNA replication	32
Pentose and glucuronate interconversions	13	Base excision repair	26
Butanoate metabolism	11	Homologous recombination	24
C5-Branched dibasic acid metabolism	4	Mismatch repair	20
		Non-homologous end-joining	8
<b><i>Amino acid metabolism</i></b> <b>235</b>		<b><i>Transcription</i></b> <b>159</b>	
Arginine and proline metabolism	33	Spliceosome	100
Glycine, serine and threonine metabolism	29	Basal transcription factors	31
Cysteine and methionine metabolism	29	RNA polymerase	28
Alanine, aspartate and glutamate metabolism	27		
Phenylalanine, tyrosine and tryptophan biosynthesis	22	<b><i>Translation</i></b> <b>146</b>	
Valine, leucine and isoleucine degradation	19	Ribosome	121
Tyrosine metabolism	17	Aminoacyl-tRNA biosynthesis	25
Phenylalanine metabolism	14		
Histidine metabolism	11	<b><i>Cellular processes</i></b>	
Valine, leucine and isoleucine biosynthesis	10	<b><i>Transport and catabolism</i></b> <b>162</b>	
Lysine biosynthesis	10	Endocytosis	35
Lysine degradation	7	Peroxisome	34
Tryptophan metabolism	7	Lysosome	29
		Phagosome	26
<b><i>Lipid metabolism</i></b> <b>197</b>		Protein export	26
Fatty acid metabolism	56	Regulation of autophagy	10
Glycerophospholipid metabolism	31	ABC transporters	2
Glycerolipid metabolism	22		
Steroid biosynthesis	15	<b><i>Cell cycle</i></b> <b>131</b>	
Fatty acid biosynthesis	13	Protein processing in endoplasmic reticulum	74
alpha-Linolenic acid metabolism	11	Cell cycle	53
Biosynthesis of unsaturated fatty acids	11	Apoptosis	4
Sphingolipid metabolism	11		
Fatty acid elongation	7	<b><i>Signalling pathways</i></b> <b>66</b>	
Ether lipid metabolism	7	Plant hormone signal transduction	36
Arachidonic acid metabolism	6	Phosphatidylinositol signaling system	16
Linoleic acid metabolism	4	MAPK signaling pathway	8
Synthesis and degradation of ketone bodies	3	Calcium signaling pathway	6
<b><i>Energy metabolism</i></b> <b>189</b>		<b><i>Environmental cellular process</i></b> <b>44</b>	
Oxidative phosphorylation	68	Plant-pathogen interaction	25
Photosynthesis	39	Circadian rhythm - plant	19
Carbon fixation in photosynthetic organisms	25		
Methane metabolism	21	<b><i>Structural component regulation</i></b> <b>39</b>	
Nitrogen metabolism	15	Regulation of actin cytoskeleton	15
Sulfur metabolism	11		

Photosynthesis - antenna proteins	10	Tight junction	10
		Focal adhesion	5
<b><i>Nucleotide metabolism</i></b>	<b>151</b>	Gap junction	5
Purine metabolism	80	Adherens junction	4
Pyrimidine metabolism	71		
<b><i>Secondary products metabolism</i></b>	<b>148</b>		
<i>Terpenoids</i>	<b>84</b>		
Terpenoid backbone biosynthesis	29		
Ubiquinone and other terpenoid-quinone biosynthesis	18		
Carotenoid biosynthesis	16		
Diterpenoid biosynthesis	6		
Monoterpenoid biosynthesis	4		
Sesquiterpenoid and triterpenoid biosynthesis	4		
Brassinosteroid biosynthesis	4		
Limonene and pinene degradation	2		
Geraniol degradation	1		
<i>Phenylpropanoids/Flavonoids/Alkaloids</i>	<b>60</b>		
Phenylpropanoid biosynthesis	14		
Flavonoid biosynthesis	10		
Tropane, piperidine and pyridine alkaloid biosynthesis	8		
Isoquinoline alkaloid biosynthesis	7		
Stilbenoid, diarylheptanoid and gingerol biosynthesis	5		
Aminobenzoate degradation	4		
Flavone and flavonol biosynthesis	3		
Benzoate degradation	3		
Caffeine metabolism	2		
Indole alkaloid biosynthesis	1		
Betalain biosynthesis	1		
Glucosinolate biosynthesis	1		
Fluorobenzoate degradation	1		
<i>Polyketides</i>	<b>4</b>		
Zeatin biosynthesis	2		
Biosynthesis of ansamycins	1		
Polyketide sugar unit biosynthesis	1		
<b><i>Metabolism of cofactors and vitamins</i></b>	<b>101</b>		
Porphyrin and chlorophyll metabolism	29		
Pantothenate and CoA biosynthesis	16		
Nicotinate and nicotinamide metabolism	10		
Folate biosynthesis	10		
One carbon pool by folate	10		
Biotin metabolism	7		
Thiamine metabolism	6		
Riboflavin metabolism	6		
Vitamin B6 metabolism	5		
Lipoic acid metabolism	2		
<b><i>Glycan metabolism and biosynthesis</i></b>	<b>90</b>		
N-Glycan biosynthesis	56		
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	21		
Other glycan degradation	8		
Glycosphingolipid biosynthesis - globo series	3		

Glycosphingolipid biosynthesis - ganglio series 2

<b><i>Metabolism of other amino acids</i></b>	<b>48</b>
Glutathione metabolism	16
beta-Alanine metabolism	13
Selenocompound metabolism	9
Cyanoamino acid metabolism	7
Taurine and hypotaurine metabolism	3
<b><i>Other metabolic products</i></b>	<b>31</b>
Drug metabolism - other enzymes	12
Metabolism of xenobiotics by cytochrome P450	6
Chloroalkane and chloroalkene degradation	3
Styrene degradation	3
Naphthalene degradation	2
Polycyclic aromatic hydrocarbon degradation	2
Chlorocyclohexane and chlorobenzene degradation	1
Toluene degradation	1
Bisphenol degradation	1

### 3.3.3 Transcript abundances

A transcript expression, or abundance, matrix was prepared by Slava at the University of Lethbridge using RSEM software (Li and Dewey, 2011). This matrix was received as a delimited text file which was opened in Excel for analysis. The matrix contained raw read counts which were already RPKM-normalized (output of RSEM software) for each transcript, including redundant transcripts. In cases where redundancy occurred counts were summed for all those transcripts with greater than 97% homology. Differential transcript abundance was assessed via  $\log_2$  ratios. A negative  $\log_2$  ratio represents down-regulation, a positive value, up-regulation and a ratio of 0 is indicative of no differential expression between two samples.

#### 3.3.3.1 Isoprenoid biosynthetic genes

The differential transcript abundance data for each of the DXP and MVA pathway genes as well as the related prenyl transferases and terpene synthases/cyclases are represented in Table 3.5 and Table 3.6. The expression data in these tables are represented as RPKM-normalized counts. The fold change (differential transcript abundance) is represented by  $\log_2$  ratios.

**Table 3.5** Transcript abundance of genes involved in the DXP and MVA pathways, as well as genes encoding prenyl transferases across three stages of coriander seed development: S1-early, S2-mid and S3-late. Differential transcript levels are shown in bold.

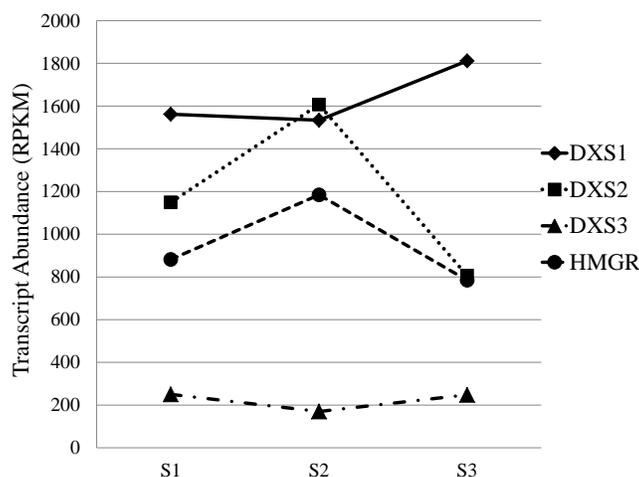
	RPKM-Normalized Counts					
	S1	S2	S3	$\log_2(S2/S1)$	$\log_2(S3/S1)$	$\log_2(S3/S2)$
<b>DXP Pathway</b>						
DXS1	1562	1534	1811	-0.03	0.21	0.24
DXS2	1149	<b>1607</b>	<b>806</b>	0.48	-0.51	<b>-1.00</b>
DXS3	250	170	248	-0.56	-0.01	0.54
DXR	727	672	598	-0.11	-0.28	-0.17
MCT	1020	1231	931	0.27	-0.13	-0.40
CMK	184	261	183	0.50	-0.01	-0.51
MDS	1368	1641	1242	0.26	-0.14	-0.40
HDS	2456	2954	2886	0.27	0.23	-0.03
HDR1	1345	1345	1359	0.00	0.02	0.01
HDR2	919	1002	1097	0.12	0.26	0.13
<b>MVA Pathway</b>						
HMGS	1599	2080	1440	0.38	-0.15	-0.53
HMGR	882	1184	785	0.42	-0.17	-0.59
MK	474	501	438	0.08	-0.11	-0.19
PMK	328	381	316	0.22	-0.06	-0.27
MDC	641	881	600	0.46	-0.09	-0.55
<b>Prenyl Precursors</b>						
IPPI	1030	1111	925	0.11	-0.16	-0.26
FPPS	822	1206	881	0.55	0.10	-0.45
GPPS	383	520	390	0.44	0.03	-0.42
GGPPS1	716	1057	751	0.56	0.07	-0.49
GGPPS2	192	256	303	0.42	0.66	0.24

**Table 3.6** Transcript abundance for genes involved in coriander isoprenoid biosynthesis across three stages of seed development (S1, S2, S3). Differential transcript levels shown in bold.

	RPKM-Normalized					
	Counts			log <sub>2</sub>	log <sub>2</sub>	log <sub>2</sub>
	S1	S2	S3			
<b>Monoterpene biosynthesis</b>						
(S)-Linalool synthase	<b>3015</b>	<b>7317</b>	<b>1350</b>	<b>1.279</b>	<b>-1.160</b>	<b>-2.4387</b>
mTPS2	4110	3829	3962	-0.10	-0.05	0.05
mTPS3	<b>502</b>	<b>255</b>	<b>847</b>	<b>-0.98</b>	0.75	<b>1.73</b>
γ-Terpinene synthase	<b>664</b>	<b>1645</b>	<b>715</b>	<b>1.31</b>	0.11	<b>-1.20</b>
<b>Sesqui- and Triterpene biosynthesis</b>						
Squalene monooxygenase1	333	419	350	0.33	0.07	-0.26
Squalene monooxygenase2	158	209	183	0.40	0.21	-0.19
Squalene monooxygenase3	589	551	503	0.50	0.03	-0.47
Squalene synthase	750	1061	764	-0.10	-0.23	-0.13
sTPS1	<b>1127</b>	<b>3177</b>	<b>631</b>	<b>1.50</b>	-0.84	<b>-2.33</b>
sTPS2	578	<b>1056</b>	<b>466</b>	0.87	-0.31	<b>-1.18</b>
Brassinosteroids	438	694	579	0.66	0.40	-0.26
<b>Diterpene biosynthesis</b>						
ent-Kaurene acid hydroxylase	502	348	494	-0.53	-0.02	0.51
Gibberellin 2-oxidase1	<b>201</b>	<b>37</b>	<b>174</b>	<b>-2.44</b>	-0.21	<b>2.23</b>
ent-Kaurene oxidase	191	275	147	0.53	-0.38	-0.91
ent-Kaurene synthase	150	180	136	0.26	-0.14	-0.40
Gibberellin 2-oxidase2	<b>91</b>	<b>409</b>	<b>104</b>	<b>2.17</b>	0.20	<b>-1.97</b>
Gibberellin 3-β-dioxygenase	<b>493</b>	<b>609</b>	<b>226</b>	0.30	<b>-1.12</b>	<b>-1.43</b>
<b>Tetraterpene biosynthesis</b>						
9-cis- Epoxycarotenoid dioxygenase	<b>108</b>	<b>20</b>	<b>44</b>	<b>-2.44</b>	<b>-1.30</b>	<b>1.14</b>
Lycopene ε-cyclase	468	385	323	-0.28	-0.54	-0.25
Prolycopene isomerase	472	457	471	-0.05	0.00	0.04
ζ-Carotene isomerase	<b>85</b>	<b>171</b>	105	<b>1.01</b>	0.30	-0.70
Xanthoxin dehydrogenase	390	472	349	0.28	-0.16	-0.44
(+)-Abscisic acid 8-hydroxylase	54	<b>105</b>	<b>48</b>	0.96	-0.17	<b>-1.13</b>
ζ-Carotene desaturase	663	799	907	0.27	0.45	0.18
Zeaxanthin epoxidase	716	<b>570</b>	<b>1312</b>	-0.33	0.87	<b>1.20</b>
Violaxanthin de-epoxidase	265	262	455	-0.02	0.78	0.80
Lycopene β-cyclase	254	332	418	0.39	0.72	0.33
15-cis-Phytoene desaturase	813	950	1035	0.22	0.35	0.12
Abscisic-aldehyde oxidase	342	275	411	-0.31	0.27	0.58

$\beta$ -carotene 3-hydroxylase	202	<b>109</b>	<b>280</b>	-0.90	0.47	<b>1.37</b>
Phytoene synthase	422	270	499	-0.64	0.24	0.89

Three *DXS* genes were found in coriander and their individual transcript levels as well as transcript abundance for *HMGR*, across three developmental stages of coriander seeds, are shown in Figure 3.3. The *DXS2* gene exhibited a 2-fold down-regulation from S2 to S3 (Figure 3.3 and Table 3.5) while *DXS1*, *DXS3* and *HMGR* all displayed relatively constant transcript levels throughout coriander seed development.

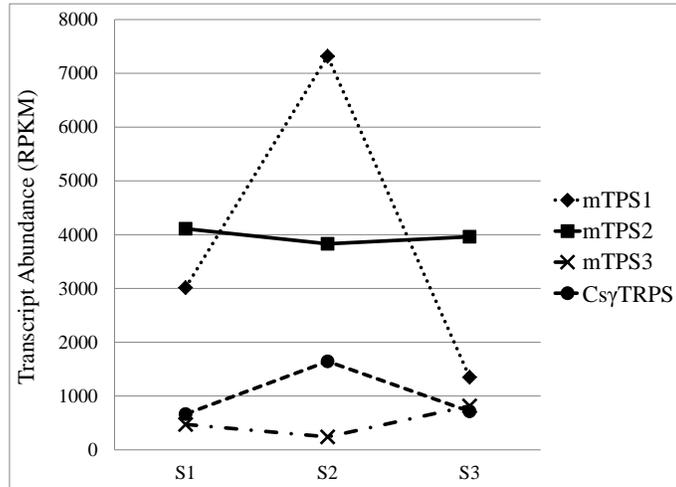


**Figure 3.3** Transcript abundances for key regulatory genes of the 1-deoxy-D-xylulose 5-phosphate (DXP) and mevalonate (HMGR) isoprenoid precursor biosynthetic pathways.

Upon analyzing the presence of terpene synthase genes it was found that there were 4 monoterpene, 7 sesqui- and triterpene, 6 diterpene and 14 tetraterpene biosynthetic genes present in the *C. sativum* seeds (Table 3.6). Two sesquiterpene synthase genes were identified that, according to GO annotations, encoded for enzymes responsible for the production of  $\beta$ -caryophyllene,  $\alpha$ -humulene, and germacrene D. Both genes were KEGG annotated as part of the  $\beta$ -caryophyllene and  $\alpha$ -humulene biosynthetic pathways.

Four mTPS candidate genes were identified, *CsLINS* (*mTPS1*), *mTPS2*, *mTPS3* and *CsyTRPS* (see Chapter 4 for details). *CsLINS* (*mTPS1*) has a 2-fold up-regulation from S1 to S2 followed by a 5-fold down-regulation from S2 to S3. The *mTPS2* gene maintains

relatively constant transcript levels throughout the three stages of seed development. There is a slight dip (at least 2-fold) at S2 for *mTPS3* while *CsyTRPS* displayed a slight peak (at least 2-fold) at S2 (Figure 3.4 and Table 3.5).



**Figure 3.4** Transcript abundances for monoterpene synthases in coriander at three developmental stages.

### 3.3.3.2 Other genes

The genes which encode necessary proteins for active photosynthesis (photosystems I and II, cytochrome b6/f complex, photosynthetic electron transport and F-type ATPases) were all present in coriander's transcriptome (Table 3.7).

**Table 3.7** Photosynthetic genes identified in *C. sativum* seed tissue from RNA-seq data.

KEGG Code	Gene Description	Transcript Library IDs
<b>Photosystem II</b>		
PsbA	Photosystem II reaction center protein A	L64
PsbC	Photosystem II reaction center protein C	L5588; L19579
PsbB	Photosystem II reaction center protein C	L33094; L36651; L39297; L45482
PsbF	Photosystem II reaction center protein C	L23304
PsbK	Photosystem II reaction center protein K	L5901
PsbO	Oxygen-evolving enhancer protein 1	L1767; L1801; L2305; L2555; L3056 L12338; L12454; L13968; L14827; L2123; L2239; L24281; L28291; L3234; L40289; L4051; L4236; L639;
PsbP	Photosystem II reaction center protein P	L7196; L7680
PsbQ	Photosystem subunit Q	L17771; L435; L9681
PsbR	Photosystem subunit R	L1407; L2901; L3379; L36228; L4301
PsbS	Chlorophyll A-B binding family protein Photosystem II reaction center protein	L1533
PsbW	W	L2516; L8357
PsbY	Photosystem II core complex protein	L1536
Psb27	Photosystem II 11-kDa protein Photosystem II reaction center Psb28 protein	L10979; L14995 L3648
<b>Photosystem I</b>		
PsaD	Photosystem I reaction center subunit D	L1511; L39967; L9083
PsaE	Photosystem I reaction center subunit E	L2211
PsaF	Photosystem I reaction center subunit F	L4140
PsaG	Photosystem I reaction center subunit G	L2432
PsaH	Photosystem I reaction center subunit H	L4171
PsaK	Photosystem I reaction center subunit K	L1661
PsaL	Photosystem I reaction center subunit L	L72
PsaN	Photosystem I reaction center subunit N	L924
PsaO	Photosystem I reaction center subunit O	L1155
<b>Cytochrome b6/f complex</b>		
PetB	Photosynthetic electron transfer B	L30444; L35676; L36154
PetD	Photosynthetic electron transfer D	L7736
PetA	Ubiquinol-cytochrome C reductase iron-sulfur subunit	L12311; L1259; L2350
PetG	Cytochrome b6f complex subunit petG	L44087
<b>Photosynthetic Electron Transport</b>		
PetE	Plastocyanin 1	L1094
PetF	Ferredoxin (glutamate synthase)	L1175; L1911; L269
PetH	Ferredoxin NADP oxidoreductase	L1019; L13916; L7562
PetJ	Cytochrome c6	L15226; L36565
<b>F-type ATPase</b>		
alpha/beta	ATPase (beta/alpha) subunits	L1786; L25143; L26178; L37052;

---

		L5888; L602; L94
		L16093; L17713; L18340; L1864;
		L20415; L24077; L251; L33138;
gamma	ATPase (gamma) subunit	L38321; L41586
delta	ATPase (delta) subunit	L10086; L20212; L2668; L28; L8101

---

Genes involved with fatty acid biosynthesis and metabolism (desaturation and elongation) found in the coriander transcriptome are given in Table 3.8.

