TERPENE AND ISOPRENOID BIOSYNTHESIS IN CANNABIS SATIVA

by

Judith Booth

B.Sc., The University of King's College, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Genome Science and Technology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Terpene and Isoprenoid Biosynthesis in *Cannabis sativa*

submitted by Judith Booth in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Genome Science and Technology

Examining Committee:

Dr. Joerg Bohlmann, Professor, Michael Smith Laboratories, UBC
Supervisor

Dr. Anne Lacey Samuels, Professor, Botany, UBC
Supervisory Committee Member

Dr. Murray Isman, Professor Emeritus, Land and Food Systems, UBC
University Examiner

Dr. Corey Nislow, Professor, Biochemistry and Molecular Biology, UBC
University Examiner

Additional Supervisory Committee Members:

Dr. Jonathan Page, Adjunct Professor, Botany, UBC
Supervisory Committee Member

Dr. Reinhard Jetter, Professor, Professor, Botany, UBC
Supervisory Committee Member

Abstract

Cannabis sativa (cannabis, marijuana, hemp) is a plant species grown widely for its psychoactive and medicinal properties. Cannabis products were made illegal in most of the world in the early 1900s, but regulations have recently been relaxed or lifted in some jurisdictions, notably Canada and parts of the United States. Cannabis is usually grown for the resin produced in trichomes on the flowers of female plants. The major components of that resin are isoprenoids: cannabinoids, monoterpenes, and sesquiterpenes. Terpene profiles in cannabis flowers can vary widely between cultivars. My research addresses the genomic underpinnings and biochemical mechanisms of terpene and cannabinoid biosynthesis in cannabis, and patterns of terpene accumulation between organs, developmental stages, and cultivars. Using metabolite profiling, I demonstrated that terpenes accumulate in floral trichomes over the course of development, and that terpene profiles in trichomes differ based on tissue and developmental stage. In this thesis, I describe the terpene profiles of seven cannabis cultivars. I identified and characterized 29 terpene synthase (TPS) genes and their encoded enzymes and describe the relationship between TPS expression and metabolite profiles. I describe trichome-specific transcriptomes for five cultivars and identify highly expressed genes common to cannabis trichomes. I also identified and describe an aromatic prenyltransferase responsible for biosynthesis of cannabigerolic acid, the branch-point intermediate in cannabinoid biosynthesis. Collectively, this thesis comprises a broad and detailed characterization of specialized isoprenoid biosynthesis in cannabis. The results provide new insights into mechanisms of terpene and cannabinoid biosynthesis, and the roles of different enzymes in determining the metabolite complement of cannabis trichomes.

Lay Summary

Cannabis sativa, commonly known as cannabis, marijuana, or hemp, is a plant species grown around the globe. The most valuable component of cannabis is its resin, which is produced on female flowers. The most abundant components of the cannabis resin are two related types of molecules: terpenes and cannabinoids. Cannabinoids have medicinal and psychoactive properties, while terpenes are responsible for the distinctive aromas of cannabis. Cannabis produces hundreds of different terpenes, and different types of cannabis contain different terpenes. My research focuses on enzymes that produce terpenes and cannabinoids, with a special focus on the terpene synthases. This thesis describes the discovery and characterization of 29 different terpene synthases and a cannabinoid synthase. I also describe the variety of terpenes found in cannabis, their distribution within the plant, and how they change as the plant develops.

Preface

Chapter 1: Literature review and overall thesis introduction - Enzymes and systems of isoprenoid biosynthesis

Portions of chapter 1 have been published. Sections 1.2, 1.3, 1.4, and 1.5, Figures 1.1 and 1.2, and Table 1.1 were published: Booth, J. K., & Bohlmann, J. (2019). Terpenes in Cannabis sativa – From plant genome to humans. *Plant Science*. The chapter was conceived and designed by Judith Booth and Dr. Jörg Bohlmann and written by Judith Booth. The manuscript was reviewed and revised by Dr. Jörg Bohlmann.

Chapter 2: Terpene synthases from *Cannabis sativa*

A version of Chapter 2 has been published: Booth, J. K., Page, J. E., & Bohlmann, J. (2017). Terpene synthases from Cannabis sativa. *PLOS ONE, 12*(3), e0173911. The chapter was conceived and designed by Judith Booth, Dr. Jörg Bohlmann, and Dr. Jonathan Page. Experiments and data analysis were performed by Judith Booth. The manuscript was written by Judith Booth, and reviewed and revised by Dr. Jörg Bohlmann with input from Dr. Jonathan Page.

Chapter 3: Terpene synthases and terpene variation in *Cannabis sativa*

A version of Chapter 3 has been accepted to the journal *Plant Physiology* with the following author list: Judith K Booth, Macaire MS Yuen, Sharon Jancsik, Lufiana L Madilao, Jonathan E. Page, and Jörg Bohlmann. The chapter was conceived and designed by Judith Booth and Dr. Jörg Bohlmann, with input from Dr. Jonathan Page. Experiments were designed and performed by Judith Booth. Macaire Yuen performed transcriptome assembly and assisted with transcriptome data analysis and statistics. Sharon Jancsik assisted with all aspects of data generation and bench work. Lufiana Madilao assisted with GC/MS experiments and data analysis. The manuscript was written by Judith Booth, and reviewed and revised by Dr. Jörg Bohlmann.

Chapter 5: Synthetic Biology of Cannabinoids and Cannabinoid Glycosides in *Nicotiana benthamiana* and *Saccharomyces cerevisiae*

A version of Chapter 6 has been accepted to the *Journal of Natural Products* with the following author list: Thies Gülck, Judith K. Booth, Ângela Carvalho, Christoph Crocoll, Mohammed Saddik Motawie, Birger Lindberg Møller, Jörg Bohlmann, Nethaji J. Gallage. This part of the thesis was a collaborative work: Judith Booth performed experimental design, construct design and cloning in tobacco, phylogenetic analysis, pathway expression in tobacco, subcellular localization of GFP-fusion proteins in tobacco, shotgun proteomics, LC-MS sample preparation, LC-MS data analysis. Thies Gülck performed experimental design, construct design and cloning for tobacco and yeast experiments, pathway expression in tobacco, transcriptomics interpretation, pathway expression in yeast and cannabinoid glycosylation in yeast in vivo, LC-MS sample preparation, LC-MS data analysis, manuscript writing. Ângela Carvalho performed construct design and cloning for tobacco

expression and UGT screening, UGT-screening in vitro. Christoph Crocoll performed LC-MS Q-TOF and tQuad performance and analysis, writing of the LC-MS method description. Mohammed Saddik Motawie performed chemical synthesis of OA-glucoside, writing of the method description of the chemical synthesis and reaction schemes. Jörg Bohlmann provided mentoring and discussion throughout the work and contribution to finalization of the manuscript**.** Birger Lindberg Møller provided mentoring and discussion throughout the work and contribution to finalization of the manuscript. Nethaji J. Gallage was responsible for experimental design, mentoring and discussion throughout the work and contribution to finalization of the manuscript.

Appendix: Cannabis glandular trichomes alter morphology and metabolite content during gland maturation

The data in this appendix is drawn from: Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D. C. J., Rensing, K. H., Laflamme-Yonkman, J., Castellarin, S. D., Bohlmann, J., Page, J. E., & Samuels, A. L. (2020). Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. *The Plant Journal*, *101*(1), 37–56. The appendix represents the portions performed by Judith Booth.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAE – acyl-activating enzyme

AK – Afghan Kush

ANOVA – analysis of variance

aPT – aromatic prenyltransferase

BC – Blue Cheese

CaMV – cauliflower mosaic virus

CBD(A) – cannabidiol(ic acid)

CBDAS – CBDA synthase

CBG(A) – cannabigerol(ic acid)

CHIL – Chalcone isomerase-like

Choc - Chocolope

CMK - 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase

CPM – counts per million transcripts

CSH – CBD Skunk Haze

DMAPP – dimethylallyl diphosphate

DOXP - 1-deoxy-D-xylulose 6-phosphate

DXR - DOXP reductoisomerase

DXS – DOXP synthase

FAD – fatty acid desaturase

FN – ‘Finola’

FPP – farnesyl diphosphate

FPPS – farnesyl diphosphate synthase

GFP – green fluorescent protein

GGPP – geranylgeranyl diphosphate

GGPPS – geranylgeranyl diphosphate synthase

glc – glycoside

glu - glucoside

GPP – geranyl diphosphate

GPPS – geranyl diphosphate synthase

GT – glandular trichome

HDR – HMP-PP reductase

HDS – HMB-PP synthase

HMB-PP - 4-hydroxy-3-methyl-but-2-enyl diphosphate

HMG-CoA - 3-hydroxy-3-methylglutaryl-CoA

HMGR - HMG-CoA reductase

HMGS – HMG-CoA synthase

IDI – IPP isomerase

IPP – isopentenyl diphosphate

LS – Lemon Skunk

LTP – lipid transfer protein

MCT - 2-C-methyl-D-erythritol cytidyltransferase

MEP – methylerythritol phosphate

MEV - mevalonate

MK – mevalonate kinase

MPDC - mevalonate-5-phosphate decarboxylase

OA – olivetolic acid

OAC – olivetolic acid cyclase

OLS/PKS– olivetol synthase/polyketide synthase

PK – Purple Kush

PMK – phospo-mevalonate kinase

RI – retention index

RT – retention time

THC(A) – Δ^9-tetrahydrocannabinol(ic acid)

THCAS – THCA synthase

TPS – terpene synthase(s)

UGT – uridine diphosphate glycosyl-transferase

Acknowledgements

First and foremost, I would like to acknowledge my lab family: Dr. Jörg Bohlmann, Dr. Carol Ritland, Karen Reid, Christine Chiu, Justin Whitehill, Angela Chiang, Katrin Geisler, Andreas Gesell, Maria Diaz, Chris Keeling, Reed Clements, Philipp Zerbe, Hannah Henderson, Suzanna Ivamoto, Jose Celedon, Chris Roach, Melissa Mageroy, Sandra Irmisch, Kristina Kshatriya, Omnia Gamal, Jenny Jo, Mack Yuen, Tal Shalev, Fred Sunstrum, and Lina Madilao for their help and support. Extra special appreciation goes to Sharon Jancsik, for her hard work and her unwavering patience and kindness. Additional thanks to my students over the years: Grant, Morton, and K'sana.

Many thanks to my committee, Dr. Anne Lacey Samuels, Dr. Reinhard Jetter, and Dr. Jonathan Page for their support. They elevated the science and always kept me thinking critically.

Thank you to my collaborators at UBC, Anandia, and University of Copenhagen: Samuel Livingston, Teagan Quilichini, Eva Chou, Samantha Mishos, Thies Gülck, Dr. Birger Lindberg Møller, and Dr. Nethaji Gallage.

Funding agencies that made this possible were NSERC, Genome BC, and the Carlsberg Foundation.

It was a crowd that helped me through every day, and I could not have done it without all of you.

This thesis is dedicated to Kai Miller. His support was the light that kept me going. I could never have even begun without Kai.

For Angela and Justin, who became like family. I could not imagine better friends.

Finally, for my parents. They have always believed in me, and I will never stop trying to make them proud.

Chapter 1: Literature Review and Overall Thesis Introduction - Enzymes and Systems of Isoprenoid Biosynthesis in *Cannabis sativa*

1.1 Summary

Cannabis sativa (cannabis) produces a resin that is valued for its psychoactive and medicinal properties. Despite being the foundation of a multi-billion dollar global industry, scientific knowledge and research on cannabis is lagging behind compared to other high-value crops. This is largely due to legal restrictions that have prevented many researchers from studying cannabis, its products, and their effects in humans. Cannabis resin contains hundreds of different terpene and cannabinoid metabolites. Many of these metabolites have not been conclusively identified. As a consequence, there is concern about lack of consistency with regard to the terpene and cannabinoid composition of different cannabis 'strains'. Terpenes are produced by terpene synthases (TPS) which are often promiscuous in their product specificity and encoded in large gene families. The branch-point intermediate in cannabinoid biosynthesis, cannabigerolic acid, is produced by aromatic prenyltransferases (aPTs). Our understanding of the genomic and biosynthetic systems of terpenes and cannabinoids in cannabis, and the factors that affect their variability, is rudimentary. Likewise, claims of some of the medicinal properties attributed to cannabis terpenes would benefit from thorough scientific validation. This thesis focuses on the TPSs and aPTs in cannabis with the overarching goal of understanding (i) isoprenoid biosynthesis in cannabis and (ii) variation of terpene profiles in different cannabis 'strains'.

1.2 Introduction

Cannabis sativa (cannabis) is thought to have originated from central Asia, and has been domesticated for over 5,000 years (Li, 1973; Hanuš et al., 2016). Cannabis varieties that are low in psychoactive cannabinoids are used for the production of fiber and oilseed. However, the most valuable cannabis product today is the terpene- and cannabinoid-rich resin with its various psychoactive and medicinal properties. The resin is produced and accumulates in glandular trichomes that densely cover the surfaces of female (pistillate) inflorescences and, to a lesser degree, the foliage of male and female plants (**Figure 1.1**). In total, more than 150 different terpenes and approximately 100 different cannabinoids (Hanuš et al., 2016) (**Figure 1.2**) have been identified in the resin of different cannabis types (**Table 1.1**). The predominant cannabinoids in cannabis grown for medicinal or recreational use are $^{\Delta}$9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA). While cannabinoids are the primary psychoactive and medicinal components of cannabis resin, volatile terpenes (monoterpenes and sesquiterpenes) contribute many of the different fragrance attributes that influence consumer preferences.

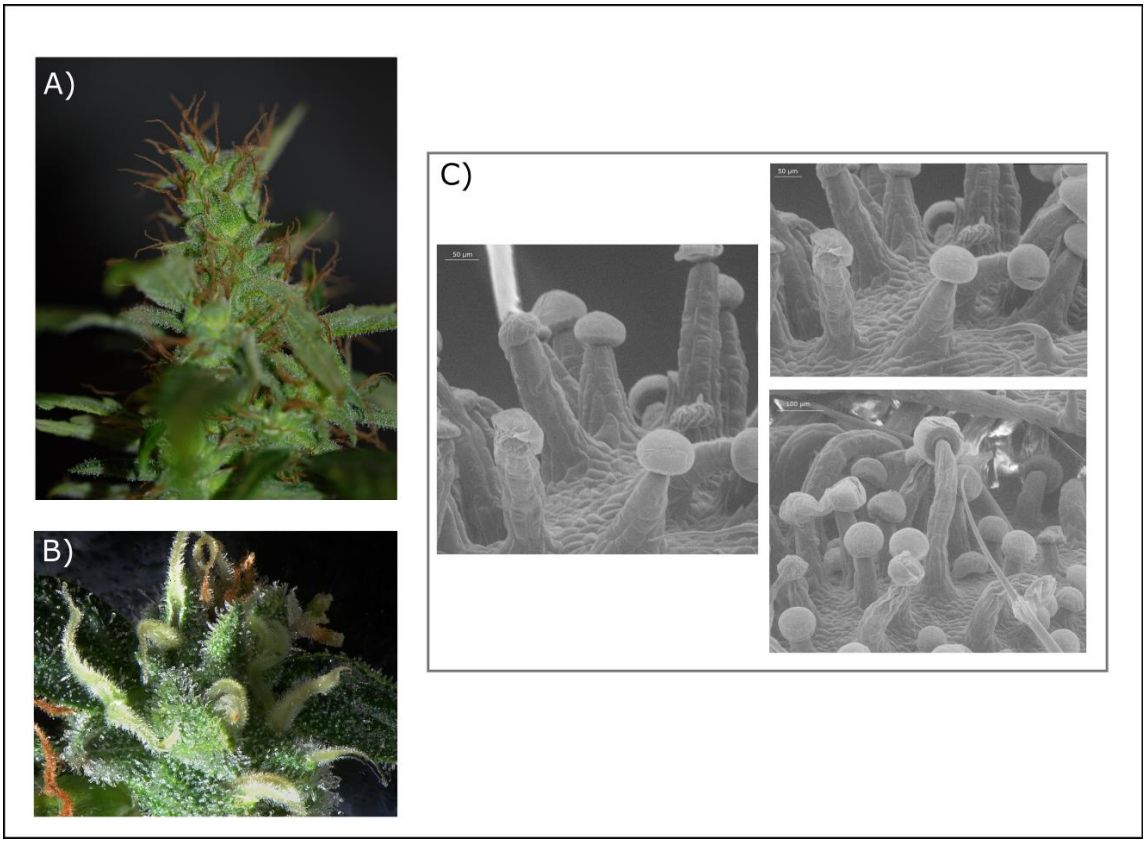


Figure 1.1 Cannabis inflorescence and stalked glandular trichomes.
A) Apical inflorescence from the strain Purple Kush, eight weeks post onset of flowering. B) Floret cluster from the strain Lemon Skunk, five weeks post onset of flowering. C) Stalked glandular trichomes on the surface of 'Finola' pistillate flowers. Scanning electron microscopy and image credit for C) thanks to Samuel Livingston, UBC, Department of Botany.

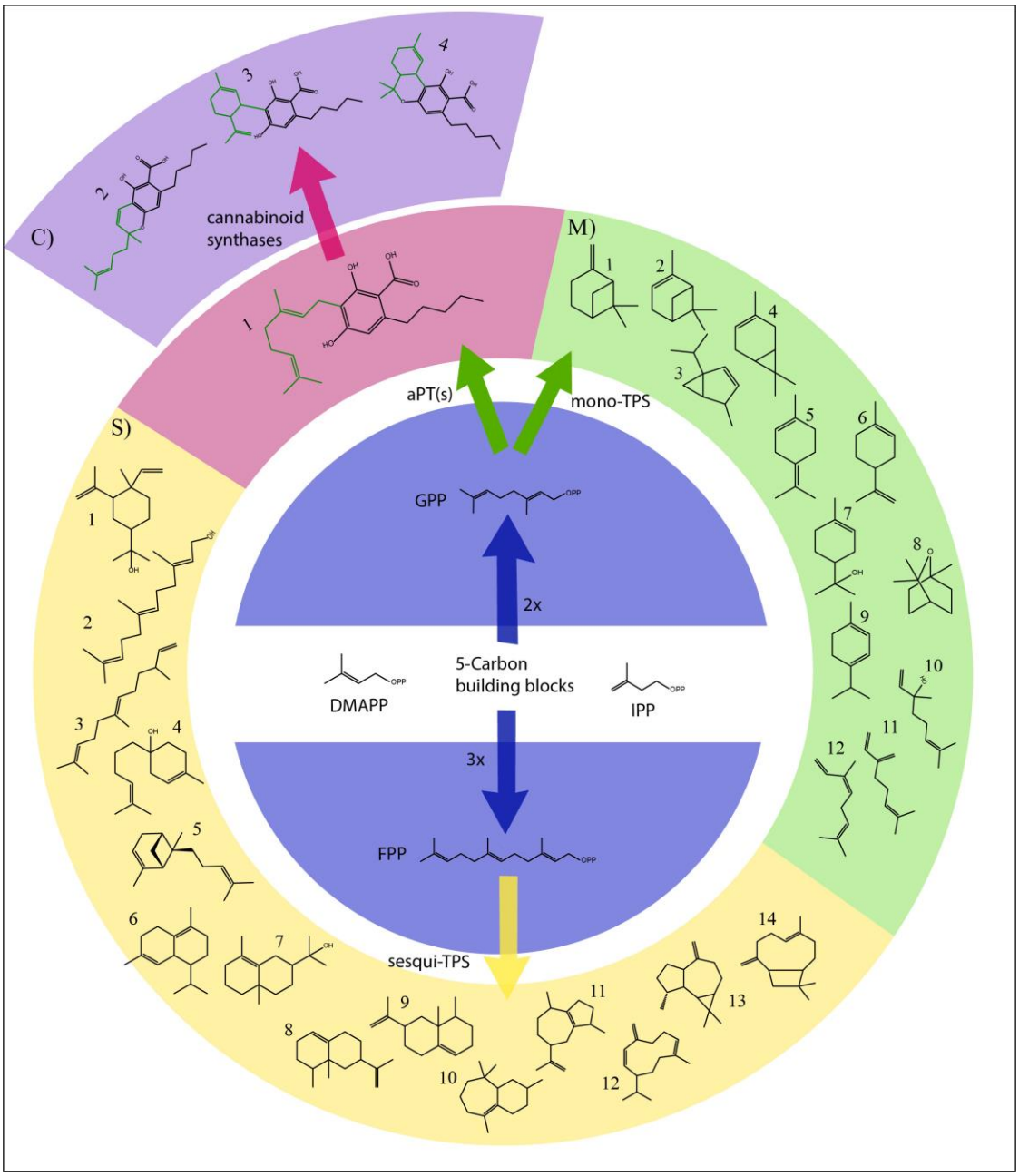


Figure 1.2 Schematic of terpene and isoprenoid biosynthesis in cannabis
5-Carbon isoprenoid building blocks isopententyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are condensed to form geranyl diphosphate (GPP) (C10) or farnesyl diphosphate (FPP) (C15). Terpene synthases (TPS) convert GPP or FPP into terpenes. Aromatic prenyltransferases (aPTs) condense GPP with olivetolic acid to form cannabigerolic acid (CBGA), which is cyclized by cannabinoid synthases to produce cannabinoids. Cannabinoids: C1: cannabigerolic acid, C2: cannabichromenic acid, C3: cannabidiolic acid, C4: tetrahydrocannabinolic acid. Monoterpenes: M1: β-pinene, M2: α-pinene, M3: β-thujone, M4: 3-carene, M5: terpinolene, M6: limonene, M7: terpineol, M8: 1,8-cineole, M9: α-terpinene, M10: linalool, M11: myrcene, M12: (Z)-β-ocimene. Sesquiterpenes S1: α-elemol, S2: (E)-β-farnesol, S3: (E)-β-farnesene, S4: bisabolol, S5: (+)-α-bergamotene, S6: δ-cadinene, S7: γ-eudesmol, S8: valencene, S9: eremophilene, S10: β-himachalene, S11: α-guaiene, S12: germacrene D, S13: alloaromadendrene, S14: β-caryophyllene.

Different cannabis types and their derived consumer products are commonly referred to with ‘strain’ names. These names often relate to fragrance attributes conferred, at least in part, by terpenes (Fischedick, 2017). Different ‘strains’ may be distinguished by morphological features or differences in the chemical composition of the resin. However, due to a history of largely illicit cannabis production, cannabis ‘strains’ are often poorly defined genetically. ‘Strains’ may lack reproducibility with regard to profiles of terpenes and cannabinoids (Elzinga et al., 2015; Sawler et al., 2015). The species encompasses large genetic diversity, with most strains having high levels of heterozygosity and genetic admixture (Sawler et al., 2015; Lynch et al., 2016). Cannabis is wind-pollinated, which also contributes to variability of cannabis metabolites. As a result, many cannabis ‘strains’ lack the level of standardization that producers and consumers are accustomed to with other crop plants, such as genetically and phenotypically well-defined grapevine varieties. In the absence of proper genetic or genomic characterization, some attempts have been made at chemotaxonomic classification of cannabis ‘strains’ based on terpenes, and cannabis plants have also been described as belonging to different chemotypes (**Table 1.1**). However, the complexity of terpene biosynthetic systems, and the many different sources of terpene variation, often renders these efforts futile. In general, concepts of chemotaxonomy have been made obsolete by genome sciences, and chemotypes cannot reliably substitute for properly genotyped plants.

# of terpenes identified	Origin of plant material	Purpose of analysis	Reference
25	Wild-grown in Kashmir	Plant Biology	Marchini et al., 2014
50	Forensic samples	Classification	Brenneisen et al., 1988
66	Grown by researchers	Plant Biology	Ross & ElSohly, 1996
48	Breeders, researchers, law enforcement	Classification	Hillig, 2004
16	Grown by researchers	Plant Biology	Potter, 2009
27	Bedrocan BV	Classification	Fischedick et al., 2010
49	Grown by researchers outdoors	Metabolite survey	Bertoli et al., 2010
28	Grown by researchers	Metabolite survey	Casano et al., 2011
20	Coffee shops in the Netherlands and Bedrocan BV	Classification	Hazekamp & Fischedick, 2012
12	Bedrocan BV	Industrial	Romano & Hazekamp, 2013
13	Grown outdoors	Industrial	Da Porto et al., 2014
27	Indoor cultivator in California	Industrial	Giese et al., 2015
28	Submissions from medical patients	Classification	Elzinga et al., 2015
28	Grown by researchers	Plant Biology	Aizpurua-Olaizola et al., 2016
17	Bedrocan BV	Industrial	Hazekamp, 2016
50	Bedrocan BV	Classification	Hazekamp et al., 2016
16	Submitted by dispensary	Classification	Fischedick, 2017
14	Licensed producers in Canada	Classification	Jin et al., 2017
20	Indoor cultivator in New Mexico, assorted growers	Classification	Richins et al., 2018
21	Dispensary in California	Medical	Blasco-Benito et al., 2018
45	Grown outdoors	Medical	Gallily et al., 2018

Table 1.1 Publications listing cannabis terpene profiles.
Purpose refers to the stated objective of the study. Origin of plant material indicates what the authors stated as the source of their cannabis or extracts. Number of terpenes identified includes all named or numbered compounds listed by the authors, including those not identified using authentic standards. Publications are listed in order of date published, from earliest to most recent.

With the lifting of some of the legal restrictions on cannabis research in Canada, and in some other jurisdictions, there is now an opportunity to build stronger scientific knowledge of the genomic, molecular and biochemical properties that define terpene and cannabinoid profiles in different cannabis 'strains'. This in turn can support the development of a larger number of well-defined cannabis varieties. another aspect that requires new research are the various effects that are attributed to cannabis terpenes in humans. While some of the effects of the cannabinoids have been scientifically explained, there is a great deal of uncertainty about the effects of cannabis terpenes in humans beyond fragrance perception. Although it is not a part of this thesis to study cannabis terpene effects in humans, claims of such effects provide some of the underlying motivation for this thesis on terpene biosynthesis and terpene variation and are therefore briefly reviewed in this chapter.

1.3 Terpene diversity and variation in cannabis differs between strains

Terpene composition is a phenotypic trait that shows great variation across different cannabis cultivars. (**Table 1.1**). The majority of terpenes found in cannabis are hydrocarbons, which are the direct products of terpene synthases (TPS) (Chen et al., 2011; Booth et al., 2017), as opposed to more complex terpenoids that require modification by other enzymes such as cytochrome P450s. Therefore, the chemical diversity of cannabis terpenes reflects the diversity of TPS enzymes encoded in the cannabis (Cs)TPS gene family.

The monoterpene myrcene as well as the sesquiterpenes β-caryophyllene and α-humulene appear to be present in most cannabis cultivars. Other common compounds include the monoterpenes α-pinene, limonene, and linalool as well as the sesquiterpenes bisabolol and (*E*)-β-farnesene. It is important to note that some terpenes, in particular sesquiterpenes, remain difficult

to identify due to the lack of authentic standards for many of these compounds. As a result, reports of terpene profiles in cannabis may include unknown compounds, rely on tentative identification, or present incomplete profiles of selected compounds. Stereochemistry is also not consistently described, or is often ignored, in reports on cannabis terpenes. These issues make it difficult to fully assess the diversity of terpenes in cannabis using the available data and make comparing different studies problematic.

To resolve issues of poor reproducibility of terpene profiles in cannabis, it will be essential to perform rigorous studies with a diversity of cannabis genotypes grown under controlled environmental conditions and analyze terpene profiles quantitatively and qualitatively over the course of plant development. This would need to include organ-, tissue- and cell-type specific terpene analysis, and would have to include controlled experiments to assess effects of environmental conditions such as light, irrigation, and nutrients. Such experiments should include not only terpene metabolite analysis, but also a comprehensive transcriptome profiling of CsTPS gene expression. The results of such a study would enable much needed proper assignment of reproducible terpene profiles to different 'strains' and support the standardization of cannabis varieties and derived consumer products.

1.4 Effects attributed to terpenes in cannabis

Arguably, the only unquestionable effect of cannabis terpenes on humans is the fragrance attributes of different mono- and sesquiterpene volatiles and their mixtures. Depending on the variable composition of cannabis terpene profiles, different cultivars elicit different fragrance impressions, which may affect consumer preference (Gilbert and DiVerdi, 2018). However,

other attributes assigned to terpenes in cannabis products, including medicinal properties, remain for now in the realm of ongoing research.

The so-called ‘entourage effect’ is a popular idea. It suggests a pharmacological synergy between cannabinoids and other components of cannabis resin, in particular terpenes (Russo, 2011; Nuutinen, 2018). Putative aspects of the entourage effect include the treatment of depression, anxiety, addiction, epilepsy, cancer, and infectious diseases. The anecdotal notion of a synergistic effect appears to stem from the perception among cannabis users that different cultivars have different physiological effects. There is no doubt that the large chemical space of thousands of plant terpenes and terpenoids includes many biologically active molecules. Some terpenoids, such as the anticancer drug taxol, are potent and highly valuable pharmaceuticals, the effects of which are supported by the full range of pharmacological and clinical studies (Gershenzon and Dudareva, 2007). In one of the few examples of the entourage effect being tested, terpenes were found not to contribute to cannabinoid-mediated analgesia in rats (Rousseau and Sabol, 2018). No molecular mechanism has been demonstrated to explain a potential synergy of terpenes with cannabinoids. One potential explanation for the effects attributed to cannabis terpenes is revealed in a recent review (Gertsch, 2018), pointing out that the placebo effect is partially mediated through the endocannabinoid system, which may explain some of the perceived effects of cannabis products.

The sesquiterpene β-caryophyllene is prominent in many cannabis cultivars and products. The molecule binds to the mammalian CB_2 cannabinoid receptor, which may provide a plausible mechanism for interaction with cannabinoids and a starting point for future research (Gertsch et al., 2008). β-caryophyllene is one of the least variable terpene components of cannabis (**Table 1.1**), which would suggest that it cannot explain ‘strain’-specific effects in humans. The

proposed synergistic effects of terpenes in the effects of cannabis in humans is an area that will require careful research, which will now be possible in those jurisdictions in which some of the legal restrictions have been lifted.

1.5 Claims of anticancer effects of cannabis and cannabis terpenes may do more harm than good

Certain monoterpenes have been shown to block tumor formation or inhibit cell cycle progression *in vivo* and in rats (Karlson et al., 1996; Burke et al., 1997; Gould, 1997). However, the amounts of terpenes required to produce anti-proliferative effects in rats are excessively high with up to 10% of the animals' diet (Gould, 1997). Similarly, cannabinoids may inhibit tumor formation in animal models of cancer (Blaquez et al., 2003). Laboratory studies such as these may have led to the suggestion that cannabis extracts, with their combination of cannabinoids and terpenes, have anti-cancer properties (Russo, 2011; Nuutinen, 2018). However, to our knowledge, there is no conclusive evidence to support claims of anticancer activity of terpenes consumed with cannabis products. While the ethanolic extract of cannabis flowers has higher antitumor activity than pure THC, this effect was not attributed to any of the five most abundant terpenes (Meehan-Atrash et al., 2019).

In general, it is important to remember that cannabis is often consumed by smoking or as a vapor. This includes cannabis consumption by young adults. Consumer habits such as inhaling combusted or vaporized cannabis products must be considered a health risk, including the potential risk of causing cancer or other health issues (Aldington et al., 2008; Meehan-Atrash et al., 2017; Meehan-Atrash et al., 2019), before promoting unsupported claims of anti-cancer effects of cannabis.

1.6 Isoprenoid biosynthesis in trichomes

Cannabis resin accumulates in glandular trichomes (GTs) that cover the surface of the flowers. These trichomes occur on the foliage of female and male plants, but they are most abundant on pistillate (female) flowers (Turner et al., 1980). Three types of GTs have been described in cannabis: small bulbous trichomes, sessile trichomes resting on the surface of the epidermis, and stalked trichomes that are elevated by their stalks above the epidermis (Hammond and Mahlberg, 1973). Resin is produced in a disc of secretory cells in each GT, and then exported to an apical extracellular cavity where it is stored (Kim and Mahlberg, 1991).

Terpene biosynthesis in plants can conceptually be divided into three stages (**Figure 1.2**). First, C_5 isoprenoids isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are formed via one of two pathways: the plastidial methylerythritol phosphate (MEP) pathway from pyruvate and glyceraldehyde-3-phosphate, or the cytosolic mevalonate (MEV) pathway from acetyl-CoA (Lichtenthaler et al., 1997; Tholl, 2015). In the second stage, isoprenoid diphosphates are elongated by *trans*-prenyltransferases, which condense IPP and DMAPP into longer-chain isoprenoid diphosphates. Geranyl diphosphate (GPP), a C_{10} isoprenoid, is synthesized by plastidial GPP synthase (GPPS). GPPSs are either homo- or heterodimeric. In the case of heterodimeric GPPS systems, the ratio of large to small subunits likely controls GPPS activity (Burke and Croteau, 2002; Orlova et al., 2009). The GPPS complex in *Humulus lupulus* (hop), a close relative of cannabis, is heterodimeric. It consists of a large subunit (GPPS.lsu) with C_{20} isoprenoid diphosphate synthase activity, and a small subunit (GPPS.ssu) that complexes with the lsu and modifies its activity to produce GPP (Wang and Dixon, 2009). The C_{15} isoprenoid (*E*,*E*)-farnesyl diphosphate (FPP) is the product of cytosolic FPP synthase (FPPS).

Monoterpenes are produced from GPP, and sesquiterpenes from FPP (**Figure 1.2**). In the third stage, isoprenoid precursors are dephosphorylated and rearranged by TPS to form cyclic or acyclic terpenes.

1.7 Terpene synthase phylogeny and function

TPS are encoded by gene families in plants that vary in size and diversity of functions among species (Chen et al., 2011). They are involved in both primary and specialized metabolism. The plant TPS gene family has been classified into subfamilies, which correlate with their general biochemical functions and taxonomy (Chen et al., 2011). In vascular plants, two subfamilies TPS-c and TPS-e/f contain enzymes that produce the diterpenes of primary metabolism as well as members that have evolved functions in specialized diterpene metabolism. In angiosperms, the subfamily TPS-a generally contains sesqui-TPS, and TPS-b generally contains mono-TPS and hemi-TPS. Enzymes of the TPS-g clade accept substrates of different chain length and typically produce linear terpenes (Dudareva, 2003).

Angiosperm TPS are generally 50-100 kDa α-helical proteins. They feature two aspartate-rich motifs that coordinate divalent metal ions, specifically Mg^{2+} or Mn^{2+} (Bohlmann et al., 1998; Christianson, 2006). In the active site of a TPS, isoprenoid diphosphate substrates are dephosphorylated to form a carbocationic intermediate (**Figure 1.3**), which undergoes isomerization and rearrangement before deprotonation or water-capture to yield the final terpene product. Many TPS are multiple-product enzymes. The size of the active site cavity is a predictor of substrate chain-length preference (Gao et al., 2012). While residues that control substrate specificity, isomerization, and cyclization have been identified (Greenhagen et al., 2006; Salmon et al., 2015; Srividya et al., 2015), the specific products of a TPS cannot be predicted from its

amino acid sequence. Recent efforts have used machine learning or directed evolution to computationally predict and alter the products of sesqui-TPS (Salmon et al., 2015; Durairaj et al., 2019). Biochemical characterization remains the most reliable method to determine the products of TPS.

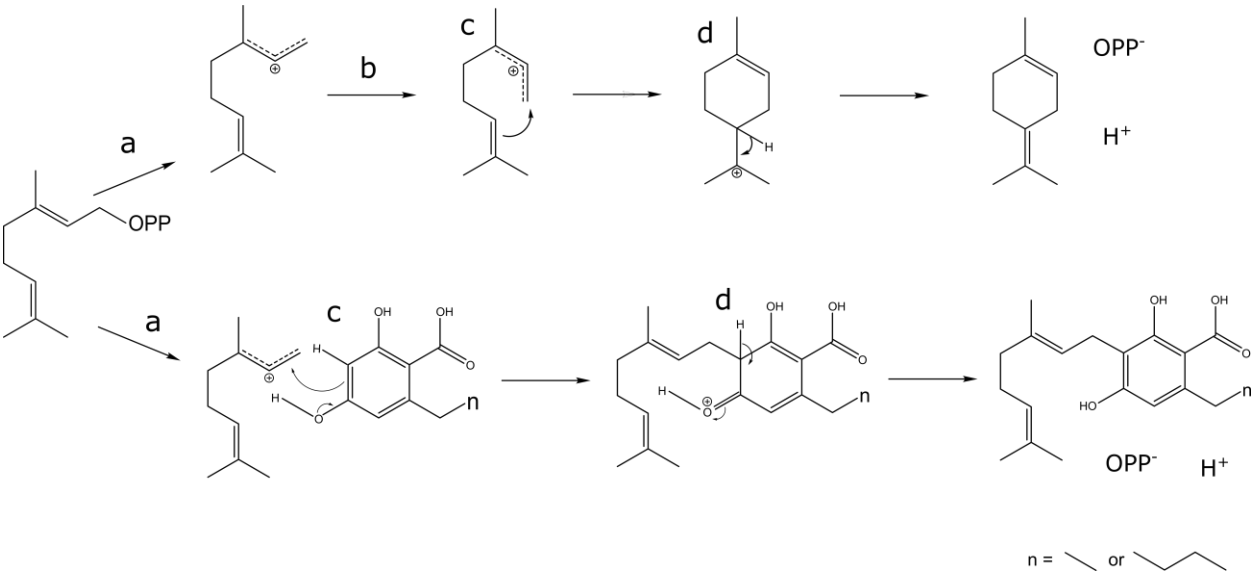


Figure 1.3 Carbocations in cannabis resin biosynthesis
Similar processes occur in terpene cyclization (top) and prenyl transfer to an aromatic substrate (bottom). Geranyl diphosphate (far left) is dephosphorylated (a) to form an allylic carbocation. In terpene biosynthesis, the carbocation is isomerized (b). The carbocation is condensed (c) internally in terpene biosynthesis, or with an aromatic substrate in prenyl transfer, leading to a tertiary carbocation. The final product is formed by deprotonation (d), with PPi and H+ as byproducts.

CsTPS responsible for many of the different terpenes found in cannabis are still unknown. Most of the characterized CsTPS are multi-product enzymes that generate several different terpene structures from either GPP or FPP. The multi-product nature of CsTPS can explain why different terpenes, such as α-humulene and β-caryophyllene, commonly co-occur in the terpene profiles of different cannabis samples.

Variation of the CsTPS gene family at the level of cultivars and in CsTPS gene expression is likely to explain observed variations of terpene profiles across the species.

However, the level of variation of the size, composition and expression of the CsTPS gene family, and factors that influence CsTPS gene expression, are mostly unknown. In other plant species, variation of terpene biosynthesis at the genome, transcriptome, proteome and biochemical levels accounts for phenotypic intra-specific variation of terpene profiles (eg. Drew et al., 2015; Hall et al., 2011). Terpene profiles may also substantially change as a result of differential CsTPS gene expression over the course of plant development or in response to environmental factors. In addition, developmental or tissue-specific expression of CsTPS may affect variation of terpene profiles in cannabis products. None of these factors of terpene variation, which may affect reproducibility of terpene composition, have been systematically studied in cannabis.

Variation of the composition of the *TPS* gene family, or variation of *TPS* gene expression, within a given plant species has been linked to variation of terpene profiles in a number of different systems. This includes both cultivated and non-cultivated plants, as well as angiosperms and gymnosperms. For example, in grapevine (*Vitis vinifera*), members of a large *VvTPS* gene family are differentially expressed between tissues, developmental stages, and cultivars, leading to differences in terpene profiles depending on the specific combination of *TPS* genes that are expressed during flowering and fruit ripening (Martin et al., 2009; Martin et al., 2010; Drew et al., 2015; Smit et al., 2019). In rice (*Oryza*), lineage-specific blooms of similar *TPS* genes contributed to variation of terpene defenses between different rice species (Chen et al., 2020). Similarly, in corn (*Zea mays*), variation of expression of *ZmTPS* encoding β-caryophyllene synthase is central to the variation of terpene-mediated indirect defense against corn borer (Köllner et al., 2008). In Sitka spruce (*Picea sitchensis*), a gymnosperm, copy number variation and variation of expression of *SsTPS* genes encoding (+)-3-carene synthase caused

differences in monoterpene composition associated with insect resistance (Hall et al., 2011; Roach et al., 2014).