**Table 3.8** Unigenes in *C. sativum* seeds involved in fatty acid biosynthesis.

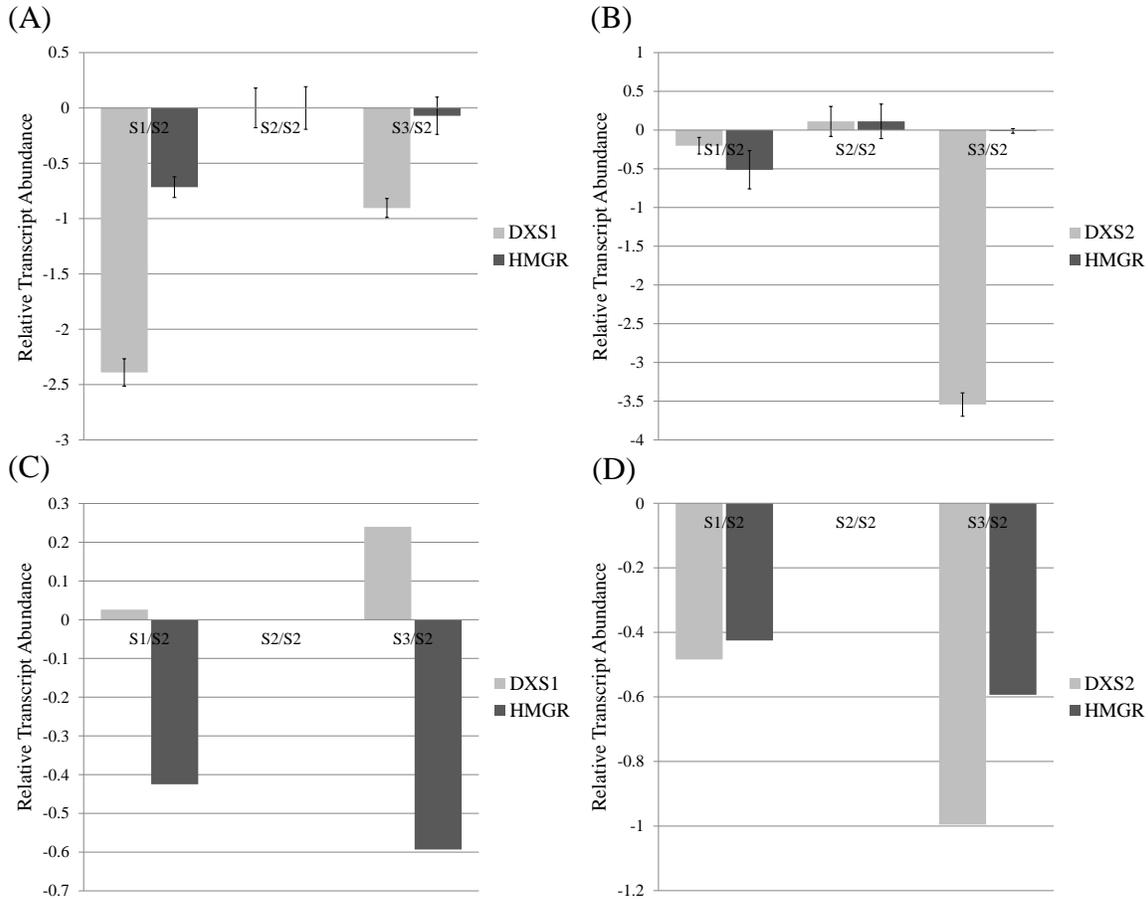
KEGG Code	Gene Description	Transcript Library IDs
<b>Fatty acid biosynthesis</b>		
ACC/acc		
C	Acetyl-CoA carboxylase	L4672/L6405/L25177
KAS I	$\beta$ -Ketoacyl-ACP synthetase I	<b>L117/L46502</b>
KAS II/FabF	$\beta$ -Ketoacyl-ACP synthetase II	L6266
KAS III/FabH	$\beta$ -Ketoacyl-ACP synthetase III	L2906 L132/L25600/L34317/L27021/L20898/L37921/L24615/L39582
KAR	$\beta$ -Ketoacyl-ACP reductase	4615/L39582
HAD/FabZ	Hydroxyacyl-ACP dehydrase	L2669
EAR/FabI	Enoyl-ACP reductase	L43
OAT	Oleoacyl-ACP thioesterase	L12944/L16261
FatA/FatB	Acyl-ACP thioesterase A/B	L475/ <b>L12150</b>
FabG	3-oxoacyl-[acyl-carrier protein] reductase	L20913
accA	acetyl-CoA carboxylase carboxyl transferase subunit alpha	L2785
accB	acetyl-CoA carboxylase biotin carboxyl carrier protein	L14753
ACAC	acetyl-CoA carboxylase / biotin carboxylase	L6755
FabD	[acyl-carrier-protein] S-malonyltransferase	L5494
<b>Fatty acid desaturation</b>		
AAD/DE		
SA1	Palmitoyl- $\Delta^{4/9}$ -ACP desaturase	L1727/ <b>L6822</b> L8034/L7126/L20266/L17581/L38578/L20132/L5100/L28192/L28381/L30189/L49040/L27218/L34502/L3204/L16286/L33876/L18457/L21175/L31983/L38207/L41280/L2663
$\Delta^{12}D$	$\Delta^{12}(\omega^6)$ -Desaturase	83/L38207/L41280/L2663
<b>Fatty acid elongation</b>		
CHAD	3-Hydroxyacyl-CoA dehydrogenase	L8921/L13304 L1252/L19534/L6293/L11094/L11627/L26638/L42044/L2515/L10374/L15640/L19544/L14315/L8921/L37338/L8226/L7050/L8226/L6694/L4400/L20583/L13304/L7823/L8097/L6805/L13304/L6694
ECH	Enoyl-CoA hydratase	0583/L13304/L7823/L8097/L6805/L13304/L6694
NADH		
TER	Trans-2-enoyl-CoA reductase	L4494/L3333
ACOT1_2_3_4	Palmitoyl-CoA hydrolase	L466/L9845

PPT	palmitoyl-protein thioesterase	L16781
KCS	3-ketoacyl-CoA synthase	L23464/L4463/L5198/L5510
HSD17B 12	17beta-estradiol 17- dehydrogenase / very-long- chain 3-oxoacyl-CoA reductase	L6757
PHS1	very-long-chain (3R)-3- hydroxyacyl-[acyl-carrier protein] dehydratase	L8286/L8873

### 3.3.3.3 *DXS* and *HMGR* relative transcript abundance by qPCR

Relative transcript abundances of selected EO biosynthetic genes, *DXS1*, *DXS2* and *HMGR* as analyzed via quantitative RT-PCR, are summarized in Figure 3.5 A and B. In the same Figure, relative transcript abundances for these genes as determined from the RNA-seq data are given for comparison (Figure 3.5 C and D). The RNA-seq and qPCR data are each presented as relative expression ratios using S2 as control stage (denominator).

Real-time PCR data were normalized to reference gene, actin-related protein 5 (transcript library ID L2604). There were no replicates in the RNA-seq experiment, however, the qPCR experiment was carried out with three biological and two technical replicates.



**Figure 3.5** Comparison of relative transcript abundances for *DXS1*, *DXS2* and *HMGR* genes across three coriander seed developmental stages (S1, S2 and S3) between quantitative RT-PCR and RNA-seq data. Error bars on the qPCR data displays SEM of three biological replicates and two technical replicates for each column. No replicates, therefore no error values available for RNA-seq data. Relative (A) *DXS1* and *HMGR* via qPCR, (B) *DXS2* and *HMGR* via qPCR, (C) *DXS1* and *HMGR* via RNA-seq and (D) *DXS2* and *HMGR* via RNA-seq.

### 3.4 Discussion

#### 3.4.1 Quality of *C. sativum* Transcriptome Dataset

Of the three alternative k-mer length transcript libraries assembled, the 21 k-mer library was selected as the best quality, due to the length distribution which looked better than the other libraries, as it had the smoothest negative binomial distribution; the longest transcript length for the most abundant number of transcripts, as well as the widest range

of transcript lengths. Additionally, the alignment results in the 21-mer library had the highest number of transcripts aligned with a 1E-10 cutoff e-value (Table 3.2).

Transcript sequences with no BLASTx hits (no significant homology to other sequences) are likely novel and are involved in functions specific to *C. sativum*. The study of these genes may lead to uncovering evolutionary or species-specific processes including adaptation and speciation. Figure 3.1 demonstrates how the proportionality of the *C. sativum* transcriptome compares to those of two reference databases, *Medicago truncatula* and *Arabidopsis thaliana*. This comparison was made to determine whether bias occurred among the mRNA's sequenced in coriander. Based on the GO percent distributions of *C. sativum* to those of the two reference plants, the coriander transcriptome is proportionally representative of the mRNA transcripts.

The cDNA synthesis is considered to be sufficiently robust as previous research has shown that RNA fragmentation prior to cDNA synthesis greatly improves the uniformity of sequence coverage over all the transcripts by reducing the amount of secondary structures in the RNA template, which can cause shielding of certain transcript portions while favoring others (Mortazavi, 2008). In this study the RNA was fragmented chemically before reverse transcription using both random hexamers and oligo dT primers. The hexamers easily anneal throughout the RNA molecule, including regions where secondary structures are present. However, there is a tendency for over-estimation of copy-number to occur when using hexamers alone. Oligo dTs will only anneal to the poly A tail and thus reverse transcription is always initiated at the 3' end of the RNA template. In this case secondary structure in the RNA molecule may lead to incomplete cDNA synthesis. When both primer types are used in conjunction the cDNA's quality is improved as the reverse transcription gains the benefits of both primer types. For example, poly A tail primers can represent low abundance transcripts which are under-represented by hexamer primers.

To determine the quality of the Velvet short read assembly, three of the next generation sequencing (NGS)-derived transcript sequences were aligned against Sanger sequenced known coriander genes from Genbank,  $\Delta^4$ -palmitoyl-acyl carrier protein desaturase ( $\Delta^4$ -palmitoyl-ACP desaturase),  $\beta$ -ketoacyl-ACP synthetase I and oleoyl

(petroselinoyl) -ACP thioesterase. Each of these known coriander nucleic acid sequences aligned to transcripts in the coriander transcript library with percent identities above 71% (Table 3.9). The  $\Delta^4$ -palmitoyl-ACP desaturase aligned with 100% identity to transcript ID L6822; both nucleic acid sequences had open reading frames of 1158 base pairs. The  $\beta$ -ketoacyl-ACP synthetase I gene aligned with 93% identity to transcript ID L46502 with open reading frames of 1458 and 552 base pairs, respectively. The  $\beta$ -ketoacyl-ACP synthetase I aligned with 72% identity to the transcript ID L117 which had an open reading frame closer in size to the known coriander synthetase gene, 1305 base pairs. Regarding oleoyl (petroselinic) -ACP thioesterase, both L12150 and L12944 transcript IDs aligned with 99% and 80% identity, respectively. L12150 and the known coriander thioesterase shared a 717 base pair reading frame while L12944 was a bit longer, 816 base pairs. These percent identity values show that short read assembly was of sufficient quality to enable the identification of transcripts within the library which correspond to known coriander genes currently available on Genbank.

**Table 3.9** Percent identities for alignments between coriander transcript library against three known coriander fatty acid nucleotide sequences.

Known Gene	Transcript Library ID	% Identity
$\Delta^4$ -palmitoyl-ACP desaturase	L6822	100.0
$\beta$ -ketoacyl-ACP synthetase I	L46502	92.96
	L117	71.68
Oleoyl (petroselinoyl) -ACP thioesterase	L12150	99.58
	L12944	79.50

The KEGG database contains reference pathways against which the coriander transcriptome was compared. These KEGG annotations served to identify active biological pathways in coriander. Among those active biological pathways are photosynthesis (Table 3.7), isoprenoid (including specialized products) biosynthesis (Table 3.5 and 3.6) and fatty acid biosynthesis and metabolism (Table 3.8) (Xia, 2011).

### ***3.4.2 Implications of Gene Expression of Isoprenoid Biosynthetic Genes***

The data were analyzed using Reads Per Kilobase per Million mapped reads (RPKM), which is a normalization method that combines between and within sample normalization

by normalizing a transcript's read count with the length of that transcript, as well as the total number of reads in the experiment.

For a gene to be considered differentially expressed between two samples or two conditions (treatments) there must be an absolute log<sub>2</sub>-fold change greater than 0.5 (Marioni, 2008). Transcripts for DXS, which is considered a key regulatory enzyme in isoprenoid biosynthesis (Munoz-Bertomeu, 2006; Rodriguez-Concepcion, 2010) were more abundant (2.7-fold) than those of HMGR, the key regulatory gene for the MVA pathway. The presence and modest peak pattern at S2 of DXS, HMGR and all other isoprenoid prenyl transferase genes suggests that a slight peak in isoprenoid production in *C. sativum* seeds occurs during the mid-developmental stage.

In a study by Lane et al, 2010, it was found that *L. angustifolia* flowers exhibit clear differential expression of the DXS and HMGR genes with DXS expressing 7-fold more than HMGR, leading to the conclusion that the flower terpene content was primarily produced via the DXP pathway (Lane, 2010). The seeds of *C. sativum* exhibit a relatively constant and constitutive pattern of both DXS and HMGR gene expression, with DXS being very slightly up-regulated (2.7-fold) over HMGR. This suggests that, like in *L. angustifolia* flowers, in *C. sativum* seeds, terpene content of the EOs are likely primarily produced through the DXP pathway. However, the differential expression between DXS and HMGR genes was not as pronounced in coriander seeds as was the case in the lavender flowers.

According to the literature, linalool content is initially at ~40% total EO, then slightly increases at S2 to ~45% and finally rises to ~75% by late development (S3) (Msaada, 2009b). The initially constant linalool synthase expression correlated with the relatively constant linalool EO content at both S1 and S2. As the linalool content spiked at the end of seed development, so did the putative linalool synthase mRNA (*mTPS3*) expression (Figure 3.4). However, *mTPS1* (CsLINS) has a 5-fold down-regulation in transcript abundance from S2 to S3 (Figure 3.4) This trend does not at all parallel that of the linalool EO content from S2 to S3. It is likely that the primary reason for the large amount of linalool in coriander EO at S3 is due to accumulation of the monoterpene over time rather than the increased linalool synthase expression at S3.

In the case of  $\gamma$ -terpinene synthase, transcript abundance also peaked at S2 and did not parallel  $\gamma$ -terpinene oil content, which dips lowest at mid development. This may be due to a slightly different EO composition in the coriander individuals grown for this experiment, as compared to those plants used for Msaada 2009b EO composition studies. The low  $\gamma$ -terpinene content could also result from the post-transcriptional regulation of the related genes due to developmental or environmental factors. In this context, it is known that EO content is sensitive to a plant's growing environment.

Another mTPS transcript (*mTPS2*) identified in this study corresponded to a putative myrcene/ocimene synthase gene, with relatively constant transcript abundance throughout seed development (Figure 3.4). Given that over 35 monoterpenes can be identified in coriander EO (Msaada, 2009b), and only 4 monoterpene synthase genes were identified in this study, it is likely that as with many known plant TPSs (Jones, 2008; Wise, 1998), coriander terpene synthases are multiproduct enzymes that can produce several monoterpene products from a single GPP substrate.

Two sesquiterpene synthase genes were identified that, according to GO annotations, encoded for enzymes responsible for the production of  $\beta$ -caryophyllene,  $\alpha$ -humulene, and germacrene D. Both genes were KEGG annotated as part of the  $\beta$ -caryophyllene and  $\alpha$ -humulene biosynthetic pathways. Given that coriander EO contains all three sesquiterpenes, it is likely that one of those genes predominantly converts FPP to  $\beta$ -caryophyllene and  $\alpha$ -humulene as observed for a known tomato sesquiterpene synthase (Schillmiller, 2010), while the other exclusively converts FPP to germacrene D. More work is required to conclusively establish this. The diterpene biosynthetic genes in coriander such as, ent-kaurene acid hydroxylase and gibberellin 3- $\beta$ -dioxygenase, are directly involved in gibberellin biosynthesis, diterpene-derived plant hormones which play roles in fruit/seed senescence. These diterpene biosynthetic genes exhibited constant levels of transcription throughout seed development.

The FPPS transcripts were more abundant than those of GPPS, and it is likely that FPPS feeds the biosynthesis of large amounts of non-EO related metabolites such as triterpenes, which are also derived from FPP, whereas GPPS feeds the biosynthesis of monoterpenes (EO metabolites). All plants possess triterpenes, many of which are

precursors to important plant sterols and hormones, e.g.,  $\beta$ -sitosterol, stigmasterol and brassinosteroids (Benveniste, 2004; Clouse, 1998). Triterpene biosynthetic gene transcription in coriander exhibited constant levels across seed development with genes specific to brassinosteroid biosynthesis found (L15352 and L17385). These plant steroids play important roles in plant cell growth or elongation, resulting in processes such as pollen tube growth or the unrolling of leaves (Buchanan, 2002).

The high GGPPS transcript expression suggests that coriander seeds also produce tetraterpenes, which are commonly found in plant seeds as precursors to important growth hormones (e.g., abscisic acid), photoprotective quenching compounds and as accessory pigments in the photosynthetic system (Howitt, 2006; Maluf, 1997). The four most abundantly transcribed tetraterpene genes in coriander were, 15-cis-phytoene desaturase and zeaxanthin epoxidase,  $\zeta$ -carotene desaturase and prolycopene isomerase. This indicates that carotenoid biosynthesis in coriander is more actively transcribing genes involved with the  $\beta$ -carotene, rather than the  $\alpha$ - branch of carotenoid biosynthesis. The  $\beta$ -carotene branch gives rise to zeaxanthin xanthophylls rather than lutein xanthophylls. It is the  $\beta$ -carotene branch which goes on to yield abscisic acid. This plant hormone has been shown to accumulate during seed development due to an association with seed maturation, an increased tolerance to desiccation and the suppression of embryo development (Buchanan, 2002). The key abscisic acid biosynthesis regulatory gene, 9-cis-epoxycarotenoid dioxygenase, exhibited a 5-fold reduction in transcript abundance from S1 to S2, followed by a 2-fold increase from S2 to S3 (Table 3.6). It is known that expression of this rate limiting enzyme correlates with the amount of abscisic acid biosynthesized, and transcription of this gene occurs in response to water stress (Qin, 2002).

Relative transcript abundance of DXS1, DXS2 and HMGR was studied by real time PCR in order to determine whether primers designed from the RNA-seq derived sequence information would yield any results in a qPCR experiment, using different individuals of the same coriander species. A brief glimpse into the amount of biological variability found within the *Coriandrum sativum* species was observed. In the plots shown in Figure 3.5, everything is related to S2. Thus, from the denominator (S2) to numerator there is a

down-regulation or up-regulation if the bar has a negative or positive value, respectively. At first glance it is clear that the relative transcript abundance trend for DXS1, DXS2 and HMGR as determined by qPCR do not match the trends as determined from the RNA-seq experiment. These two datasets, from the qPCR and the RNA-seq are not comparable as the biological replicates used in the qPCR experiment differ from those used in the RNA-seq experiment. From these two datasets, the great amount of biological variation found between individuals of the same species can be seen. In the qPCR analysis, differential transcript abundance occurs as a 5-fold up-regulation from S1 to S2 for DXS1. In the case of DXS2 there is an 11-fold down-regulation from S2 to S3 (Figure 3.5 A and B). The RNA-seq data tell a different story, where DXS1 exhibits a 2-fold down-regulation from S2 to S3, while DXS1 maintains relatively constitutive levels of transcript abundance throughout the three stages of seed development (Figure 3.5 C and D).

It should be noted that the real-time PCR and transcriptome sequencing data only suggest changes in the transcript abundance, and do not necessarily represent “protein levels” since translation to active protein may be post-transcriptionally and post-translationally regulated (Barrett et al., 2005; Valasek, 2005).

Transcripts for all the genes necessary for photosynthesis to occur were found in the coriander transcript library (Table 3.7). When coupled with the fact that coriander seeds remain green tissues throughout development as well as the observation of stomata on the seed surface (via SEM, Figure 2.1 C), these findings indicate that as coriander seeds develop they are actively photosynthesizing, producing their own energy.

Three coriander fatty acid biosynthetic genes known today, palmitoyl- $\Delta^4$ -acyl carrier protein-desaturase (accession: M93115.1) [L6822],  $\beta$ -ketoacyl-acyl carrier protein-synthase I (accession: AF263992.1) [L117, L46502] and petroselinoyl-acyl carrier protein thioesterase (accession: L20978.1) [L12150] were BLASTn aligned against the coriander transcript library and the transcripts which had the smallest e-values and highest conserved identities (79-100%) are shown in square brackets, as transcript IDs, and are bold in Table 3.8. All three genes are involved with the biosynthesis of petroselinic acid which occurs in the plastids (Mekhedov, 2001). Firstly, there is the  $\Delta^4$  desaturation of palmitoyl-ACP by palmitoyl- $\Delta^4$ -acyl carrier protein-desaturase (Cahoon,

1992) and secondly the elongation of  $\Delta^4$  hexadecenoyl by  $\beta$ -ketoacyl-acyl carrier protein-synthase I (Mekhedov, 2001). Finally, petroselinoyl-ACP thioesterase terminates the chain elongation, yielding petroselinic acid (18:1  $\Delta^6$ ) (Dormann, 1994).

In conclusion, a coriander transcript library was produced via Illumina technology and the isoprenoid biosynthetic transcript abundance was analyzed across three *C. sativum* seed developmental stages. Select prenyl precursor genes, e.g., DXS2, were found to exhibit differential transcript abundance between the three developmental stages studied. Additionally, three of the four genes encoding mTPSs were found to be differentially expressed, with varying trends, across the three developmental stages. Upon analysis of relative transcript abundance for the DXS1, DXS2 and HMGR genes via real time PCR, the great biological variability which occurs between individuals of the same species was observed as different relative transcript abundance trends were found via qPCR than with the Illumina sequencing, for the same genes. All genes necessary for active photosynthesis and fatty acid biosynthesis/metabolism were also identified in the coriander transcript library using the BLASTx alignment results and the KEGG and GO annotations.

## CHAPTER 4: MOLECULAR CLONING AND FUNCTIONAL CHARACTERIZATION OF $\gamma$ -TERPINENE SYNTHASE, AND CLONING AND ATTEMPTED PROTEIN EXPRESSION OF PUTATIVE (S)-LINALOOL SYNTHASE.

### 4.1 Synopsis

Using sequence homology and searching for the presence of specific amino acid motifs, conserved between known monoterpene synthase genes, together with GO and KEGG annotations, four monoterpene synthase and two sesquiterpene synthase candidate genes were identified from the coriander transcript library. A phylogenetic analysis of these candidate genes revealed that three of the monoterpene synthase candidates clustered into the terpene synthase TPSb subfamily while the fourth fit into the TPSg subfamily. The two sesquiterpene synthase candidates clustered into the TPSa subfamily.

Two of the coriander monoterpene synthase genes, *Cs $\gamma$ TRPS* and *mTPS3*, were cloned and transformed into *Escherichia coli* to produce the corresponding recombinant proteins. One of those, *Cs $\gamma$ TRPS* was found to encode a 66 kDa protein with  $\gamma$ -terpinene synthase activity while *mTPS3* was never successfully expressed as an active enzyme. After purification by Ni-NTA affinity chromatography, characterization of *Cs $\gamma$ TRPS*'s kinetics properties demonstrated that this enzyme's  $V_{max}$ ,  $K_m$  and  $k_{cat}$  were  $2.2 \pm 0.2$   $\mu\text{kat}/\text{mg}$ ,  $66 \pm 13$   $\mu\text{M}$  and  $1.476 \times 10^{-4}$   $\text{s}^{-1}$ , respectively. *Cs $\gamma$ TRPS* exhibited a preference for manganese as catalytic cofactor however it performed optimally when both magnesium and manganese were present at concentrations of 50 mM and 1 mM, respectively.

### 4.2 Materials and Methods

#### 4.2.1 Monoterpene Synthase Candidate Selection

##### 4.2.1.1 Sequence Homology Search and Phylogentic Analysis

Four monoterpene synthase (mTPS) candidate genes (*mTPS1*, *mTPS2*, *mTPS3* and *Cs $\gamma$ TRPS* – *Coriandrum sativum*  $\gamma$ -terpinene synthase) were chosen based on sequence homology to known mTPS genes, especially the presence of certain conserved

motifs shared by all known TPS genes, DDXXD, (N,D)D(L,I,V)X(S,T)XXXE and RRX<sub>8</sub>W, as well as one partially conserved motif, LQLYEASFLL. [Note: *mTPS1/CsLINS* was discovered in the transcript library by Lukman Sarker (See Preface, and Galata et al., 2013)]. Those transcripts with gene ontology (GO) annotation “monoterpene biosynthetic process” as well as transcripts with BLASTx hits to known mTPS genes having an e-value less than 10E-60 were selected as coriander mTPS gene candidates. Sesquiterpene synthase candidates were selected in the same fashion. Protein sequence alignments were made between the four coriander mTPS candidates against *Citrus limon*  $\gamma$ -terpinene synthase, *Lavandula angustifolia* linalool synthase, *Salvia fruticosa* 1,8-cineole synthase, *Cannabis sativa* limonene synthase, *L. angustifolia*  $\beta$ -phellandrene synthase and *Salvia officinalis* sabinene synthase were performed with ClustalW2.

Thirty-nine known plant terpene synthase genes were selected (see Accession Numbers section for details) for construction of the phylogenetic tree; their sequences were obtained in FASTA format from Genbank. The multiple protein sequence alignment and phylogeny were performed with Clustal W2 and the tree was assembled using Treeview software via the neighbor-joining method.

## 4.2.2 Cloning of Full Length Monoterpene Synthases

### 4.2.2.1 $\gamma$ -Terpinene Synthase

The N-terminal signal peptide sequence of *CsyTRPS* was predicted using ChloroP (v.1.1). The open reading frame of *CsyTRPS* excluding the N-terminal signal peptide was sticky end PCR amplified (Zeng, 1998) by Kapa HiFi DNA polymerase ready-mix (Kapa Biosystems Inc., Woburn, MA, USA). Sticky end primers, forward (5' TATGTC GAA TGT TAG AAG ATC CGG AAA TTA TCC 3' and 5' TGT CGA ATG TTA GAA GAT CCG GAA ATT ATC C 3') and reverse (5' TTA GAC TAA ACT CTA TAG GTA TGG GGT CAA CA 3' and 5' AGC TTT AGA CTA AAC TCT ATA GGT ATG GGG TC 3'), were designed to create EcoR1 and Xho1 (New England Biolabs, Whitby, ON, Canada) overhang regions for use during ligation of *CsyTRPS* into pET41b(+) bacterial expression vector (EMD Chemicals, Darmstadt, Germany). Amplifications were carried out in 25  $\mu$ l volumes

containing 1 X Kapa HiFi ready mix, 0.3  $\mu$ M each of forward and reverse primers and ~1  $\mu$ g template. The PCR program used was as follows: 95 °C for 5 min, followed by 35 cycles of 98 °C for 20 sec, 60 °C for 15 sec and 72 °C for 1 min, and a 5 min final extension at 72 °C.

The amplicons were purified with an E.Z.N.A DNA Purification Kit (Omega Bio-Tek Inc, Norcross, GA, USA), and final DNA concentration and purity was determined by UV-Vis spectrophotometer (Nanodrop ND-100, Thermo Scientific, Wilmington, DE, USA).

To facilitate downstream protein purification the amplicon was fused to a C-terminal eight histidine moiety in the pET41b(+) vector. Two thousand five hundred nanograms of the pET41b(+) vector were digested with 6 U of NdeI and 7 U of HindIII restriction enzymes (New England Biolabs, Ipswich, MA, USA) in a 50  $\mu$ l digestion volume buffered with 1X NEB buffer 2. Digestion product as well as sticky-end PCR products were all loaded onto a 1.5 % agarose gel and excised before gel extraction via the Omega E.Z.N.A gel extraction kit. Each of primer set 'a' and 'b' gel extracted products (436.5 ng of each) were added together and denatured at 95 °C for 5 minutes before being left to cool slowly at room temperature for ~10 minutes. One hundred nanograms of digested vector and 145 ng of gene of interest were ligated to each other in a 20  $\mu$ l reaction volume with 3 U of T4 DNA ligase and 1X T4 DNA ligase buffer (New England Biolabs) at 16 °C overnight (12-16 hours).

One microliter of resulting plasmid was transformed into 25  $\mu$ l of chemically competent DH10B *Escherichia coli* bacterial cloning cells. Cells were treated with a 42 °C heat shock for 45 seconds followed by two minutes on ice before the addition of 975  $\mu$ l of SOC (Super Optimal Broth with Catabolite repression) recovery media. Transformed cells were incubated at 37 °C and 150 rpm and then plated onto solid Luria-Bertani media supplemented with 30  $\mu$ g/ml kanamycin. Plates were incubated at 37 °C overnight. Digested (linearized) pET41b(+) vector was transformed into DH10B *E. coli* as a negative control.

Five isolated colonies were used to inoculate 5 ml liquid LB cultures supplemented with 30  $\mu$ g/ml kanamycin and grown at 37 °C and 180 rpm overnight. Plasmids were

extracted using Omega E.Z.N.A Plasmid kit, spin protocol and concentrations and plasmid purity were determined with via the Nanodrop spectrophotometer. Plasmids were sent for sequencing to the National Research Council - Plant Biotechnology Institute (NRC-PBI, Saskatoon, SK, Canada). A diagnostic restriction enzyme digestion was also performed and products were visualized on a 1.0 % agarose gel. Fifteen percent glycerol stocks were prepared for each of the five colonies from which plasmids were extracted. Glycerol stocks were flash frozen in liquid nitrogen and stored at -80 °C.

#### **4.2.2.2 Other *mTPS* candidates**

Each of *mTPS2* and *mTPS3* were cloned into expression vectors. Three sets of primers were prepared for *mTPS3*, two for cloning into pET41b(+) and the third for cloning into pGEX-4T-1. It should be noted that cloning into pET41b(+) bacterial expression vector was attempted first, with the initial primer set including the signal peptide as part of the full length amplicon product while the second primer set excluded the signal peptide. A signal peptide of 78 nucleotides was identified using ChloroP publicly available software. The third *mTPS* (*mTPS1/CsLINS*), was also cloned into the pET41b(+), expressed in bacteria, and functionally characterized by Lukman Sarker.

#### **First *mTPS3* clone:**

Cloning of *mTPS3* into pET41b(+) with the signal peptide as part of the full length amplicon was performed first. The putative (S)-linalool synthase gene was amplified from cDNA which was prepared as previously described in the cloning of  $\gamma$ -terpinene synthase, using 1 U iProof high fidelity DNA polymerase (Bio-Rad Laboratories Ltd., Mississauga, ON, CAN), 3.5 mM MgCl<sub>2</sub>, 0.1 mM ea. dNTPs, 0.5  $\mu$ M forward (5' - GCT CAG CAT ATG GAA AAG GTT AAG CGT GAA C - 3') and reverse (5' - GAT CTC GAG GGG GGG TAA AAC AAA CAT G - 3') primers, 1X thermopol buffer and ~1  $\mu$ g template. The NdeI and XhoI restriction enzyme sites were designed into each of the forward and reverse primers, respectively (underlined). PCR program was as follows: 95 °C for 5 minutes followed by 35 cycles of 95 °C for 30 seconds, 60 °C annealing

temperature for 30 seconds and 72 °C extension for 1 minute 30 seconds. There was a final extension at 72°C for 10 minutes.

Amplicon was purified using the PCR purification protocol in the Omega E.Z.N.A Gel Extraction kit. The *mTPS3* amplicon and pET41b(+) vector were each digested with 22 U NdeI and 166 U XhoI restriction enzymes. The 50 µl digestions included 1X NEB buffer 4, 1X BSA from NEB and 5 µg DNA template. Digestions were incubated at 37 °C for 1.5 hours followed by enzyme inactivation at 65 °C for 20 minutes. Digested products were loaded onto a 1.5% agarose gel from which products of interest were excised and gel extracted via Omega E.Z.N.A Gel Extraction kit. Gel extracted 90 ng of *mTPS3* and 75 ng of pET41b(+) were ligated together, transformed in DH10B *E. coli* and plasmid extracted as described above.

### **Second *mTPS3* clone:**

Cloning of truncated *mTPS3* (signal peptide removed) was performed next. The unigene, was amplified from cDNA via the same protocol described above for the full length *mTPS3*, with a 66 °C annealing temperature. The forward (5'-TAC AGT CAT ATG GTT CAA AGG TTA GGG ATT GAT TA-3') and reverse (5'-GCT ACT AAG CTT GGG GGG TAA AAC AAA CAT G-3') primers had NdeI and HindIII restriction enzyme sites designed into each primer, respectively (underlined). A non-specific product of about 2000 base pairs in size was removed by gel excision of the gene of interest for use as template in a second PCR amplification, using 2µl of purified gel extracted amplicon as template.

Truncated *mTPS3* was restriction enzyme digested with 29.5 U of NdeI and 34.5 U of HindIII, as well as 1X NEB buffer 2 and 3000 ng of template. The pET41b(+) vector was digested as previously described for full length *mTPS3* cloning and same digestion protocol were used. Digestion products were gel extracted via the Omega E.Z.N.A gel extraction kit, and eluted in 30 µl volumes. Ligation was carried out with 145 ng of *mTPS3* and 100 ng of pET41b(+). Concentrations of all other ligation reagents and protocol were the same as for full length *mTPS3* into pET41b(+).

Plasmid was transformed into DH10B *E. coli* as previously described and plasmid extracted from three isolated colonies using the Omega E.Z.N.A plasmid mini kit. A diagnostic restriction enzyme digestion using NdeI and HindIII was performed on extracted plasmids and results were visualized on a 1.0 % agarose gel. Plasmid extracted from colonies 2 and 3 were sent for DNA sequencing.

### **Third *mTPS3* clone:**

The third cloning of *mTPS3* also excluded the signal peptide. Sticky end primers, forward (5'-AAT TCA GGC AAC ACA ACT ACT ATG T-3' and 5'-CAG GCA ACA CAA CTA CTA TGT GT-3') and reverse (5'-GTA AAA CAA ACA TGT CCT TGA TGT ACT C -3' and 5'-TCG AGT AAA ACA AAC ATG TCC TTG ATG-3'), were used for this trial. The cDNA used for sticky-end PCR amplification (Zheng, 1998) of *mTPS3* was prepared in the same manner as for the first two *mTPS3* clones. The sticky-end PCR recipe was the same as for *CsyTRPS*, and the protocol began with 95 °C for 5 minutes, followed by 35 cycles of 98 °C for 20 seconds, 60 °C annealing temperature for 15 seconds, and 72 °C for 1 minute. There was a final extension at 72 °C for 5 minutes.

The pGEX-4T-1 vector was restriction enzyme digested with 19.5 U of EcoRI, 97 U of XhoI, 2500 ng of template and 1X NEB Eco buffer (New England Biolabs). Digestion reagents were incubated at 37 °C for 2 hours followed by enzyme inactivation at 65 °C for 20 minutes. Digestion product and sticky-end PCR products denatured and re-annealed as described for *CsyTRPS*.

Ligation was carried out in a 20 µl volume with 100 ng of vector, 128.4 ng of gene of interest, 1X T4 DNA ligase buffer and 3 U T4 DNA ligase (New England Biolabs). Ligation reagents were incubated at 16 °C overnight. Resulting plasmids were transformed into electrically competent DH10B *E. coli* cloning cells. One microliter of plasmid was electroporated into 40 µl of cells (Micropulser electroporator, Bio-Rad). After addition of 960 µl of SOC media, transformation mixture was incubated at 37 °C and 150 rpm for 1 hour. Cells were plated onto solid LB media supplemented with 100 µg/ml ampicillin and incubated overnight at 37 °C.

Cloned plasmids were extracted from DH10B using the Omega E.Z.N.A plasmid mini kit and a diagnostic PCR amplification was performed using 0.5  $\mu$ M each of primer set 'a', 1X thermopol buffer (New England Biolabs), 2.5 mM MgCl<sub>2</sub>, 0.1 mM each dNTPs, 1  $\mu$ l of extracted plasmid template and 1 U *Taq* DNA polymerase. PCR program began with 95 °C for 5 minutes followed by 35 cycles of 95 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 1 minute and 30 seconds. There was a final extension at 72 °C for 10 minutes.

#### **The *mTPS2* clone:**

The *mTPS2* gene, including the signal peptide, was amplified using iProof HiFi DNA polymerase and cloned into pET41b(+) as described for the first *mTPS3* clone. The following primers were used: forward (5' TAC AAT CAT ATG GGG TTG CTC AGC TTG TAT G 3') and reverse (5' GCT CTC GAG AAG AGT AAA AGG TTC CAC C 3') and the annealing temperature was 56 °C for 30 seconds. Two thousand five hundred nanograms of amplicon were restriction enzyme digested with 14.88 U Nde1 and 104 U Xho1 in 10X NEB buffer 4 supplemented with 1X NEB BSA solution. Ligation reaction and transformation into DH10B *E. coli* was carried out the same as for *mTPS3* clone. Diagnostic restriction enzyme digestion was performed on ten plasmids extracted from isolated *E. coli* cells after transformation.

### **4.2.3 Recombinant Protein Expression, Crude Enzyme Assay and Purification**

#### **4.2.3.1 $\gamma$ -Terpinene synthase**

The *Cs $\gamma$ TRPS* expression construct was transformed via heat shock into *E. coli* Rosetta (DE3) pLysS expression cells (Novagen, Darmstadt, Germany). Successful transformants were selected on solid Lauria-Bertani (LB) medium containing 30  $\mu$ g/ml of kanamycin and 34  $\mu$ g/ml of chloramphenicol. Single colonies were inoculated in 5 ml LB media with the same antibiotics and grown at 37 °C and 190 rpm for 14-16 hours. The 5 ml cultures were transferred to 95 ml LB media with antibiotics and incubated at 30 °C and 190 rpm until OD<sub>600</sub> of ~0.8 was reached. Protein expression was induced by the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) to a final concentration of 0.5

mM, and incubation at 18 °C and 190 rpm for 14-16 hours. Induced cells were split into two 50 ml portions and pelleted at 4 °C and 4000 rpm for 20 minutes before storage at -80 °C to lyse the cells. A non-induced culture (with no IPTG added) and induced culture with cells transformed with empty pET41b(+) vector were included as controls for basal protein expression from host bacteria and protein expressed from the empty vector, respectively.

For initial activity determination, frozen induced pellets were resuspended in 1 ml crude assay buffer (25 mM TRIS-Cl, 5% glycerol, 1 mM dithiothreitol (DTT), 10 mM MgCl<sub>2</sub> and 1 mM MnCl<sub>2</sub> at pH 7.5) (Crowell and Williams, 2002) supplemented with 1 mM protease inhibitor, phenylmethanesulfonylfluoride (PMSF). Cells were sonicated on ice for six 10 sec intervals with 1 min cool down periods using a Sonic Dismembrator Model 100 (Fisher Scientific, Ottawa, ON, Canada). The lysate was centrifuged at 14,000 rpm at 4°C for 11 min to obtain the soluble protein fraction. One hundred microliters of crude soluble lysate was added to 2.9 ml of crude assay buffer with 25 µM geranyl diphosphate (GPP) as substrate. The assay mixture was overlaid with 1 ml pentane to trap the volatile products and incubated in a 30 °C water bath for 2 hours. Following incubation, the assay was mixed vigorously and pentane overlay was removed for product analysis.

For CsγTRPS purification, *E. coli* pellets containing CsγTRPS were resuspended in 5 ml Ni-NTA binding buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl and 10 mM imidazole at pH 8.0) supplemented with 1 mM PMSF. Cells were sonicated on ice for eight 15 sec intervals with 1 min cool down periods in between. The soluble fraction was collected by centrifugation at 10,500 rpm and 4 °C for 15 min. The His-tagged protein was then purified from the soluble lysate by Ni-NTA agarose affinity chromatography (EMD Chemicals, Germany) according to the manufacturer's protocol. Pure, soluble, and insoluble protein as well as total protein from non-induced *E. coli* cells was resolved with 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by Coomassie Brilliant Blue staining (Supplementary Figure S4). Total protein concentration was determined via the Bradford photometric method and CsγTRPS quantitative analysis was performed by densitometry using Kodak 1D software (v. 3.6).

#### **4.2.3.2 *mTPS3* (putative (*S*)-linalool synthase)**

##### **First *mTPS3* attempt:**

Three microliters of the expression construct for *mTPS3* (with signal peptide) was transformed into 50  $\mu$ l Rosetta (DE3) pLysS expression *E. coli* cells (Novagen) via heat shock protocol described above. Transformed cells were plated onto solid LB media supplemented with 34  $\mu$ g/ml chloramphenicol and 30  $\mu$ g/ml kanamycin and incubated at 37 °C overnight. Single colonies were used to inoculate 5 ml LB broth supplemented with same antibiotics as plates and grown at 37 °C and 180 rpm overnight. The 5 ml cultures were added to 45 ml fresh LB medium with antibiotics and growth was continued at 37 °C and 180 rpm until OD<sub>600</sub> was equal to ~0.8. The 50 ml culture was split into two 25 ml cultures (one destined for induction at 18 °C for 14-16 hours and the other at 37 °C for 3.5 hours). One millimolar IPTG was added to cultures and protein induction was carried out at aforementioned temperatures with 180 rpm shaking speed.

After induction cells were pelleted and stored at -80 °C overnight. Each ~0.2 g pellet was resuspended in 7 ml/g of 1X phosphate-buffered saline (PBS). Bacterial cell disruption was completed by sonication as described for C<sub>s</sub>TRPS. Sonicates were centrifuged at 12,000 rpm for 15 minutes to separate the soluble and insoluble fractions. Each of pellet, soluble, non-induced control and pET41b(+) control were resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (5% stacking and 12% running gel) and visualized with coomassie brilliant blue staining.

A second protein induction was carried out on full length *mTPS3* at 23 and 30 °C with 1 mM IPTG, overnight. Culture and corresponding pellets sizes were the same as described above for 18 and 37 °C inductions.

##### **Second *mTPS3* attempt:**

The pET41b(+) construct harboring the truncated *mTPS3* was transformed into chemically competent Rosetta (DE3) pLysS expression *E. coli* cells as previously

described for full length mTPS3. Single colonies were used to inoculate liquid LB media and 100 ml cultures were grown to  $OD_{600} \sim 0.8$  before addition of 1 mM IPTG and incubated overnight at 18 °C and 180 rpm. All media contained both chloramphenicol and kanamycin antibiotics. Protein was extracted via the same protocol as for full length mTPS3 and visualized by SDS-PAGE.

### **Third mTPS3 attempt:**

The mTPS3\_pGEX-4T-1 construct was transformed into Rosetta (DE3) pLysS expression *E. coli* cells as previously described. Seven single colonies were used to inoculate 5 ml liquid LB media supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol antibiotics. One colony transformed with empty pGEX-4T-1 vector was also inoculated into liquid LB media as a vector control. Cultures were grown at 37 °C and 180 rpm until  $OD_{600} \sim 0.8$  before addition of 0.5 mM IPTG. Induction was carried out at 37 °C and 180 rpm for 3 hours. Protein was extracted as previously described and visualized on SDS-PAGE.

Colony one was used to inoculate a 1 liter culture which was induced with 0.5 mM IPTG at 18 °C and 180 rpm overnight. Cells were pelleted at 4000 rpm and 4 °C for 20 minutes before decanting media and storage at -80 °C. Protein was extracted by pellet thawing and sonication as previously described on ice. Sonicate was centrifuged at 10,500 rpm and 4 °C for 15 minutes to separate insoluble and soluble fractions. Insoluble fraction was kept for analysis by SDS-PAGE while soluble fraction was purified by GST affinity batch purification. In addition to purification of mTPS3, a crude enzyme assay was conducted for mTPS3 with GPP and FPP as described for C $\gamma$ TRPS.

Ten milliliters of soluble fraction were incubated with 100 µl of GST resin pre-equilibrated with 10X GST (140 mM NaCl, 27 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) wash buffer at 4 °C and end over end rotation at medium speed for 1 hour. Lysate/resins were centrifuged at 4 °C and 2200 rpm for 6 minutes, then supernatant was decanted and resins were washed with 1 ml of 10X GST wash buffer; this was repeated for 3 washes. Bound protein was eluted by incubation with 100 µl of GSH elution buffer (50 mM TRIS-Cl, 10 mM reduced glutathione, pH 8.0) at 4 °C for 30

minutes with occasional tapping to stir up resins. Resins were centrifuged at 4 °C and 2200 rpm for 7 minutes and supernatant containing the eluted protein was collected. This was repeated with two more elutions to ensure the resins were clean. Protein extracted from pGEX-4T-1 vector control was also GST purified alongside protein of interest.

#### **4.2.4. Enzyme Assays and Product Analysis**

##### **Cs $\gamma$ TRPS:**

Assays were performed in 200  $\mu$ l of the assay buffer (50 mM TRIS-Cl, 5% glycerol, 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 1 mg/ml Bovine Serum Albumin (BSA) and 1 mM DTT at pH 6.8) containing 25  $\mu$ M of the GPP substrate (Echelon, Salt Lake City, UT, USA) and 5  $\mu$ g of the semi-pure protein (75.90 pmol Cs $\gamma$ TRPS). To study the linear kinetic properties of Cs $\gamma$ TRPS, six time course points, 2.5, 5.0, 10, 15, 30 and 45 minutes were chosen. Optimum temperature and pH were determined by assaying at six temperatures, 25, 27, 30, 32, 35 and 37 °C, and six pH levels. The 2-(N-Morpholino) ethanesulfonic acid (MES) buffer maintained pH's 5.5, 6.0 and 6.5 while 3-Morpholino-2-hydroxypropanesulfonic acid (MOPSO) buffer maintained pH's 7.0, 7.5 and 8.0. Divalent cation preference and optimums were determined by assaying five different MnCl<sub>2</sub> (Mn<sup>2+</sup>) and MgCl<sub>2</sub> (Mg<sup>2+</sup>) concentrations, 0.0025, 0.005, 0.25, 1.0, 5.0 mM and 2.0, 10.0, 50.0, 100.0, 250.0 mM, respectively. Negative control assays contained affinity purified protein extracted from Rosetta (DE3) pLysS *E. coli* cells which were transformed with “empty pET41b(+) vector” rather than construct containing gene of interest. For all assays two biological replicates and two technical replicates were prepared.