The oxygen functionality of simple terpene alcohols found in cannabis such as linalool or bisabolol may result from two major reactions. The direct enzymatic activity of CsTPS can add a hydroxyl group to simple terpene skeletons when a water molecule is trapped in the active site cavity during catalysis (Tholl et al., 2004; Booth et al., 2017; Zager et al., 2019). Cytochrome P450s oxidise terpenes to alcohols or ketones in other plants species, resulting in compounds such as diterpene resin acids, aromatic santalols, menthol, and artemisinin (Lange et al., 2000; Teoh et al., 2006; Zerbe et al., 2012; Celedon et al., 2016), a process that could also be active in cannabis resin biosynthesis. Other terpene derivatives detected in cannabis may arise non-enzymatically due to oxidation or due to thermally or UV-induced rearrangements during processing or storage, such as caryophyllene oxide, β-elemene, or derivatives of myrcene (Marchini et al., 2014; Booth et al., 2017; Zager et al., 2019). These non-enzymatic modifications may add a level of variation that is independent of the plant genome and plant biochemistry. When terpene analysis is performed with dried plant material, variable quantitative losses of terpenes, especially the more volatile monoterpenes (Ross and ElSohly, 1996), may be another cause of terpene variation.

1.8 Biosynthesis of cannabinoids

Compared to terpene biosynthesis, cannabinoid biosynthesis has been a priority of the limited research on metabolite biosynthesis in cannabis to date. Cannabinoids are a combination of polyketide and terpene precursors. Much of the core cannabinoid biosynthetic pathway has been characterized (Taura et al., 1996; Gagne et al., 2012; Stout et al., 2012; Luo et al., 2019).

Terpenes and cannabinoids are biochemically related through a common precursor, GPP. Cannabinoids are initially formed by the creation of a C-C bond between GPP and the 3` position of an aromatic compound of fatty acid origin, olivetolic acid (OA) (**Figure 1.3**). This reaction leads to the formation of CBGA, the branch-point intermediate of the canonical cannabinoid biosynthetic pathway. Less commonly, propyl cannabinoids can be formed by the geranylation of divinaric acid, a propyl side-chain variant of OA, leading to cannabigiveric acid (**Figure 1.3, bottom**). The portion of cannabigerolic acid (CBGA) that is of isoprenoid origin is cyclized by cannabinoid synthases, leading to the formation of further cannabinoids including THCA, CBDA, and cannabichromenic acid (CBCA). Canonical cannabinoid synthases are FAD-linked oxidocyclases structurally related to Berberine bridge-like enzymes (Taura et al., 1996). Biosynthesis of CBGA is likely catalyzed by the UbiA family of aPTs, named for its involvement in ubiquinone biosynthesis in vertebrates (Mugoni et al., 2013). Two UbiA family cannabis enzymes with CBGA synthase (CBGAS) activity have been described, CsPT1 and CsPT4 (Page and Boubakir, 2012; Luo et al., 2019). aPTs are alpha-helical, integral membrane proteins that catalyze the transfer of prenyl groups to phenolic acceptors. aPTs include enzymes involved in the biosynthesis of quinones, chlorophylls, and tocopherols (reviewed in: Li, 2016). In specialized metabolism, the aPT family contributes to the formation of compounds that serve as flavours, pigments, and toxins. UbiA family aPTs have been characterized in different species including hop (*Humulus lupulus*), soybean (*Glycine max*), lemon (*Citrus limon*), and lupus (*Lupinus albus*) (Sasaki et al., 2011; Shen et al., 2012; Munakata et al., 2014; Li et al., 2015). These enzymes depend on Mg^{2+} ions to catalyze dephosphorylation of the prenyl donor (**Figure 1.3**). Crystal structures suggested that UbiA aPTs accept polar substrates from their cytosolic face, and release hydrophobic products directly into the membrane (Yazaki et al., 2009; Cheng

and Li, 2014). Similar to cannabis, hop also produces prenylated aromatic compounds in glandular trichomes on pistillate flowers. In hop, the prenylations are performed by a pair of aPTs, HlPT1 and HlPT2, aided by accessory chalcone-isomerase like (CHIL) proteins (Tsurumaru et al., 2012; Li et al., 2015b; Ban et al., 2018).

In cannabis, the aPTs that have been reported to have CBGA activity all perform reactions similar to that of the hop bitter acid prenylation complex. Both CsPT1 and CsPT4 are closely related to HlPT1 and 2 (Rea et al., 2019). The activities of CsPT1 and CsPT4 were demonstrated in yeast systems, *Pichia pastoris* and *Saccharomyces cerevisiae*, respectively (Page and Boubakir, 2012; Luo et al., 2019). A further cannabis aPT, CsPT3, catalyzes the formation of prenylated flavonoids known as cannaflavins (Rea et al., 2019) Both hop and cannabis aPTs catalyze the formation of a C-C bond between an aromatic compound and an isoprenoid moiety. Bitter acids in hop and cannabinoids in cannabis are both formed by the attachement of DMAPP or GPP to an aromatic structure of fatty acid origin (Page and Boubakir, 2012; Vickery et al., 2016; Luo et al., 2019). Similarly, hop contains prenylated naringenins that are cousins to cannflavins, the prenylated flavonoids producted by CsPT3 (Stevens and Page, 2004; Rea et al., 2019)h.

While both CsPT1 and 4 exhibit CBGAS activity, neither has been characterized in the context of cannabinoid biosynthesis *in planta*. Both CsPTs contain predicted plastidial targeting sequences, but their localization to particular organelles or tissues has not been established. The potential roles of cannabis genes homologous to the hop CHIL proteins have also not been investigated. In addition to cannabinoids, cannabis produces unique prenylated flavonoids known as cannflavins (Radwan et al., 2008). CsPT3, which catalyzes the prenylation of chrysoeriol to produce cannflavins, is closely related to CsPT1 and 4 (Rea et al., 2019). The same report also

showed that the cannabis genome encodes at least eight members of the UbiA family, associated with both primary and specialized metabolism. Most of these enzymes have yet to be characterized.

The mechanisms that facilitate transport within and export from the secretory cells to the storage cavity are not known, either in cannabis or in model species like mint and sage (Lange and Turner, 2013). Isoprenoid biosynthesis in glandular trichomes is a complex process involving different subcellular compartments, intracellular transport and eventually secretion, as has been shown for example for the biosynthesis of menthol in glandular trichomes of mint (*Mentha spicata*) (Turner and Croteau, 2004). In cannabinoid biosynthesis, the fatty acid-derived olivetolic acid (OA) is synthesized in the cytosol and ER (Gagne et al., 2012; Stout et al., 2012). The isoprenoid precursor geranyl diphosphate (GPP) is produced in plastids (Fellermeier et al., 2001), and the enzyme that produces the branch-point cannabinoid, cannabigerolic acid (CBGA), is also likely plastidial. Downstream cannabinoid cyclases that act on CBGA such as THCA synthase (THCAS) and CBDA synthase (CBDAS) are localized to the ER and possibly exported to the storage cavity, where they may be active (Taura et al., 1996a; Sirikantaramas et al., 2004; Rodziewicz et al., 2019). Cannabis terpenes are likely to be produced in plastids of glandular trichomes, specifically the monoterpenes, and in the cytosol for sesquiterpenes and are transported into the extracellular cavity by mechanisms that are not known.

1.9 Cannabis genomics

Genomics has been slow to reach cannabis, largely due to legal restrictions on funding agencies and researchers. A first reference-quality cannabis genome was published in 2019 (Laverty et al., 2019), enabling the genome-wide analysis of metabolic systems in cannabis. More genotyping

and sequencing studies are required to encompass the full diversity of the species. A special emphasis is needed on Eurasian and African landraces, which have been under-sampled. Critical tools for functional genomics of metabolic systems, and ultimately crop improvement, such as genetic transformation or genome editing, are not yet established for cannabis research in the public domain. Beyond the genes that encode enzymes for the biosynthesis of terpenes and cannabinoids in cannabis, research is needed to elucidate the factors that control expression of these biosynthetic systems. This would include, for example, the regulation of cell-type-specific gene expression in the context of the development of glandular trichomes, plant architecture, and onset of female flowering.

Specialized metabolism in trichomes has been investigated in various species including tomato (*Solanum lycopersicum*), artemisia (*Artemesia annua*), as well as mint (*Mentha*) and sage (*Salvia*) species. Glandular trichomes can be removed intact from the surface of the plant, facilitating the generation of trichome-specific transcriptomes. Trichome transcriptomes have aided biosynthetic pathway elucidation in other species (Ma et al., 2012; Zerbe et al., 2013; Ali et al., 2017), and transcriptome sequencing has been useful in identifying cannabinoid and terpene biosynthetic genes in cannabis (Marks et al., 2009; Taura et al., 2009; van Bakel et al., 2011; Gagne et al., 2012; Stout et al., 2012)(Gunnewich et al., 2007; Zager et al., 2019).

A cannabis reference genome sequence exists for the cultivar Purple Kush (Laverty et al., 2019), which improved an earlier draft version (van Bakel et al., 2011). More fragmented whole-genome assemblies have also been produced for six other cultivars, and genotyping by sequencing (GBS) data for over 300 samples is available online (Vergara et al., 2016). These resources have led to the characterization of cannabichromenic acid synthase (CBCAS), have clarified the role of copy-number variation in determining chemotype, clarified relationships

between cultivars, and will likely lead to a deeper comprehensive understanding of cannabis genomics (Sawler et al., 2015; Laverty et al., 2019; Vergara et al., 2019).

Future biochemical and functional work on biosynthetic systems in cannabis would benefit from a focused community effort to produce and archive a complete and reproducible set of metabolite and genomic data for one or a few genotypes that will serve as a reference framework. In parallel, a larger number of cannabis types need to be properly genotyped and phenotypically characterized (e.g. with regard to their metabolites) to overcome current issues with inconsistencies in what is referred to as 'strains'. The goal would be to establish reproducible cannabis varieties for use in research and in the industry, comparable to the well-defined grapevine varieties that are used in viticulture. Moving from 'strains' to varieties will require the cooperation of cannabis researchers, breeders and growers. To my knowledge, so far, no industry association has taken a lead to set community standards and practices or define community-accessible varieties. Researchers and industry in Canada, as the first developed nation to have fully legalized cannabis, are uniquely positioned to lead this effort.

1.10 Thesis objectives and significance

As restrictions on research on cannabis relax, it is likely to become a more popular research organism both for the gain of basic knowledge and for industry applications. Cannabis is a useful system for terpene research as it produces a large volume of a diverse terpene-rich resin on its trichome-covered surfaces. The abundance and size of its glandular trichomes make it a useful system for research in cell specialization and regulation of terpene and cannabinoid metabolism. When this thesis was initiated, of the dozens of terpene metabolites that had been identified in cannabis, the biosynthesis of only two had been investigated on the level of the corresponding

CsTPS genes. A draft genome and transcriptome for the model hemp cultivar ‘Finola’ was available. The goals of my thesis are:

1. To describe the distribution of isoprenoid biosynthesis in cannabis tissues, including:
    a. Describing terpene profiles in cannabis flowers and other tissues, and
    b. Describing the expression of genes related to isoprenoid and cannabinoid biosynthesis in various tissues.
2. To characterize the enzymes responsible for the formation of specialized metabolites from their isoprenoid diphosphate precursors by:
    a. Creating and mining transcriptome data for various cannabis tissues and cultivars,
    b. Identifying candidate TPS genes and characterizing their products, and
    c. Identifying candidate genes involved in cannabinoid biosynthesis, characterizing their products and expression profiles.
3. Explain terpene profiles in cannabis, and possibly variations thereof, based on new knowledge of the cannabis CsTPS gene family and CsTPS functions, by:
    a. Exploring evolutionary relationships between cannabis TPS,
    b. Correlating TPS products with terpene profiles in cannabis flowers, and
    c. Drawing connections between TPS representation in genome and transcriptome data with cannabis terpene profiles.

Chapter 2 addresses the first two thesis objectives in the cannabis cultivar ‘Finola’. I used the available genomic resources to identify candidate CsTPS and enzymes involved in isoprenoid biosynthesis. I characterized CsTPS that are expressed in ‘Finola’ floral trichomes, and used

qPCR to investigate the expression patterns of isoprenoid biosynthesis genes throughout the plant.

Chapter 3 expands objective 2 into six further cultivars. I developed transcriptome resources to identify isoprenoid biosynthetic genes and explain the genetic basis for terpene variation between the cultivars. I comprehensively characterized the CsTPS gene family and the TPS enzymes responsible for producing the major terpenes identified in the different cultivars.

Chapters 3 and 4 address the third objective. In Chapter 4, I used two TPS for targeted mutagenesis and modelling approaches to explain mechanistic differences between two similar β-ocimene synthases identified in 'Finola'.

Finally, in Chapter 5 I characterized the aPTs involved in CBGA biosynthesis. Working with researchers from Dr. Birger Lindberg Møller's group, we used a bioinformatics approach to find candidate genes for CBGAS. We then characterized an aPT that produced CBGA when expressed in *Nicotiana benthamiana* and *Saccharomyces cerevisiae*. We used this enzyme to develop a platform for the biosynthesis of novel glycosylated cannabinoids.

Chapter 2: Terpene Synthases from *Cannabis sativa*

2.1 Summary

The medicinal and psychoactive plant *Cannabis sativa* produces a terpene-rich resin in its glandular trichomes. Bouquets of different monoterpenes and sesquiterpenes are important components of cannabis resin as they define some of the unique flavors and may also influence medicinal properties of different cannabis strains. Transcriptome analysis of trichomes of the cannabis variety 'Finola' revealed sequences of all stages of terpene biosynthesis. Nine terpene synthases (TPS) were identified in subfamilies TPS-a and TPS-b. Functional characterization identified mono- and sesqui-TPS, whose products collectively comprise most of the terpenes of 'Finola' resin. Transcripts associated with terpene biosynthesis are highly expressed in trichomes compared to non-resin producing tissues. Knowledge of the TPS offers opportunities for improving terpene profiles in cannabis.

2.2 Introduction

Cannabis sativa, referred to here as cannabis, has been used for millennia as a medicine and recreational intoxicant (Li, 1973; Russo et al., 2008). The genus *Cannabis* comprises both marijuana and hemp, which are functional groups within the species *Cannabis sativa* (Gilmore et al., 2003; Sawler et al., 2015; Weiblen et al., 2015). *C. sativa* is highly valued for its pharmacologically active cannabinoids, a class of terpenophenolic metabolites unique to cannabis. These compounds are primarily found in resin produced in the glandular trichomes of pistillate (female) cannabis flowers. Cannabis resin further contains a variety of volatile monoterpenes and sesquiterpenes (**Figure 2.1**), which are responsible for much of the scent of

cannabis flowers and contribute characteristically to the unique flavor qualities of cannabis products. Similarly, terpenes in hop (*Humulus lupulus*), a close relative of cannabis, are a critical flavoring component in the brewing industry. Differences between the pharmaceutical properties of different cannabis strains have been attributed to interactions (or an 'entourage effect') between cannabinoids and terpenes (ElSohly, 2007; Russo, 2011). For example, the sesquiterpene β-caryophyllene interacts with mammalian cannabinoid receptors (Gertsch et al., 2008). As a result, medicinal compositions have been proposed incorporating blends of cannabinoids and terpenes (Wagner and Ulrich-Merzenich, 2009). Terpenes may contribute anxiolytic, antibacterial, anti-inflammatory, and sedative effects (Russo 2011).

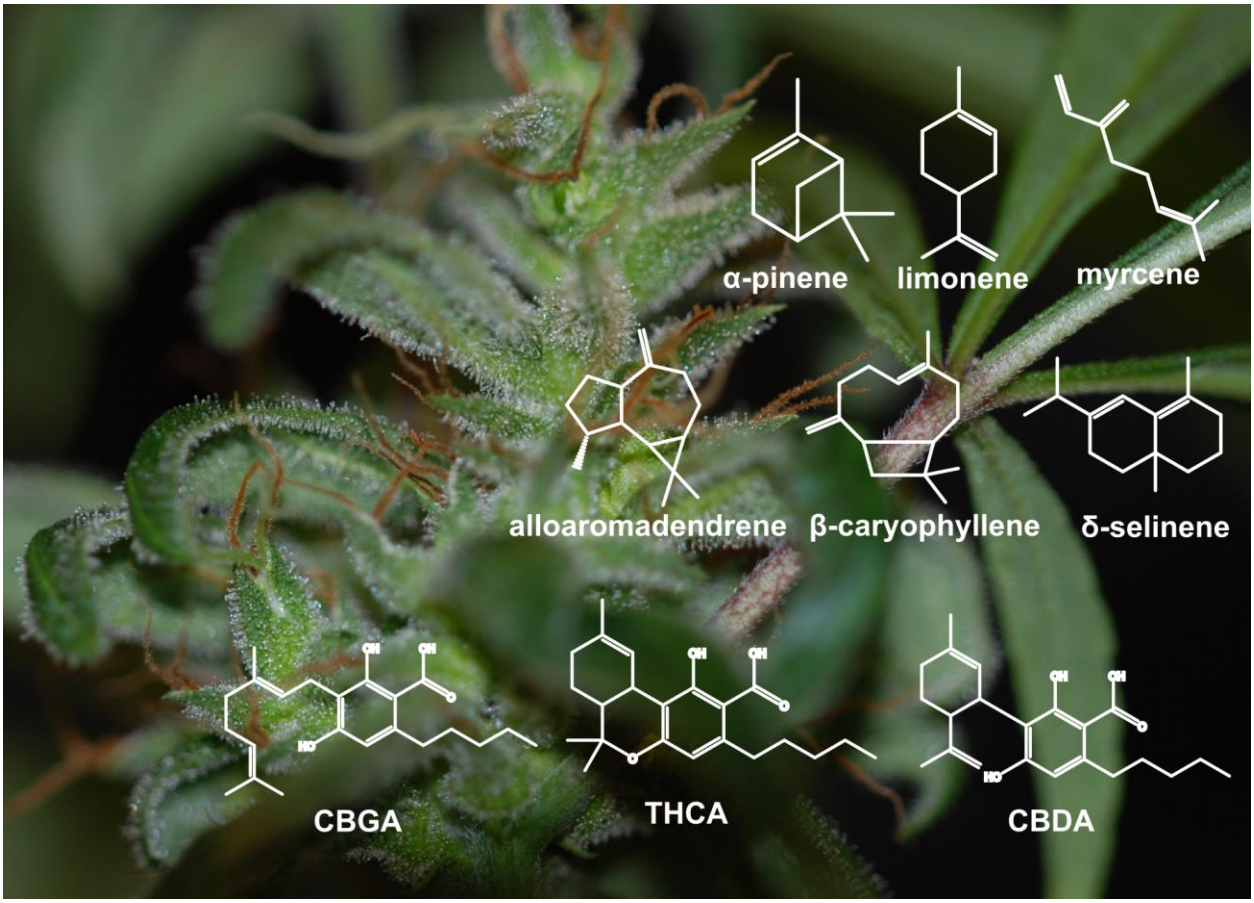


Figure 2.1 Glandular trichomes on the surface of pistillate flowers and inflorescence leaves of *Cannabis sativa* 'Finola'.
The inflorescence (left) with high density of glandular trichomes was at five weeks post onset of flowering. Non-inflorescence leaves (right) have lower density of glandular trichomes. Structures of representative cannabis resin components are shown in white: monoterpenes (top row), sesquiterpenes (middle row), and cannabinoids (bottom row).

Plants use two compartmentalized pathways to produce isoprenoid diphosphates, the precursors of all terpenes. Geranyl diphosphate (GPP), the 10-carbon precursor of monoterpenes, is generally produced via the plastidial methylerythritol phosphate (MEP) pathway. GPP is also a building block in the biosynthesis of cannabinoids (Fellermeier et al., 2001; Gagne et al., 2012). The MEP pathway is comprised of seven steps that convert pyruvate and glyceraldehyde-3-phosphate into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (**Figure**

2.2). Enzymes thought to be critical for flux regulation through this pathway include the first two and final two steps: 1-deoxy-D-xylulose 5-phosphate synthase, 1-deoxy-D-xylulose 5-phosphate reductase, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase, and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (Tholl et al., 2004; Lange et al., 2015). Farnesyl diphosphate (FPP), the 15-carbon precursor of sesquiterpenes, is usually a product of the cytosolic mevalonate (MEV) pathway. The MEV pathway converts three units of acetyl-CoA to IPP, which is then isomerized to DMAPP by IPP isomerase. A rate-limiting step in this six-step pathway is 3-hydroxy-3-methylglutaryl-CoA reductase, which produces mevalonate (Chappell et al., 1995).

IPP and DMAPP are condensed into longer-chain isoprenoid diphosphates by prenyltransferases, which include GPP synthase (GPPS) and FPP synthase (FPPS). GPPS and FPPS condense one unit of IPP and one or two units of DMAPP to form 10- and 15-carbon linear *trans*-isoprenoid diphosphates, respectively. GPPSs exist as homo- or heterodimeric enzymes. In hops, the closest known relative of cannabis, heterodimeric GPPSs can produce both GPP and the 20-carbon geranylgeranyl diphosphate (GGPP), with the ratio of large to small G(G)PPS subunits controlling the product outcome (Burke and Croteau, 2002; Orlova et al., 2009; Wang and Dixon, 2009). Linear isoprenoid diphosphates are substrates for terpene synthases (TPS), which diversify these universal precursors into thousands of different terpenes.

TPS genes are typically found in large and diverse gene families in plants (Chen et al., 2011), where they contribute to both general and specialized metabolism. The plant *TPS* gene family has been annotated with six subfamilies, which correlate to some degree with the evolution of functions of mono-, sesqui- and di-TPS. In angiosperms, the subfamily TPS-b is generally comprised of mono-TPS, which convert GPP into monoterpenes. Angiosperm TPS-a enzymes are generally sesquiterpene synthases, which form sesquiterpenes from FPP. TPS

produce cyclic and acyclic terpenes via carbocationic intermediates, formed by divalent metal co-factor dependent elimination of the diphosphate. The reactive cationic intermediate can undergo cyclization and rearrangements until the reaction is quenched by deprotonation or water-capture (Christianson, 2006). Many TPS form multiple products from a single isoprenoid diphosphate precursor.

The terpene composition of cannabis resin varies substantially based on genetic, environmental, and developmental factors (Ross and ElSohly, 1996; Hillig, 2004; Fischedick et al., 2010; Hazekamp and Fischedick, 2012). Concentrations and ratios of cannabinoids are relatively predictable for each strain, but terpene profiles are often unknown or unpredictable (Hillig 2004; Fischedick et al. 2010). To improve cannabis strains with desirable terpene profiles, it is necessary to identify genes responsible for terpene biosynthesis, which can be accomplished by harnessing cannabis transcriptome and genome resources. Draft genomes and transcriptomes for the marijuana strain Purple Kush and the hemp variety 'Finola' have previously been published (van Bakel et al., 2011). We used these resources to explore the expression of genes involved in all stages of terpene biosynthesis. We identified 11 *TPS* gene models in the 'Finola' transcriptome. *TPS* genes and gene transcripts in the MEP and MEV pathways were highly expressed in floral trichomes. We identified biochemical functions of *TPS* that are highly expressed in 'Finola'. The TPS enzymes characterized account for most of the terpenes found in 'Finola' resin.

2.3 Materials and Methods

2.3.1 Plant Materials

Seeds were obtained from Alberta Innovates Technology Futures (www.albertatechfutures.ca). All plants were grown indoors in a growth chamber under a Health Canada license. Seeds were germinated on filter paper, then transferred to 4:1 Sunshine Mix #4 (www.Sungro.com):perlite. Daylight length was 16 h under fluorescent lights, and ambient temperature 28℃. About two weeks after germination, seedlings were transferred to larger pots. After repotting, all plants were fertilized weekly with Miracle-Gro all-purpose plant food (24-8-26) (www.miraclegro.com) according to manufacturer's instructions.

2.3.2 Terpene extraction

Pistillate flowers were collected and trimmed of leaves and stems. All flowers from each individual plant were pooled. Tissue samples of ~0.2 g were weighed to determine fresh weight. Three rounds of extraction in 1 ml of pentane were performed for 1 hour each at room temperature with gentle shaking. Isobutyl benzene was added as an internal standard. After three extractions, no terpenes were identified in a fourth solvent extraction. Floral tissue was then dried overnight and weighed to determine dry weight. All three pentane extracts were combined for a total volume of 3 ml for analysis.

2.3.3 Trichome isolation

The heads of glandular trichomes were isolated from whole inflorescences as previously described (Gershenzon et al. 1992), but excluding XAD-4 and including 5 mM aurintricarboxilic

acid in the isolation buffer. In place of a cell disruptor, floral tissue was vortexed with glass beads in a Falcon tube (**Figure 2.2**).

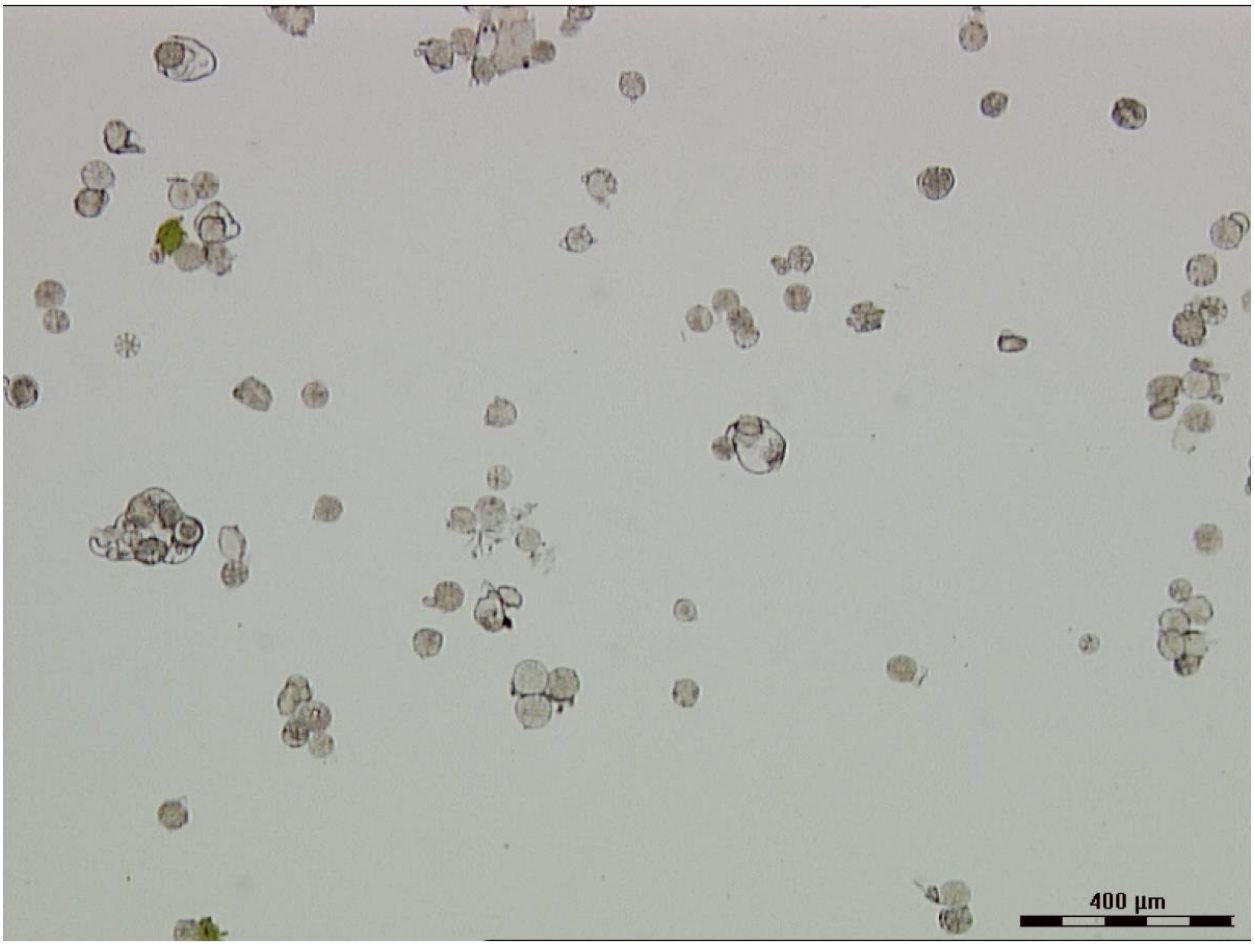


Figure 2.2 Isolated glandular trichome heads.

2.3.4 Metabolite analysis

Gas chromatography (GC) analysis of floral extracts was performed on an Agilent (www.chem.agilent.com) 7890A GC with a 7683B series autosampler and 7000A TripleQuad mass spectrometer (MS) detector at 70 eV with a flow rate of 1 ml min^{-1} He and electrospray ionization. The column was an Agilent VF-5MS or DB-5MS (30 m, 250 µm internal diameter, 0.25 µm film). The following temperature program was used: 50℃, then increase 15℃ min^{-1} to 320℃, hold for 5 minutes. Injection was pulsed splitless at 250℃. Compounds were identified by

comparison of retention index and mass spectra to authentic standards. Standards were available for all monoterpenes and the following sesquiterpenes: β-caryophyllene, α-humulene, farnesol, valencene, germacrene D, and alloaromadendrene. Tentative identifications for all other sesquiterpenes were made by comparison of retention index and mass spectra to National Institute of Standards and Technology (NIST) MS library. Identifications of bergamotene, δ-selinene, and farnesene were strengthened by comparison to essential oils of *Citrus bergamia* (Bergamot) and *Pimenta racemose* (Bay) (www.lgbotanicals.com). TPS assay products were analyzed by the same procedure described above for plant extracts, but with the following temperature program: 50℃ for 3 minutes, then increase 15℃ min^{-1} to 280℃, hold for 2 minutes. Assay products were analyzed using Agilent HP-5 and DB-Wax columns (30 m length, 250 μm internal diameter, 0.25 μm film). For cold injection of sesqui-TPS assay products, the following program was used on a DB-Wax column: 40℃ for 3 minutes, then increase 10℃ min^{-1} to 230℃, hold for 7 minutes. Injection was at 40℃ with a 1:1 split ratio.

2.3.5 cDNA cloning and characterization of TPS genes

Total RNA was isolated from 'Finola' flowers, leaves, stem, and roots using Invitrogen PureLink Plant RNA reagent (www.thermofisher.com). RNA quality and concentration was measured with a Bioanalyzer 2100 RNA Nanochip assay (www.agilent.ca). cDNA was synthesized with the Superscript III reverse transcriptase kit (Thermo Fisher). Full length and N-terminally truncated cDNAs without transit peptides where applicable were amplified from cDNA using gene-specific primers (**Table 2.1)** designed from published transcriptomic data (van Bakel et al. 2011). N-terminal transit peptides were predicted based on sequence alignments (Bohlmann et al. 1998) and using the TargetP and ChloroP servers (Emanuelsson et al. 2000). PCR amplified 'Finola' cDNAs were ligated into pJET vector (www.clontech.com) for sequence verification, and

subcloned into expression vectors pET28b+ (www.endmillipore.ca) or pASK-IBA37 (www.iba-lifesciences.org) in the case of CsTPS5FN.

Gene ID	F primer	R primer
CsTPS2FN	ATGCATTGCATGGCT	TTATAAAGGAATAGGGTTAATAAT
CsTPS3FN	TGTAGTTTGGCCAAAAGCC	TTATTTAGGAATATTAATTGGAGTAAT
CsTPS4FN	GGTGTATTTTTTAGACCAAAATT	TTATGTATATAGGGGAATAGGTTC
CsTPS5FN	ATGTCACTATCAGGACTAATCTCCACT	TCAAATGGGAATGGAAGTGAAGA
CsTPS6FN	ATGTCCACTCAAATCTTAGC	TTATGGAATTGGATCAATGA
CsTPS7FN	ATGTCTAGTCAAGTGTTAGCTTC	CTATAATGGGATGGGATCTA
CsTPS8FN	TCATCTCAATTAAGTGACAAAA	TTAATATGGGATTGGATCTATAAG
CsTPS9FN	ATGTCATATCAAGTTTTAGCCTCAT	TCATGGGATTTGATCTATAAGTAAC

Table 2.1 Primers used to clone *TPS* genes

High-confidence full-length *TPS* cDNA candidates from Purple Kush (CsTPS13PK, CsTPS30PK, and CsTPS33PK) were synthesized by GenScript (www.genscript.com) into pET28b+. For this purpose, putative *TPS* sequences from Purple Kush transcriptome data were verified by comparison to genomic sequences (van Bakel et al. 2011).

Plasmids were transformed into *E. coli* strain BL21DE3-C43 for heterologous protein expression, as previously described (Roach et al., 2014). Recombinant protein was purified using the GE healthcare His SpinTrap kit (www.gelifesciences.com) according to manufacturer's instructions. Binding buffer for purification was 20 mM HEPES (pH 7.5), 500 mM NaCl, 25 mM imidazole, and 5% glycerol. Cells were lysed in binding buffer supplemented with Roche complete protease inhibitor tablets (lifescience.roche.com) and 0.1 mg ml^{-1} lysozyme. Elution buffer was 20 mM HEPES (pH 7.5), 500 mM NaCl, 350 mM imidazole, and 5% glycerol. Purified protein was desalted through Sephadex into TPS assay buffer. *In vitro* assays were performed in 500 µl volume by incubating purified protein with isoprenoid diphosphate substrates (Sigma) as previously described (O'Maille et al., 2004), except that the TPS assay buffer was 25 mM HEPES (pH 7.3), 100 mM KCl, 10 mM $MgCl_2$, 5% glycerol, and 5 mM DTT.

Isoprenoid diphosphate substrates were dissolved in 50% methanol and added to the assay at final concentrations of 32 µM (GPP) and 26 µM (FPP). Enzyme concentrations were variable ranging from 20 to 100 µg per 500 µl assay volume. Assays were overlaid with 400 µl hexane or pentane, with 2.5 µM isobutyl benzene as internal standard.

2.3.6 *Nicotiana benthamiana* transformation and transient expression

The CsTPS5FN coding sequence was inserted into the Golden Gate plant expression vector pEAQ-GG, which contains a CaMV 35S promoter. This construct and the suppressor-of-silencing gene p19 were transformed into *Agrabacterium tumefasciens* strain AGL1. For infiltration, *A. tumefasciens* was grown overnight as previously described (Sparkes et al., 2006), then pelleted and resuspended in 10 mM 2-(N-morpholine)-ethanesulphonic acid (MES) buffer, pH 5.8, 10 mM $MgCl_2$, 20 µM acetosyringone to OD_{600} 0.5. Equal volumes of bacteria, 25 ml each, containing TPS5 and p19 were infiltrated into the abaxial side of 4-week-old *N. benthamiana* plants. Infiltrated plants were grown for three days in the dark. Infiltrated leaves were harvested and ground in TPS assay buffer, and enzyme activity assays were conducted as above.

2.3.7 RT-qPCR analysis of transcript abundance

cDNA for qPCR was synthesized using the Maxima First Strand cDNA synthesis kit (Thermo Fisher) according to manufacturer's instructions. qPCR reactions were done in 15 µl volumes with SsoFast EvaGreen supermix (Bio-Rad), 4 µl template (2 ng), and 0.3 µM primers. Primers (**Table 2.2**) were designed using Primer3 software (Untergasser et al., 2012). Reference genes were chosen by geNorm (Vandesompele et al., 2002), analyzed with qBase+ software

(www.biogazelle.com). Reference genes used for RT-qPCR of early isoprenoid biosynthesis across different plant organs were *actin* and *CDK3*. For RT-qPCR of *TPS* transcripts in trichomes, reference genes were *CDK3* and *GAPDH*. RT-qPCR analyses were done with four biological and two technical replicates for the early isoprenoid biosynthetic transcripts in different organs. For *TPS* transcript analysis in trichomes, three biological and three technical replicates were performed. Gene expression was analyzed using qBase+. Statistical analysis was performed by ANOVA on log-transcript abundance, with Bonferroni correction.

Gene ID	Probe 1	Probe 2
CDK3	CCTTTTCTGCAGGGATCTAGTG	TCCACGTAATCAGGCAGGAA
GAPDH	GCACCTATGTTCGTGGTTGG	ACGGTCTTCTGTGTTGCTGT
Actin	AATGGTCAAGGCTGGGTTTG	TCCATGTCATCCCAGTTGCT
CsTPS1FN	ATGGTTGTCTGTGGGAGGAC	AGGAACATCGCCTCTTTTCA
CsTPS2FN	AGAACGCCTCGCTTTCAATA	AGGAACATCGCCTCTTTTCA
CsTPS3FN	CTTCCGCTTCATTGGAGAAT	TACAATCCCTCCACCACCTC
CsTPS5FN	TTGGTAACGACAATGGACGA	TAGCACGTCATACGCCATTT
CsTPS6FN	AGCACCGGTTCTAATTGTGC	AATTCCTCCGATGATGTTGC
CsTPS9FN	ACTATTGCGCCAACATGGAT	TGAGTGGTGGTGAAAGCAAG
CsTDXS1	TATCATGGGGTGGCCAAGTT	ACATCTTGTTGGGAAACGGC
CsDXS2	TGCTTTGATGTGGGGATTGC	TCCAATGCAAAACGGACAGG
CsDXR	ACTGGACATTGTGGCAGAGA	AGGCCTCTTTCAGTTCACCA
CsHDR	ACGGCAGCAGATTCTGATTT	CATATGCTTCGGCCAATTTT
CsGPPS.ssu	AGTTCACGAGGCCATGTACA	AAGGGAGGTGGTCATGAGTG
CsHMGR1	GGCCAAGATCCAGCACAAAA	TGCTAACACAGAACCGGCTA
CsHMGR2	ATCCTGCTCAGAACGTGGAA	AACATGCTGACTGTGATGCC
CsIDI	AAGCCCTTGTCTCCTCTCAC	ACTGAAAGCCCTGTGTAGCA
CsFPPS1	GAACTATGGCAAGGCAGACC	ACTTCAACACTGCTTGCACA
CsFPPS2	GTTACACGTCGCGGTCAAC	TGTCCTGAGGCAGTTTGGAA

Table 2.2 RT-qPCR primers

2.3.8 TPS gene prediction and phylogeny

'Finola' genome and transcriptome assemblies (van Bakel et al., 2011) were downloaded from the cannabis genome browser (http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway). These assemblies were used as the subject of a tblastn search using 72 *TPS* genes (**Table 2.3**) downloaded from GenBank and Phytozome. Gene and splice site prediction was performed on scaffolds containing regions with similarity to *TPS* sequences using the Exonerate gene prediction algorithm (Curwen et al., 2004). A preliminary Purple Kush genome assembly based on PacBio (www.pacb.com) sequencing data was also used. Predicted genes were manually curated against earlier Purple Kush sequence data, and examined to establish open reading frames, start codons, and stop codons. A phylogeny was built using phylogeny.fr (Dereeper et al., 2008). The alignment used for input was built using the MUSCLE algorithm with all translated amino acid sequences from the predicted *TPS* gene models from cannabis and the 79 published *TPS* sequences listed above.