Volatile assay products were concentrated by exposure to a stream of pure helium gas, identified and quantified by GC-MS using a Varian GC 3800 Gas Chromatographer coupled with a Saturn 2200 Ion Trap mass detector as previously described (Demissie, 2011). The GC was equipped with a 30 m  $\times$  0.25 mm capillary column coated with a 0.25  $\mu$ m film of acid-modified polyethylene glycol (ECTM 1000, Altech, Deerfield, IL, USA); as well as a CO<sub>2</sub> cooled 1079 Programmable Temperature Vaporizing (PTV) injector (Varian Inc., USA). Two microliter concentrated samples were injected at 40 °C. The

oven temperature was initially maintained at 40 °C for 3 minutes, followed by a two-step temperature increase, first to 130 °C (ramp rate of 10 °C per minute) and then to 230 °C (ramp rate of 50 °C per minute), and held at 230 °C for 8 minutes. The helium carrier gas flow rate was 1 ml per minute (Demissie, 2011). Assay products were identified by comparison of their mass spectra to those in the NIST library database (NIST MS Search v.2.0), and by comparing their retention times and mass spectra to those of authentic standards (Sigma-Aldrich, Oakville, ON, Canada). Terpene assay products were quantified by addition of 100 µl of 1 ppm camphor internal standard and integration of the peak areas for each product. All volatile enzyme assay extracts were GC analyzed with two technical replicates (two separate 2 µl injections).

#### **Truncated mTPS3 expressed in pGEX-4T-1:**

Twenty microliters of semi-pure protein was added to 180 µl of enzymatic assay buffer as described for CsyTRPS. Purified protein from the pGEX-4T-1 vector was added to assay buffer as a negative control. Aqueous portion was overlaid with ~ 500 µl of pentane and assays were incubated in a 30 °C water bath for 1 hour. One hundred microliters of 1 ppm camphor internal standard was added to assays and reaction was stopped by immediate storage at -80 °C. Volatile products were collected and analyzed via GC-MS as for CsyTRPS.

### **4.3 Results**

#### **4.3.1 Monoterpene Synthase Candidates**

Transcripts with the GO annotation “monoterpene biosynthetic process” as well as transcripts with BLASTx hits to known mTPS genes having a cutoff e-value less than 10E-60 were selected as putative coriander mTPS gene candidates (Table 4.1).

<b>Table 4.1</b> Coriander essential oil terpene synthase candidate genes.			
Candidate	Transcript Library ID	GO Annotation	Top BLASTx hits
mTPS1 (SLINS)	CL2235-1_L3610	Myrcene synthase activity; $\beta$ -ocimene synthase activity; monoterpene biosynthetic process	Limonene synthase; $\alpha$ -terpineol synthase
mTPS2	Locus303_T15	Response to wounding; cytosol; sesquiterpenoid biosynthetic process; response to herbivore; response to jasmonic acid; myrcene synthase activity; $\beta$ -ocimene synthase activity; $\alpha$ -farnesene synthase activity	Linalool synthase; sesquisabinene B synthase; monoterpene/bisabolene synthase
mTPS3	CL360-2_L5634	(S)-linalool synthase activity; monoterpene biosynthetic process	(S)-linalool/nerolidol synthase; linalool synthase
Cs $\gamma$ TRPS	Locus4307_T7	Protein binding; sesquiterpene synthase activity; myrcene synthase activity; $\beta$ -ocimene synthase activity; sabinene synthase activity; limonene synthase activity; pinene synthase activity; monoterpene biosynthetic process	$\alpha$ -terpineol synthase; sesquiterpene synthase; limonene synthase; $\beta$ -ocimene/myrcene synthase
sTPS1	Locus_6713_T1	Cellular component; $\beta$ -caryophyllene synthase activity; sesquiterpenoid biosynthetic process; $\alpha$ -humulene synthase activity; response to herbivore; sesquiterpene biosynthetic process	Vetispiradiene synthase; germacrene D synthase
sTPS2	Locus10645_T3	Cellular component; $\beta$ -caryophyllene synthase activity; sesquiterpenoid biosynthetic process; $\alpha$ -humulene synthase activity; response to herbivore; sesquiterpene biosynthetic process	Germacrene D synthase; $\beta$ -caryophyllene synthase; sesquiterpene cyclase

Three of the four mTPS candidates, *Cs $\gamma$ TRPS*, *mTPS1 (CsLINS)*, and *mTPS2*, exhibited sequence homology to known mTPS genes, especially the presence of conserved motifs shared by all known TPS genes, DDXXD, (N,D)D(L,I,V)X(S,T)XXXE and RRX<sub>8</sub>W, as well as one partially conserved motif, LQLYEASFLL. The mTPS3 protein sequence lacked the RRX<sub>8</sub>W motif but contained the other three motifs. Multiple protein sequence alignments between coriander mTPS genes and *Citrus limon*  $\gamma$ -terpinene synthase, *Lavandula angustifolia* linalool synthase, *Salvia fruticosa* 1,8-cineole synthase, *Cannabis sativa* limonene synthase, *L. angustifolia*  $\beta$ -phellandrene synthase and *Salvia*

*officinalis* sabinene synthase were performed with ClustalW2 (Figure 4.1). Sequence alignment via the BLASTx algorithm against NCBI non-redundant protein sequences demonstrated *CsγTRPS* to share 68% conserved identity with  $\gamma$ -terpinene synthase from *Citrus unshiu*, 66% with (-)- $\beta$ -pinene synthase from *Citrus limon*, 63% with  $\alpha$ -thujene synthase from *Litsea cubeba*.

```

γ-Terpinene_synthase_C.limonon     ---MALNLLSS-----LPAACNFRTRLSLP---LSSKVNQFVFP 31
Limonene_synthase_C.sativa         MQCIAFHQFASSSSLPIWSSIDNRFTPKTSITSISKPKPKLKSNSLKR 50
CsyTRPS_C.sativum                  ---MALSILNLG---LNSSSIRSMKPKTSFTSVNPS--STARKEVFEITR 41
CsLINS_C.sativum                    -----MAAITIFPLSYSIKFRRSSPCNPKD 25
R-Linalool_synthase_L.angustif     -----MSININMPAAAVLRP-- 15
mTPS2_C.sativum                     ---MSLKGLSSN-----LLFTALCRSSFP LARNHSHKIFASKR 34
mTPS3_C.sativum                     -----
S-Linalool_synthase_A.polygama     -----MASFHRFCVSSLLVPNN 17
β-Ocimene_synthase_A.majus         -----MAFCISYLGAVLPFSL 17

MOTIF: RRX6W
γ-Terpinene_synthase_C.limonon     PITQVQYPMMAASTSS--IKPVDQTIRRSAYGFPIWSFDYIQSLDSKYK 79
Limonene_synthase_C.sativa         SRSSTCYSIQCTVVDNPSSTITNNSDRRSANYGFPIWSFDVQSLPIQYK 100
CsyTRPS_C.sativum                  IRCSSHTDAVVSRGNVSDEVSN--VRRSYYPSWWYDYDFQSLSSNFK 89
CsLINS_C.sativum                    VTACKSVIKSVTGMTKVPVPEPIVRRSANYKPIWSNDNDFLQSLKTEYT 75
R-Linalool_synthase_L.angustif     -FRCSQLHVDET-----RRSANYRPSWSNSNYIQSLNSQYK 50
mTPS2_C.sativum                    PIQHIKSKATRINHD-----HGATLRRNANYSPSWSWYDNFVKLSLSDYS 78
mTPS3_C.sativum                    -----MNGECF---FKQV 10
S-Linalool_synthase_A.polygama     SPQISNAYRAPAVPSMPTTQKWSITEDLAFISNPSKQHNHQTG---YRTF 64
β-Ocimene_synthase_A.majus         PRTKFAIFHNTSKHAAYKTRCWNIPRDVGSTPPPSKLHQALCLNAHSTSC 67

γ-Terpinene_synthase_C.limonon     GESYARQLEKLEKQVSAMLQQDNKVVLDLPHLQLELIDNLHRLGVSYHFE 129
Limonene_synthase_C.sativa         GESYTSRLNKLEKDVKRMLIG-----VENSLAQLELIDTIQRLGISYRFE 145
CsyTRPS_C.sativum                  GEICSKHASELKENVRMLLNKE---DLDSLHKLELVDTIQRLGVSYHFQ 135
CsLINS_C.sativum                    GEAINARASEMKEEVRMIFNN-----VVEPLNQLELIDQLQRLGLDYHFR 120
R-Linalool_synthase_L.angustif     EKCKQLTRLEGLIEQVKELKGT-----KMEAVQQLELIDDSQNLGLSYFYQ 95
mTPS2_C.sativum                    EEKYARQVDELKDYVKRLIHAK---TDVPLAKLELLDTVQRLGLNRYRQ 124
mTPS3_C.sativum                    QEMEKVKRELMVKNIG-----HNPKDLIIVDVVQRLGIDYAFK 49
S-Linalool_synthase_A.polygama     SDEFYVVKREKLKDVRRALREV---EETPLEGLVMIDTLQRLGIDYHFQ 110
β-Ocimene_synthase_A.majus         MAELPMDYEGKIQTRHLLHLKD---ENDPIESLIFVDATQRLGVNHHFQ 114
. . . . . : : * . * : : *

γ-Terpinene_synthase_C.limonon     DEIKRTLDR-----IHNKNTNKSLYARALKFRLLRQYGYKTPVKETF 171
Limonene_synthase_C.sativa         NEIISILKEK---FTNNDNPNPNYDLYATALQFRLLRQYGFVFP-QEIF 191
CsyTRPS_C.sativum                  DEIKRILEAM---YS--TNEKLHSHDLNAVSLKFRLLRQHSFDVP-EEIF 179
CsLINS_C.sativum                    DEINHTLKNVHNGQK---SETWEKDLHATALEFRLLRQHGHYIS-PEGF 165
R-Linalool_synthase_L.angustif     DKIKHILNLIYNDHKYFYDSEAEGMDLYFTALGFRLFRQHGFKVVS-QEVF 144
mTPS2_C.sativum                    NDVKQAVDVIYN---NSTDAWLSDDLYSTALFRFRILREHGYTVS-QDVF 169
mTPS3_C.sativum                    EEIEQVLKR---QNMEMDQIVKKNDLFVSVCFRLFRQHNYYVS-ADAF 94
S-Linalool_synthase_A.polygama     GEIGALQK---QQRKSKCDYPEHDLFEVSTRFRLLRQEGHNVNP-ADVF 155
β-Ocimene_synthase_A.majus         KEIEEILRKSY--ATMKSPSICKYHTLHDVSLFFCLMRQHGRYVS-ADVF 161
. . . . . : : * : * : : * . . : *

MOTIF: VSLYEASYLS
γ-Terpinene_synthase_C.limonon     SRFMDEK-GSFKLSSHSDECKGMLYEAAYLSVEEESSIFRDAIRFTTA 220
Limonene_synthase_C.sativa         NNFKNHKTGEFKAN-ISNDIMGALGLYEASYLSHGKKGES-ILEEARIFTTK 239
CsyTRPS_C.sativum                  KEFMDES-GKFKAS-LSKDMKGIVSLYEASYLSIKNEK-IMDEAQGFATK 226
CsLINS_C.sativum                    KRFT-ENGSFNKGI--RADVRGLLSLYEASYLSTEGES-LMEEAWSFTSN 211
R-Linalool_synthase_L.angustif     DRFKNENGTYFK---HDDTKGLLYEASYLSVREGEE-TLEQAREFATK 189
mTPS2_C.sativum                    GRFKDET-GNFKAN-LCEDVKGLLSLYEASYLSYFGIKGED-IIDKAKSFSRE 216
mTPS3_C.sativum                    DIFVNMKRKLDRG--ESNE-ALSVYEASYLSQLRMEGED-VLDEAEFLSRQ 140
S-Linalool_synthase_A.polygama     NHFRDKKGRFKSEL--SRDIRGLLSLYEASYLSIQGED-ILDQAADFSSQ 202
β-Ocimene_synthase_A.majus         NNFKGESGRFKKEEL--KRDTRGLLYEASYLSFEGER-ILDEAENFSRQ 208
. . . . . : . . : : * : * : * : :

γ-Terpinene_synthase_C.limonon     YLKEWVAKHDIDKNDNEYLCTLVKHALELPLHWRMRRLEARWFID-VYES 269
Limonene_synthase_C.sativa         CLKKYKLMSSNNNMTLISLLVNHALEMPLQWRITRSEAKWFIEETYER 289
CsyTRPS_C.sativum                  HMKDY--ISDINNTKDNLIKLVSHALEFPQHWREPRQEARWFID-FYET 273
CsLINS_C.sativum                    ILKEC--LEN---TIDLDLMQVRHALELPLQWRIPRFDAKWYIN-LYQR 255
R-Linalool_synthase_L.angustif     SLQRK--LDEDGDGIDANIESWIRHSLEIPLHWAQRLEARWFLD-AYAR 236
mTPS2_C.sativum                    HLENL-----VQGKLSPNMARKVNHAIDMPLHWKLPRLEAVWYIN-TYEQ 260
mTPS3_C.sativum                    ILEER-----MKFLDHQAITIRKTLAHPHKSFARITEKHLISNIING 184
S-Linalool_synthase_A.polygama     LLSGW-----ATNPHHQARLVRNALTHPYHKSLATFTARNFHYDCK-G 245
β-Ocimene_synthase_A.majus         ILHGN-----LASMEDNLRRSVGNKLRYPFHKSIARFTGINYDDDLG-G 251
. . . . . : . . : * :

```

**Figure 4.1** Multiple protein sequence alignment of *Coriandrum sativum* γ-terpinene synthase (*CsyTRPS*), (S)-linalool synthase (*CsLINS*), and two other putative *C. sativum* monoterpene synthases (*mTPS2* and *mTPS3*) against five known plant monoterpene synthases. Conserved motifs are underlined. Black blocking signifies identical residues, while grey signifies residue similarity. Find accession numbers in designated section. (continued on next page)

γ-Terpinene\_synthase\_C.limon GP---DMNPILLELAKVDYNIQAVHQEDLKYVSRWKKKTGLGEKLN FAR 316  
 Limonene\_synthase\_C.sativa KQ---DMNPITLLEFAKLDNFNMLQSTYQEEELKVLRSRWWKDKSLGEKLPFVR 336  
 CsyTRPS\_C.sativum SAPAEDYVSNFLYFAKLDYNNMQSIYQDDLKYLKSKWKKDTEWGEKLG FAR 323  
 CsLINS\_C.sativum SG---DMI PAVLEFAKLDNFNRQALNQEELKDLRSRWSRLDMGEKLP FAR 302  
 R-Linalool\_synthase\_L.angustif RP---DMNPVIFELAKLNFNIVQATQQEELKALSRWSSSLGLAEKLPFVR 283  
 mTPS2\_C.sativum EQ---NMNSLLKLAKLDFNSVQSVHQREVSKLASWLDLGL-DKMT FAR 306  
 mTPS3\_C.sativum KD---ISGKALQELAILDLAVMRTIHERESSVSRWNNELGLAQELKLV R 231  
 S-Linalool\_synthase\_A.polygama QN---GWNNLQELAKMDLTVVQSMHQEVLQVQSWKDRGLANELKLV R 292  
 β-Ocimene\_synthase\_A.majus MY---EWGKTLRELALMDLQVERSVMQEEELQVSKWNNELGLYKKT LAR 298  
 . : \* : : : : : \* : : : \*

MOTIF: DDXXD

γ-Terpinene\_synthase\_C.limon DRVVENFFWTVGDIFEPQ-FGYCRMSAMVNCLLTSIDDDYDVTGLDEL 365  
 Limonene\_synthase\_C.sativa DRLVECFWLQVGVRFEPQ-FSYFRIMDTKLYVLLTIIDDDHDIYGTLEEL 385  
 CsyTRPS\_C.sativum ERLMECFYVSVGYNSHPQ-FSYGRKVLAAITAFITTTDDIYDVYASLDEL 372  
 CsLINS\_C.sativum DRLVTSFFWSLGITGEPH-HRYCREVLTKIIEFVGVYDDYDVTGLDEL 351  
 R-Linalool\_synthase\_L.angustif DRLVESYFWAIPLEFEPHQ-YGYQRKVATKIITLITSIDDDYDIYGTLEEL 332  
 mTPS2\_C.sativum DRLVEHYFWCNGMVS DPE-YSAFRDMGTVKVICLITTTDDIYDIYGSLEEL 355  
 mTPS3\_C.sativum DQPLKWMYCTTALLTDPS-FSEERIELAKPISLIYIIDDIDLYGTINEL 280  
 S-Linalool\_synthase\_A.polygama NQPLKWMYMPMAALTDPR-FSEERVELTKPISFIYIIDDIDVTGLLEEL 341  
 β-Ocimene\_synthase\_A.majus NRPFEFYMWSMVILTDYINLSEQRVELTKSVAFIYLIIDDIIDVTGLDEL 348  
 . : . : : \* : : : \*\* : \* : : : \*

γ-Terpinene\_synthase\_C.limon ELFTDAVERWDATTEQLPYMKLCFHALYNSVNMGFIALRDQEVGMII 415  
 Limonene\_synthase\_C.sativa QLFTNALQRWDLKELDKLPDYMKTAIFYFTYNFTNELAFDVLQEHGF-VHI 434  
 CsyTRPS\_C.sativum EVLTKLTKRWDAEALDQLPDYVKVCFVTFYNDINEVANVAERKHGV-SIL 421  
 CsLINS\_C.sativum ELFTNVVKKRWDTNAMKELPDYMKLCFSLINMVNETTYDILKDHNI-DTL 400  
 R-Linalool\_synthase\_L.angustif QLFTNLIFERWDNASIGRLPEYLQLFYFAIHNHFVSEVAYDILKEKGF-TSI 381  
 mTPS2\_C.sativum ELFTDYIDRWIDITEIDKLPNMIKTVLLAMFNTNEIGYWTIRERDF-NII 404  
 mTPS3\_C.sativum TLFTEAVNRWDIAATKKLPDYMQKCFSSLHKITNEIGYKIYKYG L-NPI 329  
 S-Linalool\_synthase\_A.polygama TLFTDAVNRWELTAVEQLPDYMKVCFKALYDITNEIAYKIYKKGW-NPI 390  
 β-Ocimene\_synthase\_A.majus IIFTEAVNKWDYSATDTPDNMCMCYMTLLDTINGTSQKIYKYG H-NPI 397  
 \* . : : : \* . : : : \* . : : :

γ-Terpinene\_synthase\_C.limon PYLKKAWADQCKSYLVEAKWYNSGYIPTLQEYMENAWISVTAPVMLLHAY 465  
 Limonene\_synthase\_C.sativa EYFKKLMVELCKHHLQEAQWYFSGYKPTLQEVVENGWLSVGGQVILMHAY 484  
 CsyTRPS\_C.sativum PFFQKVVWTDLFDAYLVEAKWYHSGYKPSLSEYLDKAWISISGPVILTHSY 471  
 CsLINS\_C.sativum PHQRKWFNDLFEFERYIVEARWYNSGYQPTLEEYLNKGFVSI GGPVGLVSY 450  
 R-Linalool\_synthase\_L.angustif VYLQRSWVDLLKGYLKEAKWYNSGYTPSLEEYFDNAFMTIGAPPVLSQAY 431  
 mTPS2\_C.sativum PYLSKQWANLCKAYLTEAKWYHSGHKPALEEYLQNAVVSIAAPIMLFCA Y 454  
 mTPS3\_C.sativum DYLLKLSWAKLCNAFLSESKWFALEHLPKAEYLN TGI IISGVHVVLVHLF 379  
 S-Linalool\_synthase\_A.polygama DSLRRMWASLCNAFLVEAKWFAAGHLPKAEYLNKGI IISGMHVVTVHMF 440  
 β-Ocimene\_synthase\_A.majus DSLKTTWKSLSCAFLVEAKWSAGSLPSANEYLENEKVS SGVYVVLHLF 447  
 . \* : \* \* . \*\* . : :

MOTIF: (N, D) D (L, I, V) X (S, T) XXXE

γ-Terpinene\_synthase\_C.limon AFTANPITKEALEFLQD-SPDIIRISSMIVRLEDDLCTSSDELKRGDVPK 514  
 Limonene\_synthase\_C.sativa FAFTNPVTKEALECLKDGHPNIVRHASII LR LADDLGTSDELKRGDVPK 534  
 CsyTRPS\_C.sativum FVKANSMNHEDFQSLMT-YPNIRLSATILRLADDTATSSSEMERGDNPK 520  
 CsLINS\_C.sativum ICTEDPIKKEDELFIED-LPDIIVRLTCEIFRLTDDYGTSSAEKRGDVPS 499  
 R-Linalool\_synthase\_L.angustif FTLGSSMEKPIIESMYE-YDNILRVSGMLVRLDDDLGTSSEMERGDVPK 480  
 mTPS2\_C.sativum FLTAEKITVEALDYIDK-LPSIMWCFSMVLR LTN DDLGTSSEDELARGDNLK 503  
 mTPS3\_C.sativum FLIGDGSATERARLMNS-NASILSYPAAILRLWDDLGSARDENQKGDGS 428  
 S-Linalool\_synthase\_A.polygama FLLGGCFTEBSVNLVDE-HAGITSSIATILRLSDDLGSARDHQDQDGDGS 489  
 β-Ocimene\_synthase\_A.majus FLMGLGGTNRGSIELND-TRELMSSIAIIVRIWDDLGCARDEHQNGKDG S 496  
 : \* : \* : \* : \*

γ-Terpinene\_synthase\_C.limon SIQCYMHET-GVSEDEAREHIRDLIAETWMKMNSARFGNPP--YLPDVFI 561  
 Limonene\_synthase\_C.sativa SIQCYMHD T-GASEDEAREHIKYLISESWKEMNEDGNINS--FFSNEFV 581  
 CsyTRPS\_C.sativum AIQCYMNDS-GVSEEEAREHIKYLITETLKE LNEE--SARS--SFSK PFI 565  
 CsLINS\_C.sativum SIYCYMSDT-GVTEEVSRRKHMNLRKQWQIN KLRFSKEYNPNLSWSFV 548  
 R-Linalool\_synthase\_L.angustif SVQLYMKET-NATEEEAVEHVRFLNREAWKMMNTAEAGDS--PLVSDVV 527  
 mTPS2\_C.sativum AVQCYMNHT-GESEQVARNYVDNLVHETWKILNKDLLGSYF--FNEPFL 549  
 mTPS3\_C.sativum YLACYMKEHQEVSVETARKHVQNMISDTWKRLNKECLSPNP--YSKT-FI 475  
 S-Linalool\_synthase\_A.polygama YVEYLLKDKHGSSVENAREEVIRMISDAWKRLNEECLSPNP--FSAT-FR 536  
 β-Ocimene\_synthase\_A.majus YLDCYKHEINLTAQVHEHALELVAIEWKRLNKE SFNLNH--DSVSSFK 544  
 : \* . : : : : \*

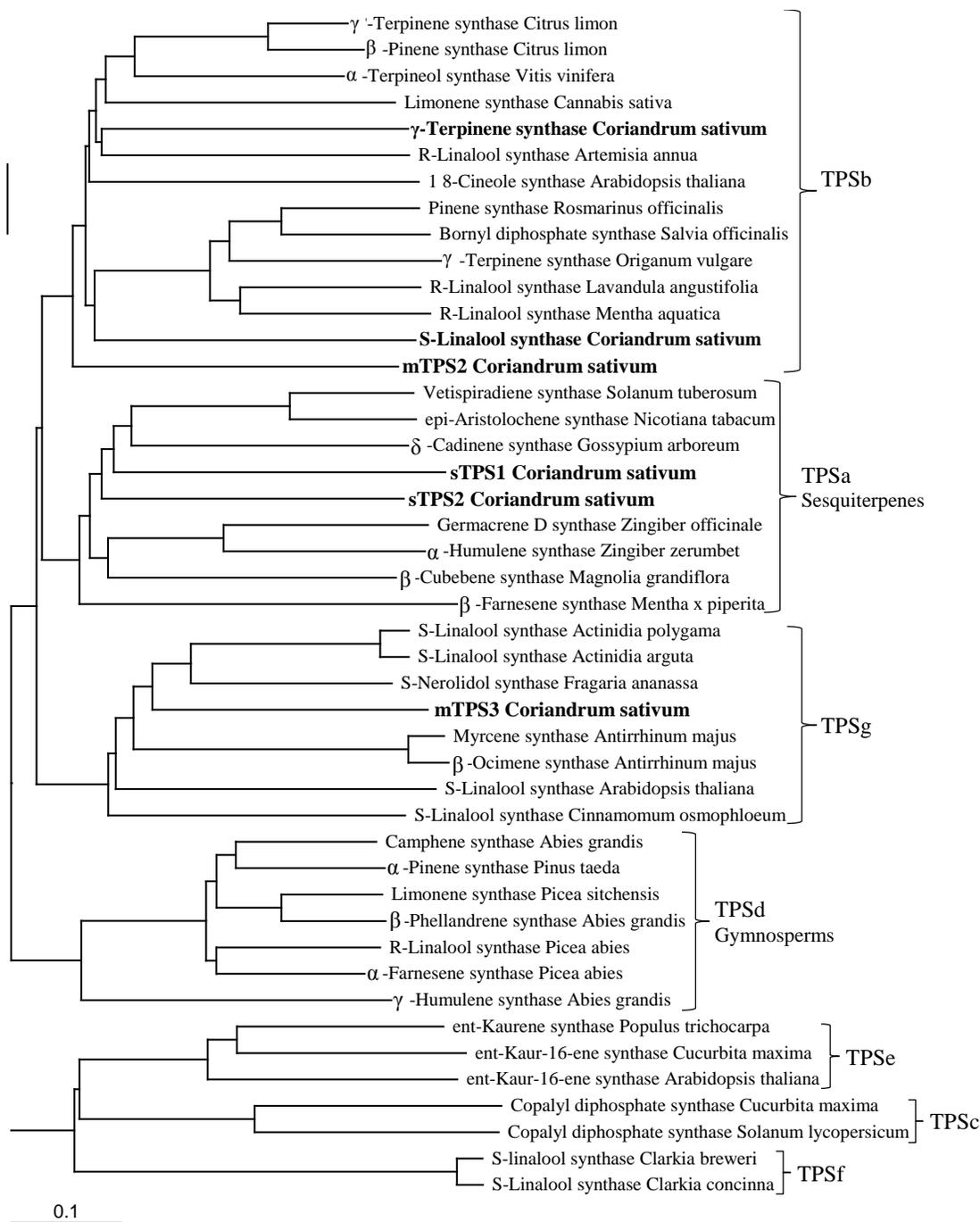
Figure 4.1 (continued on next page)

$\gamma$ -Terpinene_synthase_C.limon	GIAMNLVRMSQCMVLYGDG-HGVQEN-TKDRVLSLFIDPIP-----	600
Limonene_synthase_C.sativa	QVCQNLGRASQFIYQYGDG-HASQNNLSKERVGLIITPIPM----	622
CsyTRPS_C.sativum	DSCLNLARMSLNLYLYGDG-HGAPTLKDKERSTYLFVDPIPIEFSL	610
CsLINS_C.sativum	DIMLNIIRAAHFLYNTGDDGFGVEDVAVEATLVSLLEVEPIPL----	590
R-Linalool_synthase_L.angustif	AVAANLGRAAQFMYFDGDG----NQSSLQQWIVSMLFEPYA-----	564
mTPS2_C.sativum	TANPNLARTQTIFYQYGDG-HGIPQHWTKDHLTSLLEVEPFTL----	590
mTPS3_C.sativum	NGCLNLARMVPLMYSYDDN---QSLPLLEEYIKDMFVLPP-----	512
S-Linalool_synthase_A.polygama	KGCLNIARMVPLMYSYDDN---HNLPLLEEEMKAMLYDSSS-----	574
$\beta$ -Ocimene_synthase_A.majus	QAALNFARMVPLMYSYDNN---RRGFVLEEYVKFMLSDE-----	579
	: * . * . . . : :	

**Figure 4.1**

### 4.3.2 Phylogenetic Tree

Terpene synthases are organized phylogenetically according their amino acid relatedness, into seven subfamilies (TPSa through TPSg) (Bohlmann, 1998; Dudareva, 2003). All TPSs in plants share a common evolutionary origin and the first point of bifurcation separate TPSs involved in primary metabolism from those involved in specialized metabolism. Those TPSs involved with primary metabolism include copalyl diphosphate synthase (CDP)- and kaurene synthases (Bohlmann, 1998). Phylogenetic analysis revealed that three of coriander's mTPS candidates (*CsLINS*, *CsyTRPS* and *mTPS2*) are clustered with the TPSb subfamily, while the fourth mTPS candidate (*mTPS3*) was clustered with the TPSg subfamily. The two sTPS candidates were clustered with the TPSa subfamily (Figure 4.2).

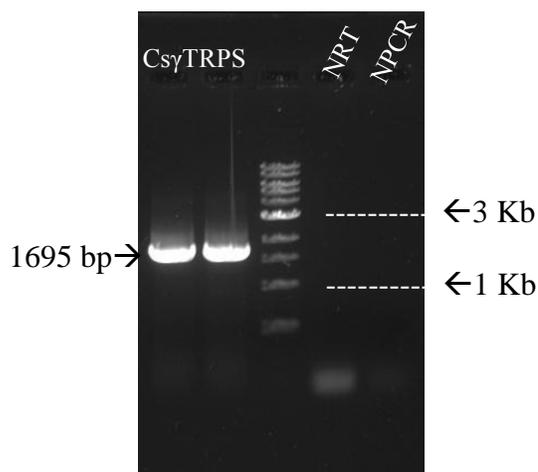


**Figure 4.2** Phylogenetic tree showing the evolutionary relationships between plant terpene synthases, including *Cs* $\gamma$ TRPS and *Cs*LINS. Multiple sequence alignments were performed using ClustalW2 and the phylogenetic tree was constructed in Treeview (Page, 1996) using the neighbour-joining method; the tree was rooted to the 'TPS c, e and f' outgroup. The seven terpene synthase classes, TPSa to g, are clearly shown. The scale denotes 0.1 amino acid substitutions per site.

### 4.3.3 Cloning of Full Length Monoterpene Synthases

#### 4.3.3.1 $\gamma$ -Terpinene Synthase

A 1695 bp product corresponding to the truncated *CsyTRPS* with the eight histidine tag was PCR amplified from coriander seed cDNA (Figure 4.3). Negative controls, NRT (no reverse transcriptase control) and NPCR (no template control) yielded no product. The amplicon was cloned into pET41b(+). Plasmid sequencing results confirmed that the full length *CsyTRPS* gene was cloned in frame into pET41b(+) vector as the nucleic acid alignment between the plasmid sequencing result and the *CsyTRPS* sequence from RNA-seq data were 100% homologous, excluding the absence of the signal peptide from the 5' end and the addition of the eight-histidine tag (CAC) to the 3' end of the *CsyTRPS* clone.



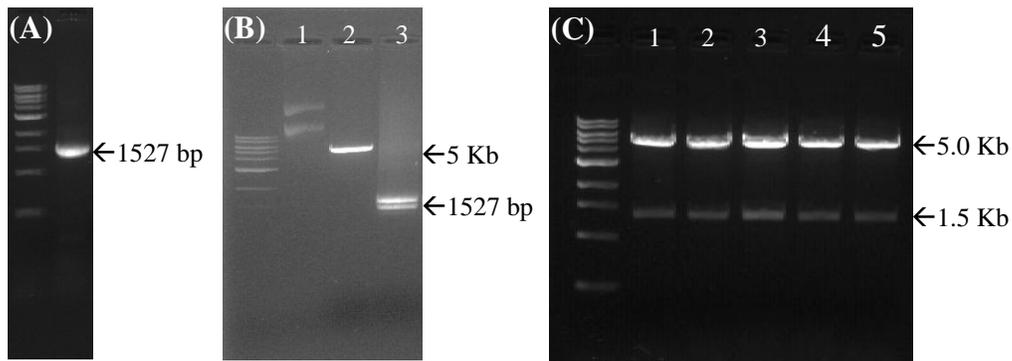
**Figure 4.3** PCR amplification of *CsyTRPS*.

#### 4.3.3.2 Other *mTPS* candidates

##### **First *mTPS3* clone (full length with signal peptide):**

PCR amplification of full length *mTPS3* yielded a single product of 1527 bp, which included sequences encoding a signal peptide on the 5' terminal (Figure 4.4 A). The purified amplicon product had a concentration of 175.1 ng/ $\mu$ l and A260/280 and A260/230 ratios of 1.94 and 1.55, respectively. Gel extraction after restriction enzyme digestion yielded pET41b(+) and *mTPS3* digested DNA extract concentrations of 18.3 and 42.6 ng/ $\mu$ l, respectively and A260/280 ratios of 1.85 and 1.92, respectively and

A260/230 ratios of 2.63 and 2.03, respectively. Double band digestion products appeared in mTPS3 after restriction enzyme digestion (Figure 4.4 B). Spread plates after transformation had many (>200) colonies present with no colonies on the negative plates. Concentrations and purity of plasmids extracted from DH10B *E. coli* after cloning are shown in Table 4.2. Diagnostic restriction enzyme digestion revealed two DNA products, one 5000 bp and the other ~1500 bp in size (Figure 4.4 C).



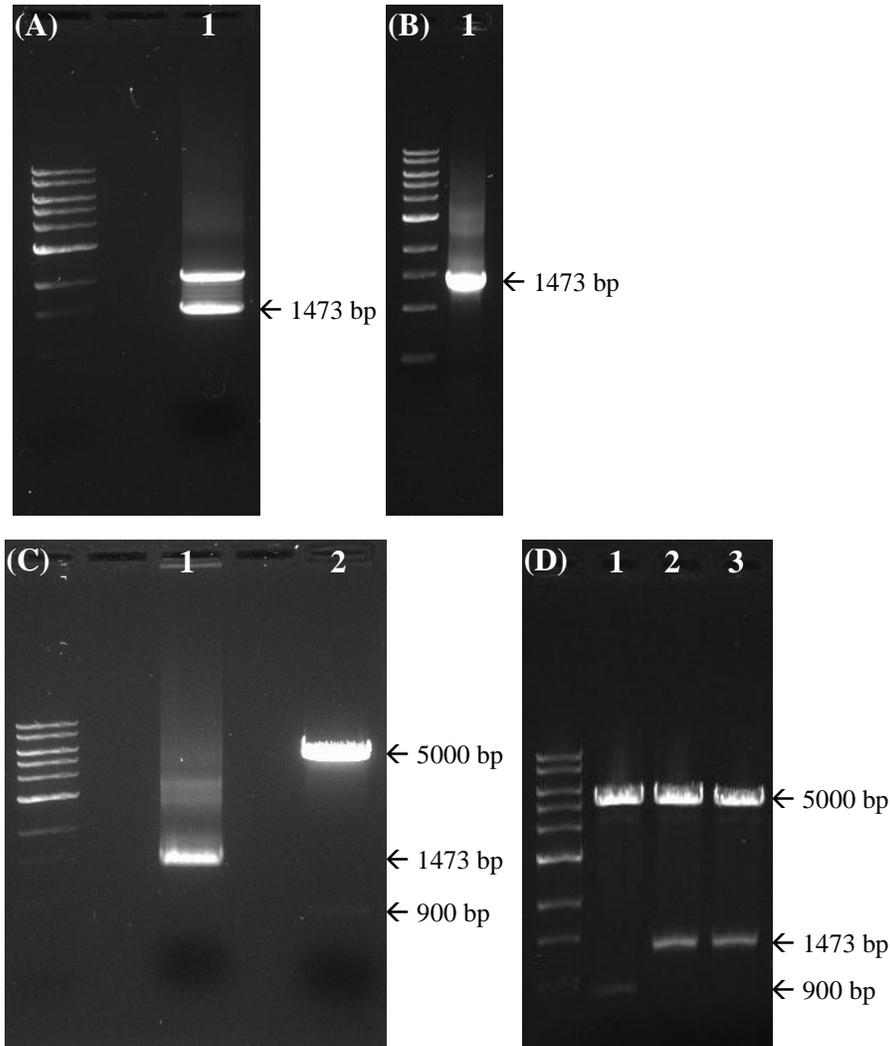
**Figure 4.4** (A) PCR amplification of full length *mTPS3*. (B) Gel extraction after restriction enzyme (NdeI and XhoI) digestion of full length *mTPS3* and pET41b(+) vector. Lane 1 - undigested pET41b(+) vector control; lane 2 - digested pET41b(+) vector; lane 3 - digested *mTPS3*. (C) Diagnostic restriction enzyme (NdeI and XhoI) digestion of full length *mTPS3\_pET41b(+)* plasmid extracted from DH10B *E. coli*. Each lane corresponds to a different colony.

**Table 4.2** Nanodrop spectrophotometer results for plasmids containing *mTPS3* with signal peptide extracted from DH10B *E. coli* cells.

Colony	Concentration (ng/μl)	A260/A280	A260/A230
1	111.3	1.82	1.17
2	69.9	1.89	1.89
3	108.0	1.93	1.95
4	70.7	1.92	1.76
5	64.4	1.94	1.69

**Second *mTPS3* clone (truncated, signal peptide removed):**

PCR amplification of truncated putative linalool synthase, *mTPS3* yielded a single product of expected size, 1473 bp after PCR amplification using gel extracted product (Figure 4.5).



**Figure 4.5** (A) PCR amplification of truncated *mTPS3* with iProof DNA polymerase. (B) PCR amplification of truncated *mTPS3* with iProof DNA polymerase from gel extracted template. (C) Gel extraction of *mTPS3* and pET41b(+) vector after restriction enzyme (NdeI and HindIII) digestion. (D) Diagnostic restriction enzyme (NdeI and HindIII) digestion of *mTPS3\_pET41b(+)* plasmid extracted from DH10B *E. coli* cloning cells.

Nanodrop results for gel extracted gene of interest were as follows, 19.9 ng/μl, A260/280 ratio of 1.71 and an A260/230 ratio of 1.75. Purification of final truncated *mTPS3* yielded 249.9 ng/μl of pure product with an A260/280 ratio of 1.91 and an A260/230 ratio of 1.94. Gel extraction of digestion products yielded the following, 26.7 ng/μl and 35.5 ng/μl of *mTPS3* and pET41b(+), respectively as well as A260/280 ratios of 1.95 and 1.97, and A260/230 ratios of 2.26 and 1.64, respectively. Concentrations for *mTPS3\_pET41b(+)* plasmids extracted from DH10B after cloning are shown in Table 4.3.

**Table 4.3** Nanodrop spectrophotometer results for plasmids containing truncated *mTPS3* extracted from DH10B *E. coli* cells.

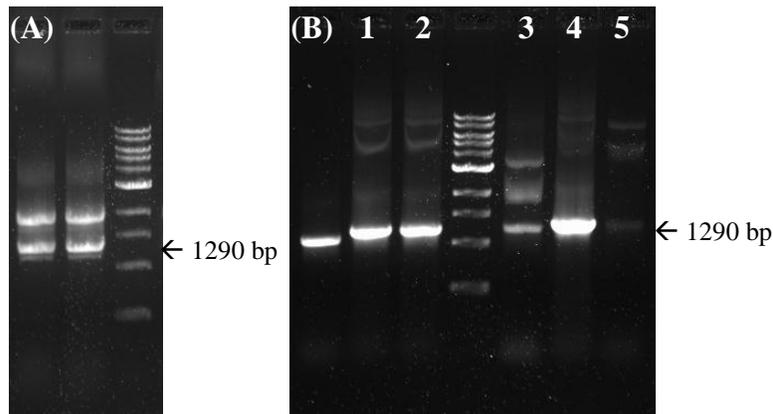
Colony	Concentration (ng/μl)	A260/A280	A260/A230
1	58.2	1.89	1.96
2	70.0	1.93	2.0
3	59.0	1.87	1.61

**Third mTPS3 clone (with pGEX-4T-1 system):**

Sticky-end PCR amplification of truncated mTPS3 yielded two amplicons of 1290 bp in size, with slightly differing overhang regions at each of the termini. A non-specific product of approximately 1800 bp also resulted (Figure 4.6 A). After denaturation and re-annealing of sticky-end PCR products the resulting concentration was 71.4 ng/μl with an A260/280 ratio of 1.89 and an A260/230 ratio of 1.80. The transformation yielded excellent colony numbers (>500) and plasmids extracted from six of those randomly selected colonies yielded the concentrations shown in Table 4.4. Diagnostic PCR amplifications from these extracted plasmids can be visualized by agarose gel in Figure 4.6 B.

**Table 4.4** Nanodrop spectrophotometer results for *mTPS3\_pGEX-4T-1* plasmids extracted from DH10B *E. coli* cells.

Colony	Concentration (ng/μl)	A260/A280	A260/A230
1	9.9	2.35	1.68
2	144.4	1.92	2.23
3	138.8	1.91	2.23
4	404.8	1.89	2.26
5	138.3	1.92	2.25
6	115.4	1.90	2.08



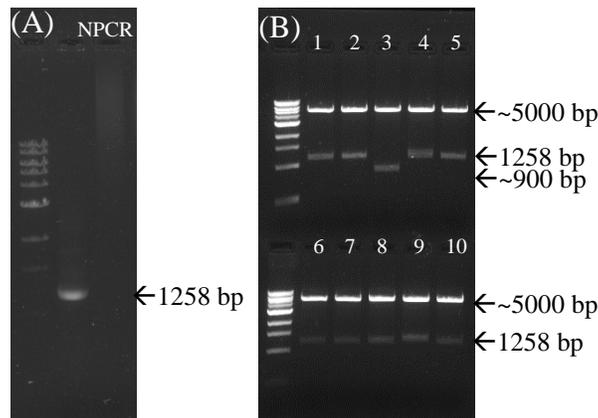
**Figure 4.6** (A) Sticky-end PCR amplification of truncated *mTPS3* using Kapa DNA polymerase. (B) Diagnostic PCR amplification of *mTPS3\_pGEX-4T-1* plasmids extracted from DH10B. Each lane represents a different colony from which plasmid was extracted.

### The *mTPS2* clone:

Amplification with iProof HiFi DNA polymerase yielded a 1258 bp specific product (Figure 4.7 A). Colony numbers were excellent after transformation via electroporation (>500 colonies). Plasmid extractions of ten isolated DH10B colonies were of good purities but had lower concentrations than previously observed for other clones (e.g., third *mTPS3* clone) (Table 4.5). The diagnostic restriction enzyme digestion of plasmids extracted from DH10B yielded nine colonies which had cloned gene of approximately 1200 bp (expected gene of interest size) and one fragment corresponded to pET41b(+) vector release (900 bp) (Figure 4.7 B). The approximately 5000 base pair fragment corresponds to the digested pET vector.

**Table 4.5** Nanodrop spectrophotometer results for *mTPS2*\_pET41b(+) plasmids extracted from DH10B *E. coli* cells.

Colony	Concentration (ng/μl)	A260/A280	A260/A230
1	52.5	1.89	1.92
2	63.6	1.82	1.94
3	74.6	1.81	1.41
4	68.5	1.85	1.99
5	62.7	1.89	1.84
6	59.8	1.86	1.72
7	53.2	1.90	1.88
8	73.4	1.84	1.71
9	63.8	1.86	1.69
10	53.2	1.81	1.86



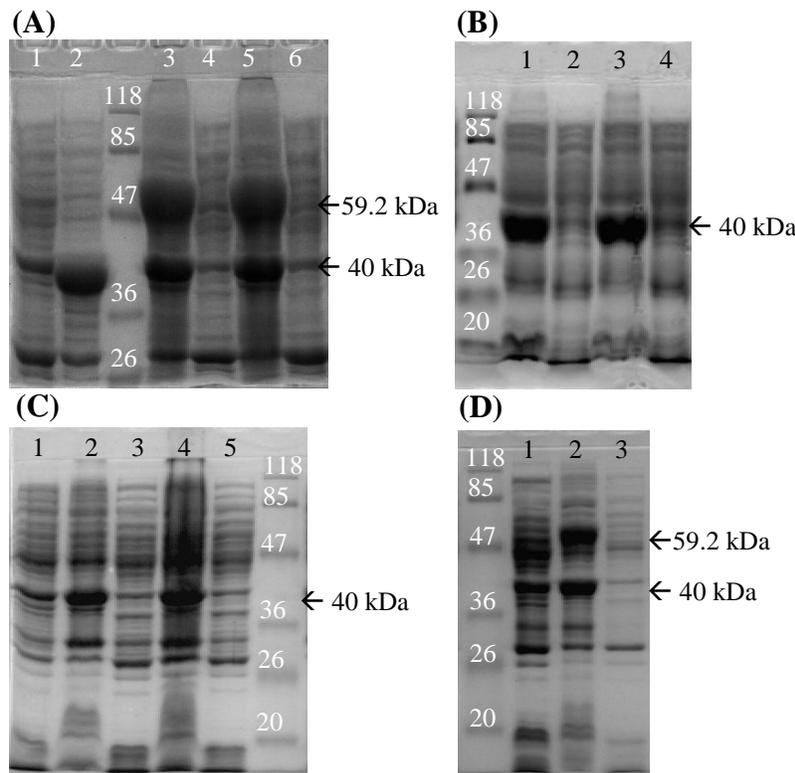
**Figure 4.7** (A) PCR amplification of *mTPS2* using iProof HiFi DNA polymerase and (B) diagnostic restriction enzyme digestion of *mTPS2* in pET41b(+) plasmids extracted from DH10B *E. coli* colonies.