ID	Accession
PpCPS/KS	BAF61135.1
Osent-CPS	AY602991.1
CmCPS	AAD04292.1
CmKS	AAB39482.1
OsKS1	AAQ72560.1
AtGeranyllinaloolS	Q93YV0.1
CbLinalool synthase	AAC49395.1
AgDelta-SelineneS	AAC05727.1
AgLimoneneS	AAB70907.1
PaLinaloolS	AAS47693.1
AgTerpinoleneS	Q9M7D0.1
AgAlpha-pineneS	O24475.1
AgBeta-phellandreneS	Q9M7D1.1
AgMyrceneS	AAB71084.1
FaNerolidolS	APB87285.1

AmMyrceneS	AAO41727.1
AmOcimeneS	AAO42614.1
Mp(E)-B-farneseneS	AAB95209.1
RcCasbeneS	EEF48743.1
CiGermacrene AS	AAM21658.1
StVetispiradieneS	AAD02223.1
Nt5-EAS	Q40577.3
LeGermacrene CS	NP_001234055.1
GaDelta-cadineneS	NP_001316949.1
VvValenceneS	ACO36239.1
VvAlpha-terpeneoleS	AAS79352.1
VvAlpha-PhellandreneS	NP_001268167.1
VvTPS09	XP_002275273.1
VvMyrceneS	NP_001268009.1
VvAlpha-bergamoteneS	ADR74195.2
VvGeraniolS	ADR74218.1
VvE-B-CaryophylleneS	ADR74193.1
VvEnt-Kaur-16-eneS	XP_010645103.1
HlSTS2	ACI32640.1
HlSTS1	ACI32639.1
Sl(E)-B-ocimeneS	NP_001308094.1
SlCamphene S	G1JUH1.1
So1,8CineoleS	AAC26016.1
SaBornylPPS	AAC26017.1
Ss3-careneS	AAM89254.1
MsLimoneneS	AGN90914.1
PfLinaloolS	ACN42013.2
PfMyrceneS	AAF76186.1
SoSabineneS	AAC26018.1
ClLimoneneS	Q8L5K3.1
HlMTS1	ACI32637.1
HlMTS2	ACI32638.1
CsTPS1	A7IZZ1.1
CsTPS2	A7IZZ2.1
ZzMyrceneS	XP_015881067
MnMyrceneS	XP_010088018
RcTPS10	NP_001310631
RoPineneS	ABP01684
LlLinaloolS	ABD77417
TcGamma-terpinene	AGK88252

OvTPS2	ADK73620
PtTPS13	AII32477
CsTPS10	XP_011650166
PsLimoneneS	ABA86248
BnGES-like	XP_013718286
BrGES-like	XP_009113235
SmCPS1	TPS9_SELML
PgKS	ACY25275
PsLinaloolS	ADZ45501
NaViridifloreneS	OIT23701
TcDelta-cadineneS	EOY12645
GhDelta-cadineneS	XP_016705995
VvValenceneS-L	NP_001268028
RcTPS2	XP_002523635
ZjTPS10	XP_015895399
PtTPS	XP_006370878

Table 2.3 Accession numbers of TPS sequences used in tblastn and to construct phylogeny

2.4 Results

2.4.1 Terpene profiles of cannabis inflorescences

We used the *C. sativa* oilseed hemp variety 'Finola' to investigate terpene profiles of pistillate flowers. 'Finola' was chosen because reference draft genome and transcriptome assemblies have been published for this variety (van Bakel et al. 2011). Pistillate flowers, which have the highest density of glandular trichomes relative to other parts of the plant (**Figure 2.1**), were sampled to cover early to mid-stage inflorescences between three and eight weeks post onset of flowering, where onset of flowering is defined as the first appearance of pistils. Independent of the stage of inflorescence, the most abundant monoterpenes were β-myrcene, (+)-α-pinene, (–)-limonene, (+)-β-pinene, terpinolene, and (*E*)-β-ocimene (**Table 2.1**). The most abundant sesquiterpenes were β-caryophyllene, α-humulene, bergamotene, and farnesene. Terpene profiles showed considerable variations between individual plants as indicated with the relatively high standard deviation (**Table 2.1**). No trends were observed for individual metabolites as a function of inflorescence development, but total monoterpenes increased compared to sesquiterpenes as inflorescences matured. Mid-stage flowers (~4 weeks post onset of flowering) had a mean monoterpene content of 389 μg g^1 DW (SE = 44, n = 9), and a mean sesquiterpene content of 34 μg g^1 DW (SE = 6.3, n = 9).

Metabolite	Percent Proportion (mean ± st. dev)
(+)-α-Pinene	47 ± 18
(+)-β-Pinene	14 ± 5.1
Myrcene	37 ± 5.1
(-)-Limonene	16 ± 15
(*E*)-β-Ocimene	5.5 ± 3.4
Terpinolene	13 ± 12
β-Caryophyllene	46 ± 13
Bergamotene	3.6 ± 3.0
Farnesene	4.4 ± 3.6
α-Humulene	19 ± 7.6

Table 2.4 Relative composition of terpene profiles in C. sativa 'Finola' pistillate flowers.
22 individuals were sampled. Contribution of individual terpenes is expressed as a proportion of the total terpenes within a given class (i.e., monoterpenes or sesquiterpenes).

2.4.2 Transcriptome mining of early isoprenoid biosynthesis genes

We queried the 'Finola' transcriptome for transcripts involved in the early stages of isoprenoid biosynthesis. We combined four transcriptome sets downloaded from the Cannabis Genome Browser (http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway), including transcripts from developing seeds, mature pistillate flowers, stamenate (male) flowers, and whole seedlings. The tBLASTn algorithm was used to search translated 'Finola' nucleotide sequences, using amino acid sequences from *Vitis vinifera* and *Arabidopsis thaliana*, and an e-value cut-off of 1^{-10}.

At least one full-length or nearly full-length (>95%) transcript was found for each of the core genes in the MEP and MEV pathways, and linear isoprenoid diphosphate prenyltransferases (**Figure 2.4**). The genes included in the analysis of the MEP pathway were 1-deoxy-D-xylulose 6-phosphate (DOXP) synthase (DXS), DOXP reductoisomerase (DXR), 2-C-methyl-D-erythritol cytidyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 4-hydroxy-3-methyl-but-2-enyl diphosphate (HMB-PP) synthase (HDS), and HMB-PP reductase

(HDR). Two versions of DXS, CsDXS1 and CsDXS2, were found, which are 62.8% identical at the amino acid level. In a phylogeny, CsDXS1 clusters with members of the DXS subfamily (**Figure 2.3**).

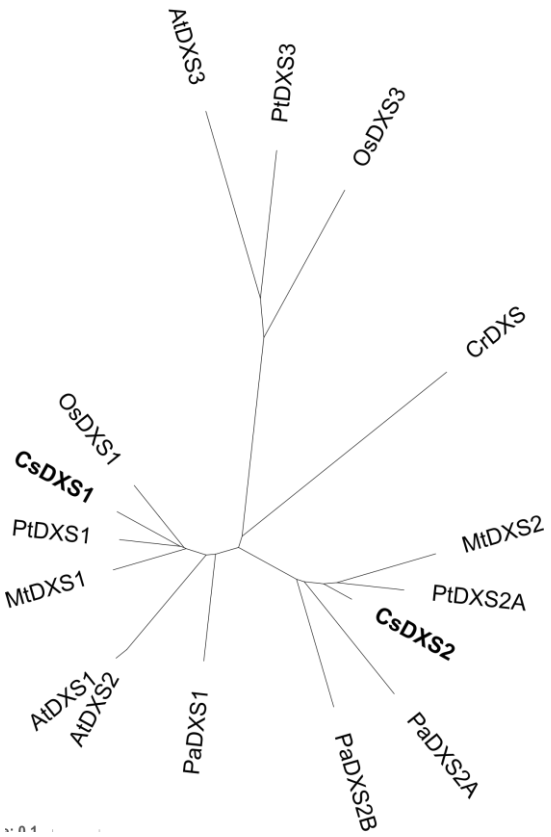


Figure 2.3 DXS phylogeny
Neighbour joining phylogeny of DXS enzymes. *Cannabis sativa* genes are in bold. DXS of other species included are from: At: *Arabidopsis thaliana*; Pt: *Populus trichocarpa*; Os: *Oryza sativa*; Cr: *Chlamydomonas reinhardtii*; Mt: *Medicago truncatula*; Pa: *Picea abies*

The genes included in the MEV pathway analysis were 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase (HMGS), HMG-CoA reductase (HMGR), mevalonate kinase (MK), phospho-mevalonate kinase (PMK), mevalonate-5-phosphate decarboxylase (MPDC), and IPP isomerase (IDI). At least one transcript was found corresponding to each enzyme. Two

transcripts were found for HMGR, HMGR1 and HMGR2, which are 72.7% identical at the amino acid level.

As candidate prenyltransferases for producing the mono- and sesquiterpene precursors GPP and FPP synthase, we found transcripts of a heterodimeric GPPS system similar to that characterized in hop (Wang & Dixon 2009), with a GPPS large subunit (GPPS.lsu) and a GPPS small subunit (GPPS.ssu). Two transcripts were identified corresponding to FPPS, 80.3% identitical to one another at the amino acid level.

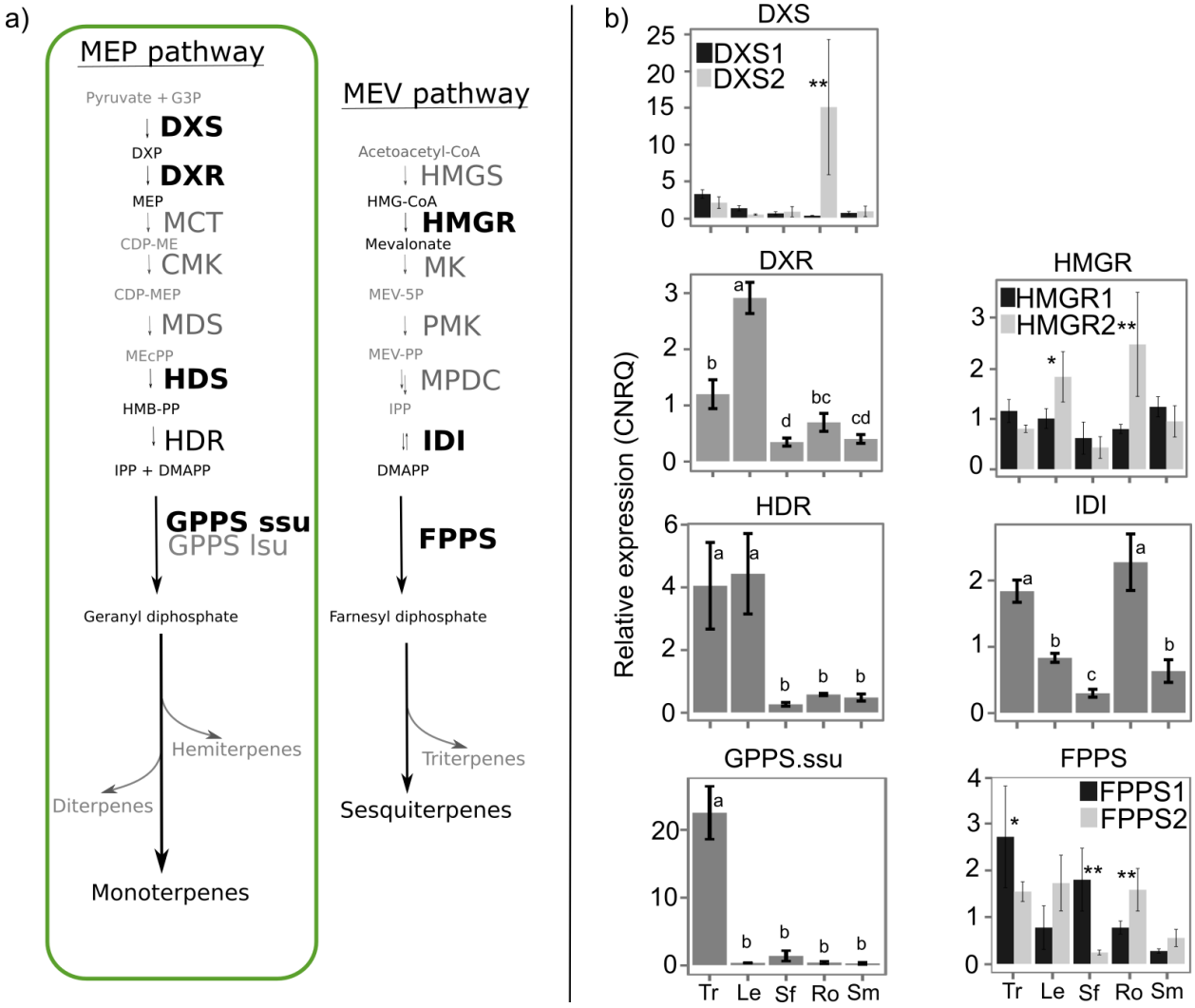


Figure 2.4 Schematic of the plastidial methylerythritol phosphate pathway (MEP) and mevalonic acid pathway (MEV) and analysis of transcript abundance in different parts of cannabis.
Steps shown in bold (a) were included in the qPCR analysis (b) of relative abundance of transcripts. Asterisks indicate significantly different mean abundance within a given part of the plant, two-sided t-test. **p<0.01, *p<0.05, with Bonferroni correction. Letters indicate significantly different means between tissues (tested within each gene), Fisher's LSD (alpha = 0.05).

2.4.3 Members of the cannabis TPS gene family

We identified nine full-length or nearly full-length (>95% of amino acid length) and six partial putative *TPS* genes (*CsTPS FN*) in the 'Finola' (FN) trichome transcriptome. A maximum likelihood phylogeny of the nine full-length *CsTPS FN* translated transcripts and representative *TPS* from other plant species placed the *CsTPS FN* most closely with each other and with *TPS*

from hop (*HlSTS1* and *HlSTS2*) indicating a recent expansion of *TPS* genes in the *Cannabaceae* (**Figure 2.5**). Five of the nine *CsTPS FN* (*CsTPS1FN*, *CsTPS2FN*, *CsTPS3FN*, *CsTPS5FN*, *CsTPS6FN*) clustered with members of the TPS-b subfamily, and the remaining four (*CsTPS4FN*, *CsTPS7FN*, *CsTPS8FN*, *CsTPS9FN*) clustered with the TPS-a subfamily. Two of the *CsTPS FN* TPS-b genes, *CsTPS1FN* and *CsTPS2FN*, encode predicted proteins that were 98.7% and 96.8% identical to CsTPS1 and CsTPS2 previously reported (Gunnewich et al. 2007) and identified there as (–)-limonene synthase (CsTPS1) and (+)-α-pinene synthase (CsTPS2) from the *C. sativa* cultivar 'Skunk'.

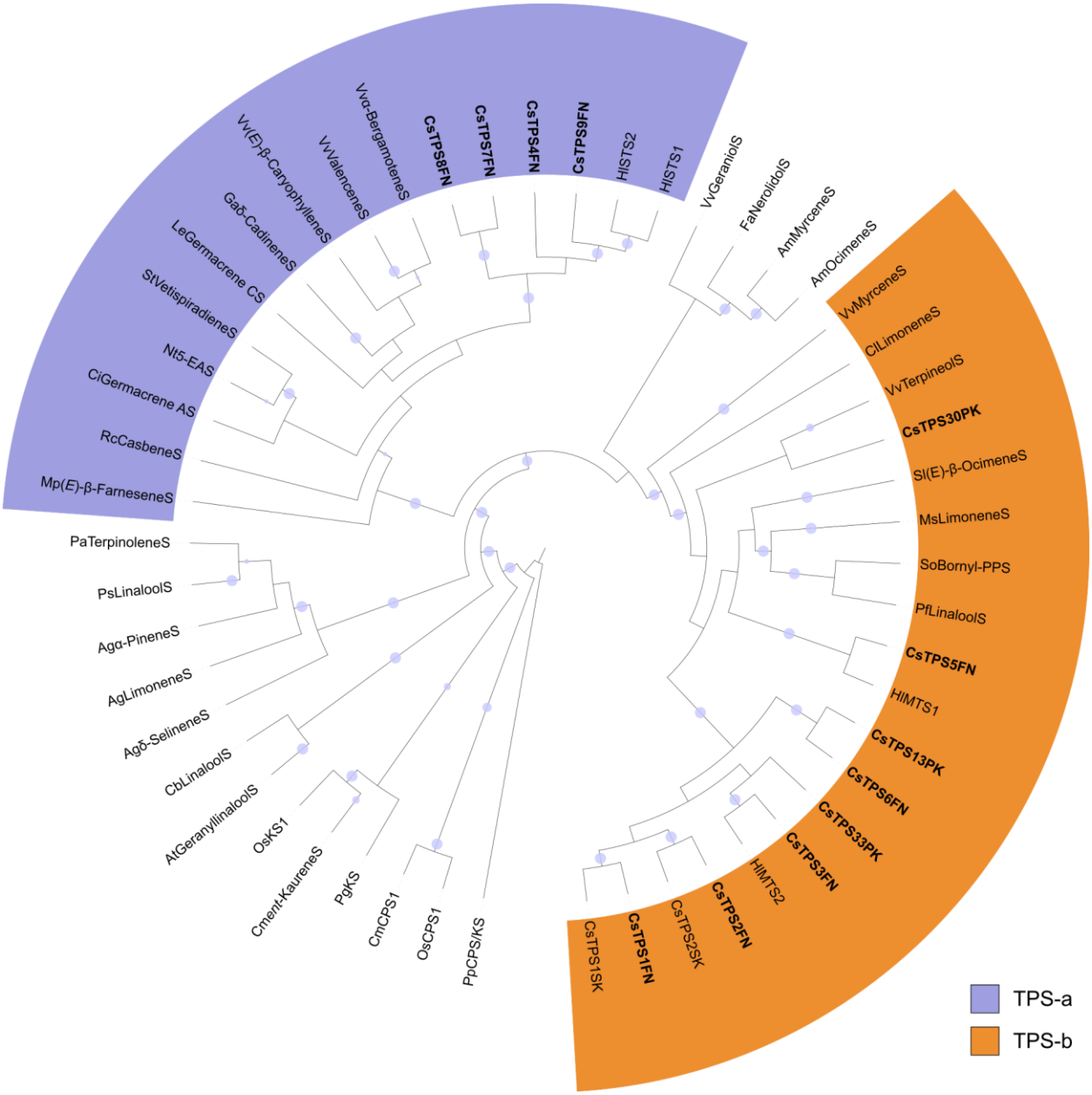


Figure 2.5 Maximum likelihood phylogeny of CsTPS.
Within the TPS-a and TPS-b subfamilies, TPS from the *Cannabaceae*, including cannabis and hops, are more closely related to one another than to TPS from other angiosperms. Cannabis TPS are in bold. The cannabis strain of origin is indicated by two letters following the TPS#: FN: 'Finola', SK: 'Skunk', PK: Purple Kush. Branches with bootstrap values >80% (100 repetitions) are indicated with a grey dot. TPS of other species included are from Pp: *Physcomitrella patens*, Os: *Oryza sativa*, Cm: *Cucurbita maxima*, At: *Arabidopsis thaliana*, Cb: *Clarkia breweri*, Ag: *Abies grandis*, Pa: *Picea abies*, Fa: *Fragaria ananassa*, Am: *Antirrhinum majus*, Mp: *Mentha x piperita*, Rc: *Ricinus communis*, Ci: *Cichorium intybus*, Sl: *Solanum lycopersicum*, Nt: *Nicotiana tabacum*, Le: *Lycopersicum esculentum*, Ga: *Gossypium arboreum*, St: *Solanum tuberosum*, Vv: *Vitis vinifera*, Hl: *Humulus lupulus*, So: *Salvia officinalis*, Cl: *Citrus limon*, Ms: *Mentha spicata*, Pf: *Perilla frutescens*. 'S' suffix = synthase.

Functional gene ID	Nearest 'PK' *TPS* gene model	Major products	Strain of origin	Activity on GPP	Activity on FPP
CsTPS1FN	CsTPS1PK	(-)-limonene	Finola	+++	-
CsTPS1SK[ŧ]	CsTPS1PK	(-)-limonene	Skunk	ND	ND
CsTPS2FN	CsTPS2PK	(+)-α-pinene	Finola	+++	-
CsTPS2SK[ŧ]	CsTPS2PK	(+)-α-pinene	Skunk	ND	ND
CsTPS3FN	CsTPS3PK	β-myrcene	Finola	+++	-
CsTPS5FN	CsTPS5PK	β-myrcene, (-)-α-pinene	Finola	+++	+
CsTPS30PK	CsTPS30PK	β-myrcene	PK	+++	+
CsTPS6FN	CsTPS6PK	(*E*)-β-ocimene	Finola	+++	-
CsTPS7FN	CsTPS7PK	δ-selinene*	Finola	+	+++
CsTPS8FN	CsTPS8PK	γ-eudesmol, valencene	Finola	+	+++
CsTPS9FN	CsTPS9PK	β-caryophyllene, α-humulene	Finola	+	+++
CsTPS13PK	CsTPS13PK	(Z)-β-ocimene	PK	+++	+
CsTPS33PK	CsTPS33PK	α-terpinene, γ-terpinene	PK	+++	-

[ŧ] Published in Gunnewich et al., 2008 *Product not compared to authentic standard. ND, not detected.

Table 2.5 Functionally characterized TPS enzymes.

2.4.4 Functional characterization of *CsTPS-FN* TPS-b subfamily members

CsTPS FN were cloned as cDNAs from 'Finola' pistillate flowers or synthesized for heterologous expression and identification of product profiles of the encoded enzymes. We cloned four TPS-b family members, *CsTPS1FN*, *CsTPS2FN*, *CsTPS5FN*, and *CsTPS6FN*, from cDNA. *CsTPS3FN* could not be cloned from cDNA and was obtained as a synthetic DNA. Three TPS-b sequences from the Purple Kush (PK) trichome transcriptome, *CsTPS13PK*, *CsTPS30PK*, and *CsTPS33PK*, were also synthesized for comparison. These five CsTPS from 'Finola' and three from Purple Kush were expressed as recombinant proteins and then tested for activity with GPP and FPP and products identified by GC-MS analysis (**Figure 2.6, Table 2.5, Figure 2.7**, and **Figure 2.8**).

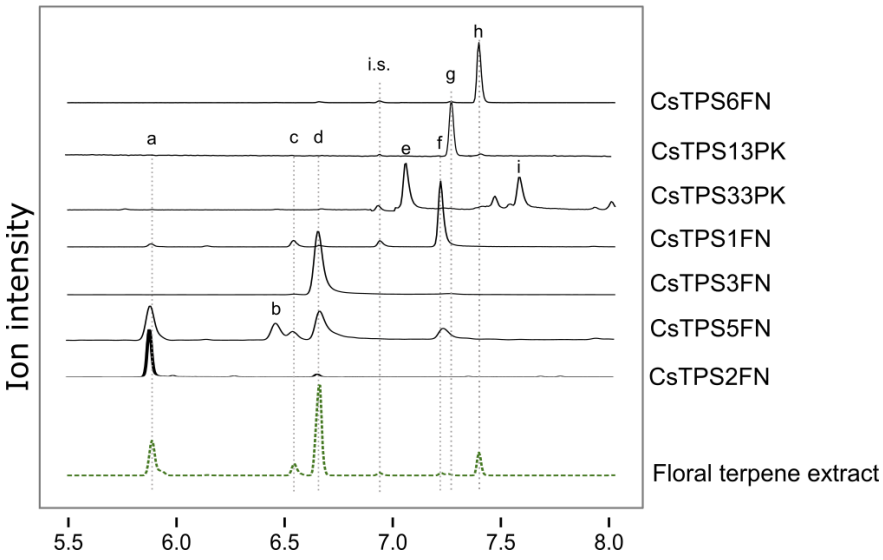


Figure 2.6 Representative GC-MS traces showing the products of CsMono-TPS.
Black traces show GC-MS total ion chromatogram (CsTPS2FN is extracted ion trace) from CsTPS assays with GPP. Green trace, dotted line, is a representative terpene profiles from `Finola' inflorescences. Peaks: a) α-pinene, b) sabinene, c) β -pinene, d) myrcene, e) α-terpinolene, f) limonene, g) (*Z*)-β-ocimene, h) (*E*)- β -ocimene, i) γ-terpinene. i.s. = internal standard.

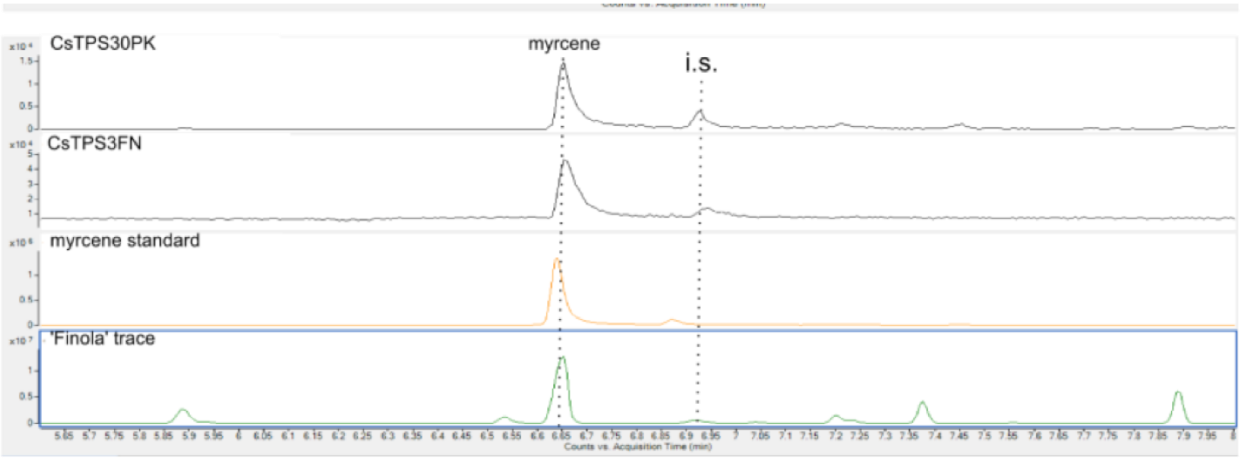


Figure 2.7 Representative GC-MS traces of myrcene synthase products

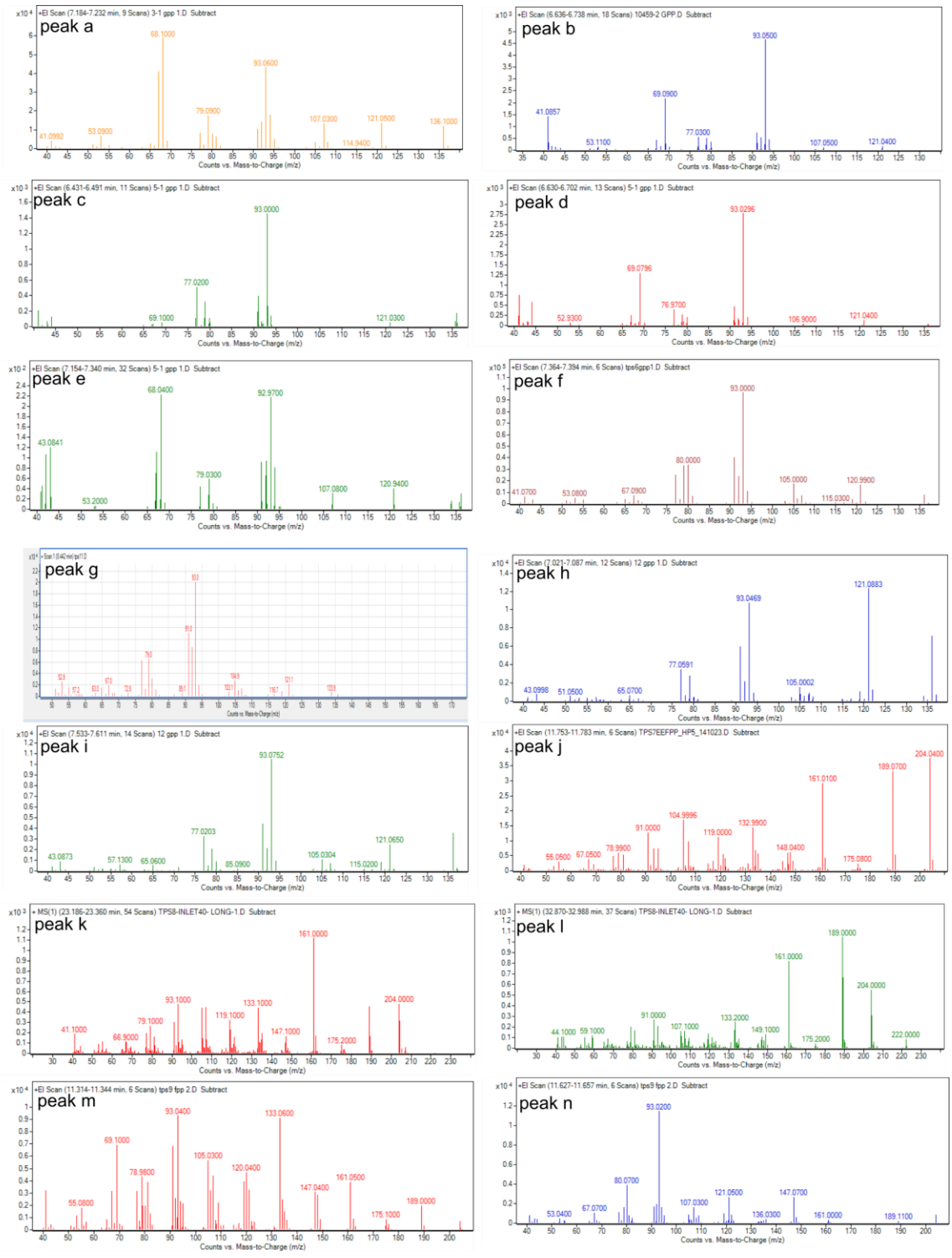


Figure 2.8 Mass spectra of TPS products

. Labels “Peak a” through “Peak n” correspond to peaks labeled in Figures 2.4 and 2.5. a) α-pinene, b) sabinene, c) B-pinene, d) myrcene, e) α-terpinolene, f) limonene, g) (*Z*)-□-ocimene, h) (*E*)-□-ocimene, i) γ-terpinene, j) β-caryophyllene, k) α-humulene, l) valencene, m) γ-eudesmol, o) δ-selinene, n) alloaromadendrene.

The major product of CsTPS1FN was (–)-limonene, with minor products of (+)-α-pinene, (+)-β-pinene, myrcene, and terpinolene. CsTPS2FN produced mostly (+)-α-pinene, with minor amounts of (+)-β-pinene, myrcene, (–)-limonene, β-phellandrene and a monoterpene tentatively identified as isoterpinolene. CsTPS3FN produced myrcene as a single detectable product when incubated with GPP. CsTPS30PK also produced only myrcene when tested with GPP (**Figure 2.7**). These two single-product myrcene synthases share only 54.5% amino acid identity. CsTPS5FN also produced myrcene as its most abundant monoterpene product (37%) (**Figure 2.6**), but unlike CsTPS3FN and CsTPS30PK, CsTPS5FN produced four additional monoterpenes (–)-α-pinene (24%), (–)-limonene (17%), (–)-β-pinene (15%), and sabinene (7%). The same product profile was identified when CsTPS5FN was transiently expressed in *N. benthamiana* (**Figure 2.9**). CsTPS5FN was somewhat unusual among TPS-b members in lacking any obvious N-terminal plastidial targeting sequence. CsTPS5FN also produced minor amounts of farnesene when incubated with FPP, making it the only member of the TPS-b subfamily to produce detectable sesquiterpenes. CsTPS6FN produced 97% (*E*)-β-ocimene with GPP, and the remaining 3% of product was (*Z*)-β-ocimene. A TPS sequence found in Purple Kush, CsTPS13PK, shares 95.5% AA identity with CsTPS6FN. CsTPS13PK produces 94% (*Z*)-β-ocimene. A third TPS from Purple Kush, CsTPS33PK, produced two different monoterpenes, α-terpinene (61%) and γ-terpinene (39%) (**Figure 2.6**).

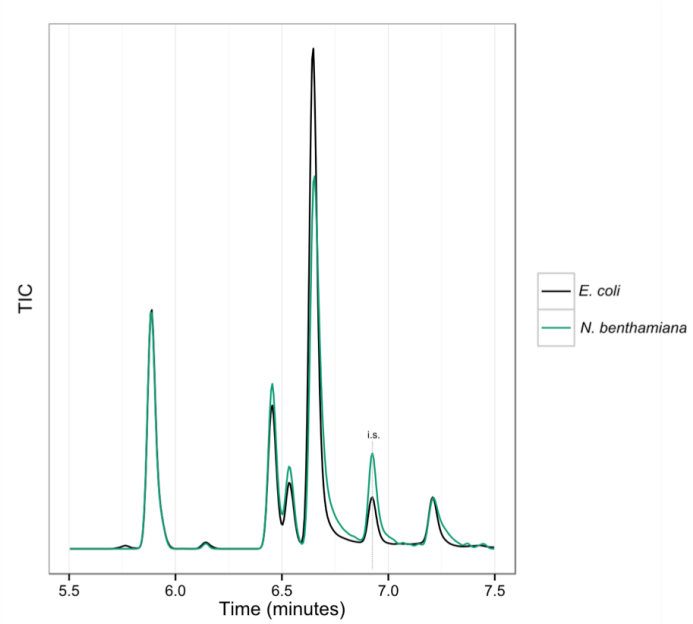


Figure 2.9 Products of CsTPS5FN expressed in *E. coli and Nicotiana benthamiana*.
Black trace represents products of recombinant enzyme expressed in *E. coli*, green trace represents products of recombinant enzyme expressed in *N. benthamiana*.

2.4.5 Functional characterization of *CsTPS-FN* TPS-a subfamily members

Four TPS-a family members cloned as cDNAs from 'Finola', *CsTPS4FN*, *CsTPS7FN*, *CsTPS8FN*, and *CsTPS9FN*, were expressed as recombinant proteins, proteins tested with GPP and FPP and products identified by GC-MS (**Figure 2.10, Table 2.5**). CsTPS4FN produced mostly alloaromadendrene (52.3% of total products) with FPP (**Figure 2.10**). The remaining products are a mixture of five sesquiterpene olefins and two alcohols, including valencene, α-humulene, and farnesol. CsTPS4FN was also active with GPP, producing minor amounts of myrcene. CsTPS7FN produced 21 sesquiterpene olefins and two sesquiterpene alcohols. Of these, products tentatively identified as δ-selinene and selina-6-en-4-ol make up 20.5% and 13.9% of the product profile, respectively. The remaining minor products each make up <10% of total sesquiterpene products. When incubated with GPP, CsTPS7FN produced low levels of sabinene, limonene, and myrcene. The most abundant product of CsTPS8FN was initially

identified as β-elemol (**Figure 2.11**), which is often an artifact of heat-induced rearrangement. Using a lower injection temperature of 40℃, the β-elemol product was no longer detected and was replaced by peaks corresponding to 11 sesquiterpene olefins and three sesquiterpene alcohols. The major products were γ-eudesmol (19.8%) and valencene (19.6%) (**Figure 2.10**).

No terpenes were detected when CsTPS8FN was incubated with GPP.

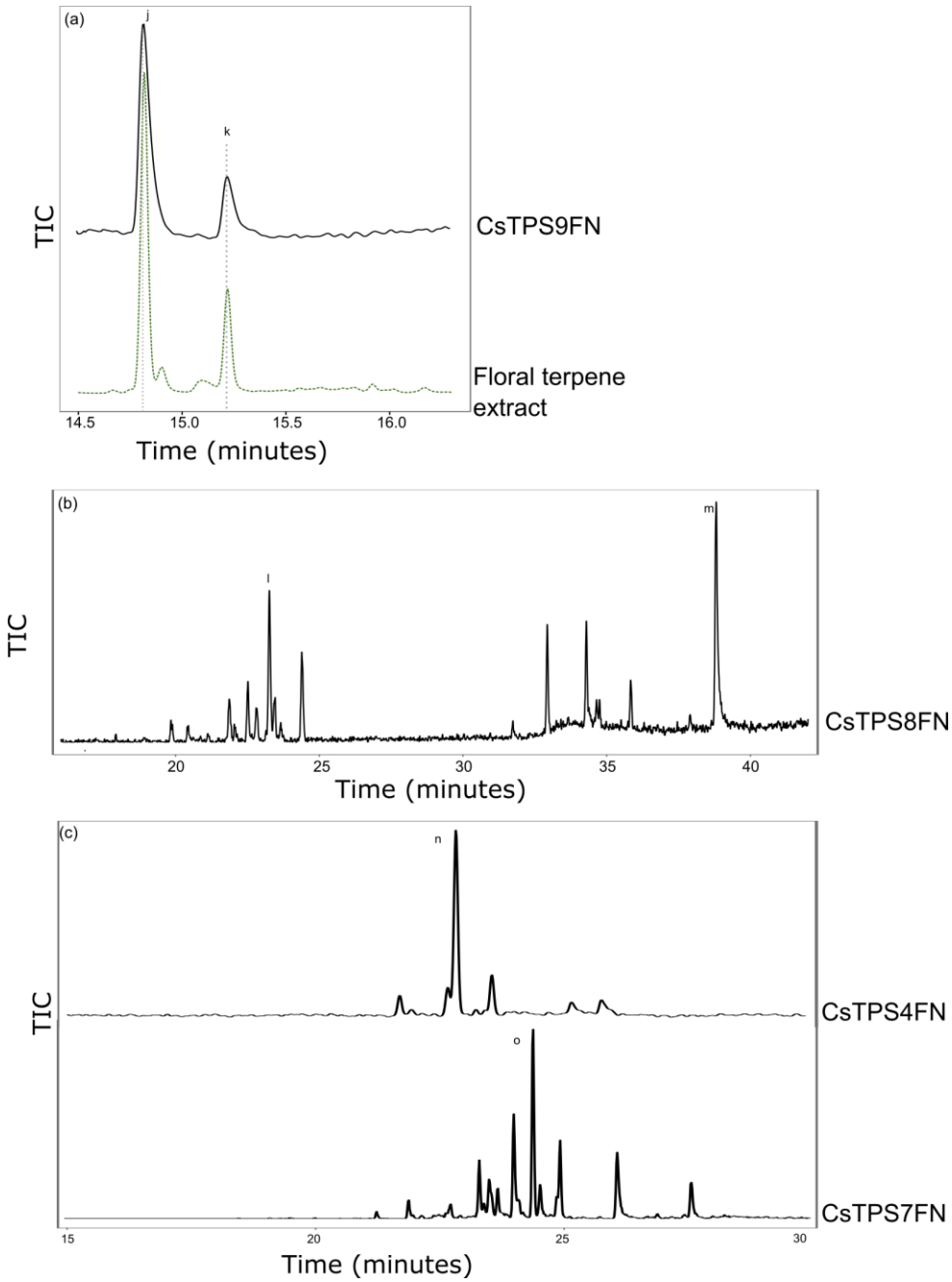


Figure 2.10 Representative GC-MS traces showing the products of CsSesqui-TPS.
Black traces show GC-MS total ion chromatogram (TIC) from CsTPS assays with FPP. Green trace, dotted line, in (a) is representative terpene profiles from `Finola' inflorescences. (b) shows the trace after cold injection (40°C inlet) onto a DB-wax column. Peaks: j) β-caryophyllene, k) α-humulene, l) valencene, m) γ-eudesmol, o) δ-selinene, n) alloaromadendrene.

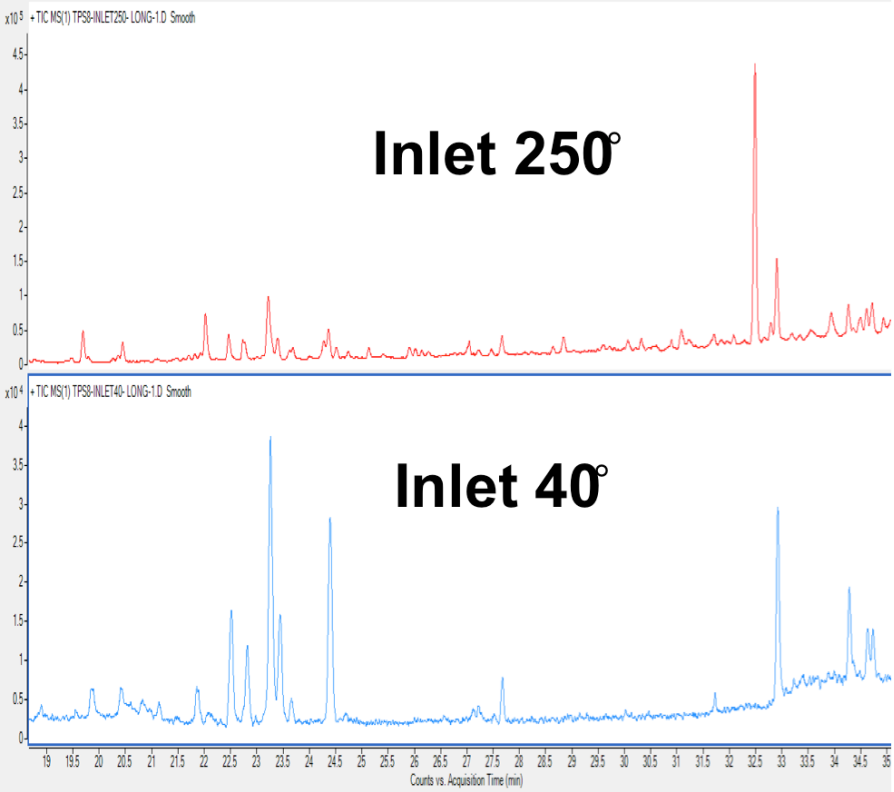


Figure 2.11 Hot vs. cold injection of CsTPS8FN products.
Top panel represents total ion chromatogram (TIC) with the injection port at 250°C on a DB-Wax column. The bottom panel represents TIC with the injection port at 40°C, using the same program and the same column.

CsTPS9FN produced β-caryophyllene and α-humulene from FPP (**Figure. 2.10**). These two terpenes are always the most abundant sesquiterpenes in cannabis resin terpene profiles. The CsTPS9FN enzyme produces these two sesquiterpenes in a ratio of approximately 2.5 to 1, which is similar to the ratio of 2.4 +/- 0.2 to 1 observed in 'Finola' terpene profiles.

2.4.6 *CsTPS* transcripts are highly abundant in pistillate inflorescences

In order to determine to what extent the Cs*TPS* genes described above contribute to the trichome terpene profile, we performed RT-qPCR on five *CsTPS* transcripts in glandular trichomes isolated from pistillate flowers. Transcript levels of *CsTPS1FN*, *CsTPS2FN*, *CsTPS3FN*, *CsTPS6FN*, and *CsTPS9FN* were examined in trichomes (**Figure 2.14**) isolated from eight

‘Finola’ individuals between two and four weeks post onset of flowering. These five *CsTPS* were chosen because they have a single product or at most two products, thus it was deemed more likely to be possible to attempt correlating metabolite abundance with transcript abundances than would be possible with the multiproduct *CsTPS*.

Of the eight individual plants, seven showed typical inflorescence terpene metabolite profiles (**Figure 2.12**). Surprisingly, one individual had no detectable inflorescence monoterpenes except for traces of (*E*)-β-ocimene, although it did contain cannabinoids and sesquiterpenes in floral trichomes. The total amount of individual terpenes was highly correlated with the total terpene content of each flower, and amounts of individual terpene were highly correlated with one another. For that reason, metabolite levels are expressed as a proportion of the total mono- or sesquiterpenes in each sample (**Figure 2.12**). *CsTPS2FN* was the most abundant of the six different *TPS* transcripts measured. Its major product, (+)-α-pinene, was the most abundant monoterpene on average in the eight plants examined. The correlation between α-pinene metabolite abundance and *CsTPS2FN* transcript level was significant (**Figure 2.12**). Transcripts of *CsTPS1FN*, *CsTPS3FN*, and *CsTPS9FN* were also abundant. However, the correlation between metabolite level and transcript abundance was not significant for any of these metabolite/transcript pairs. There was also no significant correlation between mean metabolite abundance and mean transcript abundance across all transcript/product pairs.

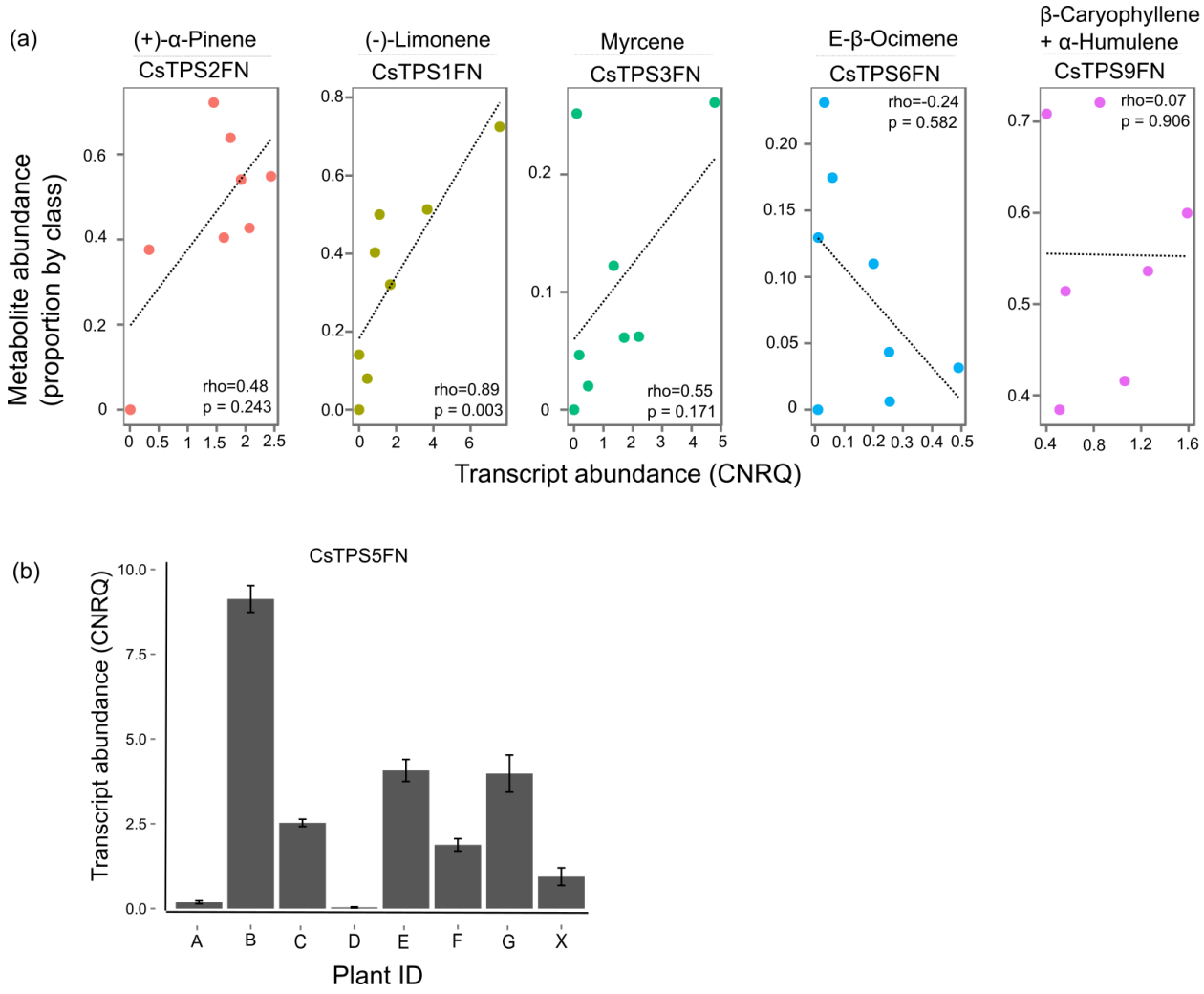


Figure 2.12 Correlation analysis of metabolite abundance in inflorescence and transcript abundance for five CsTPS in isolated trichomes.
(a) Metabolites given with their relative abundance were those that match the product of the corresponding CsTPS. Data are shown for five CsTPS/metabolite pairs each in eight 'Finola' individuals. Metabolite abundances are expressed as a proportion of the total mono- or sesquiterpenes for each individual. Transcript abundances are calibrated normalized values compared to two reference genes. rho = Spearman rank correlation between transcript and metabolite abundances. (b) Transcript abundance of CsTPS5FN in eight 'Finola' individuals.

In addition, we examined the transcript abundance of the multiproduct monoterpene synthase *CsTPS5FN*, to assess if its expression may contribute to terpene profiles in the resin. *CsTPS5FN* transcripts were highly abundant in some individuals, comparable to the highest transcript levels of any other *CsTPS* tested (**Figure 2.12**). Transcript levels of this gene did not

account for any poor correlations between the five terpene-metabolite pairs tested above. Additionally, plant X, which had no detectable monoterpenes, had moderate levels of *CsTPS5FN* transcript. It is therefore likely that *CsTPS5FN*, while highly expressed, is not contributing appreciably to terpene production in 'Finola'.

2.5 Discussion

The resin of *C. sativa* is rich in mono- and sesquiterpenes, which are of interest for their putative contributions to cannabis pharmacology (Russo 2011). Most studies of terpenes in cannabis have focused on phytochemical composition for forensics and breeding, while less research has gone into the biochemistry of terpene formation in cannabis. Knowledge of the genomics and gene functions of terpene biosynthesis may facilitate genetic improvement of cannabis for desirable terpene profiles. Using the hemp strain 'Finola' and its genome and transcriptome resources (van Bakel et al., 2011), we identified early isoprenoid pathway genes as well as specific *CsTPS* genes and enzymes involved in the biosynthesis of nearly all of the different monoterpenes identified in extracts of the cannabis inflorescences, which are densely covered with terpene and cannabinoid accumulating glandular trichomes (**Figure 2.1**). One exception is terpinolene, for which a *CsTPS* has not yet been identified. The terpene profiles of cannabis can be explained by expression of both single-product and multi-product *CsTPS*. Individual 'Finola' plants showed substantial variation in their profiles of mono- and sesquiterpenes. 'Finola' has few monoterpene alcohols or ethers, such as linalool or geraniol, which are common in some cannabis strains.

It is reasonable to expect that there are additional *CsTPS* not described in this work, such as a *CsTPS* that encodes a terpinolene synthase. A search of a new assembly of the Purple Kush genome, to which we recently had pre-publication access (Dr. Jonathan Page, personal

communication), identified a total of 33 complete *CsTPSPK* gene models and additional partial sequences (**Figure 2.13**). Purple Kush is a marijuana strain which requires special research licensing to grow. Thus, characterization of this more comprehensive set of *CsTPSPK* will have to be completed in future work as it requires synthesized genes.

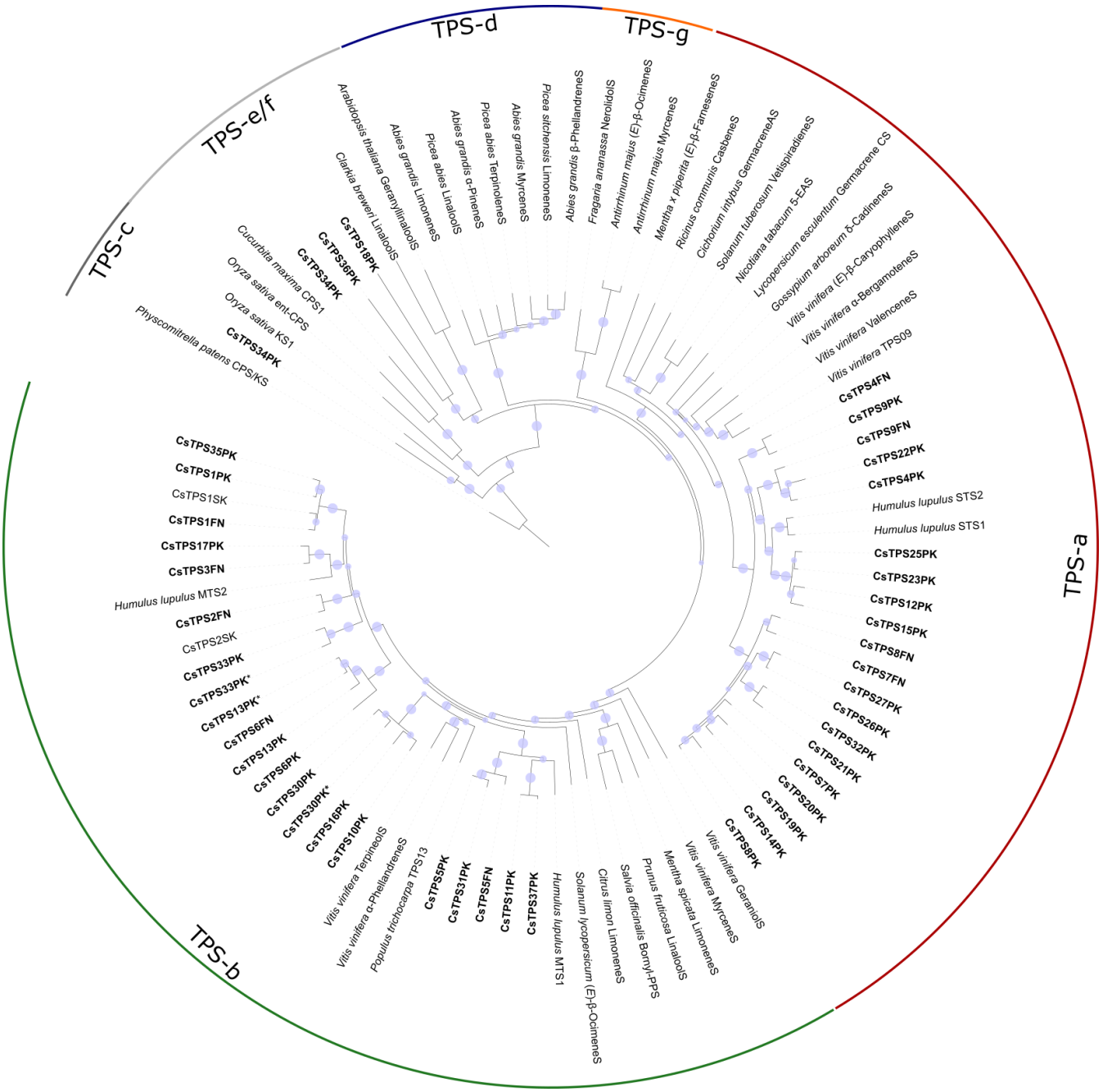


Figure 2.13 Maximum likelihood phylogeny for 33 TPS translated from gene models identified in the *Cannabis sativa* Purple Kush genomic sequences.
41 published TPS sequences from other organisms were included for comparison. Names of cannabis genes identified in this study are in bold. Gene names from Purple Kush followed bzy an asterisk (*) represent biochemically characterized enzymes from Purple Kush transcriptome data. Their nearest homologue in the genome was assigned the same gene ID when the sequences had >95% amino acid identity.

Figure 2.13 indicates of a set of putatively orthologous *CsTPSFN* and *CsTPSPK* genes, which may contribute to overlapping terpene profiles in hemp and marijuana varieties. However, some orthologous genes may have evolved different functions in different strains, and non-orthologous *CsTPS* may contribute some of the same terpene products in different cannabis strains. For example, α-pinene is a major component of strains reported as Purple Kush (Elzinga et al., 2015), but no obvious orthologue of the α-pinene synthase CsTPS2 as identified in the 'Finola' and 'Skunk' strains was found in the Purple Kush genome (**Figure 2.13**). Another example is the set of apparently non-orthologous single-product myrcene synthases, CsTPS3FN and CsTPS30PK identified in 'Finola' and Purple Kush, which only share 52.5% amino acid identity but produce the same monoterpene. Not all *CsTPS* are expected to contribute to terpene accumulation in the resin of cannabis inflorescences and some may function in a different context of the plant biology. For example, *CsTPS5FN* is expressed in inflorescences and the recombinant enzyme produces a mixture of monoterpenes, but does not contribute substantially to the terpene profile of the resin. This gene appears most closely related to *MTS1* from hops (**Figure 2.5**) where no enzyme products were detected *in vitro* (Wang et al. 2008).

Cannabis inflorescences are densely covered with glandular trichomes, which are specialized to produce and accumulate terpenes (Lange 2015). Transcripts of several *CsTPS* genes (**Figure 2.12**) are abundant in trichomes isolated from mid-stage 'Finola' inflorescences. Transcripts associated with early isoprenoid biosynthesis and especially the MEP pathway, which feeds into both monoterpene and cannabinoid biosynthesis, were also abundant in trichomes (**Figure 2.4**). Sesquiterpenes have been reported to be most abundant in early floral stages (Aizpurua-Olaizola et al., 2016), and thus MEV pathway transcripts may be more abundant at earlier stages of flower development. Different DXS and HMGR genes were

differentially expressed in roots relative to other parts of the plant. Terpenes in the roots, if present in cannabis, may contribute to defense as reported in other plant species (Tholl 2015). In plants, *DXS* genes generally fall into two clades, of which DXS I members are generally involved in primary metabolism, and DXS II members are often induced in defense responses (Walter et al., 2002; Paetzold et al., 2010; Carretero-Paulet et al., 2013). Abundance of cannabis *DXS2* transcripts, which clusters with the DXS II subfamily (**Figure 2.3**), suggests defense related terpenoids in cannabis roots and warrants future work on the cannabis root metabolome. We also observed high *FPPS* transcript abundance in stamenate flowers and roots, resembling a previous finding that Arabidopsis FPS1 was primarily expressed in flowers and roots compared to AtFPS2 (Cunillera et al., 1996).

Domestication and selective breeding can result in changes in terpene profiles and abundance. For example, domestication can lead to a decrease in the quantity or variability of terpenes (Aharoni et al., 2004; Köllner et al., 2008; McDowell et al., 2011). Cannabis, especially marijuana, has been domesticated for thousands of years for increased resin volume and potency (Li, 1973; Small, 2015) and as a result profiles and ecological roles of terpenes in ancestral (i.e., undomesticated) cannabis are unknown. While cannabinoid-free individuals have occasionally been reported (de Meijer et al., 2009), we were unable to find any reports in the literature of terpene-free cannabis. In this study, we observed a single monoterpene-free individual, which however still contained cannabinoids and sesquiterpenes. This observation implies that biosynthesis of the different classes of terpenoid metabolites are independently regulated. That terpenes have persisted throughout domestication as a substantial and diverse component of cannabis resin highlights their significance for human preferences.

2.6 Genbank accessions

GenBank accession numbers for the terpene synthases described in this paper are CsTPS1FN: KY014557, CsTPS2FN: KY014565, CsTPS3FN: KY014561, CsTPS4FN: KY014564, CsTPS5FN: KY014560, CsTPS6FN: KY014563, CsTPS7FN: KY014554, CsTPS8FN: KY014556, CsTPS9FN: KY014555, CsTPS11FN: KY014562, CsTPS12PK: KY014559, CsTPS13PK: KY014558. Accession numbers for genes in the MEP pathway are CsDXS1: KY014576, CsDXS2: KY014577, CsDXR: KY014568, CsMCT: KY014578, CsCMK: KY014575, CsHDS: KY014570, CsHDR: KY014579, CsIDI: KY014569. Accession numbers for genes in the MEV pathway are CsHMGS: KY014582, CsHMGR1: KY014572, CsHMGR2: KY014553, CsMK: KY014574, CsPMK: KY014581, CsMPDC: KY014566. Prenyltransferase accession numbers are CsGPPS.ssu1: KY014567, CsGPPS.ssu2: KY014583, CsFPPS1: KY014571, CsFPPS2: KY014580.

Chapter 3: Terpene Synthases and Terpene Variation in *Cannabis sativa*

3.1 Summary

Cannabis (*Cannabis sativa*) resin is the foundation of a multi-billion dollar medicinal and recreational plant bioproducts industry. Major components of cannabis resin are the cannabinoids and terpenes. Variations of cannabis terpene profiles contribute much to the different flavor and fragrance phenotypes that affect consumer preferences. A major problem in the cannabis industry is the lack of proper metabolic characterization of many of the existing cultivars, combined with sometimes incorrect cultivar labeling. We characterized foliar terpene profiles of plants grown from 32 seed sources and found large variation both within and between sets of plants labeled as the same cultivar. We selected five plants representing different cultivars with contrasting terpene profiles for clonal propagation, floral metabolite profiling and trichome-specific transcriptome sequencing. Sequence analysis of these five cultivars and the reference genome of the Purple Kush (PK) cultivar revealed a total of 33 different cannabis terpene synthase (*CsTPS*) genes as well as variations of the *CsTPS* gene family and differential expression of terpenoid and cannabinoid pathway genes between cultivars. Our annotation of the PK reference genome identified 19 complete *CsTPS* gene models, and tandem arrays of isoprenoid and cannabinoid biosynthetic genes. An updated phylogeny of the *CsTPS* gene family showed three cannabis-specific clades, including a clade of sesquiterpene synthases within the TPS-b subfamily that typically contains mostly monoterpene synthases. The CsTPS described and functionally characterized here include 13 that had not been previously characterized and collectively explain a diverse range of cannabis terpenes.

3.2 Introduction

Pistillate flowers of cannabis (*Cannabis sativa*) are densely covered with glandular trichomes that produce and accumulate a resin that is rich in cannabinoids as well as monoterpenes and sesquiterpenes (Turner et al., 1978; Brenneisen and elSohly, 1988; Livingston et al., 2019). Cannabinoids are responsible for the various medicinal and psychoactive properties of cannabis. The terpenes of cannabis resin, which include more than a dozen different monoterpenes and over a hundred different sesquiterpenes, account for much of the diverse organoleptic impressions of cannabis products (Fischedick et al., 2010; Casano et al., 2011; Booth and Bohlmann, 2019) (**Figure 3.1**). Cannabis is broadly categorized into three major chemotypes based on the ratio of Δ^9-tetrahydrocanabinolic acid (THCA) to cannabidiolic acid (CBDA). Type I has high amounts of THCA; type II has approximately equal amounts of THCA and CBDA, and type III is CBDA-dominant (de Meijer et al., 2003). Across these three major chemotypes, terpene profiles show much variation between different cultivars with myrcene, limonene, α-pinene, α-terpinene, or β-caryophyllene as major variable components (Fischedick et al., 2010; Fischedick, 2017; Richins et al., 2018; Reimann-Philipp et al., 2019).

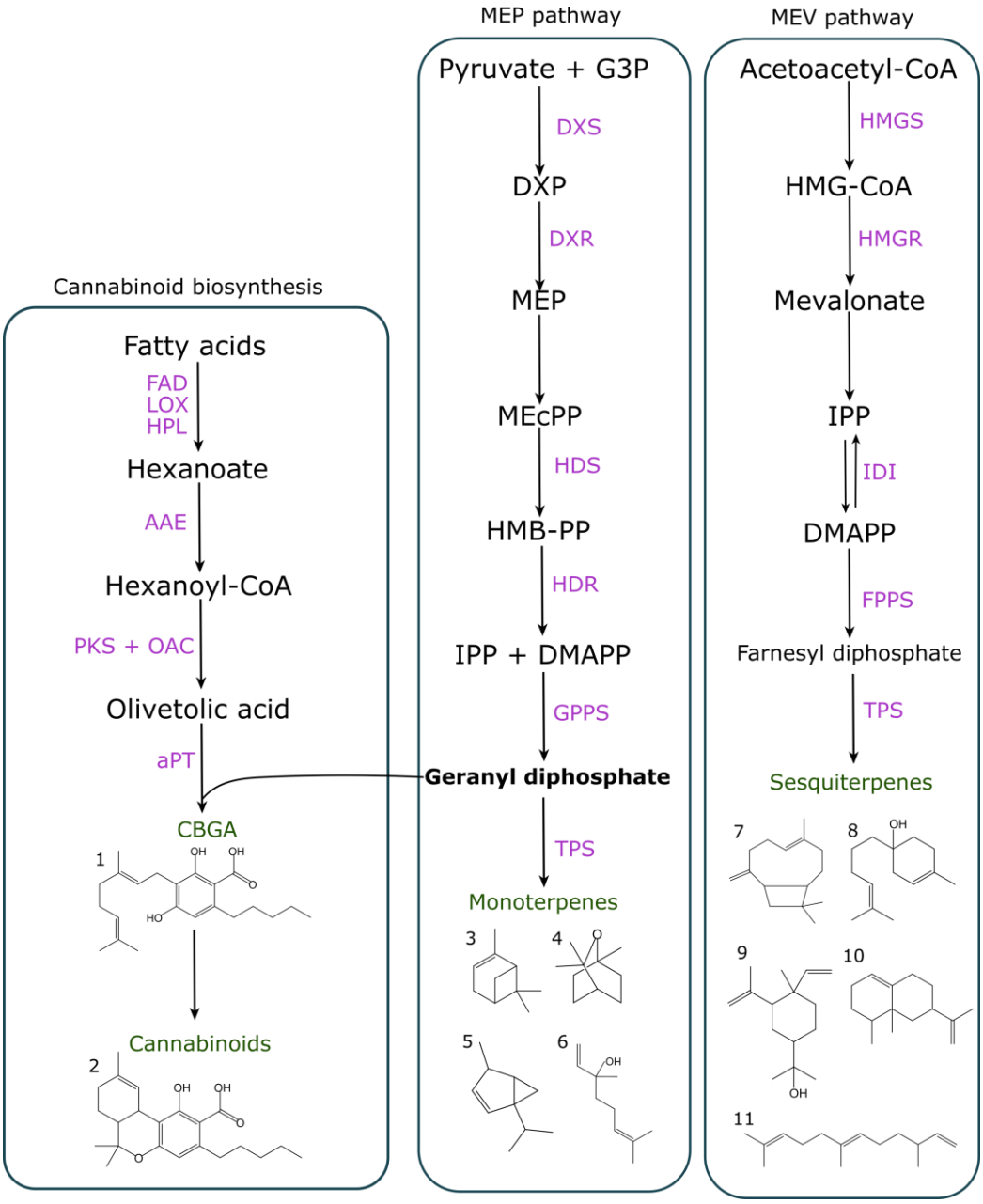


Figure 3.1 Terpene and cannabinoid biosynthetic pathways
Precursors and intermediates are shown in black, final product classes in green. Enzyme names are shown in purple. Cannabinoid pathway: FAD: Fatty acid desaturase; LOX: lipoxygenase; HPL: Hydroperoxide lyase; AAE: Acyl activating enzyme; PKS: Polyketide synthase; OAC: Olivetolic acid cyclase; aPT: Aromatic prenyltransferase. MEP (methylerythritol phosphate) pathway: DXS: 1-Deoxy-D-xylulose 5-phosphate (DXP) synthase; DXR: DXP reductase; MCT: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS: 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS: (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP) synthase; HDR: HMB-PP reductase; GPPS: Geranyl diphosphate synthase; TPS: Terpene synthase. Mevalonate (MEV) pathway: HMGS: 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; HMGR: HMG-CoA reductase; MK: Mevalonate kinase; PMK: Mevalonate-3-phosphate kinase; MPDC: Mevalonate-5-pyrophosphate decarboxylase; IDI: Isopentenyl diphosphate isomerase; FPPS: Farnesyl diphosphate synthase; IPK: Isopentenyl phosphate kinase.

Terpene synthases (TPS), which are encoded in large *TPS* gene families with several subfamilies, produce the diversity of cyclic and acyclic terpene core structures found in plants (Chen et al., 2011). In angiosperms, the TPS-a subfamily generally contains sesquiterpene synthases (sesqui-TPS), and the TPS-b subfamily contains primarily monoterpene synthases (mono-TPS) and hemiterpene synthases. Acyclic monoterpenes are also produced by members of the TPS-g subfamily. The *TPS* gene family has undergone lineage-specific expansions, leading to blooms of related TPS enzymes, as shown for example in grapevine (*Vitis vinifera*) (Martin et al., 2010), eucalyptus (*Eucalyptus grandis*) (Külheim et al., 2015), and tomato (*Solanum lycopersicum*) (Falara et al., 2011). Terpenes and cannabinoids share common isoprenoid precursors (**Figure 3.1**). The most abundant cannabinoids in most cannabis cultivars are THCA and CBDA, which are produced by cannabinoid synthases from cannabigerolic acid (CBGA) (Sirikantaramas et al., 2004; Taura et al., 2007). CBGA is formed by condensation of the monoterpene precursor geranyl diphosphate (GPP) with the aromatic polyketide olivetolic acid (OA) (Fellermeier and Zenk, 1998).

At least 55 different *CsTPS* gene models have previously been reported (**Table 3.2**), but only 14 have been functionally characterized including eight mono-TPS and six sesqui-TPS (Gunnewich et al., 2007; Booth et al., 2017; Allen et al., 2019; Livingston et al., 2019; Zager et al., 2019). The 14 functionally characterized CsTPS account for some of the major terpenes in cannabis (e.g., α-pinene, limonene, myrcene, β-caryophyllene) as well as some more rare compounds (e.g., terpinene, hedycaryol, alloaromadendrene). However, much of the terpene variation in cannabis remains to be explored. While this paper was in preparation, Zager et al. (2019) reported gene networks associated with terpenoid biosynthesis in seven different cannabis cultivars revealing relationships between gene expression and terpenoid accumulation.

Purple Kush (PK) has been established as a reference cultivar for genomic research in cannabis (van Bakel et al., 2011; Booth et al., 2017; Laverty et al., 2019). Here we report the terpene profile of PK and its genome annotation for *CsTPS* and other genes of terpenoid and cannabinoid biosynthesis. Including PK, we investigated variations in terpene profiles in flowers of six different cannabis cultivars based on metabolite analysis, trichome-specific RNA-seq transcriptome analysis, and functional characterization of CsTPS.

3.3 Materials and Methods

3.3.1 Plant material, plant growth, and clonal propagation

Cannabis seeds were provided by and plants grown at Anandia Labs, a subsidiary of Aurora Cannabis Inc. (www.auroramj.com) under a Health Canada research licence. Seeds were surface sterilized in 5% Plant Preservative Mixture (PPM) (www.plantcelltechnology.com) and placed in Petri dishes between filter paper soaked with 0.5% PPM. Germination occurred within two to ten days. Germinated seeds were planted into soil (Sunshine Mix 4, www.sungro.com) supplemented with Florikote 14-14-14 controlled-release fertilizer (www.americanhort.com). During the vegetative growth stage, plants were kept under a 18h/6h light/dark cycle under T5 HO light bulbs. Plants were fertilized twice weekly with Peter's Excel 15-15-15 water-soluble fertilizer (www.domyown.com), pH 5.6-5.8. After approximately two weeks of growth under 18h/6h light/dark cycle, plants were moved under high pressure sodium light bulbs and 12h/12h light/dark cycle to induce flowering. During the flowering stage, plants were fertilized twice weekly with MaxiBloom 15-15-14 water-soluble fertilizer (www.generallyhydroponics.ca) pH 5.6-5.8. Between fertilizations, plants were continuously watered with tap water in hydroponic chambers.

For clonal propagation of plants of five different cultivars, Lemon Skunk (LS), CBD Skunk Haze (CSH), Blue Cheese (BC), Afghan Kush (AK), and Chocolope (Choc), cuttings were taken from well-established stock plants in the vegetative stage and surface sterilized with 5% (v/v) bleach. Cut ends were dipped in 0.4% indole-3-butyric acid rooting hormone (www.valleyindoor.com) and placed in rockwool cubes soaked for 1h in pH 6 water and kept in trays under a clear plastic dome to maintain humidity and promote rooting. Rooted cuttings in rockwool cubes were moved into hydroponic chambers. Female PK plants were clonally propagated and grown as described above, but cuttings in rockwool were transferred directly into soil. All plants were grown in growth chambers (BC Northern Lights) under LED lights (BC Northern lights, 3000K 80 CRI spectrum). The plants were subjected to vegetative growth for 2-3 weeks using an 18h/6h light/dark cycle and watered with Peter's Excel® (15-5-15). To induce flower development, the light cycle was switched to 12h/12h, and plants were watered with Maxibloom (5-15-14).

3.3.2 Harvesting of leaf and flower samples

Leaves (three per plant) were removed from plants four weeks post germination with scissors and placed into 50-mL Falcon tubes. Flowers were harvested for trichome isolation by removal of entire inflorescences of plants at two stages: One week post induction of flowering, and mid-stage maturity between 51 and 60 days post induction of flowering. The time of mid-stage maturity harvest was based on three criteria: (1) All glandular trichomes had matured to have a stalk, (2) 50% of pistils had begun to brown, (3) trichome heads were translucent and had not changed color to appear amber or brown (**Figure 3.2**). Flowers were taken from several nodes along the stem. For metabolite analysis, individual florets were removed using scissors and forceps, and placed in a 1.5-mL Eppendorf tube. Fresh weight (FW) of harvested plant material

was recorded and plant material kept on ice for up to 60 minutes prior to extractions. After extraction, plant material was dried at 60°C for 16h and dry weight (DW) determined.

Figure 3.2 Stages of floral maturation
Drawing showing schematically four stages of flora lmaturation within the inflorescence. Representative photographs are of a Purple Kush (PK) inflorescence at four different stages. From youngest (1) to oldest (4), different stages are characterized by: (1) Very pale pistils, few to no stalked trichomes; (2) no browned pistils, approximately 50% stalked trichomes; (3) pistils beginning to brown, entirely stalked trichomes; (4) entirely browned pistils, brown or amber trichome heads. For this study, metabolite analyses were performed at stages (1) and (3).

3.3.3 Terpene extraction and analysis

Intact plant material was extracted with three washes with 0.5 mL pentane per 100 mg FW. For the first extraction, plant material was vortexed for 30s in pentane to disrupt trichomes and then shaken at room temperature for 4h. For the second and third extraction, the same plant material was shaken in pentane at room temperature for 1h. The three pentane extracts were combined,

centrifuged at 4,300 *x g* for 10 minutes, filtered through a 0.45 µm nylon membrane (Gelman Sciences, now Thermo Fisher Scientific) to remove precipitated waxes and starch, and used for terpene analysis.

For the initial screening of terpene profiles in foliage harvested from plants in vegetative growth at four weeks post germination, each extract was analyzed by gas chromatography-mass spectrometry (GC/MS) on an Agilent 7890A GC coupled with an Agilent 7000A triple-quad MS. An Agilent HP-5 column (5% Phenyl Methyl Polysiloxane), 30m length, 0.25mm i.d., 0.25 µm film thickness (Agilent 19091S-433HP-5MS, Folsom CA, USA) was used. The injector was operated in pulsed-splitless mode at 250°C. He gas was used as the carrier with a flow rate of 1 mL min^{-1} and pulse pressure set at 25 psi of 30s. The oven program was 50°C for 3 minutes, then increase at 10°C min^{-1} to 90°C, at 20°C min^{-1} to 120°C, at 10°C min^{-1} to 150°C, and at 15°C min^{-1} to 320°C, and held at 320°C for 5 minutes, giving a total runtime of 27.8 minutes. The mass spectrometer was operated in Electron Ionization mode at 70eV and data acquisition was made in full-scan mode with a mass range of 40-500 amu.

Analysis of terpenes in extracts from flowers and TPS assay products were done on an Agilent 6890 GC coupled with an Agilent 5973 MSD. An Agilent DB-Wax column (60m length, 0.25mm i.d., 0.25 µm film thickness, Agilent 122-7062, Folsom, CA, USA) was used. The injector was operated in pulsed-splitless mode at 250°C. He gas was used as the carrier with a flow rate of 1 mL min^{-1} and pulse pressure set at 25 psi of 30s. The initial oven temperature was 40°C, then increase at 10°C min^{-1} to 100°C, at 3°C min^{-1} to 130°C, at 30°C min^{-1} to 250°C, hold for 12 minutes. The mass spectrometer was operated in Electron Ionization mode at 70eV and data acquisition was made in full-scan mode with a mass range of 40-500 amu.

Terpenes were identified by comparison of retention indices and mass spectra using authentic standards and Wiley09 and NIST08 mass spectral libraries (http://chemdata.nist.gov/). Retention indices of terpenes were calculated by the retention time of a standard mixture of n-alkanes (C8–C20). Compounds were compared to authentic standards for the following metabolites: alloaromadendrene, bisabolol, bisabolene, cadinene, camphene, β-caryophyllene, (1,8)-cineole, citronellol, farnesene (mix of enantiomers), geraniol, germacrene D, α-humulene, limonene, linalool, myrcene, nerolidol, ocimene (mix of enantiomers), β-phellandrene, α-pinene, β-pinene, terpinen-4-ol, α-terpinene, γ-terpinene, α-terpineol, terpinolene, and valencene. Identifications of bergamotene, δ-selinene, selinane-type, and guaiane-type sesquiterpenes were supported by comparison to *Citrus bergamia* (Bergamot), *Guiaicum officinale* (guaiac wood), and *Pimenta racemose* (Bay) essential oils (www.lgbotanicals.com). Quantification was determined relative to a standard curve of authentic standards. Where no quantitative standard was available, compounds were quantified using the curve of a compound of the same terpene parent skeleton.

3.3.4 Trichome isolation

Flowers were collected at mid-stage maturity from all branches of three clonal plants for each cultivar and incubated in water containing 5 mM aurintricarboxylic (ATA) acid and 1 mM thiourea for 1-4h on ice. After incubation, tissue abrasion to remove trichomes was achieved using a BeadBeater with 30-60 g of tissue with 100 g of 1 mm diameter zirconia/silica beads and 20 g of XAD-4 in enough trichome RNA purification buffer (TRPB) to fill the BeadBeater chamber completely (total volume 350 mL). TRPB was 25 mM HEPES pH 7.3, 200 mM sorbitol, 10 mM sucrose, 5 mM DTT, 5 mM ATA, 1 mM thiourea, 0.6% methyl cellulose, 1%

polyvinylpyrrolidone 40 000 (PVP). Floral tissue was abraded 3x 15 s with a 30s rest on ice in between.

Tissue was filtered through 350 and 105 µM nylon mesh, and filtrate was collected on 40 µM mesh. Purified trichome heads were then collected in a 15-mL Falcon tube and rinsed 3x with TRPB without methyl cellulose and PVP. Purity of the trichome head preparation was determined by light microscopy (**Figure 3.11**). Trichome heads were pelleted by centrifugation at 200 g for 1 minute. Pellets were weighed and then flash-frozen in liquid nitrogen and stored at -80°C.

3.3.5 RNA isolation, transcriptome sequencing and assembly

Amounts of 200 µg of trichome pellet were used for RNA isolation. RNA was isolated using PureLink Plant RNA Reagent (Thermo Fisher), according to the manufacturer's protocol. RNA concentration, purity, and integrity were determined using an Agilent 2100 Bioanalyzer microchip. Three replicates of trichome RNA from each clone were used for RNA-seq. Total RNA in a volume of 15 µL at 100 ng μL^{-1} were used for each sample. Sequencing was performed by the McGill University and Génome Québec Innovation Centre (Montreal, Canada), who performed strand-specific library preparation without heating the samples. Sequencing was performed on Illumina HiSeq2000 platform using 100bp paired-end sequencing. All samples were pooled and sequenced on four lanes, generating approximately 1.5 billion paired-end reads in total. Quality of the sequences were assessed with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads that mapped to *Cannabis sativa* ribosomal RNA sequences, downloaded from NCBI, using Bowtie2 were removed. Adapters were trimmed with BBDuk from the BBTools software suite

(www.sourceforge.net/projects/bbmap/). To improve the contiguity of the assembly, overlapping paired-end reads were joined by BBMerge to generate longer single-end reads. All merged and unmerged reads were pooled and first assembled with Trinity (version 2.6.5) to generate 599,285 non-redundant contigs with average length of 511bp. To gain insight into cultivar-specific sequences, we re-assembled all of the unmerged sequences using RNA-Bloom (version 0.9.8) (Nip et al., 2019) to generate five separate assemblies, one per cultivar, with an average of 260,000 non-redundant contigs and average length of 1,400bp.

TransDecoder (version 5.5.0) predicted on average 170,000 open reading frames (ORF) for each assembly. Predicted peptides translated from open reading frame were clustered at 95% identity, using CD-HIT (version 4.8.1) (Fu et al., 2012) to collapse possible allelic variants. Predicted peptides from each assembly were then pooled together and clustered again at 98% amino acid (aa) identity to further reduce variations between cultivars to a total of 55,550 sequences. Salmon (version 0.14) (Patro et al., 2017) was used to quantify the level of expression on the corresponding open reading frames for downstream differential expression analysis.

3.3.6 CsTPS gene identification and genome annotation

CsTPS candidate genes were identified by using the transcriptome assemblies described above as the subject of tBLASTn search using 100 previously characterized *TPS* genes from cannabis and other plant species. The completeness of the *CsTPS* predictions was confirmed using hmmscan domain search. Gene and splice site prediction on the PK reference genome was performed using the Exonerate algorithm (Curwen et al., 2004) from a list of all characterized *CsTPS* sequences. N-terminal transit peptides were predicted using the TargetP and LOCALIZER tools (Emanuelsson et al., 2007; Sperschneider et al., 2017).