#### 4.3.3.2.1 Protein expression attempts of putative (*S*)-linalool synthase, *mTPS3*

##### First *mTPS3* expression attempt (full length clone):

Transformation of full length *mTPS3*\_pET41b(+) construct into Rosetta *E. coli* yielded many colonies (>500) on test plates and no colonies on negative control plates. Strong insoluble protein expression was seen upon induction at 37 °C but not soluble protein; no protein expression was observed upon induction at 18 °C (Figure 4.8 A, B). Later, soluble protein induction was attempted at 23 and 30 °C with all other conditions

as described for 18 and 37°C inductions. No soluble protein of interest was expressed at 30 °C and no protein of interest expression occurred at 23 °C, as was the case with 37 and 18 °C, respectively (Figure 4.8 C, D).



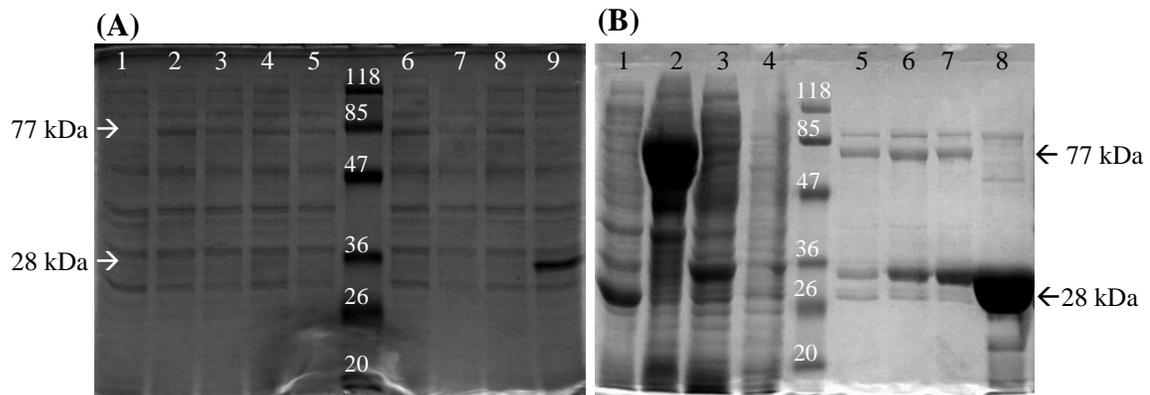
**Figure 4.8** (A) Protein expression of mTPS3 at 37 °C. Lane 1 – non-induced control; lane 2 – induced pET41b(+) control; lanes 3 and 5 – insoluble fractions of colonies ‘I’ and ‘II’, respectively; lanes 4 and 6 – soluble fractions of colonies A and B, respectively. (B) Protein expression of mTPS3 at 18 °C. Lanes 1 and 3 – insoluble fractions of colonies A and B, respectively; lanes 2 and 4 – soluble fractions of colonies ‘I’ and ‘II’, respectively. (C) Protein expression of mTPS3 at 23 °C. Lane 1 – non-induced control; lanes 2 and 4 – insoluble fractions of colonies A and B, respectively; lanes 3 and 5 – soluble fractions of colonies A and B, respectively. (D) Protein expression of mTPS3 at 30 °C. Lane 1 – non-induced control; lane 2 – insoluble fraction; lane 3 – soluble fraction. Molecular weight marker ladder represents protein fragments of known size in kDa for use as reference.

### Second mTPS3 expression attempt (truncated clone):

No protein expression of interest was seen with the truncated *mTPS3* gene in pET41b(+) vector.

### Third mTPS3 expression attempt (truncated clone in pGEX-4T-1):

Protein of expected size, 77 kDa, was expressed in the test colonies but not in the non-induced controls or controls harboring the empty vector (Figure 4.9 A). The protein purified from the pGEX-4T-1 vector had an  $A_{595}$  of 1.093 which corresponded to  $\sim 3.8$   $\mu\text{g}/\mu\text{l}$  of protein while the mTPS3 purified protein had an  $A_{595}$  of 0.152 which corresponded to  $\sim 0.35$   $\mu\text{g}/\mu\text{l}$  of protein. Non-specific products co-eluted with protein of expected size (77 kDa) (Figure 4.9 B).

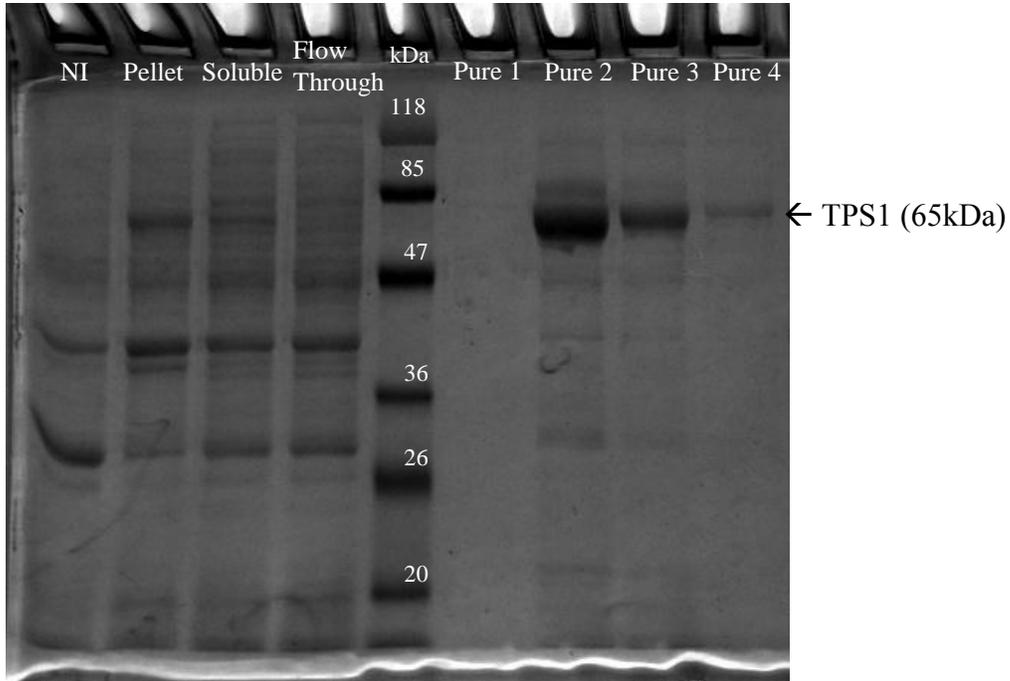


**Figure 4.9** SDS-PAGE images of protein extracted from Rosetta *E. coli* cells containing mTPS3 protein of interest. (A) Protein from 7 single colonies after 37 °C three hour induction. Lane 1 - non-induced control; lane 2-8 - colonies 1-7; lane 9 - pGEX-4T-1 control. (B) Protein affinity purification of mTPS3 and protein from pGEX-4T-1 vector. Lane 1 - non-induced control; lane 2 - insoluble fraction; lane 3 - soluble fraction; lane 4 - flow through; lane 5-7 - elutions 1-3; lane 8 - protein purified from pGEX-4T-1.

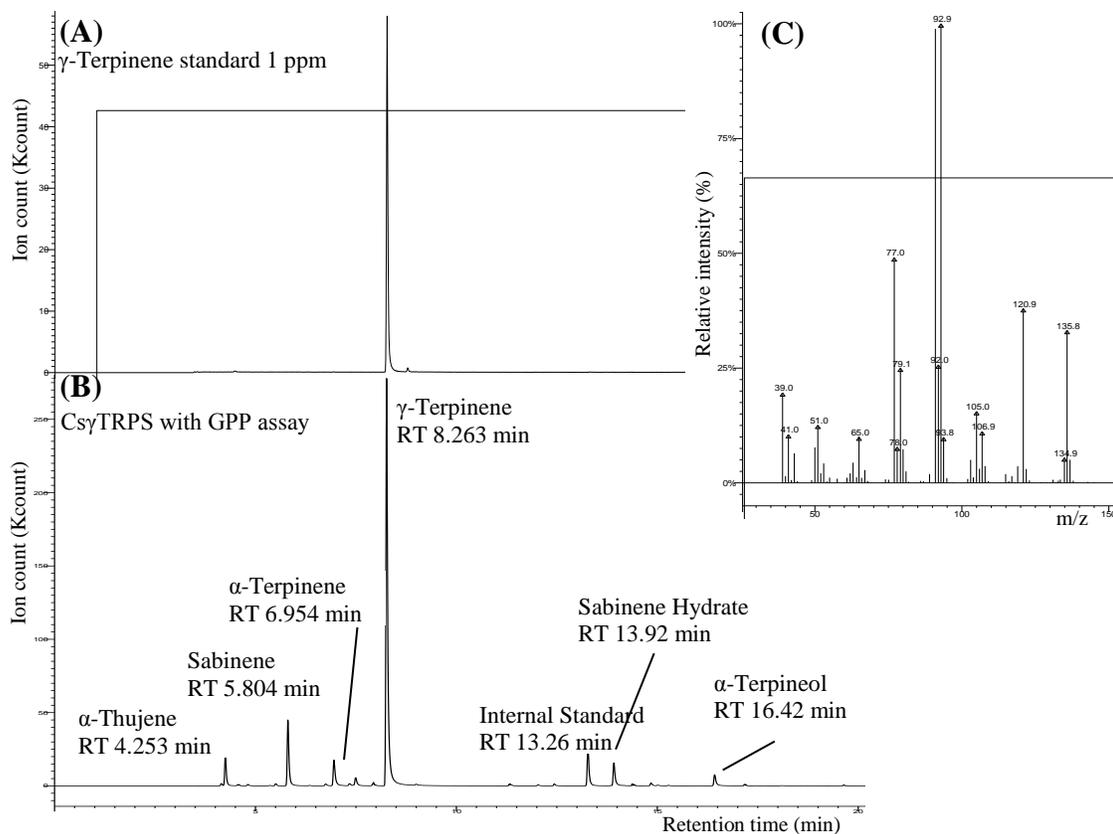
### 4.3.4 Enzyme Kinetics Data for $\gamma$ -Terpinene Synthase

The complete ORF of *CsyTRPS* was 1,833 bp, of which 186 bp corresponded to the putative signal peptide, which was removed. The truncated gene tagged with eight C-terminal histidine residues encoded a 558 amino acid protein with a predicted mass of 65.16 kDa (Figure 4.10). Incubation of the bacterially produced CsyTRPS with GPP yielded  $\gamma$ -terpinene as major product (91.1%) in addition to a number of minor products, including sabinene (6.97%),  $\alpha$ -terpinene (1.18%), terpinene-4-ol (0.533%), and  $\alpha$ -terpineol (0.246%). Incubation with NPP (data not shown) also yielded  $\gamma$ -terpinene (albeit at a lower amount of 75.7%), and a few minor products including sabinene

(11.4%),  $\alpha$ -terpinene (5.17%),  $\alpha$ -thujene (4.79%),  $\alpha$ -terpineol (2.48%) and terpinene-4-ol (0.489%) (Figure 4.11).



**Figure 4.10** SDS-PAGE of Cs $\gamma$ TRPS protein extracted from Rosetta (DE3) pLysS bacterial expression cells and purified by Ni-NTA affinity chromatography. NI - non-induced control; Pellet - insoluble fraction of extracted protein; Soluble - soluble fraction; FlowThrough - protein lysate flow-through off nickel column; Pure1-4 - first through fourth protein elution

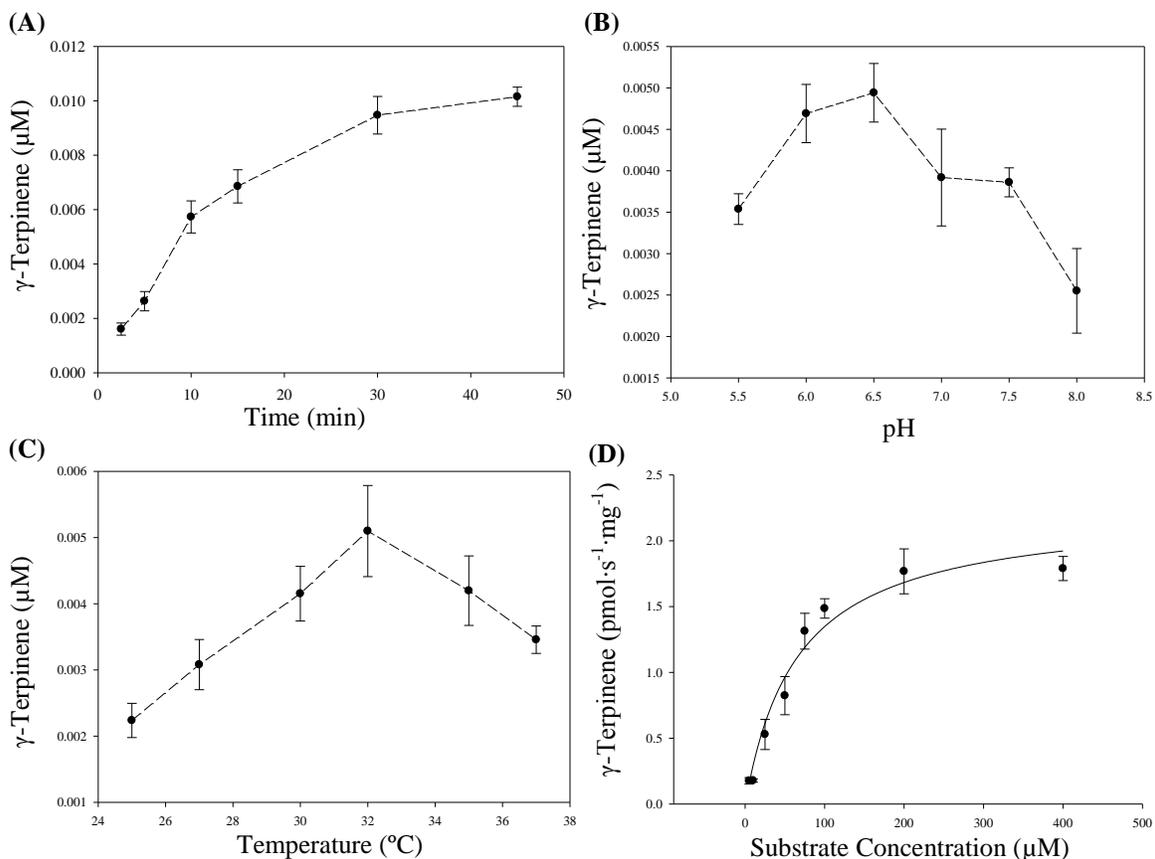


**Figure 4.11** Gas chromatograms (GC) and mass spectrometry (MS) fingerprint for volatile terpene products of Cs $\gamma$ TRPS. (A) GC for 2  $\mu$ l of 1 ppm  $\gamma$ -terpinene standard. (B) GC for 2  $\mu$ l helium-concentrated volatile products of Cs $\gamma$ TRPS with GPP assay, showing major as well as minor products. (C) MS fingerprint of major product,  $\gamma$ -terpinene.

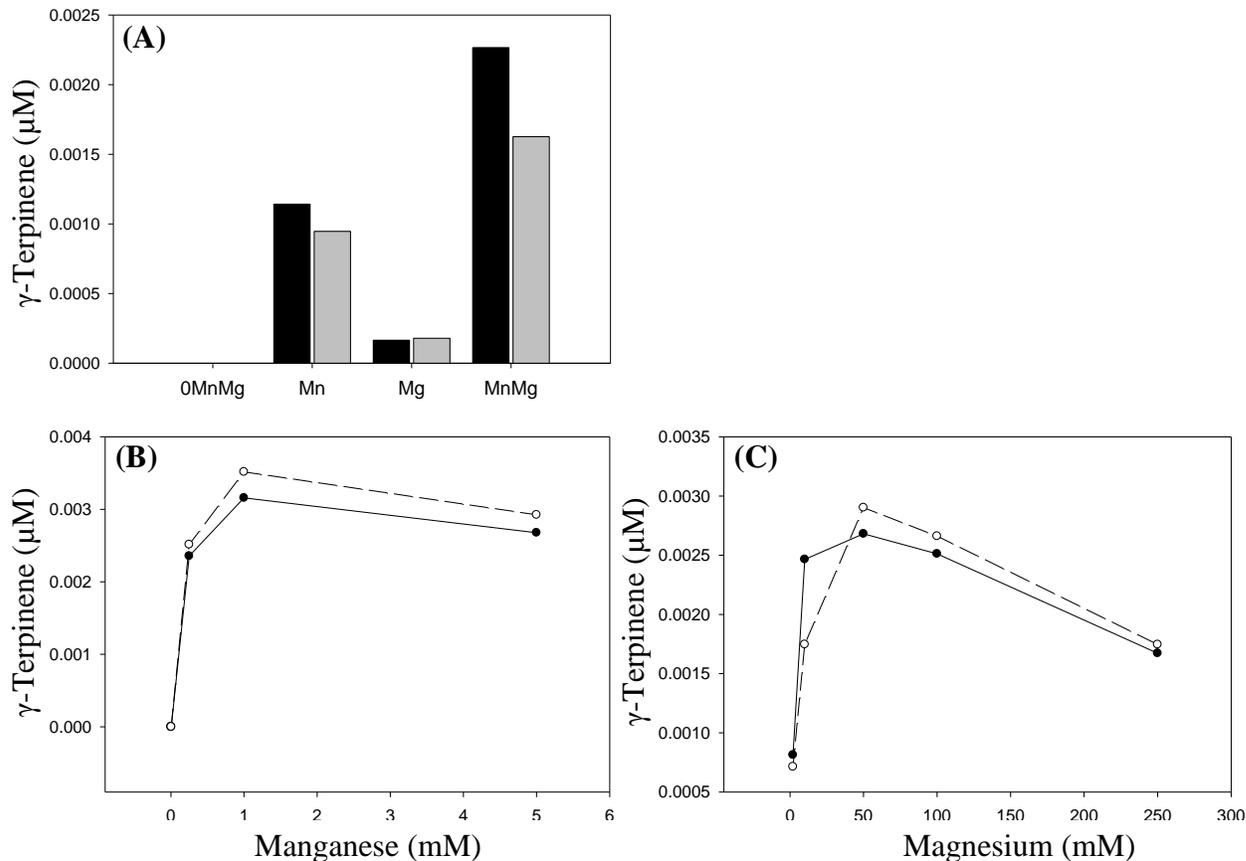
The linear rate kinetics of Cs $\gamma$ TRPS activity ranged from 2.5 to 15 minutes (Figure 4.12 A), thus 10 minutes was chosen as the length of time for all assay incubations. The optimum pH ranged from 6.0 – 6.5 (Figure 4.12 B), while the optimum temperature was 32  $^{\circ}$ C (Figure 4.12 C).

The Michaelis-Menten enzyme saturation curve was prepared using the hyperbolic enzyme analysis module in the SigmaPlot software (v.10.0) (Systat Software, Erkrath, Germany) (Figure 4.12 D). The  $K_m$ ,  $V_{max}$  and catalytic efficiency for Cs $\gamma$ TRPS were calculated to be  $66 \pm 13 \mu\text{M}$ ,  $2.2 \pm 0.2 \text{ p}_{\text{kat}}/\text{mg}$  and  $2.228 \times 10^{-6} \text{ s}^{-1} \mu\text{M}^{-1}$ , respectively. No enzymatic activity was detected upon incubation with FPP and the same products were observed upon incubation with neryl diphosphate (NPP), an isomer of GPP, although at

reduced amounts. The optimum manganese and magnesium concentrations were 1 mM and 50 mM, respectively (Figure 4.13).



**Figure 4.12** Data for *CsyTRPS* with GPP substrate. Each enzymatic assay consisted of two biological and two technical replicates. **(A)** Time course assay, **(B)** pH optimization, **(C)** temperature optimization and **(D)** enzyme velocity at increasing substrate (GPP) concentrations.



**Figure 4.13** Divalent metal cation cofactor preference and optimization of Cs $\gamma$ TRPS. Cofactor assays were performed with a single biological replicate and two technical replicates, therefore all points are shown on plots. **(A)** Manganese and magnesium cofactor preference (‘0MnMg’ indicated assay was run with neither magnesium nor manganese; ‘MnMg’ indicates both magnesium and manganese are in assay). **(B)** Manganese cofactor concentration optimization and **(C)** magnesium cofactor concentration optimization.

## 4.4 Discussion

### 4.4.1 Terpene Synthase Sequence Homology and Phylogenetic Analysis

Four mTPS and two sTPS candidate genes were selected from the coriander transcriptome library based on the presence of some sort of “terpene synthase” or “terpene biosynthetic process” gene ontology annotation as well as some KEGG annotation placing those candidates into the terpene biosynthetic processes category. Additionally these six candidates were genes which had the greatest homology with known mTPS and sTPS genes according to the BLASTx alignment results (Table 4.1).

The four motifs used to help select the mTPS candidates were DDXXD, (N,D)D(L,I,V)X(S,T)XXXE and RRX<sub>8</sub>W, as well as one partially conserved motif, LQLYEASFLL. The first two aspartate rich motifs play roles in the coordination of the divalent cation cofactors (manganese and magnesium). These motifs together with the cofactors are important for binding the GPP substrate, which is oriented so that the negatively charged diphosphate moiety interacts with the positively charged metal ion cofactors. This allows the hydrophobic moiety to be positioned into the hydrophobic catalytic pocket of a monoterpene synthase (Degenhardt, 2009). The twin arginine motif plays a role in GPP substrate cyclization by isomerizing the substrate into a cyclizable intermediate (Williams, 1998). The LQLYEASFLL motif is only partially conserved between mTPSs and is thought to be a part of the catalytic pocket (Wise, 1998).