3.3.7 *CsTPS* cDNA cloning and functional characterization

cDNA was made from trichome RNA using the Maxima First Strand cDNA synthesis kit (Thermo Fisher). cDNA was amplified using gene-specific primers, and ligated into pJET vector (Clontech). Sequences were verified by Sanger sequencing, and full-length or N-terminally truncated sequences were subcloned into expression vectors pET28b+ (EMD Millipore) or pASK-IBA37 (IBA Lifesciences), which both carry an N-terminal 6-HIS tag. Full-length CsTPS36BC synthesis was done by IDT (www.idtdna.com). Plasmids were transformed into *E. coli* strain BL21DE3 for heterologous protein expression, as previously described (Roach et al., 2014). Heterologous protein production was induced using 200 µM IPTG (pET28) or 200 ng ml^{-1} anhydrotetracycline in methanol (IBA37), and protein was expressed at 18°C overnight. Cells were harvested by centrifugation and lysed by freeze-thaw cycles, warming the pellet to 4°C then freezing in liquid N_2. Recombinant protein was purified using GE healthcare HIS SpinTrap kit (www.gehealthcare.com). Binding buffer for purification was 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethane- sulfonic acid (HEPES) (pH 7.5), 500 mM NaCl, 25 mM imidazole, and 5% v/v glycerol. Cells were lysed in binding buffer supplemented with Roche complete protease inhibitor tablets and 0.1 mg mL^{-1} lysozyme. Elution buffer was 20 mM HEPES (pH 7.5), 500 mM NaCl, 500 mM imidazole, and 5% glycerol. Purified protein was desalted through Sephadex into TPS assay buffer: 25 mM HEPES (pH 7.3), 100 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 5 mM DTT. Protein purity was determined by western blotting using mouse monoclonal anti-polyHis antibody from Sigma-Aldrich (www.sigmaaldrich.com). *In vitro* assays were performed using 50-100 µL of freshly purified protein and TPS assay buffer to a final volume of 500 µL. Isoprenoid diphosphate substrates (www.isoprenoids.com) were dissolved in 50% methanol and added to assays at a final concentration of 16 µM GPP or 13 µM

FPP. Assays were overlaid with 500 µL pentane with 1.25 µM isobutyl benzene as internal standard. Assays were shaken at 40 rpm at 30°C for 4h. Reactions were stopped and products extracted by vigorous vortexing of assay vial for 30s, and then centrifuged at 4300 *x g* for 15 minutes to separate phases. Assay products were determined using the same GC/MS equipment, program, and identification method as for floral terpene extracts above.

3.3.8 Phylogenetic analysis

ClustalW alignment of translated CsTPS and TPS sequences from other plants and maximum likelihood phylogeny construction were done in CLC Main Workbench 7. Phylogeny construction used the Neighbour Joining method, with 100 bootstrap replicates. Tree visualization and labeling was performed on iTOL (Letunic and Bork, 2019).

3.3.9 Hierarchical clustering, PCA, heatmaps, and differential expression analysis

Hierarchical clustering and PCA of the initial 32 seedlings used peak area for each compound normalized to tissue dry weight and internal standard isobutyl benzene. Clustering was performed using the R function hclust (Kaufman and Rousseeuw, 1990), with Pearson's correlation as a distance measure for metabolites (rows) and Spearman correlation for individual plants (columns). Dendrogram clusters were determined using the method 'maximum', with the number of clusters set to the maximum where inertia gain is above 1. PCA and visualization used the R package 'FactoMineR' (Lê et al., 2008) using default settings. Heatmaps were generated using the R package gplots (https://cran.r-project.org/web/packages/gplots/), with scale = 'row' and z-scores used to normalize rows. Transcript abundance was calculated as the mean normalized counts-per-million of three replicates for each clone. Read-counts estimated with

Sailfish were normalized using DESeq2 R package v. 1.6.1 for differential expression analysis (Love et al., 2014). Transcripts with a normalized CPM < 100 in three or more samples were discarded. False discovery rate was set at 5%. DE genes were defined by an adjusted $\log_2$ fold-change > 2 and a normalized P-value < 0.05.

3.4 Results

3.4.1 Annotation of the Purple Kush reference genome

As a foundation for our study of *CsTPS* genes and their role in terpenoid variation in different cannabis cultivars, we annotated *CsTPS* and other genes of isoprenoid and cannabinoid biosynthesis in the PK reference genome. We identified 19 complete *CsTPS* gene models in PK (**Figure 3.3**), including four clusters of two to five genes, which are more similar in sequence to one another than they are to any other gene model. In addition, five partial *CsTPS* genes were found in the PK genome, likely representing pseudogenes. We also located gene models for all known steps in isoprenoid and cannabinoid biosynthesis, including the plastidial methylerythritol phosphate (MEP) pathway leading to cannabinoids and monoterpenes. Many of the isoprenoid pathways genes, notably 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and 1-deoxy-D-xylulose-5-phosphate reductase (DXR) have multiple copies. Similar to the *CsTPS*, several other isoprenoid and cannabinoid pathway genes are arranged in multi-copy clusters, namely genes encoding DXS, DXR, two copies of the polyketide synthase responsible for producing olivetolic acid (Taura et al., 2009) (PKS), olivetolic acid cyclase (OAC), aromatic prenyltransferases involved in cannabinoid and cannflavin biosynthesis (Page and Boubakir, 2012; Luo et al., 2019; Rea et al., 2019) (aPTs), and cannabinoid synthases THCAS and CBDAS. None of the *CsTPS* gene models clustered with any other genes known to be related to terpenoid or cannabinoid

biosynthesis. While there are no obvious biosynthetic clusters, CBDAS, geranyl diphosphate synthase (GPPS) small subunit, and *CsTPS9* are positioned within a 10-megabase region of the PK genome.

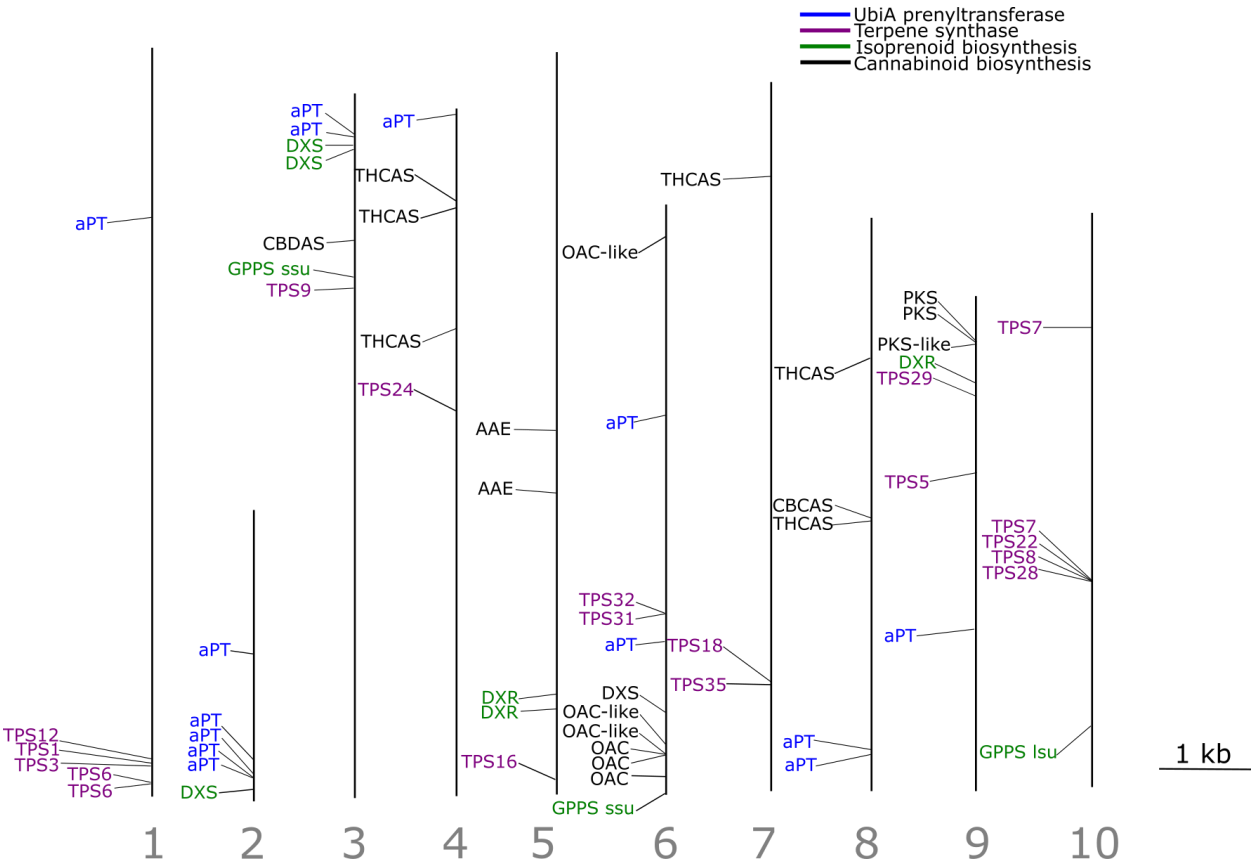


Figure 3.3 Genome locations of genes related to terpenoid and isoprenoid biosynthesis
Scaffolds are from Laverty et al. (2019). Terpene synthases (TPS) are shown in pink, UbiA family prenyltransferases in blue, MEP pathway genes in green, and cannabinoid biosynthetic genes in black.

3.4.2 Variation of foliar terpene profiles within and between cultivars

To explore variation of terpene biosynthesis in cultivars with different terpene profiles, we initially grew plants from 32 seeds, which according to the supplier's information, represented eight different cultivars named Lemon Skunk (LS), CBD Skunk Haze (CSH), Blue Cheese (BC), Afghan Kush (AK), Chocolope (Choc), Blueberry (BB), Vanilla Kush (VK), and Jack Herer (JH). The initial metabolite analysis was done with leaf samples to enable subsequent selection

of individual plants for clonal propagation. Once plants have reached the flowering stage, propagation from cuttings becomes inefficient.

In total, we detected 48 different terpene peaks in the GC/MS analysis of foliar extracts across all 32 individuals, of which 11 were annotated as monoterpenes and 37 sesquiterpenes (**Figure 3.12**). Of these, only three monoterpenes, namely myrcene, α-pinene, and limonene, as well as two sesquiterpenes, β-caryophyllene and α-humulene, were present in every individual. To select plants representing the most contrasting terpene profiles for further study, we performed a principal component analysis (PCA). Principal components (PC) 1 and 2 account for 26.4% and 19.7% of the terpene variation among the 32 plants (**Figure 3.4a**). Most plants cluster towards the lower end of PC2. All plants labeled as CSH clustered together. Only one JH seed germinated, so variability and clustering could not be assessed for this cultivar. For the other plants, there was as much variation among plants that were named as the same cultivar as there was variation between plants that were labeled as different cultivars. Five individual plants, one from each quadrant and one from near the center of the PCA plot, were selected for clonal propagation and detailed characterization, including terpene and cannabinoid analysis of flowers, floral trichome transcriptome sequencing, transcript expression analysis and *CsTPS* discovery and characterization. The selected individuals represent plants identified as belonging to the cultivars AK, BC, Choc, CSH, and LS.

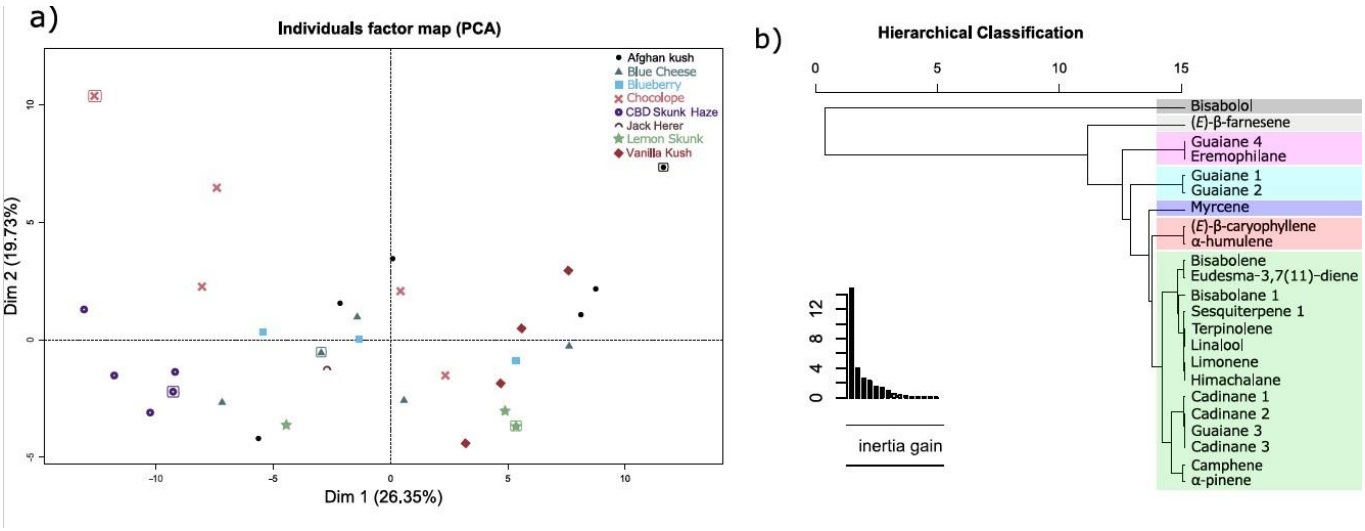


Figure 3.4 Foliar terpene profiles differentiate cannabis plants grown from seeds.
First two dimensions (Dim) of a principal component analysis (PCA) of foliar terpene profiles from 32 cannabis plants. Dim1 accounts for 26.35% of the variance between individuals, and Dim2 accounts for 19.73%. Colors indicate the names under which seeds were obtained. Circled points are individuals that were chosen for clonal propagation and further characterization. **(b)** Unsupervised hierarchical cluster analysis of 46 terpenoid peaks (x axis) in 32 cannabis seedlings. Clustering method is Ward's minimum variance. Eight clusters, indicated by colored boxes, were determined by inertia gain.

Hierarchical cluster analysis of foliar terpenes from the 32 plants was used to determine which compounds account for most of the differences between individuals, and to identify compounds that co-occur. Of the 48 total terpene peaks identified, 23 were found to account for the significant variation in seven groups. Bisabolol contributed the most to differentiation between cultivars followed by (*E*)-β-farnesene (**Figure 3.4b**). Two guaiane-type sesquiterpenes clustered together and apart from other compounds. A guaiane- and an eremophilane- type sesquiterpene also clustered together and apart from other compounds. The two sesquiterpenes β-caryophyllene and α-humulene, which are produced by the same CsTPS (Booth et al., 2017), formed a unique cluster. Myrcene was a member of this clade, but did not cluster with any other compounds. The remaining 39 compounds grouped into a larger cluster. The monoterpenes camphene and α-pinene clustered with a guaiane-type sesquiterpene and three cadinene-type compounds. The same group also included bisabolene and eudesma-3,7(11)-diene, which were

closely related to the largest cluster consisting of another bisabolane-type sesquiterpene, an unidentified sesquiterpene, terpinolene, linalool, limonene, and a himachalane-type sesquiterpene. The remaining compounds did not account for a significant proportion of the variation between terpene profiles.

3.4.3 Flower and foliar metabolite profiles from clonal plants of six cultivars

Three clonal replicates were made from each of the selected five plants, grown in a hydroponic growth chamber, and flowering was induced after five weeks of vegetative growth. Terpenes and cannabinoids were analyzed in samples from flowers and foliage of all 15 plants. For all five cultivars, terpene profiles were qualitatively similar in foliage and flower samples, but quantities of terpenes were much higher in flowers (**Figure 3.5, Table 3.3**). The five cultivars included four that are THCA-dominant and one with approximately equal amounts of THCA and CBDA (**Table 3.4**). Foliar terpenes were dominated by sesquiterpenes (**Figure 3.5**, **Figure 3.12**) with a total terpene content between 0.5 and 1.1 mg g^{-1} dry weight (DW). In contrast terpene levels in juvenile flowers (**Figure 3.2**) at 15 days post floral initiation (DPI) were between 4.9 and 7.3 mg g^{-1} DW, and between 9.3 and 13.6 mg g^{-1} DW in mature flowers (**Figure 3.5, Figure 3.12**). The proportion of monoterpene increased over the time of flower development from juvenile to mature flowers. In total, 15 different monoterpenes and 27 different sesquiterpenes were separated by GC/MS and quantified in mature flowers of the five cultivars (**Table 3.1, Figure 3.13**). In addition, several other terpenes were below the limit of quantification. Myrcene was the most abundant terpene in three cultivars, CHS, AK, and BC. In LS and Choc, the most abundant monoterpenes were (+)-α-pinene and (-)-limonene, respectively. In four of the five cultivars, β-caryophyllene was the dominant sesquiterpene. In AK, germacrene B was the dominant

sesquiterpene. (E)-β-farnesene was present in all samples and was a major component of Choc and AK.

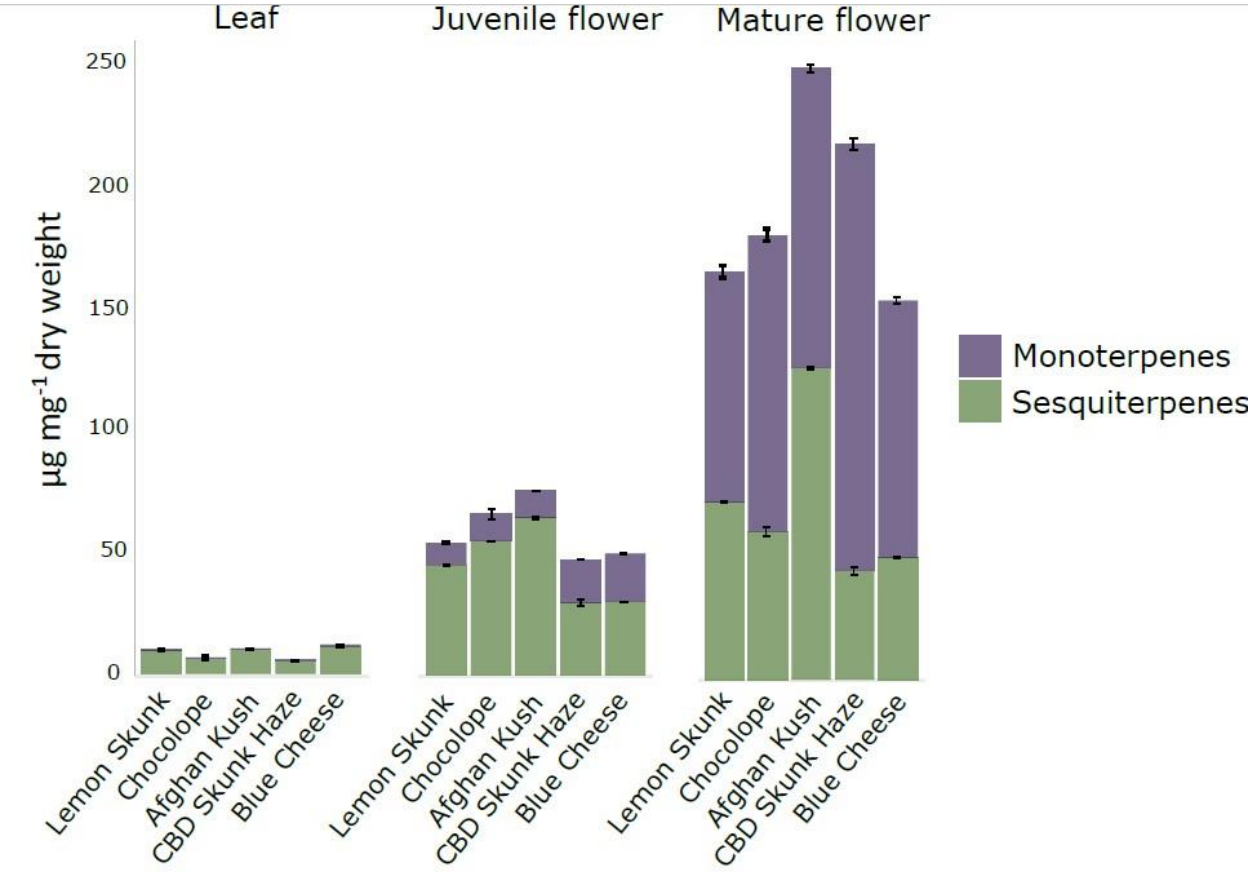


Figure 3.5 Terpene content in leaves and flowers of five different cannabis cultivars
Fan leaves were taken from flowering plants at about 14 days post induction of flowering (DPI). Juvenile flowers of stage (1) (Figure 2) were sampled in triplicate from three clones of each cultivar at the same time as the leaves. Mature flowers of stage (3) (Figure 2) were sampled in triplicate from three clones of each cultivar, between 51 and 60 DPI. Error bars represent standard error across nine samples.

		Mean **(µg g^{-1} DW)** SD					
Retention Index	**ID**	**Lemon Skunk**	**Chocolope**	**Afghan Kush**	**CBD Skunk Haze**	**Blue Cheese**	**Purple Kush**
933.8	(+)-α-pinene*	1849 768.1	198.9 134.1	n.d.	1409 111.8	1024 206.2	86 25.18
933.8	(-)-α-pinene*	139.2 57.8	98.17 60.75	93.7 74.58	n.d.	10.35 10.04	10.66 2.8
1028.07	(+)-camphene*	37.45 16.33	n.d.	n.d.	n.d.	21.16 12.6	n.d. -
1028.07	(-)-camphene*	58.6 24.5	92.35 107	18 3.11	48.3 4.46	29.8 5.15	44.04 14.05
1067.41	(+)-β-pinene*	173.9 87.64	82.86 12.96	n.d.	n.d.	57.48 11.5	n.d. -
1067.41	(-)-β-pinene*	521.7 219.1	311.7 64.81	615.3 166.3	563.26 50.05	517.3 107.8	1336.9 103.5
1123.38	myrcene*	1387 286	1369 1273	6222 1674	7680 1065	4268 275	5683 1583
1158.79	(-)-limonene*	950.8 117.6	2644 380.4	524.5 150.5	627.2 76.61	314.3 41.99	2405 725.8
1167.34	β-phellandrene*	60.36 10.65	161.7 21.29	50.5 14.85	96.5 9.21	46 26.23	n.d. -
1206.23	(*E*)-β-ocimene*	214.1 37.8	1382 1776	316.2 102.7	190.8 201.1	n.d. -	n.d.
1237.05	terpinolene*	30.56 6.98	300.4 123	32.7 23.63	34.9 5.17	12.1 1.201	35.29 10.82
1416.30	monoterpene alcohol	7.58 3.25	n.d.	tr	5 0.804	tr	n.d.
1434.28	δ-elemene	tr	tr	tr	5 1.33	tr	392.4 115
1488.23	(+)-linalool*	214.5 58.1	382.2 444	149.4 42.35	55.4 20.53	148.8 22.8	502.5 127.9
1498.09	2-pinanol	87.05 26.2	159.7 194.6	43.5 8.74	46.9 3.045	35.8 3.81	99.38 31.06
1515.91	sesquiterpene 1	5.24 5.96	14.7 16.4	18.8 1.546	tr	3.8 1.432	9.55 10.4
1523.20	fenchol*	52.64 26.23	86.7 104.7	13.8 2.124	28.4 2.715	9.1 2.19	n.d. -
1529.52	sesquiterpene 2	5.1 8.83	41.55 43.58	37.9 11.55	n.d.	n.d.	n.d.
1533.90	sesquiterpene 3	tr	3.3 4.397	12.5 4.39	n.d.	n.d.	n.d. -
1541.68	α-bergamotene	23.6 24.95	223.7 204.3	267.5 34.03	17.6 6.668	66.3 3.275	319.6 102.9
1546.06	guiane 1	116.4 44.78	41.3 33.85	tr	13.3 2.693	7.285 6.684	n.d. -
1554.82	β-caryophyllene*	270.9 166.6	313.8 211.2	742.9 128.9	155.6 23.1	242.4 51.47	347.94 68.59
1595.19	γ-elemene	153.5 67.62	104.2 94.27	265.3 52.1	68.2 2.643	141.8 34.1	316.5 70.63

1617.20	(*E*)-β-farnesene*	19.3 14.12	357.1 410	294.8 110.5	7.7 5.94	54 45.08	366.7 104
1620.13	sesquiterpene 4	3.8 6.55	tr	12.5 2.12	n.d.	1.5 1.4	n.d. -
1623.05	α-humulene*	168.5 96.21	361.5 162.6	819.5 163	139.5 27.74	219.7 45.23	108.21 21.42
1636.03	borneol	16.39 12.69	31.88 38.63	11.9 9.484	12.4 2.48	n.d.	n.d. -
1641.46	(+)-α-terpineol*	210 87.01	383.6 459.2	138.5 26.82	112.6 9.917	83.3 2.386	234.7 71.69
1664.52	guiane 2	196.3 130.8	2.4 4.072	38.13 4.054	3.9 3.58	16.6 6.863	n.d. -
1671.67	eudesmane 1	552.4 441.8	664.7 502.2	676.5 213.3	40.2 23.4	330 159.6	563.7 84.54
1677.99	α-selinene	121.3 86.87	40.4 8.648	131.3 21.75	93.7 26.12	39.7 9.94	57.11 33.23
1682.93	eudesma-3,7(11)-diene	91.6 71.19	24.24 24.78	203.8 32.57	49.7 14.75	89.1 12.33	334.3 96.23
1696.70	α-farnesene	n.d.	n.d.	n.d.	1.6 2.63	23.3 11.44	58.44 42.88
1700.81	sesquiterpene 5	488.9 320.7	189.8 114.3	644.9 81.28	475 206.8	281.2 48.99	90.67 64
1713.80	valencene	n.d.	n.d.	n.d.	n.d.	1.31 0.24	37.31 10.77
1714.22	δ-selinene	136.9 94.77	109.7 101.9	359.1 43.05	89.5 16.19	153.8 37.17	100.5 28.94
1721.49	cyclounatriene	130.1 149.0	19.6 5.604	42.1 9.52	85.9 29.63	29.33 12.97	n.d. -
1724.94	sesquiterpene 6	45.2 35.16	23.9 25.27	154.8 34.7	41.7 14.31	31.3 10.7	n.d.
1729.54	sesquiterpene 7	488.9 320.7	189.8 114.3	644.9 81.28	475 206.8	281.2 48.99	72.07 29.32
1773.22	germacrene B*	764.1 314.7	488.5 421.6	1268 275.2	343.5 23.99	711.4 178.6	246.5 70.63
1931.60	caryophyllene oxide	1.74 2.563	n.d.	3.4 1.54	n.d.	n.d.	n.d. -
2067.90	guiaol	n.d. -	492 116.1	n.d. -	505.6 219.1	n.d. -	n.d. -
2136.82	γ-eudesmol	77.48 50.66	383.4 225	59.04 8.622	412.1 219.2	101.2 58.63	42.95 31.7
2158.70	bisabolol*	234.8 177.8	n.d.	24 35.32	327.4 17.95	10.11 10.52	93.07 25.99
2108.30	bulnesol	n.d. -	353.8 372	n.d. -	311.8 151.2	n.d. -	n.d. -

Table 3.1 Amounts of terpenes in mature flowers of different cannabis cultivars

Values are the mean of three replicates from three clones of each cultivar. Metabolites marked with (*) have been verified using authentic standards; others were identified based on mass spectrum and retention index. n.d.: not detected. tr: trace (<1 µg). Compounds below the limit of quantification across all samples are omitted. Retention index was calculated on a DB-Wax GC column. Mean (top) is derived from three inflorescences from three clones per cultivar. SD (bottom) = standard deviation of nine samples. Note: Purple Kush data were obtained from plants grown for a separate study, and data may not be directly comparable with those of the other five cultivars.

Gene ID	Cultivar	Main products	Reference	NCBI Accession Number
CsTPS1	Skunk	(-)-limonene	Gunnewich et al., 2007	ABI21837.1
CsTPS2	Skunk	(+)-α-pinene	Gunnewich et al., 2007	ABI21838.1
CsTPS3	Finola	Myrcene	Booth et al., 2017	KY014561
CsTPS4	Finola	alloaromadendrene	Booth et al., 2017	KY014564
CsTPS5	Finola Purple	Myrcene	Booth et al., 2017	KY014560
CsTPS5	Kush	Myrcene/Bisabolol	This paper	MN967481
CsTPS6	Finola	(E)-β-ocimene	Booth et al., 2017	KY014563
CsTPS7	Finola	δ-selinene	Booth et al., 2017	KY014554
CsTPS8	Finola	γ-eudesmol	Booth et al., 2017	KY014556
CsTPS9	Finola Purple	β-caryophyllene, α-humulene	Booth et al., 2017	KY014555
CsTPS11	Kush Purple	Not tested	Allen et al., 2019	KY014562
CsTPS12	Kush Purple	Not tested	Allen et al., 2019	KY014559
CsTPS13	Kush	(Z)-β-ocimene	Booth et al., 2017	KY014558
CsTPS14	Canna Tsu Purple	(-)-limonene	Zager et al., 2019	MK801766
CsTPS14	Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS15	Canna Tsu Cherry	Myrcene	Zager et al., 2019	MK801765
CsTPS16	Chem Purple	Germacrene B	Zager et al., 2019	MK131289
CsTPS17	Kush Afghan	Not tested	Allen et al., 2019	Not deposited
CsTPS17	Kush Purple	Myrcene	This paper	MN967470
CsTPS18	Kush	Not tested	Booth et al., 2017	KY624356
CsTPS18	Valley Fire	Nerolidol/Linalool	Zager et al., 2019	MK801764
CsTPS18	Chocolope Black	Linalool	This paper	MN967473
CsTPS19	Lime	Nerolidol/Linalool	Zager et al., 2019	MK801763
CsTPS20	Canna Tsu Purple	Hedycaryol	Zager et al., 2019	MK801762
CsTPS20	Kush Afghan	Not tested	Allen et al., 2019	Not deposited
CsTPS21	Kush Purple	Hedycaryol	This paper	MN967483
CsTPS22	Kush Purple	Himachalane	This paper	MN967477
CsTPS23	Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS23	Chocolope Purple	Myrcene	This paper	MN967480
CsTPS24	Kush	Not tested	Allen et al., 2019	Not deposited

CsTPS25	Lemon Skunk Purple	(*E*)-β-farnesene	This paper	MN967472
CsTPS26	Kush Purple	No products found	This paper	MN967479
CsTPS28	Kush Purple	β-elemene	This paper	MN967482
CsTPS29	Cheese Blue	Linalool	This paper	MN967468
CsTPS30	Kush Purple	Myrcene	Booth et al., 2017	KY624367
CsTPS31	Kush Purple	Terpinolene	This paper	MN967474
CsTPS32	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS32	Kush Purple	Geraniol/Himachalane	This paper	MN967484
CsTPS33	Kush Purple	α-terpenene, γ-terpenine	Booth et al., 2017	KY624371
CsTPS34	Kush Purple	Predicted diterpene synthase	Booth et al., 2017	KY624373
CsTPS35	Lemon Skunk Purple	Linalool/Nerolidol	This paper	MN967475
CsTPS35	Kush Purple	Predicted diterpene synthase	Booth et al., 2017	KY624375
CsTPS36	Kush	Not tested	This paper	MN967471
CsTPS37	Finola Purple	Terpinolene	Livingston et al., 2019	MK614216
CsTPS37	Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS38	Finola Purple	(E)-β-ocimene	Livingston et al., 2019	MK614217
CsTPS38	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS40	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS41	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS42	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS43	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS44	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS46	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS47	Kush Purple	Not tested	Allen et al., 2019	Not deposited
CsTPS48	Kush	Not tested	Allen et al., 2019	Not deposited

CsTPS49	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS50	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS51	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS52	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS53	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS55	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS58	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS59	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS60	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS61	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS62	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS63	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS64	Purple Kush	Not tested	Allen et al., 2019	Not deposited
CsTPS65	Purple Kush	Putative copalyl diphosphate synthase	This paper	MT295505
CsTPS66	Purple Kush	Putative ent-kaurene synthase	This paper	MT295506

Table 3.2 Cannabis TPS (CsTPS) previously published or reported here
Gene ID: number assigned to the enzyme by the authors; Cultivar: the cultivar or 'strain' in which the sequence was identified; Main products: the major product of each TPS, where tested, as reported by the authors; Reference: paper in which the enzyme was reported; NCBI accession number: the accession under which the sequence was deposited.

µg/g dry weight

Compound number	Compound ID	Cultivar name	Lemon Skunk			Jack Herer		Blueberry			Chocolope					Afghan Kush			
		Individual plant ID	x.156	x.157	x.158	x.159	x.160	x.161	x.162	x.164	x.165	x.166	x.167	x.168	x.169	x.170	x.171	x.172	x.173
1	camphene		tr.	4.4	1.9	n.d.	tr.	tr.	1.8	1.7	tr.	n.d.	tr.	2.1	n.d.	n.d.	tr.	n.d.	n.d.
2	α-pinene		3.3	2.7	tr.	n.d.	tr.	tr.	1.9	2	2.7	3.1	2.2	2.6	tr.	1.5	2.2	2.2	tr.
3	myrcene		3.5	tr.	3.4	tr.	6.2	3.3	85	tr.	1.6	tr.	tr.	1.6	3.3	4.1	25.8	5.6	4.3
4	green-leaf volatile		9.1	5.9	8.3	6.4	8.7	6.4	8.7	6	9.7	6.4	8.5	6.6	7.4	6.4	8.1	6.9	7.7
5	limonene		33.2	13.7	5.3	tr.	3.9	5.8	11.6	tr.	25.6	16	18.8	tr.	3	4.7	14	6.2	2.6
6	(E)-β-ocimene		tr.	2.4	2.4	11	5.9	n.d.	7	n.d.	2.1	tr.	tr.	n.d.	5.3	3.8	n.d.	n.d.	n.d.
7	linalool		2.4	tr.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.
8	terpinolene		tr.	7.4	5	n.d.	tr.	1.5	tr.	1.9	tr.	8	3.6	1.7	1.7	tr.	2.7	2.5	tr.
9	Monoterpene 1		tr.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	tr.	tr.	tr.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.
10	Monoterpene alcohol		2	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	tr.	tr.	n.d.	n.d.
11	bergamotene		n.d.	n.d.	tr.	tr.	1.9	tr.	1.6	tr.	2.5	tr.	tr.	1.5	n.d.	3	2.5	1.9	tr.
12	bisabolane 1		n.d.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	2.2	4.7	tr.	2.3	2.1	n.d.	4.8	3.5	3.1	2
13	β-caryophyllene		75.8	19.4	21.8	8.1	57.6	15.4	55.1	9.5	18.5	6.7	14.9	6.8	6.2	25.8	26.4	13.3	9.7
14	γ-elemene		11.5	21.7	13.1	25	27.9	24.1	15.4	21	19.9	11.1	12.3	37.5	4.7	66.9	29.4	42.3	14.2
15	sesquiterpene 1		tr.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
16	(E)-β-farnesene		tr.	5.3	18.4	tr.	38.8	46.2	27.7	90	46.8	24.9	25.4	93.3	5.7	231	50	146.9	205
17	α-humulene		49.2	7.8	18.1	7.4	55.6	13	31.9	6.9	19.8	3.8	12.5	6.2	6	25.1	24.4	11.1	12
18	cadinane 1		n.d.	n.d.	n.d.	n.d.	5.4	tr.	2	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	2.8	5.7	2	n.d.
19	cadinane 2		tr.	tr.	tr.	1.6	1.8	1.7	tr.	tr.	tr.	tr.	tr.	2.7	n.d.	5.5	2.1	4.1	tr.
20	sesquiterpene 2		n.d.	tr.	tr.	tr.	tr.	tr.	tr.	1.7	tr.	n.d.	tr.	2.2	n.d.	4.5	2.5	3.3	tr.
21	B-eudesmene		9.9	2.5	3	2.5	3.4	2.7	1.7	tr.	2	5.2	tr.	2.9	1.8	6.2	11.2	9.5	2.2
22	a-guaiene		5.2	6.2	1.8	9.7	2.8	9.4	n.d.	n.d.	tr.	7.7	n.d.	4.6	tr.	25.8	107	19.4	1.6
23	bisabolane 2		12.6	16.4	26.7	6.3	39.9	19.8	27.7	6.2	8.6	3.5	5.2	6.7	6.2	17.7	46.1	11.4	5.4
24	sesquiterpene 3		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.3	3	tr.	tr.	2.3	tr.	6.1	n.d.	3.4	1.6
25	sesquiterpene alcohol 1		1.9	2.7	n.d.	n.d.	2.2	2	n.d.	1.9	1.8	tr.	tr.	2.7	n.d.	n.d.	2.5	n.d.	tr.
26	himachalane		n.d.	3.6	7.2	6.1	15.5	6.9	9.9	7.7	109	3.4	7.5	103	2.5	23.3	18.2	14.7	7.2
27	sesquiterpene 4		n.d.	n.d.	n.d.	2.4	n.d.	2.6	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	tr.	5.1	2.3	4.3	tr.
28	unknown compound		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.8	n.d.	n.d.	n.d.
29	Guaiane 1		42.8	66.4	15.1	32.9	25.3	35.3	15.2	tr.	4.7	23.3	2.9	13.7	4.7	40.2	31	29.2	15.9
30	eudesma-3,7(11)-diene		14	12.6	16.9	12.2	21.5	12.4	13.3	n.d.	7.8	5.6	4.5	7.1	4.7	25	25.6	18.7	15.6
31	α-farnesene		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2	6.5	tr.	3.2	6.8	tr.	2.7	n.d.	1.8	n.d.
32	sesquiterpene 5		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5.4	tr.	5.1	n.d.	3.2	3.9	2.2	4.4	12	4.3	n.d.
33	Guaiane 2		13.6	58.6	13.5	35.7	30.5	38.3	9.9	n.d.	11	10.5	6.3	57.9	4.8	58.8	32.7	33.5	13.7
34	sesquiterpene 6		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	1	1.5	n.d.	n.d..5	2.6	n.d.	2.8	n.d.	1.3	tr.
35	caryophyllene oxide		1.9	n.d.	1.5	tr.	n.d.	n.d.	2.1	1.6	8.1	tr.	2.9	3.4	tr.	2.6	n.d.	1.6	2.9
36	guaiol		n.d.	n.d.	n.d.	n.d.	tr.	n.d.	8.4	8.8	25.6	4.1	10.8	15.6	n.d.	n.d.	n.d.	n.d.	6.7
37	cedrol		n.d.	n.d.	n.d.	n.d.	n.d.	n.d..4	2.3	3.3	6.7	1.5	2.8	n.d.	n.d.	1.1	n.d.	tr.	2.2
38	sesquiterpene alcohol 2		1.6	n.d.	n.d.	n.d.	2.3	n.d.	n.d.	n.d.	7.6	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	2.8
39	sesquiterpene 7		tr.	1.8	tr.	2	2.1	1.6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.6	2.7	1.6	n.d.
40	cadinane unknown		n.d.	n.d.	tr.	tr.	n.d.	tr.	12.8	18.2	36.6	8.1	15.6	31.7	tr.	3.3	2.9	2.4	12.2
41	compound		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5.8	5.6	19	2.6	8.1	8.1	n.d.	1.9	n.d.	n.d.	6
42	sesquiterpene alcohol 3		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	6.1	11.5	27.1	5.3	10.5	18.1	n.d.	n.d.	n.d.	n.d.	8.1
43	Guaiane 4		2.8	tr.	2.9	tr.	2.1	tr.	27.8	24.7	88.3	12	38.5	37.8	tr.	n.d.	4.9	1.4	27.8
44	Eremophilane		tr.	n.d.	n.d.	tr.	n.d.	n.d.	29.6	27	104	10.7	43.4	41.6	tr.	3.5	n.d.	2.6	27.9
45	bisabolol		249.4	87.7	66.4	6.8	115.6	31	130.3	8.1	31.3	40.9	12.8	12.6	29.7	4.6	121.9	3.5	29.9
46	selinane		n.d.	n.d.	tr.	1.8	1.9	1.8	1.9	n.d.	5.7	2.2	1.5	3.3	tr.	4.4	2.2	2.6	2.9
47	guaiane 5		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	tr.	2.9	tr.	tr.	3.1	n.d.	n.d.	n.d.	n.d.	1.9
48	bulnesol		n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	2	4.1	tr.	1.3	4.4	n.d.	n.d.	n.d.	n.d.	n.d.