All four motifs were present in three of the four mTPS candidates (*mTPS1/CsLINS*, *mTPS2* and *CsyTRPS*). The absence of this motif from *mTPS3* is reasonable since that was a putative linalool synthase gene which would encode a terpene synthase that ultimately gives rise to linear monoterpene products, thus a cyclization motif would be unnecessary. Initially, before *mTPS1* was found by Lukman Sarker, *mTPS3* was chosen as the best (S)-linalool synthase candidate because it had the greatest sequence homology (55% identity and 71% positives) with known (S)-linalool synthases (from *Actinidia polygama* and *Actinidia arguta*) as well as the fact that it was missing the twin arginine motif at the N-terminal of its protein sequence, as was the case for those two *Actinidia* (S)-linalool synthases (Chen, 2010). Additionally *mTPS3* had several GO annotations such as “(S)-linalool synthase activity”. Interestingly, *mTPS1* (or *CsLINS*) did not have any annotations or BLASTx hits indicating any linalool synthase activity. Perhaps there are two genes responsible for encoding terpene synthases with linalool synthase activity in coriander. Protein expression efforts for *mTPS3* however, failed to yield any active protein when using GPP, FPP or NPP as substrate. It was thought that this may be occurring if the transcript sequence for *mTPS3* from the transcript library is not full length, however, upon analyzing the conserved motifs between *mTPS3* and known linalool synthases, those necessary for catalytic activity are present in *mTPS3*. It is also unlikely that the choice of bacterial host is affecting the protein activity. Sometimes

prokaryotic hosts can alter the folding of the recombinant protein and thus its catalytic pocket is modified or the recombinant protein is stored as insoluble inclusion bodies (Gasser, 2008). In future perhaps *mTPS3* could be expressed in a eukaryotic host such as yeast or fungi. It should be noted however, that known mTPSs have been expressed in bacterial cells with no problems to date (Chen, 2010; Dudareva, 2003; Landmann, 2007).

All the members within a TPS subfamily share at least 40% protein sequence identity (Chen, 2010). The mTPS3 gene clusters with TPSg subfamily, identified by Dudareva in 2003. This subfamily is characterized by the absence of the RRX<sub>8</sub>W motif, necessary for cyclization of GPP substrate (Dudareva, 2003). As mentioned before, the mTPS3 protein sequence is missing the RRX<sub>8</sub>W motif. Terpene synthases clustered in this subfamily give rise to linear monoterpene products, for example, AtTPS14 from *Arabidopsis thaliana* or Ama1e20 from *Antirrhinum majus* which both yield the linear (S)-linalool from GPP substrate and are clustered into TPSg (Chen, 2003; Nagegowda, 2008). This finding also contributed to the thought that mTPS3 would encode a terpene synthase with (S)-linalool synthase activity.

CsγTRPS, CsLINS and mTPS2 are all clustered with the TPSb subfamily. The presence of the RRX<sub>8</sub>W motif is a characteristic element of the TPSb and TPSd subfamilies (Dudareva, 2003). All three coriander genes clustered with TPSb have the RRX<sub>8</sub>W motif in their N-terminal regions. The TPSb subfamily contains angiosperm mTPS genes which are distinct from those angiosperm sTPS genes in the TPSa subfamily (Bohlmann, 1998). Coriander's sTPS1 and sTPS2 genes are clustered in the TPSa subfamily. According to the KEGG and GO annotations each of these candidates were classified as having putative sesquiterpene synthase activity, germacrene D synthase and β-caryophyllene/α-humulene synthase activities.

#### **4.4.2 Truncated vs. full length mTPS3**

The N-terminal signal peptide of TPSs, which is necessary for the pseudo-mature TPS to be transported to the plastid, where it becomes mature mTPS, has been found to render bacterially expressed TPSs insoluble, thus inactive. Therefore the signal peptides are generally eliminated from mTPS gene sequences during cloning work (Vonheijne and

Steppuhn, 1989). When *mTPS3* was searched for the presence of a signal peptide using SignalP 4.0 publicly available software, no such signal peptide was identified. This finding together with previous reports that linalool synthase genes have very short signal peptides (Landmann, 2007) pointed to the possibility of *mTPS3* being a putative linalool synthase gene which did not possess a complete signal peptide. Thus, the signal peptide's inclusion in the first primer set seemed inconsequential; at least until soluble protein expression proved difficult at downstream stages. Those initial attempts to express soluble protein from full length *mTPS3* (with signal peptide included) in pET41b(+) yielded little to no soluble protein when induction was carried out at 37 °C, even when the induction temperature was reduced to 18 °C, no expression of protein of interest was observed (Figure 4.8 A, B). Later, induction of soluble protein expression was attempted at 23 and 30 °C with no improvement. Induction at 23 °C yielded the same result (no protein of interest expression) as induction at 18 °C, and induction at 30 °C yielded the same result (only expression in the pellet or insoluble fraction) as induction at 37 °C (Figure 4.8 C, D). It became increasingly apparent that a signal peptide, although shorter than most signal peptides found in known mTPSs, was indeed present in the putative linalool synthase gene and was preventing soluble protein expression of mTPS3.

During protein expression of truncated *mTPS3* (no signal peptide) in the pET41b(+) vector, inconsistent expression of soluble protein of interest led to attempts with the pGEX-4T-1 bacterial expression vector since it has been reported in literature that placing an N-terminal tag on recombinant proteins improves their solubility (Harper, 2011; Madan, 2008). Additionally, fusing a protein to a soluble protein tag has been shown to improve recombinant protein soluble expression (Novagen, 2003). The pGEX-4T-1 expression system places an N-terminal GST-tag on the recombinant protein while the pET41b(+) expression vector can be used to easily place a C-terminal histidine-tag on recombinant protein.

#### **4.4.3 Functional Characterization of $\gamma$ -Terpinene Synthase**

The complete ORF of *Cs $\gamma$ TRPS* was 1,833 bp, of which 186 bp corresponding to the putative signal peptide, which was removed to improve protein solubility during

expression. The ORF of *CsyTRPS*, excluding the transit peptide, was expressed in bacterial cells and the recombinant protein purified and assayed for activity with typical monoterpene substrates GPP and neryl diphosphate (NPP), an isomer of GPP. Recombinant *CsyTRPS* yielded the cyclic monoterpene  $\gamma$ -terpinene as major product (91.1%) when incubated with GPP and NPP substrates. This  $\gamma$ -terpiene product makes up between 0.24-0.50 %, 15%, or 2.15% of the coriander EO according to Msaada , 2009b, Bhuiyan, 2009 and Sriti, 2009, respectively.  $\gamma$ -Terpiene was made up 2.66% (Table 2.1) of the EO in the coriander seeds used for this project, placing this monoterpene among the five most abundant monoterpenes in this EO.

The minor products produced from *CsyTRPS* incubation with GPP, sabinene (6.97%),  $\alpha$ -terpinene (1.18%), terpinene-4-ol (0.533%), and  $\alpha$ -terpineol (0.246%), differ from those side products reported by Suzuki, 2004,  $\alpha$ -pinene and limonene. Although the same assay buffer (TRIS-Cl) was used in both experiments, the pH in Suzuki's study was 7.8 while in this study it was 6.8. The ratio of magnesium to manganese cofactors was the same between studies, but double the amount of each cation was used by Suzuki. Additionally, Suzuki used a substantially greater amount of GPP substrate, 767  $\mu$ M, as opposed to the 25  $\mu$ M used in this experiment. While assays in this study were supplemented with BSA, no such stabilizing protein was added to assays by Suzuki. Also the assay lengths differed between studies, 15 minutes in this experiment versus 2-4 hours by Suzuki. These differing assay conditions may be contributing to the difference in minor products yielded by incubation of  $\gamma$ -terpinene synthase with GPP. However, these two enzymes are from plants belonging to completely different species, *Coriandrum sativum* versus *Citrus unshiu*, which may also contribute to the difference in minor products, since the protein sequence differences between these enzymes (47.18% identity) results in slightly different catalytic sites at the C-terminals. It has been suggested in previous studies that the minor side products often observed in TPS proteins expressed in bacterial hosts may be due to some proteolytic action by the host on the enzyme which can result in altered substrate and/or intermediate binding (Wise, 1998). Thus, the fact that *C. unshiu*  $\gamma$ -terpinene synthase was expressed in BL21 (DE3) *E. coli*

cells versus C $\gamma$ TRPS, which was expressed in Rosetta (DE3) *E. coli* cells may also be contributing to the different minor products of these bacterially expressed enzymes.

The linear kinetics of C $\gamma$ TRPS ranged from 2.5 to 15 minutes (Figure 4.12 A), thus 10 minutes was chosen as the length of time for all assay incubations. The optimum pH ranged from 6.0 – 6.5 (Figure 4.12 B), while the optimum temperature was 32 °C (Figure 4.12 C). The pH optimum here is similar to that found for  $\gamma$ -terpinene synthase from *Citrus limon*, pH 7, by Lucker, 2002, using the same MOPSO buffer.

The Michaelis-Menten enzyme saturation curve which was prepared using the hyperbolic enzyme analysis module in the SigmaPlot software exhibited a plateau towards the higher substrate concentrations which indicates this C $\gamma$ TRPS was successfully saturated during enzymatic characterization (Figure 4.12 D). The  $K_m$ ,  $V_{max}$  and catalytic efficiency for C $\gamma$ TRPS were calculated to be  $66 \pm 13 \mu\text{M}$ ,  $2.2 \pm 0.2 \text{ p}_{\text{kat}}/\text{mg}$  and  $2.228 \times 10^{-6} \text{ s}^{-1} \mu\text{M}^{-1}$ , respectively. From these values it can be seen that C $\gamma$ TRPS has a low affinity for its substrate, GPP, and is saturated at a low substrate concentration. Turnover constant for C $\gamma$ TRPS was  $k_{cat} = 1.476 \times 10^{-4} \text{ s}^{-1}$ . No enzymatic activity was detected upon incubation with FPP. The  $K_m$  calculated by Lucker, 2002, was  $2.7 \mu\text{M}$ . This indicates that  $\gamma$ -terpinene synthase from *C. limon* has a greater affinity for its GPP substrate than C $\gamma$ TRPS.

Upon testing the divalent metal ion cofactor preference of C $\gamma$ TRPS, results demonstrated a preference for manganese over magnesium as was the case with  $\gamma$ -terpinene synthase in *C. limon* (Lucker, 2002); however, optimal enzymatic performance occurred when both cofactors were present (Figure 4.13 A). When each cofactor was tested individually with C $\gamma$ TRPS it was found that optimum manganese and magnesium concentrations were 1 mM and 50 mM, respectively (Figure 4.13 B, C). The manganese concentration optimum here is only 0.4 mM greater than the optimum concentration found for  $\gamma$ -terpinene synthase in *C. limon* (Lucker, 2002).

In conclusion, the transcript library constructed for coriander was used to obtain the complete transcript sequence of C $\gamma$ TRPS and a putative linalool synthase (*mTPS3*). C $\gamma$ TRPS was cloned into pET41b(+) bacterial expression system and recombinant protein expressed and purified via Ni-NTA affinity chromatography. Enzymatic assays

were performed using purified Cs $\gamma$ TRPS protein and GPP substrate; the assays products ( $\gamma$ -terpinene and some minor monoterpene products) were quantified by GC-MS and these quantities were used to calculate Cs $\gamma$ TRPS's kinetic properties. The same was attempted for *mTPS3* (cloning into pET41b(+), then pGEX-4T-1), always unable to obtain consistent protein expression in the soluble fraction of protein extract. In future, perhaps efforts could be made to re-solubilize protein of interest (mTPS3) from inclusion bodies.

## CHAPTER 5: CONCLUSIONS

### 5.1 Summary of Research

In this MSc project, the production of essential oil in *C. sativum* plant seeds was investigated. The anatomy of *C. sativum* seeds was studied using scanning electron microscopy and stomata were found on the seed surface as well as four vittae (secretory canals) in the seed cross-section. Compositional analysis on the coriander EO extracted by steam distillation was performed via GC-MS. Seventeen monoterpenes were identified with linalool making up the vast majority of the EO terpene content (78%). Previous literature has found that  $\gamma$ -terpinene is one of the major constituents of coriander EO, making up anywhere from 0.2-14.42 % total EO terpene content (Bhuiyan, 2009; Misharina, 2001; Msaada, 2009b; Sriti, 2009). The EO analysis performed here found  $\gamma$ -terpinene to make up 2.66% of coriander EO, making this monoterpene the fifth most abundant in the EO of the coriander individuals used for this project.

RNA extracted from *C. sativum* seeds, collected at three maturity levels, was sent to the University of Lethbridge (ULeth) for transcriptome sequencing via an Illumina sequencing platform (GAIIx sequencer). The 33, 330, 312 raw single-end reads were assembled into 65, 306 transcript sequences. The transcripts were GO and KEGG annotated, aligned against the sequences in the NCBI non-redundant, Uniprot, and TAIR databases. Analysis of the read count matrix generated at ULeth revealed that the isoprenoid biosynthetic gene, DXS2, exhibited a 2-fold down-regulation in transcript abundance from mid- to late stage seed development. Three of four monoterpene synthase genes studied demonstrated varying levels of fold change between two, or all three, of the seed maturity stages analyzed.

From the transcript library, using homology searches and looking for transcripts with GO annotations such as “monoterpene biosynthetic process”, four monoterpene synthase candidate genes were identified. Three of these genes, *CsyTRPS*, *mTPS2* and *mTPS3* were cloned and *CsyTRPS* was functionally characterized. The kinetics properties of *CsyTRPS* were calculated as  $2.2 \pm 0.2 \text{ p}_{\text{kat}}/\text{mg}$ ,  $66 \pm 13 \mu\text{M}$  and  $1.476 \times 10^{-4} \text{ s}^{-1}$  for  $V_{\text{max}}$ ,  $K_m$  and  $k_{\text{cat}}$ , respectively.

## 5.2 Research Novelty

Coriander is an underutilized specialty crop which produces high yields of the monounsaturated fatty acid, petroselinic acid, useful in the production of detergents and nylon polymers (Msaada, 2009a). Coriander EO is extremely rich in (S)-linalool (75-80% total EO terpene content). This monoterpene alcohol is used in the culinary industry as a flavouring agent as it lends food a citrusy flavour (Msaada, 2009b; Pichersky, 1995). Linalool's fresh flower scent makes it a popular fragrance ingredient in household cleaning and personal hygiene products (e.g., shampoo, shower gel and soap) (Buckley, 2007; Christensson, 2010).

This work has provided a transcriptomic resource for this non-model plant which will aid future efforts to improve this crop plant via metabolic engineering and breeding research. This may be of particular interest in specialized fatty acids research as multiple genes encoding palmitoyl- $\Delta^4$ -acyl carrier protein desaturase,  $\beta$ -ketoacyl-acyl carrier protein-synthase I, and petroselinoyl-acyl carrier protein thioesterase. All three genes are directly involved in biosynthesis of petroselinic acid and to date only a single nucleic acid sequence is publicly available (Genbank) for each gene (Cahoon, 1992; Dormann, 1994; Mekhedov, 2001).

To our knowledge, terpene synthases from coriander, or from any plant seeds, have not been reported prior to this investigation. The fact that coriander produces solely (S)-linalool in its EO is of interest as the linalool synthase, characterized subsequent to this master's project by Lukman Sarker, yields only the (S)- isomer of linalool when fed with geranyl diphosphate substrate. It is known that in other plants (e.g., lavender flowers) the linalool synthase produces only the (R)-isomer of linalool (Landmann, 2007). A study on the nucleic acid and protein sequence differences between linalool synthases which produce strictly one isomer or the other will contribute to the advancement of pure terpene science research.  $\gamma$ -Terpinene synthase was partially characterized by Lucker et al., 2002 from *Citrus limon*. Lucker calculated a  $K_m$  and determined pH, temperature and cofactor optimums. Kinetic data collected here include the  $V_{max}$ ,  $k_{cat}$ , and catalytic efficiency, in addition to  $K_m$  and pH, temperature and cofactor optimums.

### 5.3 Assumptions and Limitations

One major assumption of the RNA-seq experiment was that seed size correlated with seed maturity. In reality it may be that certain seeds never reach 4mm diameter (which was considered to be the late developmental stage), yet those seeds could be the same number of days old as another seed which is indeed 4 mm in diameter. Another method of seed age determination would have been to count the days after flowering (post-anthesis) and use that number as a measure of seed maturity. However, it may also be possible that not all seeds will mature at the same speed, because there may be uneven distribution of resources throughout the parent plant due to growing environment variables, such as direct sunlight exposure on the developing seed (some seeds are partially covered by leaves or other seeds); thus, the former method (using seed size) was chosen to describe age.

Although no defined technical replicates were prepared during transcript sequencing it has been previously shown that technical reproducibility of next generation sequencing platforms, including the Illumina genome analyzer, is excellent, to the point that technical replicates are not necessary (Marioni, 2008). Lack of defined biological replicates in this project, however, limited the ability to control for biological variability. Individuals within a species will present a great range of genetic variability (Campbell, 2005). This was observed upon comparing the DXS 1 and DXS2 transcript abundance results between the RNA-seq and qPCR data, which do not correlate well. These genetic variations arise from factors associated with sexual reproduction, coupled with environmental influences (Campbell, 2005). Environmental influences include anything in an individual's environment which places a stress on that individual's survival. In terms of plants, influences can be insufficient sunlight for a developing plant tissue, or over-abundance of water around a plant's root system. Any environmental factors will force the selection of desirable phenotypic traits which improve an organism's chances of survival in a particular environment. Thus, a population will adapt genetically to better suit a certain environment.

Another assumption, associated with the production of the transcript library, was that during cDNA synthesis prior to Illumina sequencing, all the mRNA transcripts are

uniformly reverse transcribed, thus they will be equally represented in the cDNA being sequenced. Before reverse transcription, the RNA molecules are chemically fragmented and at that point it is assumed that the fragmentation occurred randomly and in a uniform manner. This assumption has been shown to be valid by previous research (Mortazavi, 2008). Secondly, there are limited known genomic or transcriptomic resources available for coriander, thus, it was difficult to assemble the transcriptome library *de novo*, without a reference library to align the resulting transcripts against. This difficulty is one commonly encountered where limited genomic or transcriptomic resources are available for the organism being studied, and researchers are constantly working to improve *de novo* assembly software.

#### **5.4 Future Directions**

Only four monoterpene synthase candidate genes were identified in this project, yet as previously described there are up to 37 different monoterpenes in coriander's EO (Bhuiyan, 2009; Misharina, 2001; Msaada, 2009b; Sriti, 2009). This leads to the question of whether any of mTPS2 or mTPS3 produce multiple monoterpene products from GPP substrate, as was the case with C $\gamma$ TRPS, or are there mTPS genes yet to be discovered in the coriander transcript library? Perhaps there are mTPS genes in the library which have no significant homology with mTPSs known to date in other plant species. Thus, these genes would not have been identified via the sequence homology and gene annotation search methods used in this project. There may still be monoterpene synthase genes to be discovered in coriander, using different approaches than those described in this thesis.

The *de novo* transcriptome assembly produced by this research enables future efforts to identify molecular markers in coriander such as (short tandem repeats) STR's and (single nucleotide polymorphisms) SNP's. These markers are important for the improvement of oil crops such as coriander by marker-assisted selective breeding (Li et al., 2012; Li et al., 2013). Those transcripts which did not align with any known plant genes and which have GO annotations such as "putative", "predicted", "unknown" or "hypothetical" may belong to untranslated regions or other non-coding RNA (e.g., tRNA). Alternatively, these transcripts may indicate coriander-specific genes which do

not have significant homology to known plant genes. The study of these genes will lead to an increased understanding of processes such as adaptation and speciation in coriander. Ultimately, any pure biosynthetic pathway research, such as the coriander terpene synthase characterization and transcript sequencing produced in this project, will aid future work regarding metabolic engineering. As an example, pure research on the taxol biosynthetic pathways in yew trees has paved the way for increased production of this medically valuable (anti-cancer) diterpene by metabolic engineering (Engels, 2008; Hao, 2011; Khosroushahi, 2006; Shoendorf, 2006).

## 5.5 Accession Numbers

The Genbank Accession numbers of all sequences used in alignments and phylogenetic tree are as follows: R-linalool synthase, *A. annua*: AAF13357.1; S-linalool synthase, *A. arguta*: DD81294.1; camphene synthase, *A. grandis*: AAB70707.1;  $\gamma$ -humulene synthase, *A. grandis*: AAC05728.1;  $\beta$ -phellandrene synthase, *A. grandis*: Q9M7D1.1; myrcene synthase, *A. majus*: AAO41726.1;  $\beta$ -ocimene synthase, *A. majus*: AAO42614.1; S-linalool synthase, *A. polygama*: ADD81295.1; S-linalool synthase, *A. thaliana*: AAO85533.1; ent-kaur-16-ene synthase, *A. thaliana*: AEE36246.1; 1, 8-cineole synthase, *A. thaliana*: AEE77075.1; S-linalool synthase, *C. breweri*: AAC49395.1; S-linalool synthase, *C. concinna*: AAD19839.1;  $\beta$ -pinene synthase, *C. limon*: Q8L5K2;  $\gamma$ -terpinene synthase, *C. limon*: Q8L5K4.1; ent-kaur-16-ene synthase, *C. maxima*: AAB39482.1; copalyl diphosphate synthase, *C. maxima*: BAC76429.1; S-linalool synthase, *C. osmophloeum*: AFQ20812.1; limonene synthase, *C. sativa*: ABI21837.1; S-nerolidol synthase, *F. ananassa*: P0CV94.1;  $\delta$ -cadinene synthase, *G. arboreum*: CAA77191.1; R-linalool synthase, *L. angustifolia*: ABB73045.1; R-linalool synthase, *M. aquatica*: AAL99381.1;  $\beta$ -cubebene synthase, *M. grandiflora*: ACC66281.1;  $\beta$ -farnesene synthase, *M. x piperita*: AAB95209.1;  $\gamma$ -terpinene synthase, *O. vulgare*: E2E2P0.1; R-linalool synthase, *P. abies*: AAS47693.1;  $\alpha$ -farnesene synthase, *P. abies*: AAS47697.1; limonene synthase, *P. sitchensis*: ABA86248.1;  $\alpha$ -pinene synthase, *P. taeda*: AAO61225.1; ent-kaurene synthase, *P. trichocarpa*: EEE88653.1; pinene synthase, *R. officinalis*: ABP01684.1; copalyl diphosphate synthase, S.