		CBD Skunk Haze						Vanilla Kush					Blue Cheese			
Compound number	Compound ID	x.174	x.175	x.176	x.177	x.178	x.179	x.180	x.181	x.182	x.183	x.184	x.185	x.186	x.187	x.188
1	camphene	2	19.5	88.1	6.4	2.5	1.9	1.5	tr.	tr.	tr.	24.5	tr.	8.9	26.9	36.3
2	α-pinene	2.9	14	42.9	4.3	tr.	tr.	2.1	2.3	tr.	2	11.8	tr.	4.4	17.4	17
3	myrcene	9.4	48	32	tr.	2	2.6	3.1	8.6	tr.	8.3	41.9	15.2	21.3	71.3	16.1
4	green-leaf volatile	6.5	9.8	7.1	9.2	7.6	8.9	6.5	11	7.2	tr.	7.8	9.7	8.1	12.7	8.3
5	limonene	13.3	tr.	13.4	5.3	4.3	tr.	16.1	25.6	7.8	21.4	tr.	14.7	2.1	8.9	4.7
6	(E)-β-ocimene	n.d.	5.5	n.d.	4.3	tr.	11.3	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
7	linalool	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	1.6	n.d.	1.6	tr.	n.d.	n.d.	n.d.	n.d.
8	terpinolene	5.3	2.6	4.8	3.6	3.6	9.4	3.5	3.5	2	2.4	3	6	4	5.3	2.8
9	Monoterpene 1	tr.	tr.	tr.	n.d.	n.d.	tr.	tr.	tr.	n.d.	tr.	tr.	tr.	n.d.	tr.	n.d.
10	Monoterpene alcohol	tr.	n.d.	tr.	n.d.	n.d.	tr.	n.d.	tr.	n.d.	tr.	n.d.	tr.	n.d.	n.d.	n.d.
11	bergamotene	tr.	n.d.	n.d.	n.d.	tr.	tr.	tr.	2.4	tr.	2.4	tr.	1.6	tr.	tr.	tr.
12	bisabolane 1	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4.5	tr.	3.8	tr.	2.9	n.d.	n.d.	n.d.
13	β-caryophyllene	14.9	13.6	7	13.4	52.9	tr.	18.8	37.1	4	39	17.6	45.6	12.8	27.1	6.3
14	γ-elemene	31.4	6.9	4.5	2.9	n.d.	tr.	14.2	28.5	3.4	14.7	19.2	29.3	33.4	23.9	tr.
15	sesquiterpene 1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.3	2.7	n.d.	n.d.
16	(E)-β-farnesene	75.7	3.2	4.2	1.3	n.d.	6.8	54.4	43.9	104	39.5	36.4	32.4	37	19.4	15.8
17	α-humulene	14.2	9.6	3.2	6.2	31.7	25.2	12.1	22.5	3.1	29.2	109	38.6	9.9	22	5.4
18	cadinane 1	1.4	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
19	cadinane 2	2.9	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	1.8	n.d.	tr.	tr.	1.9	3	tr.	n.d.
20	sesquiterpene 2	2.3	tr.	n.d.	n.d.	n.d.	n.d..5	1.4	2.4	n.d.	2	1.2	2.1	1.9	1.5	tr.
21	B-eudesmene	3.8	tr.	n.d.	n.d.	1.5	tr.	5.3	105	tr.	8	2.9	6.6	6	4.8	tr.
22	a-guaiene	19.5	tr.	n.d.	n.d.	2.9	n.d.	13.9	6.6	2.8	5	5.5	6	11.7	2.9	1.6
23	bisabolane 2	24.3	29.2	13.8	12.4	24.8	24	5.1	10	tr.	9.9	5.9	n.d.	3.3	6.3	1.9
24	sesquiterpene 3	2.6	n.d.	n.d.	n.d.	n.d.	n.d.	1.7	2.3	n.d.	2	n.d.	24.8	15.3	1.5	tr.
25	sesquiterpene alcohol 1	3.1	tr.	tr.	tr.	3.2	2.1	tr.	1.8	n.d.	n.d.	2	n.d.	n.d.	tr.	tr.
26	himachalane	11.2	4.3	2.4	2	4.2	4.1	6.9	19.2	tr.	13.3	5.7	12.9	8.1	108	2.4
27	sesquiterpene 4	3.9	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	1.8	2.5	1.8	tr.
28	unknown compound	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	102	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
29	Guaiane 1	54.8	22.5	26.3	9.3	32.6	14.7	2.8	n.d.	tr.	2.7	34.4	33.6	41.3	23.7	12.6
30	eudesma-3,7(11)-diene	17.8	9	2.7	3	n.d.	4.6	tr.	13.6	n.d.	tr.	6.9	29.5	17.2	24.1	2
31	α-farnesene	n.d.	3.2	3.1	2.6	2.1	7.6	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	1.6
32	sesquiterpene 5	4.9	n.d.	n.d.	n.d.	n.d.	5.2	3.4	9.5	tr.	3.5	n.d.	n.d.	n.d.	n.d.	n.d.
33	Guaiane 2	50.3	11.6	9.5	3.8	n.d.	12.3	2.1	25.3	n.d.	tr.	20.2	34.5	65.1	28.5	11.2
34	sesquiterpene 6	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.
35	caryophyllene oxide	1.7	n.d.	tr.	n.d.	tr.	tr.	tr.	3.4	tr.	1.5	tr.	4.8	2.5	4	tr.
36	guaiol	n.d.	6.2	10.3	7.6	7.6	19.9	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	3.2
37	cedrol	tr.	1.5	3.3	1.8	2.1	3.6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.5
38	sesquiterpene alcohol 2	tr.	n.d.	n.d.	n.d.	tr.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	2.6	tr.	2.7	tr.
39	sesquiterpene 7	3.5	n.d.	n.d.	n.d.	n.d.	n.d.	2	5.5	tr.	2.2	2.8	3.4	4.5	8.9	n.d.
40	cadinane	2	8.9	18.4	11.3	16.4	27	tr.	2.9	n.d.	1.5	1.5	1.8	tr.	n.d.	7.6
41	unknown compound	n.d.	4.4	6.1	6.5	4.8	13.1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	n.d.	2.2
42	sesquiterpene alcohol 3	n.d.	4.6	12.7	5.4	10	13.8	n.d.	n.d.	n.d.	n.d.	n.d.	2.9	6.1	2.8	5.3
43	Guaiane 4	1.8	19.9	27.5	27.4	20.4	59.5	n.d.	3.2	n.d.	tr.	n.d.	5.9	5	4.9	8
44	Eremophilane	2.1	19.7	27.8	31	18.9	70.1	tr.	2.1	n.d.	1.7	n.d.	3.2	2.2	2.4	8.8
45	bisabolol	72.2	317.7	156.4	182.6	137.3	282.8	1.6	5.5	n.d.	3.1	32.7	106.4	12.9	29.5	30.7
46	selinane	5.5	n.d.	n.d.	n.d.	n.d.	n.d.	tr.	2.6	n.d.	tr.	3.4	3.3	5.2	4.6	n.d.
47	guaiane 5	n.d.	tr.	1.5	n.d..8	n.d.	1.7	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.
48	bulnesol	n.d.	n.d.	2.2	tr.	n.d.	2.1	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	tr.

Table 3.3 Identification and amounts of foliar terpenes in 32 different cannabis seedlings

Names: cultivar names under which seeds were obtained; Individual plant ID: codes assigned to each seedling; Compound number: sequential numbering of compounds corresponding to mass spectra shown in supplemental figure S2; Compound ID: best attempt at identifying each compound based on mass spectrum, retention index, and authentic standards where available. Values represent the absolute amount of each terpene (μg g^{-1} dry weight) identified. n.d.: not detected. tr.: trace (below level of quantification).

		Foliar cannabinoids (ng mg^{-1} dry weight)					
Individual Plant ID	Cultivar Name	CBDA	THCA	CBCA	CBGA	THCA/CBDA ratio	Percent Total Cannabinoids per dry weight
AGP 156	Lemon Skunk	1224	392561	325520	11958	320.72	5.01%
AGP 157	Lemon Skunk	**1270**	**440272**	**270328**	**7683**	**346.67**	**4.98%**
AGP 158	Lemon Skunk	1119	394400	155886	6409	352.46	3.82%
AGP 159	Jack Herer	950	310801	186744	7466	327.16	3.41%
AGP 160	Blueberry	901	298420	213843	2417	331.21	3.47%
AGP 161	Blueberry	904	219785	72492	6995	243.13	2.03%
AGP 162	Blueberry	846	265954	164164	7337	314.37	2.95%
AGP 164	Chocolope	699	196554	580328	8191	281.19	5.27%
AGP 165	Chocolope	**1009**	**311272**	**357380**	**7887**	**308.50**	**4.55%**
AGP 166	Chocolope	1100	396594	201727	7133	360.54	4.08%
AGP 167	Chocolope	1753	590660	338356	107069	336.94	6.97%
AGP 168	Chocolope	1039	299179	332110	1986	287.95	4.28%
AGP 169	Afghan Kush	1272	413632	233599	8256	325.18	4.42%
AGP 170	Afghan Kush	**1401**	**499625**	**335707**	**4485**	**356.62**	**5.68%**
AGP 171	Afghan Kush	1111	263798	98810	3118	237.44	2.48%
AGP 172	Afghan Kush	940	271198	137651	2113	288.51	2.79%
AGP 173	Afghan Kush	1045	311321	377746	6934	297.91	4.69%
AGP 174	Afghan Kush	1483	469775	276006	12682	316.77	5.12%
AGP 175	**CBD Skunk Haze**	**615061**	**217411**	**116208**	**20086**	**0.35**	**6.52%**
AGP 176	CBD Skunk Haze	507592	176174	130619	0	0.35	5.50%
AGP 177	CBD Skunk Haze	489051	189486	141970	14825	0.39	5.62%
AGP 178	CBD Skunk Haze	310565	116561	131756	4586	0.38	3.79%
AGP 179	CBD Skunk Haze	743314	308518	81020	0	0.42	7.66%
AGP 180	Vanilla Kush	1278	284293	90673	9059	222.45	2.58%
AGP 181	Vanilla Kush	1146	339108	222745	3815	295.91	3.83%
AGP 182	Vanilla Kush	869	199791	112807	5003	229.91	2.14%
AGP 183	Vanilla Kush	961	254398	142134	4695	264.72	2.70%
AGP 184	Blue Cheese	922	228838	228586	1625	248.20	3.10%
AGP 185	Blue Cheese	866	248952	356368	1104	287.47	4.10%
AGP 186	Blue Cheese	723	132592	294243	2267	183.39	2.90%
AGP 187	Blue Cheese	**1090**	**332379**	**343083**	**28286**	**304.93**	**4.74%**
AGP 188	Blue Cheese	792	243133	294710	9596	306.99	3.68%

Table 3.4 Foliar cannabinoid content in 32 cannabis seedlings

Values are mean of three replicates, in ng mg^{-1} dry weight. Bold lines represent seedlings chosen to be cloned and sequenced. Individual plant ID: codes assigned to each seedling; Cultivar name: name of the cultivar as per supplier; CBDA: cannabidiolic acid; THCA: tetrahydrocannabinolic acid; CBCA: cannabichromenic acid; CBGA: cannabigerolic acid.

A terpene profile for the PK cultivar (**Table 3.1**) was produced with three clones of the PK plant that was sequenced for the reference genome; however, these plants were grown under different conditions. The terpene content of floral trichomes of PK, induced to flower after four weeks of vegetative growth, peaked at 21 mg g^{-1} DW. In PK flowers, we detected 49 different terpenes including 15 monoterpenes and 34 sesquiterpenes. Monoterpenes were dominated by myrcene, (-)-limonene, and (+)-linalool, with lesser amounts of (+)-β-pinene, α-terpineol, (-)-α-pinene, (-)-camphene, and (*Z*)-β-ocimene. The most abundant sesquiterpene was β-caryophyllene, followed by γ-elemene and an eudesmane-type olefin.

3.4.4 Transcriptomes of floral trichomes are enriched for terpene and cannabinoid biosynthesis

We produced 15 separate trichome-specific transcriptomes from three plants for each of the five cultivars. Trichome heads were isolated from mature flowers (**Figure 3.2**) from individual clonal plants prior to signs of floral senescence. Mature flowers are characterized by apparent lack of unstalked glandular trichomes, as glandular trichomes have matured to the stalked stage (Livingston et al., 2019), and with more than 80% of pistils turning from white/green to brown. Total RNA was extracted from isolated trichome heads and used for RNA-Seq. We initially assembled sequences from all five cultivars into a single pooled transcriptome. The normalization of a pooled transcriptome allows quantitative comparison among cultivars. The pooled assembly contained 599,285 non-redundant contigs with an average length of 511bp (**Table 3.5**). The trichome transcriptome raw sequence data are deposited in the NCBI Sequence Read Archive (accession number PRJNA599437).

Assembly	Raw Filtered Reads	Read Pairs Joined	Assembler	Number of Non-Redundant Contigs	Mean Non-Redundant Contig Length (bp)
5 cultivars combined	533,578,399	659,654,911	Trinity	599,285	511
Lemon Skunk	244,501,556	231,742,033	RNA-Bloom	322,787	1,338
Chocolope	288,719,568	277,234,195	RNA-Bloom	263,610	1,281
Afghan Kush	229,534,023	223,938,684	RNA-Bloom	223,849	1,255
CBD Skunk Haze	194,830,176	190,931,237	RNA-Bloom	229,327	1,787
Blue Cheese	272,775,194	269,387,161	RNA-Bloom	257,714	1,320

Table 3.5 Assembly statistics for six transcriptome assemblies used
Assembly: origin of sequences assembled; Raw filtered reads: the number of reads after filtering for adapter sequences and ribosomal RNA; Read pairs joined: number of reads successfully joined to their read-pair; Assembler: the assembly algorithm used; Non-redundant contigs; number of contigs assembled after filtering shorter identical sequences; Mean non-redundant contig length: mean length in base-pairs of assembled contigs.

To ensure that the timepoint of floral and glandular trichome development selected for RNA isolation represented active terpene and cannabinoids biosynthesis, we examined the transcriptome for genes of these pathways. In general, the pooled transcriptome assembly included at least one full-length transcript corresponding for each known step in the cannabinoid and terpene biosynthetic pathways. The 200 most highly expressed genes in the trichome transcriptome included isoprenoid biosynthesis enzymes [HMBPP synthase (HDS), HMBPP reductase (HDR), IPP isomerase (IDI), and GPP synthase (GPPS)], six cannabinoid biosynthetic enzymes [CBDAS, PKS, OAC, THCAS, and two CBGA synthases (aPT1 and aPT4)], and seven CsTPS (**Table 3.6**). Three contigs annotated as fatty acid desaturase, which may be involved in biosynthesis of cannabinoid fatty acid precursors or cell membrane biosynthesis, were also highly expressed. Additionally, several contigs annotated as lipid transfer proteins or ABCG transporters were highly abundant.

cds	Afghan Kush	Blue Cheese	CBD Skunk Haze	Choc olope	Lemon Skunk	PFAM description	Top blast hit
DN129711_c30_g1_i1	859	248	192	146	320	Abhydrolase_1	epoxide hydrolase
DN165632_c17_g1_i1	2087	101	264	1150	5	adh_short	tropinone reductase-like
DN146504_c8_g1_i1	1829	25	521	1229	4	adh_short_C2	tropinone reductase-like
DN139341_c13_g1_i1	1292	1859	765	899	1220	ADH_zinc_N	quinone oxidoreductase
DN157558_c12_g1_i2	145	1434	158	713	107	ADH_zinc_N	alkenyl reductase
DN196707_c1_g1_i3	787	662	701	1031	1825	AgrB	protein transport protein
DN165608_c5_g1_i1	439	260	368	893	322	AP2	ethylene-responsive transcription factor
DN132620_c4_g1_i1	1398	4033	1743	3074	2481	Apolipoprotein	late embryogenesis abundant-like
DN156085_c39_g1_i1	535	456	650	862	575	Arf	ADP-ribosylation factor
DN122101_c5_g1_i1	876	601	871	1744	674	Auxin_repressed	SSR marker
DN128711_c1_g1_i1	2949	2700	1912	4240	2664	Auxin_repressed	auxin-repressed
DN160007_c0_g1_i1	**0**	**0**	**1110**	**0**	**1**	**BBE**	**CBDAS**
DN161020_c6_g2_i4	15	10	2834	20	20	BET	Transcription factor
DN130463_c1_g1_i1	18654	11900	6761	17749	212	Bet_v_1	Betv1-like protein
DN145653_c15_g1_i1	86	167	216	1340	238	Bet_v_1	MLP-like protein
DN129858_c7_g1_i2	11790	15538	14865	10081	7275	BURP	BURP domain
DN154353_c2_g1_i1	4276	523	2221	5969	31	CBP_BcsG	n/a
DN162095_c8_g1_i1	817	2290	1527	806	754	ChaC	gamma-glutanmylcyclotransferase
DN169291_c0_g1_i1	**13155**	**13257**	**10948**	**10247**	**9964**	**Chal_sti_synt_N**	**PKS**
DN166782_c5_g2_i2	2545	3434	2509	1707	2517	Chalcone_3	chalcone isomerase-like

DN145227_c4_g1_i1	894	1278	583	744	852	CHCH	coiled-coil helix domain
DN142377_c5_g1_i1	521	433	360	439	2050	Chloroa_b-bind	Chlorophyll a-b binding
DN168454_c82_g1_i1	1206	1066	787	2116	5978	Chloroa_b-bind	Chlorophyll a-b binding
DN114681_c37_g1_i1	937	565	241	517	459	Chorismate_bind	anthranylate synthase alpha subunit
DN133524_c26_g1_i1	1009	999	389	434	376	Citrate_bind	ATP-citrate synthase
DN144487_c5_g1_i2	0	0	1093	257	799	CMAS	tuberculostearic acid methyltransferase
DN163313_c6_g1_i1	1036	1060	933	1016	997	Cofilin_ADF	actin depolymerizing factor
DN160239_c28_g3_i1	806	562	526	653	922	CopG_antitoxin	ADP-ribosylation factor
DN166305_c3_g1_i1	688	910	607	586	863	COX6A	Cytochrome C oxidase subunit 6a
DN156589_c1_g1_i1	723	1270	641	893	923	COX7a	Cytochrome C oxidase subunit 7a
DN150760_c1_g2_i1	845	953	581	727	903	COX7C	Cytochrome C oxidase subunit 7c
DN184366_c20_g1_i1	1640	1069	1606	1538	1001	CP12	calvin cycle protein 12-3
DN162720_c3_g2_i1	221	77	93	177	937	Cu_bind_like	basic blue protein
DN129837_c57_g1_i1	1674	2820	1760	1562	1101	Cys_Met_Meta_PP	methionine gamma-lyase
DN116026_c6_g1_i1	332	289	453	573	1093	CYSTM	cysteine-rich and transmembrane domain-containing protein
DN132560_c1_g1_i1	1333	1726	1035	1349	978	Cyt-b5	Steroid-binding protein
DN152324_c10_g1_i2	815	1266	917	592	868	Cyt-b5	cytochrome b5
DN164168_c3_g1_i1	288	577	779	203	721	Cyt-b5	cytochrome b5
DN164196_c7_g1_i1	695	634	474	408	826	Cytochrom_C	cytochrome c2

DN129339_c48_g1_i3	**37917**	**48937**	**47472**	**44294**	**72003**	**Dabb**	**OAC**
DN147907_c10_g1_i1	1170	895	59	925	6	Dabb	Dabb
DN172946_c2_g2_i1	2201	3127	2677	9	2113	Dabb	Dabb
DN205190_c24_g1_i1	1971	1251	1135	1766	380	Dabb	Dabb
DN129456_c0_g1_i1	184	859	479	317	1843	Dehydrin	dehydrin Rab-18
DN158843_c3_g3_i2	695	1022	1022	1321	5922	DUF2052	coiled-coil helix domain
DN136667_c7_g1_i1	1159	1396	1311	968	1164	DUF3511	n/a
DN152446_c6_g1_i1	1225	1396	1078	1821	917	DUF3511	n/a
DN730_c0_g1_i1	1137	748	785	1110	1120	DUF3597	SPIRAL1-like
DN156377_c4_g1_i2	50	28	829	91	328	DUF454	mitochondrial sequence
DN108990_c0_g1_i1	1917	1564	951	1745	1872	DUF538	n/a
DN126068_c6_g1_i1	3986	4958	4417	2268	6056	DUF538	n/a
DN149102_c1_g1_i1	191	935	209	168	1046	DUF761	n/a
DN160067_c4_g2_i1	688	532	583	810	531	EF1_GNE	elongation factor 1-delta
DN111330_c0_g1_i1	867	829	614	760	778	EF-hand_1	calmodulin-7
DN145605_c2_g1_i1	1421	1544	1530	1788	1334	eIF-1a	translation initiation factor 1A
DN121789_c14_g1_i1	1204	993	883	847	973	eIF-5a	translation initiation factor 5A
DN128516_c45_g1_i1	766	1614	308	713	712	Epimerase	steroid 5-beta reductase-like
DN160364_c5_g1_i2	721	972	391	290	322	Epimerase	cinnamoyl-CoA reductase 1-like
DN157397_c14_g1_i1	2238	3665	756	2285	334	FA_desaturase	Fatty acid desaturase
DN157397_c14_g1_i2	623	1193	136	547	165	FA_desaturase	Fatty acid desaturase
DN164206_c5_g3_i1	1347	1168	1283	876	556	FA_desaturase	Fatty acid desaturase

DN146083_c39_g1_i1	**1578**	**2178**	**485**	**596**	**824**	**FAD_binding_4**	**THCAS**
DN138770_c11_g1_i1	5679	4496	3833	3588	2544	Fer2	ferredoxin
DN143484_c9_g1_i1	2395	2280	1126	1637	804	Ferritin_2	desiccation-related protein
DN154816_c5_g1_i1	681	1163	557	790	816	FKBP_C	peptidyl-prolyl cis-trans isomerase
DN151772_c107_g1_i1	**880**	**966**	**558**	**411**	**328**	**GcpE**	**HDS**
DN120482_c1_g1_i1	587	742	363	553	1041	Glutaredoxin	glutaredoxin
DN169834_c90_g1_i1	79	17	51	1023	334	Glyco_hydro_19	endochitinase
DN172860_c0_g1_i1	27508	20172	31917	30602	14936	Glyco_hydro_19	chitinase protein
DN155853_c3_g2_i1	1913	3061	1131	1598	1103	Glycolytic	fructose-bisphosphate aldolase
DN138219_c26_g2_i1	1450	1893	958	1108	849	Gp_dh_C	G3P dehydrogenase
DN125323_c105_g2_i1	6557	8797	4720	6252	100	GRP	glycine-rich protein-like
DN130463_c6_g1_i2	11472	20215	8646	13060	6947	GRP	glycine-rich protein-like
DN130463_c6_g1_i6	5369	3602	5732	12942	12762	GRP	glycine-rich protein-like
DN170644_c1_g1_i1	21094	20371	12569	22002	23234	GRP	glycine-rich protein-like
DN130748_c2_g1_i1	726	2402	791	482	1234	GST_N	glutathione S-transferase
DN157883_c15_g1_i1	1426	4140	809	745	1002	GST_N	glutathione S-transferase
DN158239_c7_g1_i2	1953	1310	1353	1463	947	GTP_EFTU	elongation factor 1-alpha
DN150608_c4_g1_i2	646	432	212	998	1298	HMA	heavy metal-associated isoprenylated plant protein
DN168069_c0_g1_i1	671	941	580	864	772	HMA	copper transport protein
DN138629_c1_g2_i1	18	16	7821	31	34	LCM	tRNA wybutosine-synthesizing protein
DN140245_c39_g1_i2	439	207	697	851	531	LEA_2	NDR1/HIN1-like protein

DN151650_c0_g1_i1	326	585	545	862	333	Linker_histone	histone H1
DN122983_c11_g1_i1	1200	988	750	842	595	LTP_2	Lipid transfer protein
DN204567_c44_g1_i1	**7801**	**9822**	**6406**	**7248**	**3738**	**LYTB**	**HDR**
DN145541_c19_g1_i1	4365	4538	4816	3216	1896	Malic_M	NADP-dependent malic enzyme
DN159391_c1_g2_i1	297	1919	681	1253	563	MAP65_ASE1	response to low sulfur 3-like
DN166116_c4_g3_i2	92	1163	240	459	520	Methyltransf_3	O-methyltransferase
DN135543_c14_g1_i1	1	0	1537	3949	3196	Methyltransf_7	carboxyl methyltransferase
DN160780_c19_g3_i1	0	0	670	1175	688	Methyltransf_7	carboxyl methyltransferase
DN168343_c0_g1_i3	696	439	374	624	855	Methyltransf_7	carboxyl methyltransferase
DN130326_c15_g1_i1	1009	1217	395	624	265	MIP	aquaporin PIP1
DN170377_c49_g1_i1	1106	887	626	614	506	Mito_carr	ADP, ATP carrier protein
DN139613_c1_g1_i8	538	40	378	61	821	MWFE	NADH dehydrogenase
DN130409_c37_g1_i1	892	597	399	252	84	NAD_binding_10	cinnamoyl-CoA reductase-like
DN169547_c1_g1_i1	915	604	518	551	529	NDK	nucleotide diphosphate kinase
DN127685_c5_g1_i1	1100	1016	845	817	1436	NDUF_C2	n/a
DN91264_c0_g1_i1	895	1091	742	672	1169	NDUFB10	NADH dehydrogenase
DN156039_c0_g1_i1	1794	193	887	811	803	NmrA	isoflavone reductase
DN89782_c0_g1_i1	**1677**	**2436**	**1379**	**1541**	**1165**	**NUDIX**	**IDI**
DN132777_c14_g1_i2	1505	867	53	470	270	p450	phenylalanine N-monooxygenase
DN157883_c56_g1_i1	1001	543	685	409	572	p450	Cytochrome P450 81
DN173164_c6_g1_i1	0	1063	0	1	324	p450	Cytochrome P450 71
DN163363_c4_g3_i2	788	51	82	1163	211	PALP	D-cysteine desulfhydrase
DN128591_c8_g1_i1	2965	1673	2766	3123	1807	PAM2	early response to dehydration

DN138064_c18_g1_i1	826	1311	785	1149	749	Peptidase_C1	cysteine protease
DN136269_c12_g1_i1	1004	759	1090	1888	463	Phi_1	exordium-like
DN132093_c34_g1_i1	**1201**	**1201**	**1143**	**697**	**950**	**polyprenyl_synt**	**GPPS ssu**
DN119145_c3_g1_i1	2498	1836	1872	1751	1444	PP-binding	acyl carrier protein
DN162708_c5_g2_i1	1358	1307	1000	893	837	PP-binding	acyl carrier protein
DN118844_c1_g1_i1	1205	1373	883	1193	936	Pro_isomerase	peptidyl-prolyl cis-trans isomerase
DN162118_c1_g1_i1	382	369	350	664	1232	PsbR	photosystem II 10 kDa polypeptide
DN160210_c0_g3_i1	959	933	776	849	764	Rad60-SLD	ubiquitin-like SMT3
DN163568_c11_g1_i1	907	1008	508	1097	579	RALF	RALF-like
DN171760_c2_g1_i2	684	916	766	981	861	RAMP4	stress-associated ER protein
DN119154_c0_g1_i1	2678	3412	1402	2256	2178	Redoxin	peroxiredoxin
DN162816_c22_g1_i2	1383	1124	656	880	539	Redoxin	peroxiredoxin
DN129337_c26_g1_i1	854	794	175	531	314	Retrotrans_gag	notch homolog
DN160836_c2_g1_i1	679	2361	1412	763	1373	Rhodanese	rhodanese-like
DN159066_c23_g1_i2	1090	854	628	1535	2561	RRM_1	RNA-binding protein-like
DN134368_c13_g1_i1	1871	1788	1814	2128	5794	RuBisCO_small	RuBisCO small chain
DN166482_c9_g1_i1	384	327	368	492	1057	RuBisCO_small	RuBisCO small chain
DN152566_c1_g1_i1	600	452	847	825	661	SAM_decarbox	SAM decarboxylase
DN149793_c3_g2_i1	100	17	8	8	989	SEO_C	sieve element occlusion B-like
DN164841_c4_g1_i3	1296	832	1420	2035	3863	Sin3_corepress	amphipathic helix protein Sin3-like
DN140073_c3_g2_i1	552	1034	312	436	765	SQAPI	cysteine protease inhibitor-like
DN167611_c16_g2_i1	2151	1567	1690	1996	1897	SQAPI	CPI-4
DN150676_c35_g1_i1	550	386	160	1165	235	SRF-TF	transcription factor cauliflower

DN126423_c21_g1_i1	693	735	787	699	1062	SUI1	protein translation factor
DN163345_c5_g1_i3	716	1077	257	713	241	SURNod19	nodulin19-like
DN152932_c2_g1_i1	1175	1100	645	1226	1058	TB2_DP1_HVA22	HVA22-like
DN129708_c2_g1_i1	3837	3244	4936	4600	4050	TCTP	translationally-controlled tumor protein
DN168454_c101_g3_i1	**0**	**332**	**1180**	**2**	**362**	**Terpene_synth**	**TPS3**
DN136607_c4_g1_i2	**992**	**884**	**534**	**1**	**915**	**Terpene_synth_C**	**TPS22**
DN139403_c0_g1_i1	**2**	**1575**	**3323**	**311**	**2625**	**Terpene_synth_C**	**TPS2**
DN144524_c85_g1_i1	**17**	**35**	**606**	**1083**	**1525**	**Terpene_synth_C**	**TPS25**
DN144524_c85_g2_i1	**921**	**422**	**288**	**195**	**80**	**Terpene_synth_C**	**TPS29**
DN144524_c85_g2_i2	**997**	**1718**	**836**	**303**	**644**	**Terpene_synth_C**	**TPS29**
DN147823_c6_g1_i1	**145**	**1**	**166**	**871**	**222**	**Terpene_synth_C**	**TPS23**
DN151772_c65_g1_i1	**849**	**1097**	**1007**	**2811**	**1211**	**Terpene_synth_C**	**TPS1**
DN163833_c5_g1_i1	**0**	**631**	**935**	**67**	**329**	**Terpene_synth_C**	**TPS2**
DN165677_c5_g1_i1	**5456**	**3455**	**2827**	**1768**	**4263**	**Terpene_synth_C**	**TPS29**
DN129268_c7_g1_i1	2448	3131	1149	2398	4224	Thioredoxin	thioredoxin H-type
DN156025_c2_g2_i1	870	991	495	646	571	Tim17	outer envelope pore protein
DN146225_c6_g1_i2	30	21	20	34	1049	TOM20_plant	mitochondrial import receptor subunit
DN135519_c1_g2_i1	906	944	756	781	485	Transferase	vinorine synthase-like
DN163532_c2_g1_i1	220	833	751	807	234	Transferase	shikimate O-hydroxycinnamoyltransferase-like
DN126421_c4_g1_i1	1338	1824	499	1225	524	Tryp_alpha_amyl	non-specific lipid-transfer protein

DN128417_c4_g1_i1	41261	46016	36969	23480	25458	Tryp_alpha_amyl	non-specific lipid-transfer protein
DN161509_c2_g1_i1	777	2699	127	3471	1547	Tryp_alpha_amyl	non-specific lipid-transfer protein
DN141479_c57_g2_i1	**465**	**951**	**165**	**223**	**158**	**UbiA**	**aPT1**
DN165097_c7_g1_i4	**1080**	**694**	**650**	**445**	**419**	**UbiA**	**aPT4**
DN166567_c10_g1_i2	**1656**	**1557**	**329**	**793**	**419**	**UbiA**	**aPT1**
DN146743_c0_g1_i3	869	659	690	1044	1040	ubiquitin	ubiquitin
DN147893_c2_g1_i3	592	1367	1327	926	480	ubiquitin	ubiquitin
DN127689_c40_g1_i1	665	321	150	1035	661	UDPGT	UDP-glycosyltransferase 71
DN158932_c6_g1_i1	37466	16656	23645	54358	92821	UIM	26S proteasome non-ATPase regulatory subunit 4
DN83372_c7_g1_i1	2546	555	342	0	280	UPF0506	n/a
DN126956_c5_g1_i1	1130	1213	1056	1346	891	UQ_con	Ubiquitin-conjugating enzyme E2
DN158066_c30_g1_i1	802	671	971	780	849	UQ_con	Ubiquitin-conjugating enzyme E2
DN155564_c7_g1_i3	344	485	772	707	985	XYPPX	glycine-rich protein
DN150670_c5_g1_i2	326	1222	589	917	1780	YfhO	n/a
DN127273_c15_g1_i1	807	871	1008	1299	674	Yippee-Mis18	yippee-like
DN158589_c1_g1_i1	1297	1482	1622	1981	1101	zf-A20	zinc finger A20 and AN1 domain-containing stress-associated
DN142910_c26_g1_i1	763	1011	652	808	818	zf-C2H2_4	methylene blue sensitivity
DN155768_c6_g1_i1	664	670	835	947	942	zf-C2H2_jaz	zinc finger protein 593

Table 3.6 Highly expressed contigs in five cannabis cultivars
cds: transcript ID. Values are mean normalized counts-per-million across three replicates for each cultivar listed. PFAM description: highest confidence PFAM domain match. Top blast hit shows the name of the best hit from blastx of the full translated cds. Entries in bold are known to be involved in resin biosynthesis.

PCA was done on the complete set of 15 trichome transcriptomes. The three replicates of each cultivar clustered together, and cultivars were well differentiated (**Figure 3.6a**). LS and Choc were the two cultivars most similar to each other, while AK had the most distance from the other cultivars. We used unsupervised cluster analysis to test for patterns of expression of contigs annotated as terpene or cannabinoid biosynthesis. Contigs were selected by mutual best tBLASTn hit against known sequences involved in isoprenoid and cannabinoid biosynthesis (**Figure 3.6b**). The 26 contigs identified as putatively involved in resin biosynthesis clustered into four groups. Contigs associated with the core MEP pathway (DXS, DXR, HDS, GPPS) cluster with cannabinoid biosynthetic genes acyl activating enzyme (AAE1), aPT1, aPT4, and THCAS. Mevalonate (MEV) pathway genes grouped into two clusters, which also included the MEP pathway gene methylerythritol phosphate cytidyltransferase (MCT), CBDAS, and an aPT4 contig. A second cluster of MEV contigs were much less highly expressed on average, and also included isopentenyl phosphate kinase (IPK) and CBDAS. The final cluster had the highest average expression levels, and included cannabinoid biosynthetic genes PKS and OAC, as well as the MEP pathway gene HDR and a version of IPK.

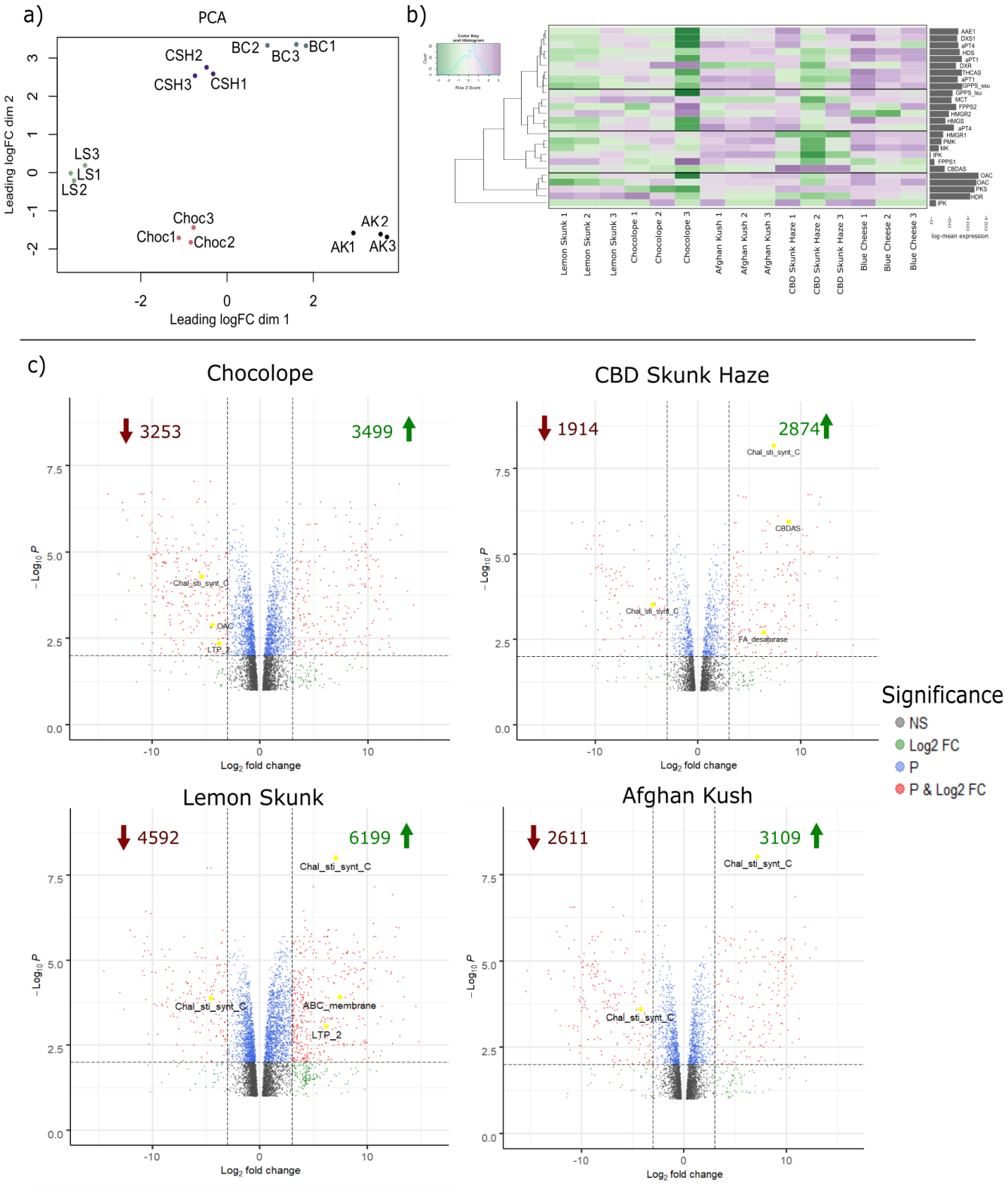

a)
PCA
CSH2
BC2
BC3
BC1
CSH3
CSH1
LS3
LS1
LS2
Choc3
Choc1
Choc2
AK2
AK1
AK3
Leading logFC dim 2
Leading logFC dim 1
b)
Lemon Skunk 1
Lemon Skunk 2
Lemon Skunk 3
Chocolope 1
Chocolope 2
Chocolope 3
Afghan Kush 1
Afghan Kush 2
Afghan Kush 3
CBD Skunk Haze 1
CBD Skunk Haze 2
CBD Skunk Haze 3
Blue Cheese 1
Blue Cheese 2
Blue Cheese 3
log-mean expression
c)
Chocolope
3253
3499
CBD Skunk Haze
1914
2874
Chal_sti_synt_C
CBDAS
FA_desaturase
OAC
LTP_2
Lemon Skunk
4592
6199
ABC_membrane
Afghan Kush
2611
3109
Log2 fold change
−Log10 P
Significance
NS
Log2 FC
P
P & Log2 FC

Figure 3.6 Gene expression in floral trichomes of five cannabis cultivars
(a) Whole transcriptome principal component analysis (PCA) of the first two dimensions (Dim). **(b)** Heatmap and expression of contigs representing genes annotated as terpene or cannabinoid biosynthesis. Colors indicate row-wise Z-score, standard deviations from the mean. Grey bars (right) show the average log2 counts-per-million (cpm) across 24 samples, eight individuals with three technical replicates. For the bar diagram, Choc 3 was treated as an outlier and not included in the log-mean expression results. **(c)** Volcano plots showing differentially expressed contigs for four cultivars compared to Blue Cheese (BC). Grey: not significant (ns); green (Log2FC): significant at a log2 fold change of 2; blue (P): significant at an adjusted p-value of 0.05; red (P & Log2 FC): significant by both fold change and adjusted p-value. Contigs labeled with names are those shown with yellow diamonds, representing transcripts that may be associated with resin biosynthesis Green numbers: the number of transcript contigs with abundance significantly higher compared to BC in each cultivar; red numbers: the number of transcript con□gs significantly lower compared to BC in each cultivar. Abbreviations are: AAE1: Acyl activating enzyme; PKS: Polyketide synthase; OAC: Olivetolic acid cyclase; aPT1 and aPT4: Aromatic prenyltransferases; DXSa: 1-Deoxy-D-xylulose 5-phosphate (DXP) synthase; DXR: DXP reductase; MCT: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; HDS: (E)-4-Hydroxy- 3-methyl-but-2-enyl pyrophosphate (HMB-PP) synthase; HDR: HMB-PP reductase; GPPS lsu: Geranyl diphosphate synthase large subunit; GPPS ssu: Geranyl diphosphate synthase small subunit; HMGS: 3-Hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; HMGR: HMG-CoA reductase; MK: Mevalonate kinase; PMK: Mevalonate-3-phosphate kinase; FPPS: Farnesyl diphosphate synthase; IPK: Isopentenyl phosphate kinase.