*lycopersicum*: AEP82766.1; bornyl diphosphate synthase, *S. officinalis*: AAC26017.1; Vetispiradiene synthase, *S. tuberosum*: BAA82092.1;  $\alpha$ -terpineol synthase, *V. vinifera*: AAS79352; germacrene D, *Z. officinale*: AAX40665.1;  $\alpha$ -humulene synthase, *Z. zerumbet*: BAG12315.1

## REFERENCES

- Adorjan, B., and Buchbauer, G.** (2010). Biological properties of essential oils: An updated review. *Flavour Fragrance J.* **25**, 407-426.
- Ahmadian, A., Ehn, M., and Hober, S.** (2006). Pyrosequencing: History, biochemistry and future. *Clinica Chimica Acta.* **363**, 83-94.
- Alba, R., Fei, Z. J., Payton, P., Liu, Y., Moore, S. L., Debbie, P., Cohn, J., D'Ascenzo, M., Gordon, J. S., Martin, G., Tanksley, S. D., Bouzayen, M., Jahn, M. M., and Giovannoni A.** (2004). ESTs, cDNA microarrays, and gene expression profiling: Tools for dissecting plant physiology and development. *Plant J.* **39**, 697-714.
- Bakkali, F., Averbeck, S., Averbeck, D., and Waomar, M.** (2008). Biological effects of essential oils - A review. *Food Chem. Toxicol.* **46**, 446-475.
- Barrero, A.F., del Moral, L.F.Q., Lara, A., and Herrador, M.M.** (2005). Antimicrobial activity of sesquiterpenes from the essential oil of *Juniperus thurifera* wood. *Planta Med.* **71**, 67-71.
- Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W. C., Ledoux, P., Rudenv, D., Lash, A. E., Fugibuchi, W., Edgar, R.** (2005). NCBI GEO: Mining millions of expression profiles - database and tools. *Nucleic Acids Res.* **33**, D562-D566.
- Bentley, D.R.** (2006). Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545-552.
- Benveniste, P.** (2004). Biosynthesis and accumulation of sterols. *Annu. Rev. Plant Biol.* **55**, 429-457.
- Bhuiyan, N., I., Begum, J., and Sultana, M.** (2009). Chemical composition of leaf and seed essential oil of *Coriandrum sativum* L. from Bangladesh. *Bangladesh J. Pharmacol.* **4**, 150-153.
- Bohlmann, J., Meyer-Gauen, G., and Croteau, R.** (1998). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4126-4133.
- Brown, G.D.** (2010). The biosynthesis of artemisinin (qinghaosu) and the phytochemistry of *Artemisia annua* L. (qinghao). *Molecules.* **15**, 7603-7698.
- Bruin, J., Dicke, M., and Sabelis, M.W.** (1992). Plants are better protected against spider-mites after exposure to volatiles from infested conspecifics. *Experientia.* **48**, 525-529.
- Buchanan, B., Gruissem, W., and Jones, R.** (2002). *Biochemistry and molecular biology of plants.* Wiley and Sons Inc, Somerset, NJ, USA. 1252-1258.

- Buckley, D.A.** (2007). Fragrance ingredient labelling in products on sale in the UK. *Br. J. Dermatol.* **157**, 295-300.
- Burdock, G.A., and Carabin, I.G.** (2009). Safety assessment of coriander (*Coriandrum sativum* L.) essential oil as a food ingredient. *Food Chem. Toxicol.* **47**, 22-34.
- Cahoon, E., Shanklin, J., and Ohlrogge, J.** (1992). Expression of a coriander desaturase results in petroselinic acid production in transgenic tobacco. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11184-11188.
- Campbell, N. A., and Reece, J. B.** (2005). *Biology*, Pearson Benjamin Cummings, SanFrancisco, California, USA, 389-390.
- Chen, X., Yauk, Y., Nieuwenhuizen, N. J., Niels, J., Matich, A. J., Wang, M. Y., Perez, R. L., Atkinson, R. G., and Beunig, L. L.** (2010). Characterisation of an (S)-linalool synthase from kiwifruit (*actinidia arguta*) that catalyses the first committed step in the production of floral lilac compounds. *Functional Plant Biology.* **37**, 232-243.
- Chithra, V., and Leelamma, S.** (2000). *Coriandrum sativum* - effect on lipid metabolism in 1,2-dimethyl hydrazine induced colon cancer. *J. Ethnopharmacol.* **71**, 457-463.
- Christensson, J.B., Matura, M., Gruvberger, B., Bruze, M., and Karlberg, A.** (2010). Linalool - a significant contact sensitizer after air exposure. *Contact Derm.* **62**, 32-41.
- Clouse, S., and Sasse, J.** (1998). Brassinosteroids: Essential regulators of plant growth and development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 427-451.
- Compeau, P.E.C., Pevzner, P.A., and Tesler, G.** (2011). How to apply de bruijn graphs to genome assembly. *Nat. Biotechnol.* **29**, 987-991.
- Croteau, R.B., Davis, E.M., Ringer, K.L., and Wildung, M.R.** (2005). (-)-Menthol biosynthesis and molecular genetics. *Naturwissenschaften.* **92**, 562-577.
- Degenhardt, J., Koellner, T.G., and Gershenzon, J.** (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry.* **70**, 1621-1637.
- Delaquis, P., Stanich, K., Girard, B., and Mazza, G.** (2002). Antimicrobial activity of individual and mixed fractions of dill, cilantro, coriander and eucalyptus essential oils. *Int. J. Food Microbiol.* **74**, 101-109.
- Dhanapakiam, P., Joseph, J.M., Ramaswamy, V.K., Moorthi, M., and Kumar, A.S.** (2008). The cholesterol lowering property of coriander seeds (*Coriandrum sativum*): Mechanism of action. *J. Environ. Biol.* **29**, 53-56.

- Dormann, P., Frentzen, M., and Ohlrogge, J.** (1994). Specificities of the acyl-acyl carrier protein (acp) thioesterase and glycerol-3-phosphate acyltransferase for octadecenoyl-acp isomers - identification of a petroselinoyl-acp thioesterase in *umbelliferae*. *Plant Physiol.* **104**, 839-844.
- Dudareva, N., Martin, D., Kish, C. M., Kolosova, N., Gorenstein, N., Faldt, J., Miller, B., and Bohlmann, J.** (2003). (E)-beta-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: Function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell.* **15**, 1227-1241.
- Engels, B., Dahm, P., and Jennewein, S.** (2008). Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards taxol (paclitaxel) production. *Metab. Eng.* **10**, 201-206.
- Fraser, P.D., and Bramley, P.M.** (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* **43**, 228-265.
- Frohman, M., Dush, M., and Martin, G.** (1988). Rapid production of full-length cDNAs from rare transcripts - amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998-9002.
- Gallagher, A., Flatt, P., Duffy, G., and Abdel-Wahab, Y.** (2003). The effects of traditional antidiabetic plants on in vitro glucose diffusion. *Nutr. Res.* **23**, 413-424.
- Ganjewala, D., Kumar, S., and Luthra, R.** (2009). An account of cloned genes of methyl-erythritol-4-phosphate pathway of isoprenoid biosynthesis in plants. *Current Issues in Molecular Biology.* **11**, 35-45.
- Garg, R., Patel, R.K., Tyagi, A.K., and Jain, M.** (2011). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* **18**, 53-63.
- Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Guiliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M. L., Mattanovich, D., Villaverde, A.** (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. *Microbial Cell Factories.* **7**, 11-32.
- Hao, D.C., Ge, G., Xiao, P., Zhang, Y., and Yang, L.** (2011). The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing. *PLoS One.* **6**, e21220.
- Harper, S., and Speicher, D., W.** (2011). Purification of protein fused to glutathione s-transferase. *Methods Mol. Biol.* **681**, 259-280.

- Hasunuma, T., Takeno, S., Hayashi, S., Sendai, M., Bambi, T., Yoshimura, S., Tomizawa, K., Fukusaki, E., and Miyake, C.** (2008). Overexpression of 1-deoxy-D-xylulose-5-phosphate reductoisomerase gene in chloroplast contributes to increment of isoprenoid production. *J. Biosci. Bioeng.* **105**, 518-526.
- Howitt, C.A., and Pogson, B.J.** (2006). Carotenoid accumulation and function in seeds and non-green tissues. *Plant Cell Environ.* **29**, 435-445.
- Iijima, Y., Davidovich-Rikanati, R., Fridman, E., Gang, D. R., Lewinsohn, E., and Pichersky, E.** (2004). The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil. *Plant Physiol.* **136**, 3724-3736.
- Illumina Inc.** (2010). Illumina sequencing technology: highest data accuracy, simple workflow, and a broad range of applications. *Technology Spotlight: Illumina® sequencing.* San Diego, California, USA, 1-5.
- Jerkovic, I., Mastelic, J., and Milos, M.** (2001). The impact of both the season of collection and drying on the volatile constituents of *Origanum vulgare* L. ssp *hirtum* grown wild in croatia. *International Journal of Food Science and Technology.* **36**, 649-654.
- Jones, C.G., Keeling, C.I., Ghisalberti, E.L., Barbour, E.L., Plummer, J.A., and Bohlmann, J.** (2008). Isolation of cDNAs and functional characterisation of two multi-product terpene synthase enzymes from sandalwood, *Santalum album* L. *Arch. Biochem. Biophys.* **477**, 121-130.
- Khosroushahi, A., Valizadeh, M., Ghasempour, A., Khosrowshahli, M., Naghdibadi, H., Dadpour, M. R., and Omid, Y.** (2006). Improved taxol production by combination of inducing factors in suspension cell culture of *Taxus baccata*. *Cell Biol. Int.* **30**, 262-269.
- Kubo, I., Fujita, K., Kubo, A., Nihei, K., and Ogura, T.** (2004). Antibacterial activity of coriander volatile compounds against *Salmonella choleraesuis*. *J. Agric. Food Chem.* **52**, 3329-3332.
- Kwok, S., Chang, S., Sninsky, J., and Wang, A.** (1994). A guide to the design and use of mismatched and degenerate primers. *PCR-Methods Appl.* **3**, S39-S47.
- Landmann, C., Fink, B., Festner, M., Dregus, M., Engel, K., and Schwab, W.** (2007). Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). *Arch. Biochem. Biophys.* **465**, 417-429.

- Lane, A., Boecklemann, A., Woronuk, G.N., Sarker, L., and Mahmoud, S.S.** (2010). A genomics resource for investigating regulation of essential oil production in *Lavandula angustifolia*. *Planta*. **231**, 835-845.
- Li, B., and Dewey, C.N.** (2011). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*. **12**, 323-338.
- Li, C., Zhu, Y., Guo, X., Sun, C., Luo, H., Song, J., Li, Y., Wang, L., Qian, J., and Chen, S.** (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and single sequence repeats in *Panax ginseng* C. A. meyer. *BMC Genomics*. **14**, 245-255.
- Li, X., Acharya, A., Farmer, A. D., Crow, J. A., Bharti, A. K., Kramer, R. S., Wei, Y., Han, Y., Gou, J., May, G. D., Monteros, M. J., and Brummer, E. C.** (2012). Prevalence of single nucleotide polymorphism among 27 diverse alfalfa genotypes as assessed by transcriptome sequencing. *BMC Genomics*. **13**, 568-579.
- Linhart, C., and Shamir, R.** (2005). The degenerate primer design problem: Theory and applications. *J. Comput. Biol.* **12**, 431-456.
- Lo Cantore, P., Iacobellis, N., De Marco, A., Capasso, F., and Senatore, F.** (2004). Antibacterial activity of *Coriandrum sativum* L. and *Foeniculum vulgare* miller var. *vulgare* (miller) essential oils. *J. Agric. Food Chem.* **52**, 7862-7866.
- Madan, L.L., and Gopal, B.** (2008). Addition of a polypeptide stretch at the N-terminus improves the expression, stability and solubility of recombinant protein tyrosine phosphatases from *Drosophila melanogaster*. *Protein Expr. Purif.* **57**, 234-243.
- Mahendra, P., and Bisht, S.** (2011). Anti-anxiety activity of *Coriandrum sativum* assessed using different experimental anxiety models. *Indian J. Pharmacol.* **43**, 574-577.
- Mahmoud, S.S., and Croteau, R.B.** (2002). Strategies for transgenic manipulation of monoterpene biosynthesis in plants. *Trends Plant Sci.* **7**, 366-373.
- Maluf, M., Saab, I., Wurtzel, E., and Sachs, M.** (1997). The viviparous12 maize mutant is deficient in abscisic acid, carotenoids, and chlorophyll synthesis. *J. Exp. Bot.* **48**, 1259-1268.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y.** (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509-1517.
- Mcgarvey, D.J., and Croteau, R.** (1995). Terpenoid metabolism. *Plant Cell.* **7**, 1015-1026.

- Mekhedov, S., Cahoon, E., and Ohlrogge, J.** (2001). An unusual seed-specific 3-ketoacyl-ACP synthase associated with the biosynthesis of petroselinic acid in coriander. *Plant Mol. Biol.* **47**, 507-518.
- Misharina, T.** (2001). Influence of the duration and conditions of storage on the composition of the essential oil from coriander seeds. *Appl. Biochem. Microbiol.* **37**, 622-628.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M.** (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, 182-185.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B.** (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods.* **5**, 621-628.
- Msaada, K., Hosni, K., Ben Taarit, M., Chahed, T., Hammami, M., and Marzouk, B.** (2009a). Changes in fatty acid composition of coriander (*Coriandrum sativum* L.) fruit during maturation. *Ind. Crop. Prod.* **29**, 269-274.
- Msaada, K., Hosni, K., Ben Taarit, M., Ouchikh, O., and Marzouk, B.** (2009b). Variations in essential oil composition during maturation of coriander (*Coriandrum sativum* L.) fruits. *J. Food Biochem.* **33**, 603-612.
- Mumm, R., Posthumus, M.A., and Dicke, M.** (2008). Significance of terpenoids in induced indirect plant defence against herbivorous arthropods. *Plant Cell Environ.* **31**, 575-585.
- Munoz-Bertomeu, J., Arrillaga, I., Ros, R., and Segura, J.** (2006). Up-regulation of 1-deoxy-D-xylulose-5-phosphate synthase enhances production of essential oils in transgenic spike lavender. *Plant Physiol.* **142**, 890-900.
- Myllykangas, S., Buenrostro, J., Ji, H. P.** (2012). Chapter 2: Overview of sequencing technology platforms. *Bioinformatics for high throughput sequencing*, 17-20.
- Nafis, T., Akmal, M., Ram, M., Alam, P., Ahlawat, S., Mohd, A., and Abdin, M. Z.** (2011). Enhancement of artemisinin content by constitutive expression of the HMG-CoA reductase gene in high-yielding strain of *Artemisia annua* L. *Plant Biotechnol. Rep.* **5**, 53-60.
- Nagegowda, D.A.** (2010). Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Lett.* **584**, 2965-2973.
- Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G., and Dudareva, N.** (2008). Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant J.* **55**, 224-239.

- Niedringhaus, T.P., Milanova, D., Kerby, M.B., Snyder, M.P., and Barron, A.E.** (2011). Landscape of next-generation sequencing technologies. *Anal. Chem.* **83**, 4327-4341.
- Novagen.** (2003). pET system manual, 10<sup>th</sup> ed. 6-8.
- Paszkiwicz, K., and Studholme, D.J.** (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics.* **11**, 457-472.
- Pfeifer, A., and Verma, I.** (2001). Gene therapy: Promises and problems. *Annu. Rev. Genomics Hum. Genet.* **2**, 177-211.
- Pichersky, E., Lewinsohn, E., and Croteau, R.** (1995). Purification and characterization of S-linalool synthase, an enzyme involved in the production of floral scent in *Clarkia-breweri*. *Arch. Biochem. Biophys.* **316**, 803-807.
- Potter, T.** (1996). Essential oil composition of cilantro. *J. Agric. Food Chem.* **44**, 1824-1826.
- Qin, X., and Zeevaart, J.** (2002). Overexpression of a 9-cis-epoxycarotenoid dioxygenase gene in *Nicotiana plumbaginifolia* increases abscisic acid and phaseic acid levels and enhances drought tolerance. *Plant Physiol.* **128**, 544-551.
- Ravi, R., Prakash, M., and Bhat, K.K.** (2007). Aroma characterization of coriander (*Coriandrum sativum* L.) oil samples. *European Food Research and Technology.* **225**, 367-374.
- Rigano, M.M., De Guzman, G., Walmsley, A.M., Frusciante, L., and Barone, A.** (2013). Production of pharmaceutical proteins in solanaceae food crops. *Int. J. Mol. Sci.* **14**, 2753-2773.
- Rodriguez-Concepcion, M.** (2010). Supply of precursors for carotenoid biosynthesis in plants. *Arch. Biochem. Biophys.* **504**, 118-122.
- Schillmiller, A.L., Miner, D. P., Larson, M., McDowell, E., Gang, D. R., Wilkerson, C., and Last, R. L.** (2010). Studies of a biochemical factory: Tomato trichome deep expressed sequence tag sequencing and proteomics. *Plant Physiol.* **153**, 1212-1223.
- Shen, Z., Liu, J., Wells, R., L., and Elkind, M., M.** (2003). Direct sequencing with highly degenerate inosine-containing primers. *Methods Mol. Biol.* **226**, 367-372.
- Schoendorf, A., Rithner, C.D., Williams, R.M., and Croteau, R.B.** (2001). Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from taxus and functional expression in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1501-1506.

- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E.** (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. **28**, 1086-1092.
- Singh, G., Kapoor, I., Pandey, S., Singh, U., and Singh, R.** (2002). Studies on essential oils: Part 10; antibacterial activity of volatile oils of some spices. *Phytother. Res.* **16**, 680-682.
- Sriti, J., Talou, T., Wannas, W.A., Cerny, M., and Marzouk, B.** (2009). Essential oil, fatty acid and sterol composition of Tunisian coriander fruit different parts. *J. Sci. Food Agric.* **89**, 1659-1664.
- Steele, C., Crock, J., Bohlmann, J., and Croteau, R.** (1998). Sesquiterpene synthases from grand fir (*Abies grandis*) - comparison of constitutive and wound-induced activities, and cDNA isolation, characterization and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *J. Biol. Chem.* **273**, 2078-2089.
- Sugiura, M., Ito, S., Saito, Y., Niwa, Y., Koltunow, A. M., Sugimoto, O., and Sakai, H.** (2011). Molecular cloning and characterization of a linalool synthase from lemon myrtle. *Biosci. Biotechnol. Biochem.* **75**, 1245-1248.
- Takeda, K., Itoh, H., Yoshioka, I., Yamamoto, M., Misaki, H., Kajita, S., Shirai, K., Kato, M., Shin, T., Murao, S., and Tsukagoshi, N.** (1998). Cloning of a thermostable ascorbate oxidase gene from *Acremonium* sp. HI-25 and modification of the azide sensitivity of the enzyme by site-directed mutagenesis. *Biochim. Biophys. Acta-Protein Struct. Molec. Enzym.* **1388**, 444-456.
- Tholl, D.** (2006). Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9**, 297-304.
- Ulubelen, A.** (2003). Cardioactive and antibacterial terpenoids from some salvia species. *Phytochemistry*. **64**, 395-399.
- Valasek, M., and Repa, J.** (2005). The power of real-time PCR. *Adv. Physiol. Educ.* **29**, 151-159.
- Vonheijne, G., Steppuhn, J., and Herrmann, R.** (1989). Domain-structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* **180**, 535-545.
- Wall, P.K., Leebens-Mack, J., Chanderbali, A. S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L. P., Hu, Y., Carlson, J. E., Ma, H., Schuster, S. C., Soltis, D. E., Soltis, P. S., Altman, N., dePamphilis, C. W.** (2009). Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics*. **10**, 347.

- Wangensteen, H., Samuelsen, A., and Malterud, K.** (2004). Antioxidant activity in extracts from coriander. *Food Chem.* **88**, 293-297.
- Williams, D., McGarvey, D., Katahira, E., and Croteau, R.** (1998). Truncation of limonene synthase pre-protein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry (N. Y. )*. **37**, 12213-12220.
- Wise, M., Savage, T., Katahira, E., and Croteau, R.** (1998). Monoterpene synthases from common sage (*Salvia officinalis*) - cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *J. Biol. Chem.* **273**, 14891-14899.
- Wu, J., and Lin, L.** (2003). Enhancement of taxol production and release in *Taxus chinensis* cell cultures by ultrasound, methyl jasmonate and in situ solvent extraction. *Appl. Microbiol. Biotechnol.* **62**, 151-155.
- Yazaki, K.** (2006). ABC transporters involved in the transport of plant secondary metabolites. *FEBS Lett.* **580**, 1183-1191.
- Xia, Z., Xu, H., Zhai, J., Li, D., Luo, H., He, C., and Huang, X.** (2011). RNA-seq analysis and de novo transcriptome assembly of *hevea brasiliensis*. *Plant Mol. Biol.* **77**, 299-308.
- Yeung, E.C., and Bowra, S.** (2011). Embryo and endosperm development in coriander (*Coriandrum sativum*). *Botany.* **89**, 263-273.
- Zaidi, M.A., El Bilali, J., Koziol A. G., Ward, T. L., Styles, G., Greenham, T. J., Faiella, W. M., Son, H. H., Wan, S., Taga, I., and Altosaar, I.** (2012). Gene technology in agriculture, environment and biopharming: Beyond bt-rice and building better breeding budgets for crops. *J. Plant Biochem. Biotechnol.* **21**, S2-S9.
- Zerbino, D.R., and Birney, E.** (2008). Velvet: Algorithms for de novo short read assembly using de bruijn graphs. *Genome Res.* **18**, 821-829.