Next, we performed a differential gene expression analysis across the five cultivars with all contigs that had expression levels of at least 100 counts-per-million (cpm). BC was used as the reference, because it placed near the center of the PCA of foliar terpene variation (**Figure 3.4a**), and the other four cultivars were compared against BC. Differential gene expression analysis was performed with an adjusted p-value cut-off of 0.05 and a $\log_2$ fold-change cut-off of 3. In total, across the cultivars 25,218 contigs were differentially expressed relative to BC; 19,987 were upregulated and 19,263 were downregulated. Contigs were identified with at least 95% identity to known enzymes involved in cannabinoid and terpene biosynthesis, but most were not significantly differentially expressed in any cultivar compared to BC (**Figure 3.6c**). Most notably, CBDAS was highly upregulated in CSH, the only cultivar to produce CBD as a major cannabinoid. CsPT1 was downregulated in CSH, and CsPT4 was downregulated in LS. HMGR, a component of the MEV pathway, was upregulated in LS. OAC was downregulated in Choc.

3.4.5 *CsTPS* gene discovery

For discovery and quantification of *CsTPS* transcripts, separate transcriptomes were assembled for each cultivar (**Table 3.5)**. While separate transcriptomes do not permit quantitative comparison between cultivars, they eliminate the risk of quantitation errors from mapping kmers from similar transcripts between cultivars. We used the RNA-Bloom assembler (Nip et al., 2019), designed for single-cell RNA-seq libraries, to capture the diversity of sequences across the five cultivars while reducing the possibility of chimeric contigs. Contigs with greater than 98% predicted amino acid sequence identity were collapsed under the longest representative sequence. These five single-cultivar transcriptomes and the previously published PK trichome transcriptome (van Bakel et al., 2011) were searched with BlastX to identify known (Gunnewich et al., 2007; Booth et al., 2017; Livingston et al., 2019; Zager et al., 2019) as well as new *CsTPS* sequences. Sequences representing all but three of the previously functionally characterized and unique *CsTPS* (a total of 18; **Table 3.2**) were present in the transcriptomes of at least one of the five cultivars (Booth et al., 2017; Allen et al., 2019; Livingston et al., 2019; Zager et al., 2019). The three missing CsTPS were *CsTPS13*, *CsTPS14* and CsTPS33. When we screened the transcriptomes of the six different cultivars, we found a total of 33 unique and apparently full-length CsTPS sequences, including two that we annotated as copalyl diphosphate synthase (CsTPS65) and *ent*-kaurene synthase (CsTPS66) of gibberellin biosynthesis.

A phylogeny of the predicted amino acid sequences of the 33 CsTPS together with TPS from other plant species placed CsTPS into the subfamilies TPS-a, TPS-b, TPS-c, TPS-e/f, and TPS-g (**Figure 3.7, Table 3.7)**. Within the TPS-a subfamily, all CsTPS fall into one cluster with TPS from hop (*Humulus lupulu*s) as the nearest non-cannabis members. Within TPS-b, the CsTPS fall into two clades, which we named the CsTPS-b1 and the CsTPS-b2 clades, with two

hop monoterpene synthases as nearest relatives. Three CsTPS fall into TPS-g, but do not cluster together. TPS-c and TPS-e/f each contain one CsTPS.

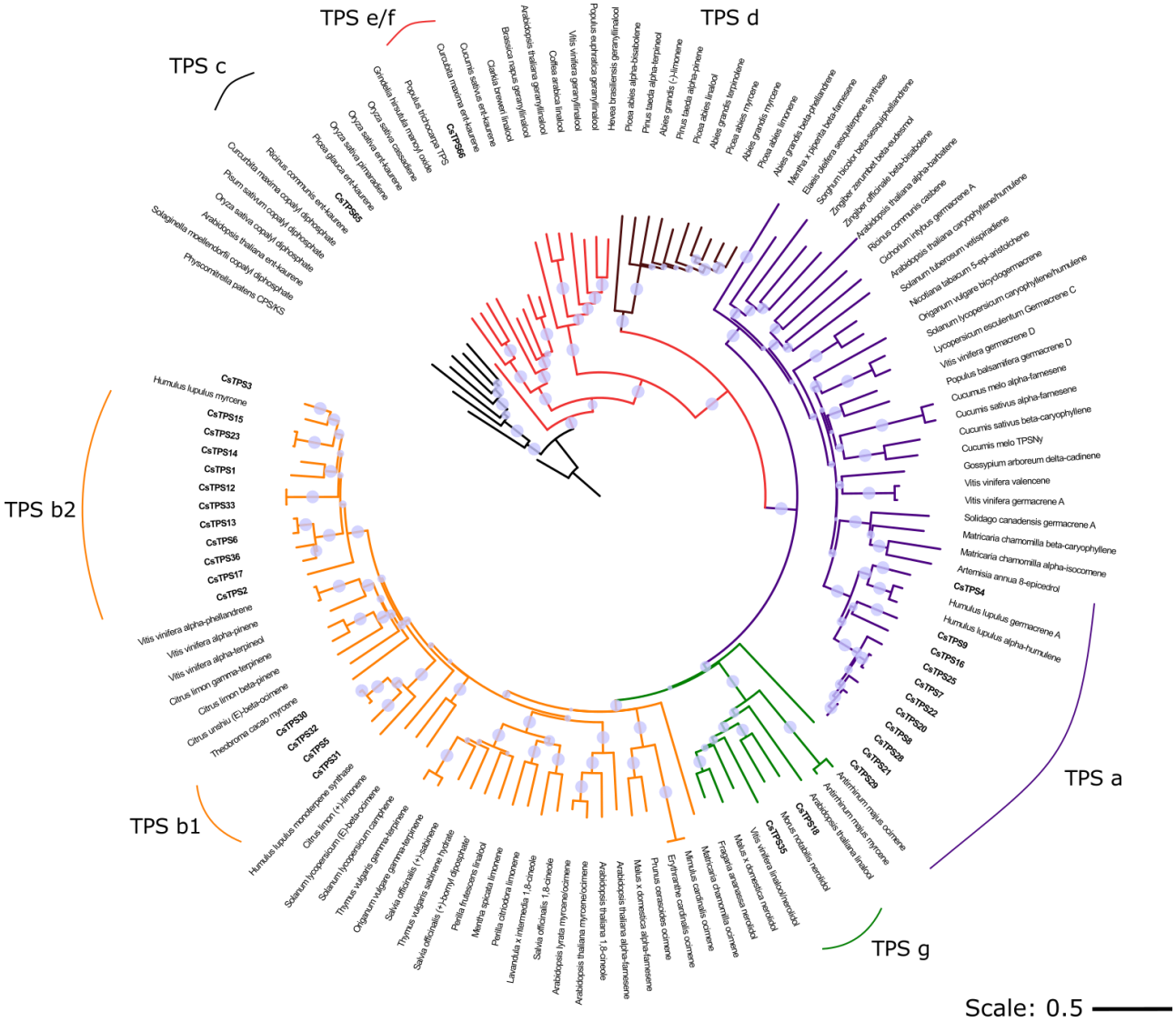


Figure 3.7 Maximum likelihood phylogeny of CsTPS relative to other plant TPS
CsTPS are in bold. The size of purple dots represents the size of bootstrap values from 100 bootstrap replicates. TPS subfamilies are color coded: TPS-a (purple), TPS-b (orange), TPS-d (brown), TPS-c (black), TPS-e/f (red), TPS-g (green). Bows show the location of CsTPS within the subfamilies. Tree scale 0.5 represents 50% sequence difference.

Description	NCBI Accession number
Abies_grandis_(-)-limonene	O22340.1
Abies_grandis_beta-phellandrene	Q9M7D1.1
Abies_grandis_myrcene	O24474.1
Abies_grandis_terpinolene	Q9M7D0.1
Antirrhinum_majus_myrcene	Q84ND0.1
Antirrhinum_majus_ocimene	Q84NC8.1
Arabidopsis_lyrata_myrcene_ocimene	XP_002878731.1
Arabidopsis_thaliana_1,8-cineole	NP_189210.2
Arabidopsis_thaliana_alpha-barbatene	NP_199276.1
Arabidopsis_thaliana_alpha-farnesene	NP_567511.3
Arabidopsis_thaliana_Ent-kaurene	NP_192187.1
Arabidopsis_thaliana_geranyllinalool	NP_564772.1
Arabidopsis_thaliana_linalool	NP_176361.2
Arabidopsis_thaliana_myrcene_ocimene	NP_179998.1
Artemisia_annua_8-epicedrol	Q9LLR9.1
Brassica_napus_geranyllinalool	CDY12887.1
Cichorium_intybus_Germacrene_A	Q8LSC3.1
Citrus_limon_(+)-limonene	Q8L5K3.1
Citrus_limon_beta-pinene	AAM53945.1
Citrus_limon_gamma-terpinene	Q8L5K4.1
Citrus_unshiu_(E)-beta-ocimene	BAD91046.1
Clarkia_breweri_linalool	Q96376.1
Coffea_arabica_linalool	XP_027111732.1
Cucumis_melo_TPSNy	NP_001284382.1
Cucumis_sativus_alpha-farnesene	NP_001267674.1
Cucumis_sativus_beta-caryophyllene	NP_001292628.1
Cucumis_sativus_ent-kaurene	NP_001292675.1
Cucumis_melo_alpha-farnesene	NP_001284384.1
Curcubita_maxima_ent-kaurene	XP_022968895.1
Curcurbita_maxima_copalyl_diphosphate	BAC76429.1
Elaeis_oleifera_sesquiterpene_synthase	AAC31570.2
Erythranthe_cardinalis_ocimene	AHI50307.1
Fragaria_ananassa_nerolidol	P0CV94.1
Gossypium_arboreum_delta-cadinene	NP_001316949.1
Grindelia_hirsutula_manoyl_oxide	AGN70886.1
Hevea_brasiliensis_geranyllinalool	XP_021650921.1
Humulus_lupulus_alpha-humulene	B6SCF5.1
Humulus_lupulus_germacrene_A	B6SCF6.1
Humulus_lupulus_monoterpene_synthase	B6SCF3.1
Humulus_lupulus_myrcene	B6SCF4.1
Lavandula_x_intermedia_1,8-cineole	AFL03421.1
Lycopersicum_esculentum_Germacrene_C	NP_001234055.1

Malus_x_domestica_alpha-farnesene	Q84LB2.2
Malus_x_domestica_nerolidol (3S,6E)-nerolidol	NP_001280833.1
Matricaria_chamomilla_alpha-isocomene	I6R4V5.1
Matricaria_chamomilla_beta-caryophyllene	I6RAQ6.1
Matricaria_chamomilla_ocimene	I6RE61.1
Mentha_piperita_beta-farnesene	O48935.1
Mentha_spicata_limonene	AAC37366.1
Mentha_x_piperita_beta-farnesene	O48935.1
Mimulus_cardinalis_ocimene	AHI50307.1
Morus_notabilis_nerolidol (3S,6E)-nerolidol	XP_010098579.1
Nicotiana_tabacum_5-EAS	Q40577.3
Origanum_vulgare_bicyclogermacrene	E2E2N7.1
Origanum_vulgare_gamma-terpinene	E2E2P0.1
Oryza_sativa_cassadiene	XP_015622683.1
Oryza_sativa_copalyl_diphosphate OsCPS1	XP_015624005.1
Oryza_sativa_ent-kaurene	Q0JA82.1
Oryza_sativa_pimaradiene	XP_015633583.1
Perilla_citriodora_limonene	AAF65545.1
Perilla_frutescens_linalool	AAL38029.1
Physcomitrella_patens_CPS_KS	XP_024380398.1
Picea_abies_alpha-bisabolene	AAS47689.1
Picea_abies_limonene	AAS47694.1
Picea_abies_linalool	AAS47693.1
Picea_abies_myrcene	AAS47696.1
Pinus_taeda_alpha-pinene	Q84KL3.1
Pinus_taeda_alpha-terpineol	Q84KL4.1
Pisum_sativum_copalyl_diphosphate	O04408.1
Populus_balsamifera_germacrene_D	AAR99061.1
Populus_euphratica_geranyllinalool	XP_011027906.1
Populus_trichocarpa_TPS	ALM22925.1
Prunus_cerasoides_ocimene	AIC76498.1
rabidopsis_thaliana_caryophyllene_humulene	NP_197784.2
Ricinus_communis_casbene	XP_002513340.1
Ricinus_communis_ent-kaurene	XP_002520733.1
Salvia_officinalis_(+)-bornyl_diposphate	O81192.1
Salvia_officinalis_(+)-sabinene	O81193.1
Salvia_officinalis_1,8-cineole	O81191.1
Solaginella_moellendorfii_copalyl_diphosphate	XP_002960350.2
Solanum_lycopersicum_(E)-beta-ocimene	NP_001308094.1
Solanum_lycopersicum_camphene	NP_001295307.1
Solanum_lycopersicum_caryophyllene_humulene	NP_001234766.1
Solanum_tuberosum_vetispiradiene	Q9XJ32.1
Solidago_canadensis_Germacrene_A	CAC36896.1

Sorghum_bicolor_beta-sesquiphellandrene	XP_002443927.2
Theobroma_cacao_myrcene	EOX92383.1
Thymus_vulgaris_gamma-terpinene	AFZ41788.1
Thymus_vulgaris_sabinene_hydrate	AGA96120.1
Vitis_vinifera_alpha-phellandrene	ADR74201.1
Vitis_vinifera_alpha-pinene	NP_001268167.1
Vitis_vinifera_alpha-terpineol	Q6PWU2.1
Vitis_vinifera_geranyllinalool	NP_001268201.1
Vitis_vinifera_germacrene_A	CBI31269.3
Vitis_vinifera_germacrene_D	NP_001268213.1
Vitis_vinifera_linalool_nerolidol	ADR74212.1
Vitis_vinifera_valencene	CBI31262.3
Zingiber_officinale_beta-bisabolene	D2YZP9.1
Zingiber_zerumbet_beta-eudesmol	B1B1U4.1

Table 3.7 Accession numbers of TPS used to construct phylogeny
Description is the organism and function of the enzyme, NCBI accession numbers are listed in the second column.

3.4.6 *CsTPS* gene expression in five different cultivars

We used the separate trichome transcriptome assemblies to determine for each cultivar expression of *CsTPS* and to correlate *CsTPS* gene expression and terpene profiles in each of the five different cultivars. The analysis was limited to predicted CsTPS sequences of 400 aa or longer to reduce quantification ambiguity, which allowed expression analysis for 18 different *CsTPS* (**Figure 3.8**). Transcript abundance was calculated within each cultivar relative to mean transcripts-per-million (tpm) values of all contigs across three clonal replicates. Each of the 18 different CsTPS was highly expressed in at least one cultivar (**Figure 3.8**). *CsTPS18*, *CsTPS29*, and *CsTPS35*, which belong to the TPS-g subfamily, were the only *CsTPS* with above-mean transcript abundance in all five cultivars.

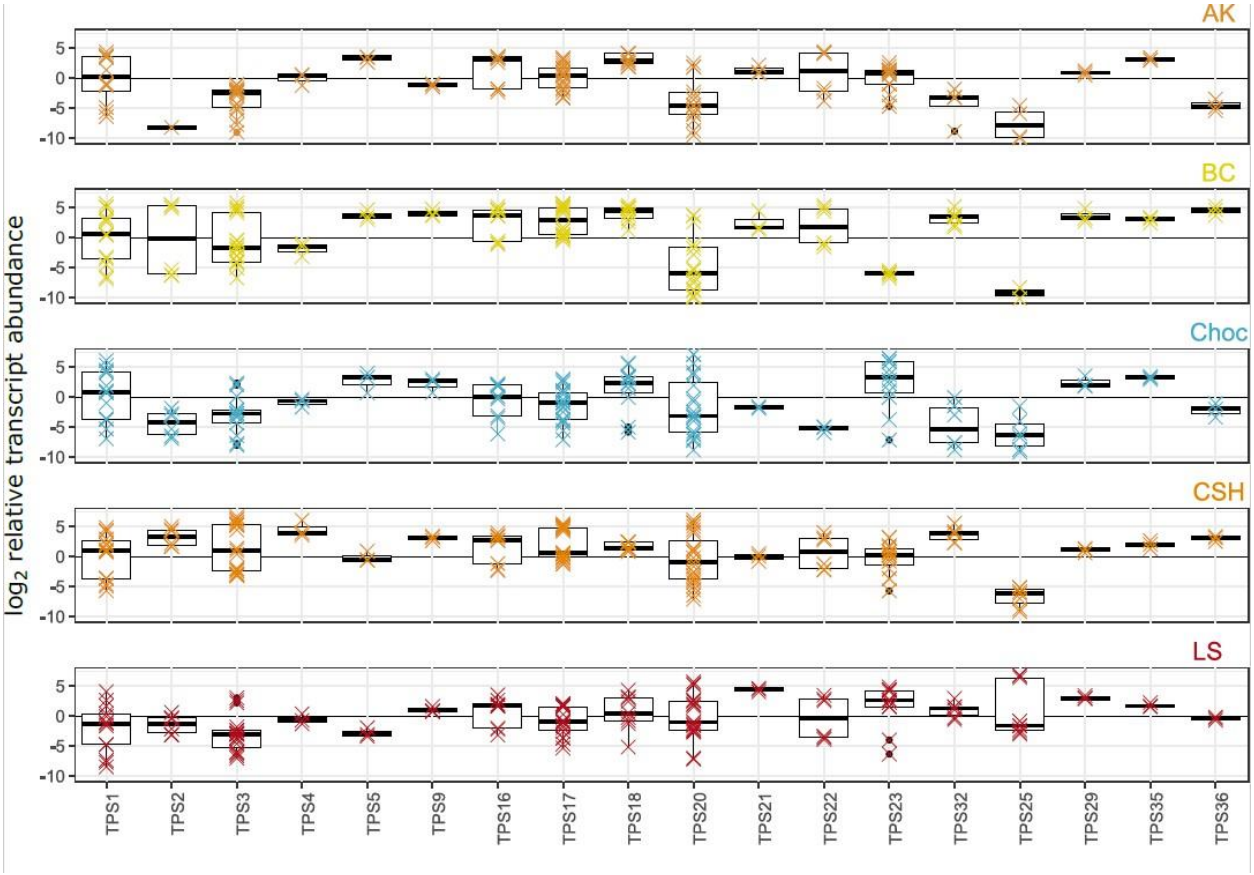


Figure 3.8 Transcript abundance of CsTPS genes in floral trichomes of different cannabis cultivars
Values are log2 fold-difference compared to average counts-per-million (cpm) for each cultivar. Colored "X" markskshow individual data points, black box plots show quartiles and outliers.

In AK trichomes, *CsTPS5*, *CsTPS16*, *CsTPS18*, and *CsTPS35* showed the highest expression, while *CsTPS2* and *CsTPS25* transcripts were barely detected. Expression of *CsTPS3*, *CsTPS4*, *CsTPS9*, *CsTPS17*, *CsTPS20*, *CsTPS21*, *CsTPS22*, *CsTPS23*, *CsTPS29*, *CsTPS32* and *CsTPS36*, were similar to the mean trichome transcript abundance, defined as within 4-fold $\log_2$ cpm of the mean. In BC, *CsTPS5*, *CsTPS36*, *CsTPS18*, and *CsTPS9* were highly expressed. *CsTPS25* transcripts were barely detected, and *CsTPS23* and *CsTPS20* transcript levels were low relative to mean trichome transcript abundance. The other 11 *CsTPS* were expressed at similar levels to the mean transcript abundance. In Choc trichomes, *CsTPS35* was the most highly expressed *CsTPS*. *CsTPS22*, *CsTPS25*, and *CsTPS32* transcripts were detected at low levels. The

remaining 14 *CsTPS* were expressed at levels similar to the mean trichome transcript abundance. In CSH, the most highly expressed *CsTPS* transcripts were *CsTPS4* and *CsTPS32*. *CsTPS25* was detected at low levels, with the remaining 15 *CsTPS* expressed similarly to the mean trichome transcript abundance. LS was the only cultivar with above-mean expression of *CsTPS25* and the only one with below-mean expression of *CsTPS1*. The most highly expressed *CsTPS* in LS was *CsTPS21*. Aside from *CsTPS21*, all *CsTPS* in LS were similar to mean transcript abundance.

3.4.7 Functions of CsTPS

The *CsTPS* phylogeny suggests that all but two, *CsTPS65* and *CsTPS66*, encode mono-TPS or sesqui-TPS enzymes (**Figure 3.7**). For functional characterization, CsTPS enzymes were produced in and purified from *E. coli*, assayed with GPP and FPP as substrates, and products identified by GC/MS (**Figure 3.9, Figure 3.14, Figure 3.15**). We identified 11 CsTPS members of the TPS-a subfamily. Of these, six were previously characterized (Booth et al., 2017; Zager et al., 2019). Zager and colleagues reported the identities of CsTPS16, a germacrene B synthase, and CsTPS20, a hedycaryol synthase. We were able to confirm the activities of both of these enzymes as germacrene B and hedycaryol synthases respectively, using CsTPS16 (cloned from PK) and CsTPS20(PK) (**Figure 3.16**). CsTPS28 is closely related to CsTPS20 (**Figure 3.7**). With FPP as a substrate, CsTPS28(PK) produced β-elemene (28%), a eudesmane-type compound (25%) with RI 1504.56, α-selinene (23%), and a cadinane-type sesquiterpene (17%) with RI 1597.67. The most diverse product profile in the TPS-a subfamily was detected with CsTPS22(PK), which produces 13 different sesquiterpenes and himachalane (20%) as the major product. Other products were a eudesmane type sesquiterpene (10%) with RI 1504.56, an unidentified sesquiterpene (10%) with RI 1528.12 and base peak 121, a cadinane type

sesquiterpene (7%) with RI 1497.67, eudesma-3,7(11)-diene (7%), cubenol (6%), a eudesmane type compound (6%) with RI 1557.15, a sesquiterpene alcohol (6%) with RI 1715.56 and base peak 206.4, γ-eudesmol (4%), an unidentified compound (4%) with RI 1552.33 and base peak 161, α-humulene (3%), nerolidol (3%), and β-elemene (2%). CsTPS25(LS) produced (*E*)-β-farnesene (56%) as its major product, as well as a cadinane type compound (22%) with RI 1494.76, (*Z,E*)-α-farnesene (15%), and nerolidol (7%). CsTPS24 clusters together with CsTPS25 and CsTPS16. No products were found in enzyme assays with CsTPS24 with either GPP or FPP.

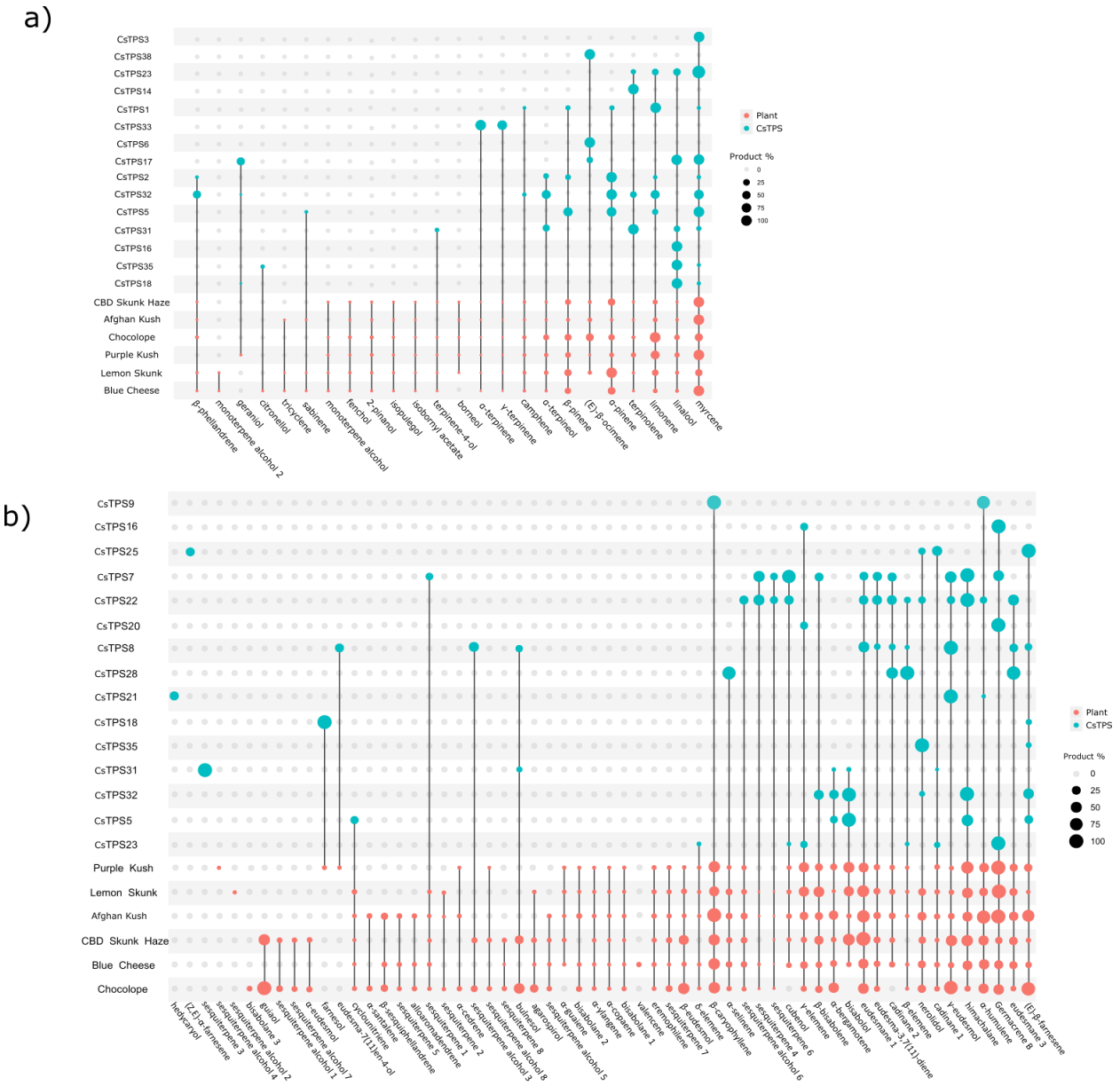


Figure 3.9 Products of functionally characterized CsTPS and their representation in cannabis floral trichome terpene profiles of different cultivars
(a) Monoterpenes. (b) Sesquiterpenes. CsTPS IDs and cultivar names are shown on the y axis; compounds on the x axis. Dot size corresponds to the percentage of each compound compared to the most abundant product of a given CsTPS (blue dots) or floral metabolite (pink dots).

Within the TPS-b subfamily, the CsTPS-b1 clade contains three members that had not been previously described, specifically CsTPS17, CsTPS23, and CsTPS36. CsTPS17(BC) was functionally characterized as a mono-TPS that produced myrcene (34%) and linalool (34%) as equal major products along with minor products geraniol (16%), (*E*)-β-ocimene (8%) and α-

terpineol (9%). CsTPS23(LS) also is a mono-TPS that produced myrcene (53%) as its major product and the minor products (20%), limonene (15%), and terpinolene (12%). We were unable to obtain a function for CsTPS36. The CsTPS-b2 clade contains four members, CsTPS5, CsTPS30, CsTPS31 and CsTPS32. Unlike clade b1 members, these CsTPS do not possess the predicted plastidial target peptides that are typical of plant mono-TPS. CsTPS5 (cultivar Finola) and CsTPS30(PK) were previously characterized as myrcene synthases (Booth et al., 2017), while CsTPS31 and CsTPS32 had not been previously characterized. Given the lack of target peptides, CsTPS5, CsTPS31 and CsTPS32 were assayed here with both GPP and FPP. CsTPS30 was previously characterized as a myrcene synthase from PK (Booth et al., 2017). In assays with GPP, the major product of CsTPS5 (PK; 96% aa identity to CsTPS5 from Finola) was α-pinene (33%) with less abundant products myrcene (18%), α-terpineol (18%), limonene (17%) and β-pinene (14%). When assayed with FPP, CsTPS5(PK) produced mainly bisabolol (46%), as well as himachalane (27%), (*E*)-β-farnesene (11%), α-bergamotene (7%), and a compound tentatively identified as a cyclounitriene (9%). CsTPS31(PK) produced terpinolene (57%) as major product with GPP, as well as α-terpineol (19%), linalool (14%), β-pinene (6%) and terpinen-4-ol (4%). Using FPP as a substrate, the major product (91%) of CsTPS31(PK) was an unknown sesquiterpene with RI 1915.54 and base peak 93. It also produced 6% bulnesol, 2% bisabolol, and trace amounts of α-bergamotene and a cadinane type sesquiterpene with RI 1494.18. CsTPS32(PK) produced eight different monoterpenes from GPP, geraniol (23%), α-pinene (20%), myrcene (16%), limonene (13%), β-phellandrene (10%), terpinolene (5%), α-terpineol (13%), and camphene (1%). With FPP, CsTPS32(PK) produced himachalane (32%), bisabolol (31%), (*E*)-β-farnesene (14%), β-bisabolene (12%), α-bergamotene (10%), and nerolidol (2%).

CsTPS18 and CsTPS19, members of the TPS-g subfamily that differ by one aa, were recently reported by Zager et al. (2019) while this work was in preparation. Here we refer to homologues of these genes (>95% aa identity) as CsTPS18. We confirmed CsTPS18 as a linalool/nerolidol synthase, with CsTPS18(Choc) producing exclusively (-)-linalool (**Figure 3.15)**. We functionally characterized TPS-g subfamily members CsTPS35 and CsTPS29. CsTPS35(LS) produced acyclic terpenes with both GPP and FPP. With GPP, it produced mostly linalool (93%), with minor amounts of citronellol (5%) and myrcene (2%). Using FPP, CsTPS35(LS) produced nerolidol (95%) and (*E*)-β-farnesene (5%). CsTPS29(BC) produced exclusively linalool from GPP. No products were detected when CsTPS29 was assayed with FPP.

3.5 Discussion

3.5.1 The *CsTPS* gene family

Previous estimates of the size of the *CsTPS* family varied from approximately 30 to 50 different genes (Booth et al., 2017; Allen et al., 2019). The present analysis of the PK reference genomes identified 19 complete and five partial *CsTPS* genes. The transcriptomes reported here and in two recent studies (Booth et al., 2017; Zager et al., 2019) cover 11 different cannabis cultivars. Screening of these cultivar-specific transcriptomes for *CsTPS* genes revealed variations of the *CsTPS* gene family, variations of *CsTPS* transcript expression, and variations of CsTPS enzyme functions with respect to their mono- and sesquiterpene products. Among the different cultivars, some *CsTPS* genes were more variable in their transcriptome representation across cultivars than others. For example, the *CsTPS*9 gene, which encodes a β-caryophyllene/α-humulene synthase, was expressed in the transcriptomes of all cultivars reported to date, including the present study.

The same is the case for *CsTPS5*, which encodes an enzyme that uses both GPP and FPP and produces respectively multiple monoterpenes and sesquiterpenes. By contrast, the *CsTPS2* gene, which encodes an α-pinene synthase, was not found in the PK genome and in the PK and AK transcriptomes, but was present in the transcriptomes of other cultivars. *CsTPS8*, which encodes a multi-product sesquiterpene synthase, was only detected in transcriptomes of Finola (FN), Choc, and CSH. Considering these variations, which are based on the analysis of 11 different cultivars, we expect that the full suite of *CsTPS* genes that differ by sequence, expression and functions, and which contribute to different terpene profiles, will be substantially larger across the many cannabis cultivars that exist around the world. In the present study, we used a conservative cut-off of 95% aa identity for assigning sequences to the same CsTPS identifier to avoid separating minor variants. However, it should be noted that even at this cut-off, minor sequence variation may result in variation of enzyme function. Assigning unique gene identifiers to transcript sequences based on 100% identity would result in a larger number of apparently different CsTPS (Allen et al., 2019; Zager et al., 2019).

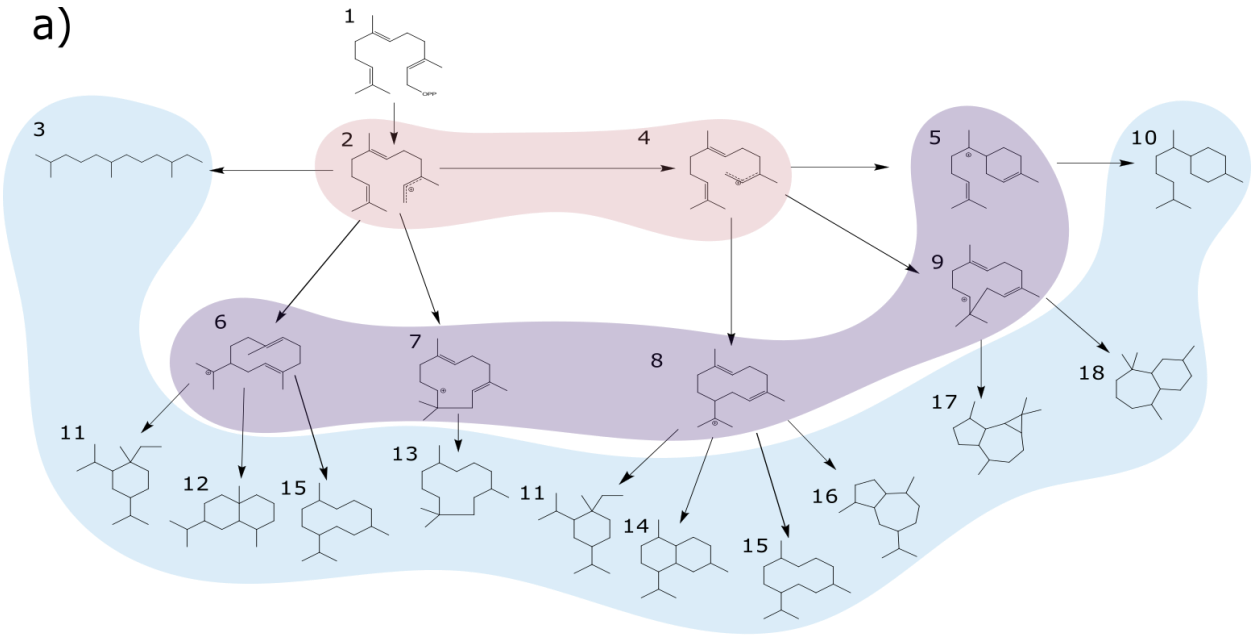
Within the plant TPS phylogeny, CsTPS of the TPS-a and TPS-b subfamilies cluster with TPS sequences of its close relative hop (**Figure 3.7**). In both subfamilies, we found cannabis-specific expansions, suggesting that the present diversity of CsTPS for mono- and sesquiterpene biosynthesis may have resulted from progressive and relatively recent multiplications of a few ancestral CsTPS. Two distinct CsTPS blooms exist in the TPS-b subfamily, identified here as CsTPS-b1 and CsTPS-b2. The CsTPS-b2 group includes four members, of which all but one (CsTPS30) lack a predicted plastid target peptide. Two of these TPS produce bisabolol, a sesquiterpene found in many cannabis cultivars, but which is not a product of any of the functionally characterized CsTPS-a group enzymes. These results suggested that indeed some

CsTPS-b members contribute to sesquiterpene production in cannabis trichomes, although the TPS-b group has been previously described as including mostly mono-TPS in other plant species (Chen et al., 2011). In sandalwood (*Santalum* spp.) a member of TPS-b also functions as a sesquiterpene synthase (Jones et al., 2011). It is striking that in both cannabis and sandalwood, these TPS-b members produce bisabolane-type sesquiterpenes, which may be due to similar routes of active site evolution from their respective monoterpene synthase ancestors (Gao et al., 2012).

Relatedness of CsTPS functions

The clades of CsTPS provided an opportunity to assess if different products of closely related CsTPS may arise through similar cyclization cascades. We tested this hypothesis with a focus on sesqui-TPS because of their commonly complex cyclization cascades. The sesquiterpenes identified in the different cannabis cultivars of this study, including PK, belong to 11 sesquiterpene parent skeletons that may originate from six central carbocationic intermediates (**Figure 3.10a**) (Degenhardt et al., 2009). The farnesane, elemane, and germacrene sesquiterpenes of the cannabis resin may be formed by CsTPS via either a farnesyl or nerolidyl cation. The eudesmane and humulane sesquiterpenes, which were abundant in the leaf and flower metabolite profiles, most likely arise from (*E*,*E*)-germacranedienyl and (*E*,*E*)-humulyl cations, respectively, which are formed by 10,1 or 11,1 closure of the farnesyl cation. Four different types of cannabis sesquiterpenes may be formed via the (*Z*,*E*)-germacranedienyl cation, the elemane and germacrene compounds, and the cadinene and guaiane skeletons. The nerolidyl cation can also cyclize into the (*Z*,*E*)-humulyl cation via 11,1 closure, leading to formation of the himachalane and aromadendrane sesquiterpenes. While these latter compounds are generally

present in cannabis terpene profiles, they were not abundant in the cultivars of this study. The bisabolane sesquiterpenes are likely formed from the bisabolane carbocation, which is generally the result of 6,1 closure of the nerolidyl cation.



b)

	Intermediates	Major product	Minor products
CsTPS35	2	3	3
CsTPS18	2	3	3
CsTPS31	5, 8, 9		10,17,14
CsTPS5	2, 4, 5, 9	10	17,18,3
CsTPS32	2, 4, 5, 9	10	18,3
CsTPS16	2, 6	12	12
CsTPS9	2, 7	13	13
CsTPS25	2, 8	3	14
CsTPS7	2, 4, 6, 9	14	12,17
CsTPS22	2, 4, 6, 9	17	15,13,14,12
CsTPS20	2, 4, 8	15	11
CsTPS8	2, 6	12	12
CsTPS28	2, 6	11	12,14
CsTPS21	2, 6	12	15

Figure 3.10 Proposed routes of sesquiterpene formation by CsTPS and correlation with CsTPS sequence relatedness

(a) Schematic of carbocation intermediates and sesquiterpene classes (according to (Degenhardt et al., 2009)) for sesquiterpenes identified in cannabis floral trichomes. Germacrane skeletons rearrange to elemane (Setzer, 2008). (b) Intermediates and major products of CsTPS described in this paper. Intermediates include all major proposed cationic intermediates, and major product is the class of the most abundant sesquiterpene product of each enzyme. 1: (*E*,*E*)-farnesyl diphosphate, 2: (*E*,*E*)-farnesyl cation, 3: farnesane skeleton, 4: nerolidyl cation, 5: bisabolyl cation, 6: (*E*,*E*)- germacranedienyl cation, 7: (*E*,*E*)-humulyl cation, 8: (*Z*,*E*)-germacranedienyl cation, 9: (*Z*,*E*)- humulyl cation, 10: bisabolane skeleton, 11: elemane skeleton, 12: eudesmane skeleton, 13: humulane skeleton, 14: cadinane skeleton, 15: germacrane skeleton, 16: guaiane skeleton, 17: aromadendrane skeleton, 18: himachalane skeleton.

We attempted to correlate CsTPS positions in the TPS phylogeny (**Figure 3.7**) with their assumed cyclization reactions (**Figure 3.10b**). CsTPS18 and CsTPS35 are related enzymes that each produce acyclic farnesane compounds. CsTPS5, CsTPS31, and CsTPS32 are related enzymes in the CsTPS-b1 group and share many of the same products and likely the same intermediates bisabolyl and (*Z*,*E*)-humulyl cations. The more closely related CsTPS5 and CsTPS32 share three of four of the same product skeletons and are likely to share the same four potential intermediates. CsTPS8(FN), CsTPS28(PK) and CsTPS21(PK) share only four products between them, but their major and secondary products could all be formed from the (*E*,*E*)-germacranedienyl cation. Similarly, CsTPS7(FN) and CsTPS22(PK), which are closely related, may also share three of four intermediate carbocations. CsTPS25(LS), which groups with CsTPS that have mostly cyclic primary products, produces predominantly acyclic sesquiterpenes. Its secondary product, however, is a cadinane sesquiterpene, which may be a result of its recent evolution from cyclic sesqui-TPS. Overall, we found that similar proposed cyclization routes are more commonly shared between closely related CsTPS than between more distantly related CsTPS.

3.5.2 Assessing *CsTPS* expression and CsTPS products to explain cannabis metabolite profiles

Of the total of 61 unique apparent CsTPS, only 14 had been functionally characterized prior to this work (**Table 3.2**) (Gunnewich et al., 2007; Booth et al., 2017; Zager et al., 2019; Livingston et al., 2020). Here we describe the functional characterization of 13 additional CsTPS and validation of functions of several others. The CsTPS described here, together with those previously reported, account for most of the terpenes identified in the cannabis cultivars of this

study. One of the objectives of this work was to explore to what extent information on *CsTPS* expression and CsTPS function can be used to predict terpene profiles in cannabis trichome extracts. We found that with current knowledge, metabolite profiles can only be partially predicted, and substantially more information is required about the CsTPS proteome, enzyme kinetics, and substrate availability. Across the different cultivars, *CsTPS* and other genes for terpene biosynthesis, as well as cannabinoid biosynthesis genes, were highly expressed in floral trichomes (**Figure 3.6b**). This observation is in agreement with previous reports on the cannabis MEP and MEV pathways and selected *CsTPS* genes previously reported (Booth et al., 2017; Braich et al., 2019; Livingston et al., 2020).

A single timepoint transcript assessment is likely to be insufficient to explain the accumulation of terpene profiles that occurs over longer periods of time. However, at a qualitative level, we found some general agreement between the presence of terpene products of CsTPS expressed in any given cultivar (**Figure 3.8**) and the metabolites that accumulate in the trichomes of that cultivar (**Figure 3.9**). For example, the cultivars AK and PK, which had no detectable transcript expression of the α-pinene synthase CsTPS2, also had the lowest proportion of α-pinene compared to the other cultivars (**Figure 3.8 and Figure 3.9**). Similarly, AK has a high proportion of nerolidol and relatively high expression of the linalool/nerolidol synthase CsTPS35. Cases where current knowledge of *CsTPS* expression and CsTPS function could not quantitatively explain metabolite profiles are, for example, the high proportion of (*E*)-β-farnesene in the metabolite profile of Choc. This cultivar did not reveal high levels of transcripts of any of the three *CsTPS* (*CsTPS5, CsTPS25, CsTPS32*) known to encode enzymes that produce (*E*)-β-farnesene as a major product. It is possible that one or more of these CsTPS may be a highly efficient enzyme, this protein may be highly stable, or additional (*E*)-β-farnesene

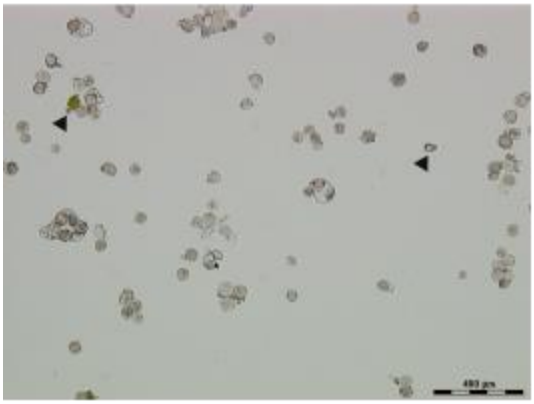
synthases may exist to account for the level of (*E*)-β-farnesene in the Choc cultivar. Similarly, the β-caryophyllene/α-humulene synthase CsTPS9 did not show particularly high transcript levels in any of the cultivars, although these two sesquiterpenes are commonly among the most abundant in cannabis. There are also several terpenes in the metabolite profiles that cannot yet be accounted for by products of known CsTPS functions. These compounds may be the products of CsTPS that remain to be characterized, or may be minor products of CsTPS that were below the detection under assay conditions, but accumulate to detectable levels in trichomes over the course of flower development. We also observed the opposite, where a CsTPS product is not found in the metabolite profile despite high transcript levels. Notably, CsTPS20 hedycaryol synthase was highly expressed across several cultivars, but hedycaryol was not observed in the metabolite profile in any of the cultivars. The labile hedycaryol may be subject to modification (Hattan et al., 2016).

3.5.3 Conclusions and Significance

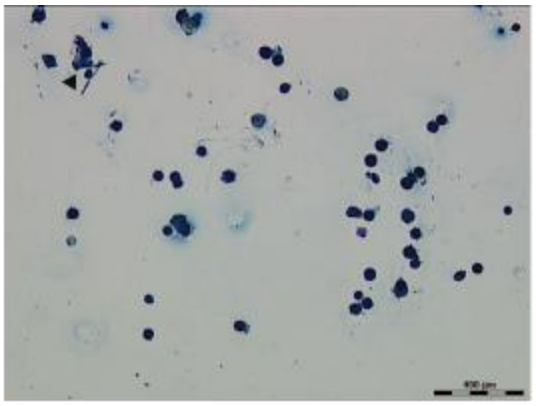
By identifying suites of *CsTPS* genes in six cannabis cultivars, we demonstrated variations of expression and functions that contribute to the different terpene profiles in cannabis cultivars. The enzymes described here, together with other recent studies on terpene biosynthesis in cannabis (Booth et al., 2017; Livingston et al., 2019; Zager et al., 2019), bring the number of characterized cannabis CsTPS to 30 across 14 cultivars.

Acknowledgements: We thank Dr. Carol Ritland and Angela Chiang for administrative and technical support. We thank Erin Gilchrist, Jose Celedon, Samantha Mishos, Eva Chou and members of the Anandia team for assistance with plant growth, access to data, and discussion.

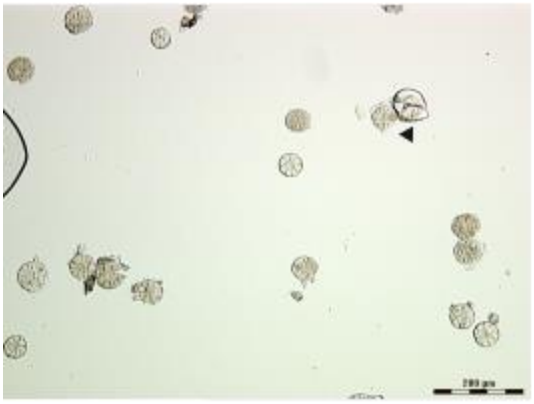
Accession numbers: Raw sequence read data associated with the trichome transcriptome sequencing is deposited in the NCBI Sequence Read Archive with the accession number PRJNA599437. TPS sequences are deposited with the following accession numbers: CsTPS5(PK): MN967481; CsTPS16: MN967478; CsTPS17: MN967470; CsTPS18: MN967473; CsTPS20: MN967469; CsTPS21: MN967483; CsTPS22: MN967477; CsTPS23: MN967480; CsTPS25: MN967472; CsTPS26: MN967479; CsTPS28: MN967482; CsTPS29: MN967468; CsTPS31: MN967474; CsTPS32: MN967484; CsTPS34: MN967476; CsTPS35: MN967475; CsTPS36: MN967471; CsTPS65: MT295506; CsTPS66: MT295505.



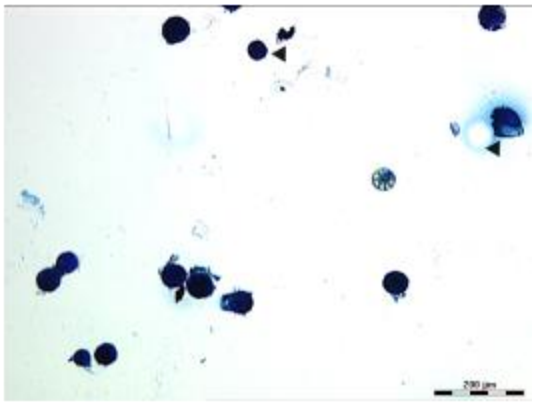
Lemon Skunk



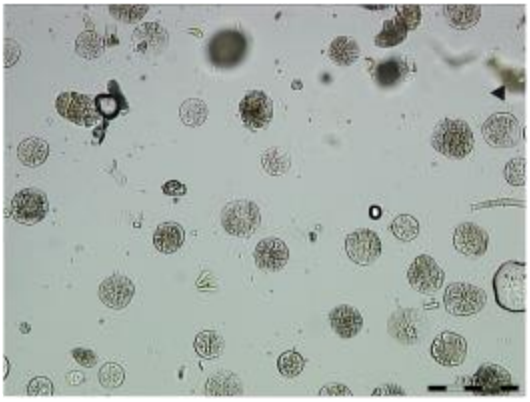
Afghan Kush



CBD Skunk Haze



Blue Cheese



Chocolope

Figure 3.11 Trichome head isolates from five cultivars

Right-hand panels are stained with toluidine blue. Black arrows indicate non-trichome head impurities. Samples were pipetted onto a microscope slide after gentle mixing of the trichome isolate pellet.

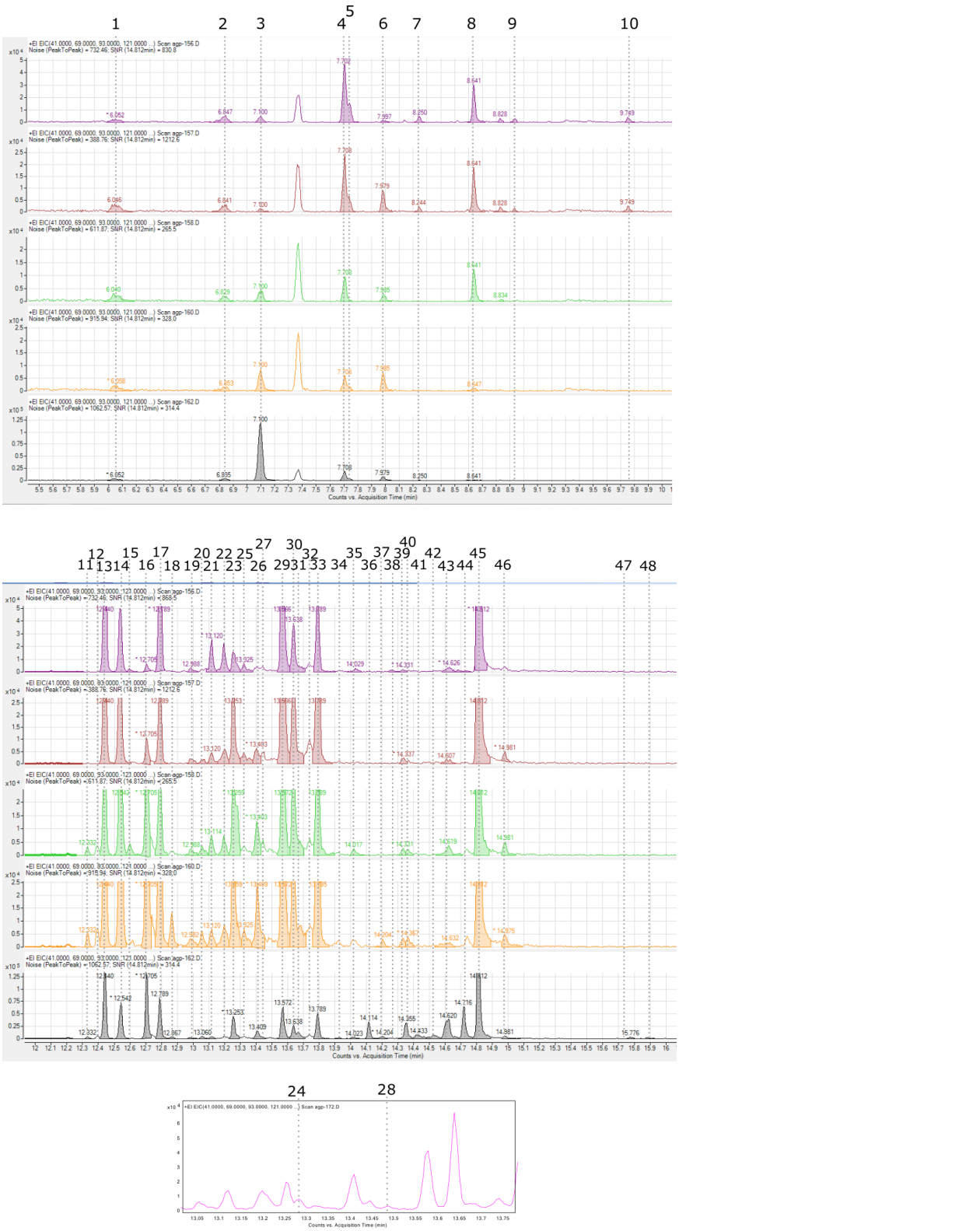


Figure 3.12 Representative extracted ion chromatograms of foliar terpene extracts from 5 cannabis plants Plants were those chosen for cloning and sequencing. Numbered dashed lines indicate peak identities, as described in Supplemental Table S3. Top panel: monoterpenes. Bottom panels: sesquiterpenes.

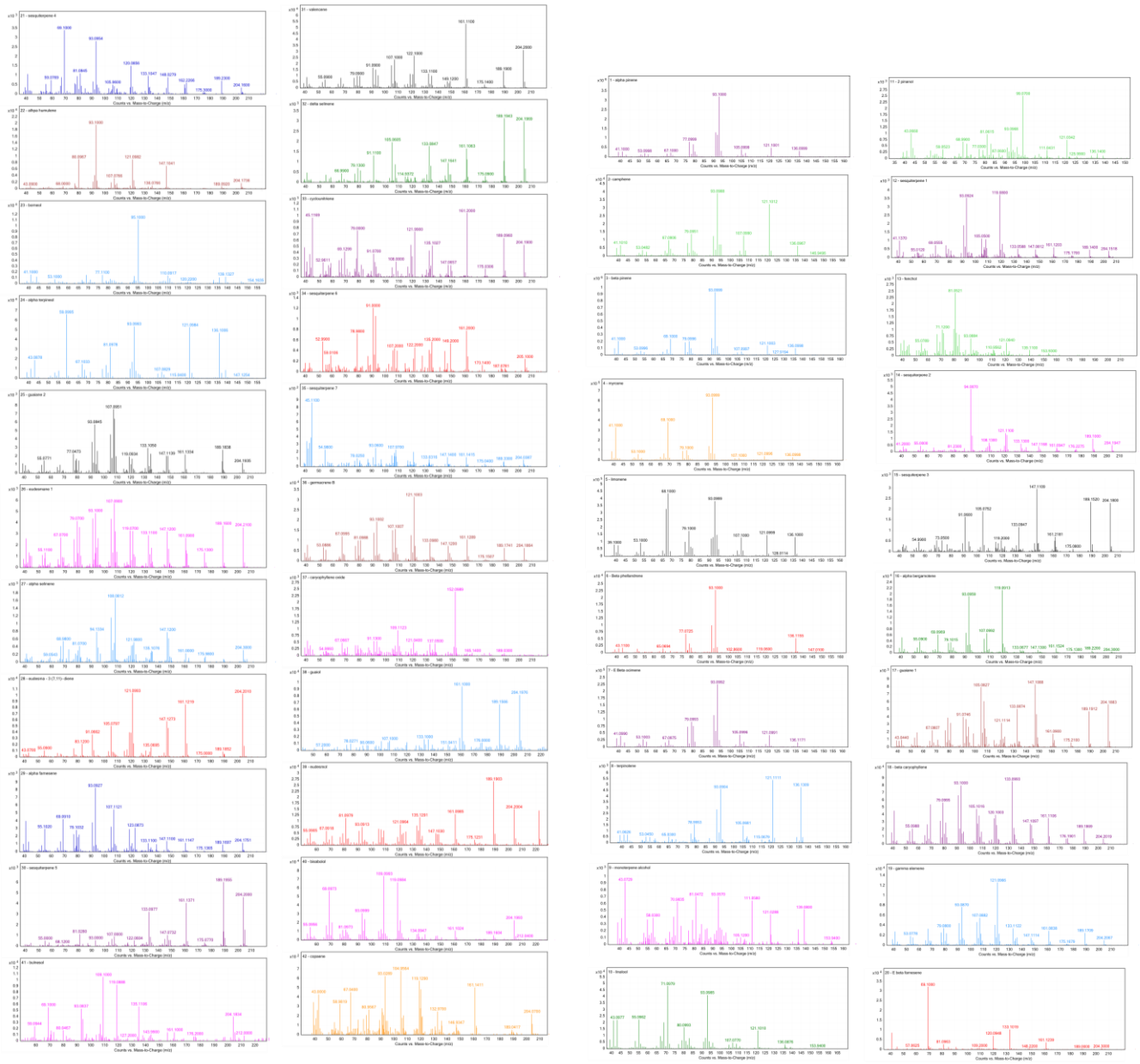


Figure 3.13 Representative mass spectra for all compounds identified in floral terpene extracts
Compounds correspond to those listed in Table 3.1. Mass spectra are background subtracted line plots taken from the cultivar with the highest peak area for each compound. Mass spectrometer does not have high resolution; mass values are accurate to three decimal places.

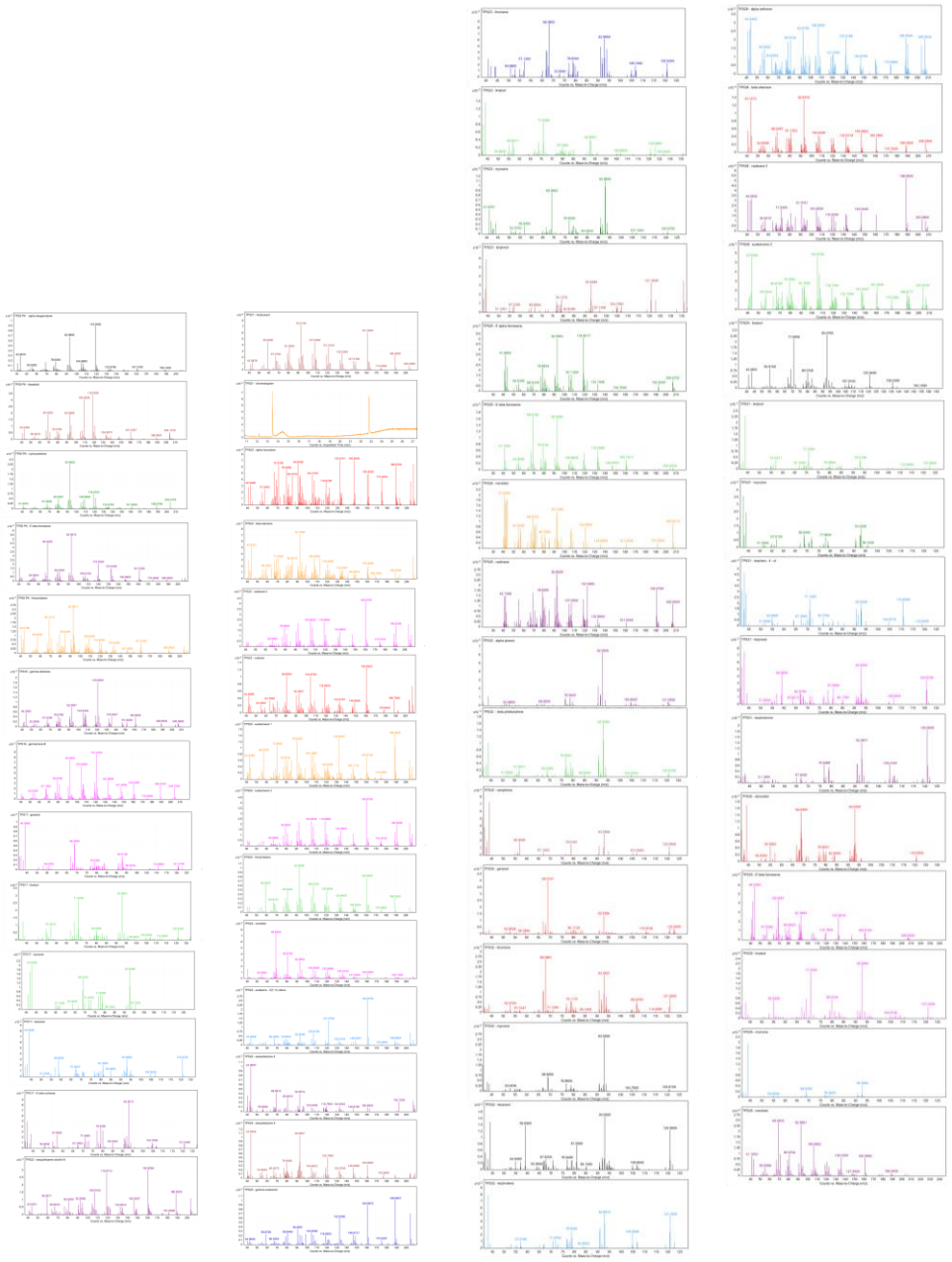


Figure 3.14 Representative mass spectra for all CsTPS products
Mass spectra are background subtracted line plots for each compound. For the CsTPS whose products undergo thermal degradation (i.e. CsTPS21) the total ion chromatogram is also shown. Mass spectrometer does not have high resolution; mass values are accurate to three decimal places.

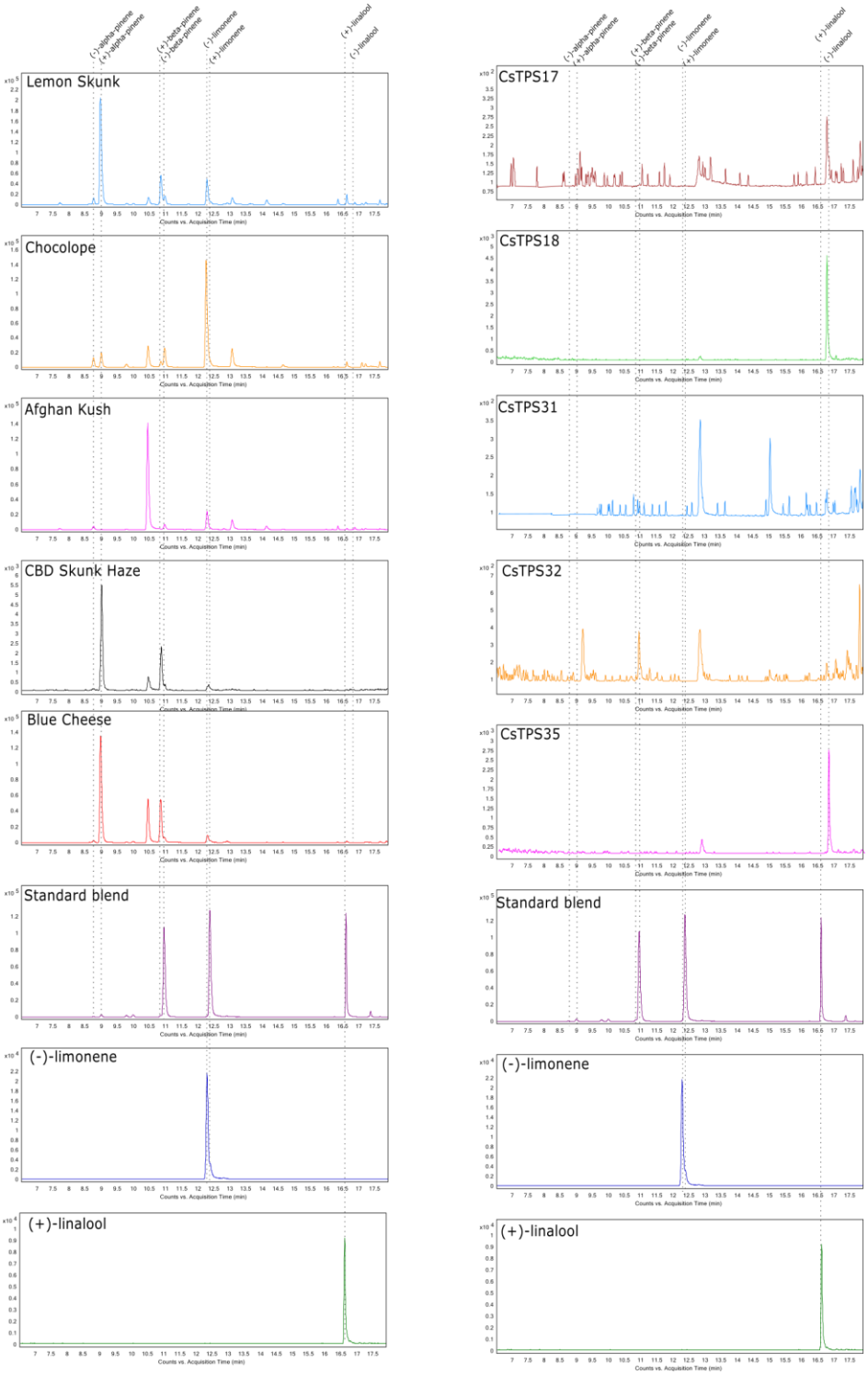


Figure 3.15 Extracted ion chromatograms showing stereochemical determination of monoterpenes in cannabis juvenile floral terpene extracts and CsTPS enzyme assays

All assays were run on Agilent Cyclodex-B 30m column, and identities verified by relative retention time and comparison to authentic standards. 'Standard blend' consisted of (+) and (-)-α-pinene, (-)-β-pinene, (+)-limonene, and (+)-linalool.

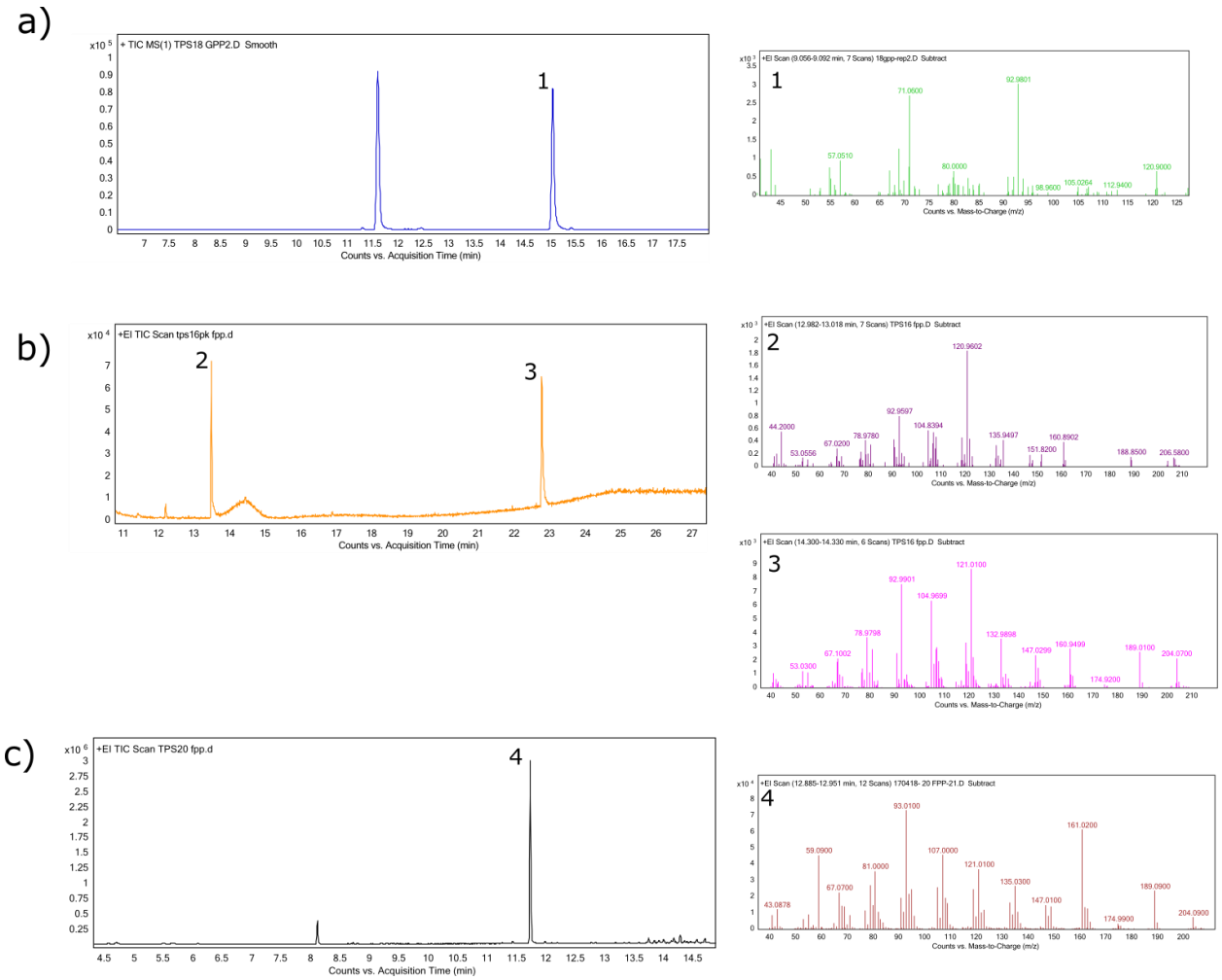


Figure 3.16 Total ion chromatograms and mass spectra for previously characterized CsTPS
Total ion chromatograms and mass spectra for a) CsTPS18, b) CsTPS16, and c) CsTPS20. 1: Linalool, 2: γ-elemene, 3: germacrene B, 4: hedycaryol. Mass spectrometer does not have high resolution; mass values are accurate to three decimal places.

Chapter 4: The Molecular Basis of Stereochemical Variation in Two Ocimene Synthases

4.1 Summary

The closely related cannabis terpene synthases CsTPS6 and CsTPS13, which share 95.5% amino acid identity, both produce the linear monoterpene β-ocimene. However, their major products are stereoisomers. TPS6 produces 97% (*E*)-β-ocimene, whereas TPS13 produces 94% (*Z*)-β-ocimene. In this chapter, I used site-directed mutagenesis to explore the mechanisms of these two enzymes that may explain their different product specificities. Homology modeling was used to predict the tertiary structures of both enzymes and to select target amino acid positions for mutagenesis. Two pairs of amino acids, Asn^{235}-Ile^{236} and Leu^{369}-Ile^{370}, were identified that, upon mutation to Lys-Val and Ile-Leu, respectively, increase the proportion of (*Z*)-β-ocimene relative to (*E*)-β-ocimene produced by CsTPS6. This analysis helps uncover how double-bond stereochemistry is determined in the active site cavity of a TPS enzyme, and highlights how diversity of terpene profiles can arise by minor variation of TPS sequences.

4.2 Introduction

In cannabis (*Cannabis sativa*), mono- and sesquiterpenes are an important component of the isoprenoids found in the resin (Chapter 1, Ross and ElSohly, 1996; Hazekamp and Fischedick, 2012; Lynch et al., 2016). These terpenes are the products of cannabis terpene synthases (CsTPS), and the diversity of CsTPS enzymes contributes to the diversity of terpene structures found in cannabis resin and cannabis aroma. The cannabis genome contains over 30 different

CsTPS genes that are variably expressed in different cannabis genotypes. The CsTPS gene family includes closely related members that produce structurally similar products (Booth et al., 2017; Zager et al., 2019). The general precursor to monoterpenes is geranyl diphosphate (GPP), a linear isoprenoid diphosphate. In the active site of monoterpene synthases (mono-TPS), GPP can undergo a series of dephosphorylation, isomerization, cyclization and rearrangement reactions, before deprotonation or water capture yield the final terpene product (Croteau, 1987; Kampranis et al., 2007; Tantillo, 2011; Christianson, 2017). Many mono-TPS are multiproduct enzymes. While amino acids important for mono-TPS substrate specificity, isomerization, and cyclization have been identified (Greenhagen et al., 2006; Roach et al., 2014; Salmon et al., 2015; Srividya et al., 2015), the specific products of a mono-TPS cannot currently be predicted from its amino acid sequence. The contours of the active site cavity constrain the rearrangement of the substrate, limiting or guiding the intermediates. In general, residues that define the shape or size of the active site cavity have stronger effects on product outcomes than residues more distant from the substrate (Salmon et al., 2015).

Mono TPS are generally 50-100 kDa α-helical proteins. They feature two domains, the N-terminal β domain and the C-terminal α domain (reviewed in: (Christianson, 2017)). The α domain forms an α-helical 'terpene synthase fold', containing the active site cavity. The β domain 'caps' the hydrophobic active site cavity during catalysis (Whittington et al., 2002). The α domain contains two conserved aspartate-rich motifs that coordinate a cluster of three Mg^{2+} ions, which are essential for catalysis (Bohlmann et al., 1998; Christianson, 2006). The first aspartate-rich motif, DDXXD, coordinates two metal ions, and the second aspartate-rich motif, NSE/DTE, coordinates the third (Aaron and Christianson, 2010). A third motif located near the N-terminus, RRX_8W, is important for cyclization (Bohlmann et al., 1998; Williams et al., 1998).

The three metal ions (usually Mg^{2+}) are required to trigger dephosphorylation of the bound substrate, leading to formation of the initial carbocation (Aaron and Christianson, 2010).

Previous work on mono-TPS mechanisms has succeeded in converting product stereochemistry. A pair of *Thymus vulgaris* sabinene synthases were interconverted by reciprocal mutation of an isoleucine and an asparagine residue (Krause et al., 2013).Similarly, exchanging four residues near the active site of a *Mentha spicata* (-)-limonene synthase resulted in the successful formation of primarily (+)-limonene products (Srividya et al., 2020).

The cannabis mono-TPS CsTPS6 and CsTPS13 each have a single β-ocimene product (Booth et al., 2017), which is a linear molecule with double bonds at the 1-2, 3-4, and 6-7 positions, where bond 3-4 may be in *E* or *Z* conformation (**Figure 4.1**). Ocimene is a common component of floral scent, and is often part of plant volatile emission in response to herbivory (Kessler and Baldwin, 2001; Dudareva, 2003). In cannabis resin, the *E* stereoisomer is more abundant than the *Z* (Booth and Bohlmann, 2019). CsTPS6 produces predominantly (*E*)-β-ocimene (97%) and only minor amounts of (*Z*)-β-ocimene (3%). CsTPS13 produces 94% (*Z*)-β-ocimene and 6% (*E*)-β-ocimene. These two enzymes share 95.5% amino acid sequence identity, except for the N-terminal transit peptide which is cleaved off in the mature enzyme. The high identity between CsTPS6 and CsTPS13 provided an opportunity to use a site-directed mutagenesis approach to explore the molecular basis of product stereochemistry. The objective of this chapter is to determine which amino acids allow these two highly similar mono-TPS to produce different stereoisomers of the same general compound.

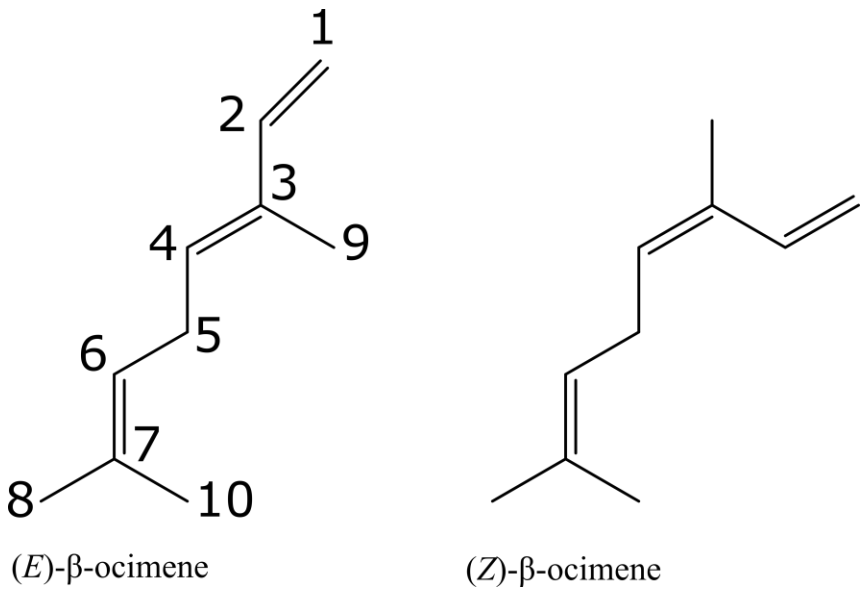


Figure 4.1 Skeletal formulas for (*E*)-β-ocimene and (*Z*)-β-ocimene

4.3 Materials and Methods

4.3.1 Homology modeling and ligand docking

Homology models for CsTPS6 and CsTPS13 were produced using the SWISS-MODEL server (Bienert et al., 2017; Waterhouse et al., 2018). The sequences used for modeling were without the predicted plastid transit peptide covering the N-terminal 35 amino acids. Models were based on the structure of citrus (*Citrus sinensis*) (+)-limonene synthase (5uv0.1). Energy minimization was done using the YASARA force field server (Krieger et al., 2009). The ligand, 2-fluorolinalyl diphosphate, was downloaded from the SwissDock target database. Docking of the ligand to the protein model was performed using CLC Drug Discovery Workbench and SwissDock. Results were visualized in PyMol (Seeliger and de Groot, 2010). Cavity volumes were calculated using the CastP server (Tian et al., 2018).

4.3.2 Cloning and site-directed mutagenesis

CsTPS6 cloned from the cannabis variety Finola was used as a template for mutagenesis. CsTPS6 and CsTPS13 were cloned into the *E. coli* expression vector pET28b+ as previously described (Booth et al., 2017). Mutations were introduced to CsTPS6 by Kwik-change PCR, using primers matching the corresponding sequences of CsTPS13. After amplification, the template was degraded by digesting with *Dpn*I at 37°C for 3 hours, followed by denaturation at 80°C for 20 minutes then 90°C for 10 minutes. Mutations were confirmed by Sanger sequencing.

4.3.3 Recombinant protein expression and enzyme assays

Plasmids were transformed into *E. coli* strain BL21DE3 for heterologous protein expression. Bacteria were grown in 50-mL cultures at 37°C with shaking at 200 rpm until the OD_{600} reached 0.6. The temperature was reduced to 18°C and cultures were induced with IPTG. Cultures were grown overnight and harvested by centrifugation. Bacterial pellets were stored at -80°C. Cells were lysed by sonication on ice for 90 s. Lysate was clarified by centrifugation for 15 minutes at 4°C. Recombinant protein was purified using the GE healthcare His SpinTrap kit (www.gelifesciences.com) according to manufacturer's instructions. Binding buffer for purification was 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) (pH 7.5), 500 mM NaCl, 25 mM imidazole, and 5% glycerol. Cells were lysed in binding buffer supplemented with Roche complete protease inhibitor tablets (lifescience.roche.com) and 0.1 mg mL^{-1} lysozyme. Elution buffer was 20 mM HEPES (pH 7.5), 500 mM NaCl, 350 mM imidazole, and 5% glycerol. Purified protein was desalted through Sephadex into TPS assay buffer (25 mM HEPES (pH 7.3), 100 mM KCl, 10 mM $MgCl_2$, 5% glycerol, and 5 mM DTT) and quantified using Bradford reagent. *In vitro* assays were performed in 500 µL volume by incubating 50 µg purified protein with isoprenoid diphosphate substrates (isoprenoids.com) at 30°C for 1 hour.

Isoprenoid diphosphate substrates were dissolved in 50% methanol and added to the assay at final concentrations of 32 µM (GPP) and 26 µM (FPP). Assays were overlaid with 500 µL pentane, with 1 µM isobutyl benzene as internal standard.

4.3.4 GC-MS analysis of enzyme products

CsTPS6 and CsTPS13 products were analyzed on an Agilent 7000A system GC with an Agilent GC sampler 80 and a triple quadrupole MS detector at 70 eV. The column was DB-WAX (30 m, 250 µm i.d., 0.25 µm film thickness); and the program was: 40°C for 5 minutes, then increase 3°C minute^{-1} to 80°C, then increase 50°C minute^{-1} to 220°C, hold for 5 minutes.

4.4 Results

4.4.1 Alignment, modeling, and site-directed mutagenesis reveal active site differences between CsTPS6 and CsTPS13

For modeling, CsTPS6 and CsTPS13 sequences were used lacking the predicted plastid-targeting motif (Δ35 amino acids) upstream of the RRX_8W motif (Bohlmann et al., 1998). The protein model used for threading was citrus (*Citrus sinensis*) (+)-limonene synthase (PDB code 5uv0.1), a monoterpene synthase whose structure had been solved to 2.3 Å (Morehouse et al., 2017). According to the model, CsTPS6 and CsTPS13 consist of two α-helical domains. The non-catalytic N-terminal domains contain 11 α-helices, and the catalytic C-terminal domains contain 13 α-helices (**Figure 4.2**). A linalyl diphosphate analog, 2-fluorolinalyl diphosphate, was used as the substrate for docking, which placed the substrate in the active site cavity. Attempts to use 2-fluorogeranyl diphosphate as the ligand did not result in the ligand docking in the active site cavity. Three Mg^{2+} ions were also docked in the active site cavity, adjacent to the DDXXD motif

(**Figure 4.2**). The diphosphate moiety of the substrate was adjacent to the Mg^{2+} ions, as has been shown for other TPS models (Aaron and Christianson, 2010; Roach et al., 2014; Morehouse et al., 2017). The cavity volumes of CsTPS6 and CsTPS13 were calculated as 656 $Å^3$ and 592 $Å^3$, respectively, and the surface areas of the active site cavity were 709 $Å^2$ and 693 $Å^2$, respectively.

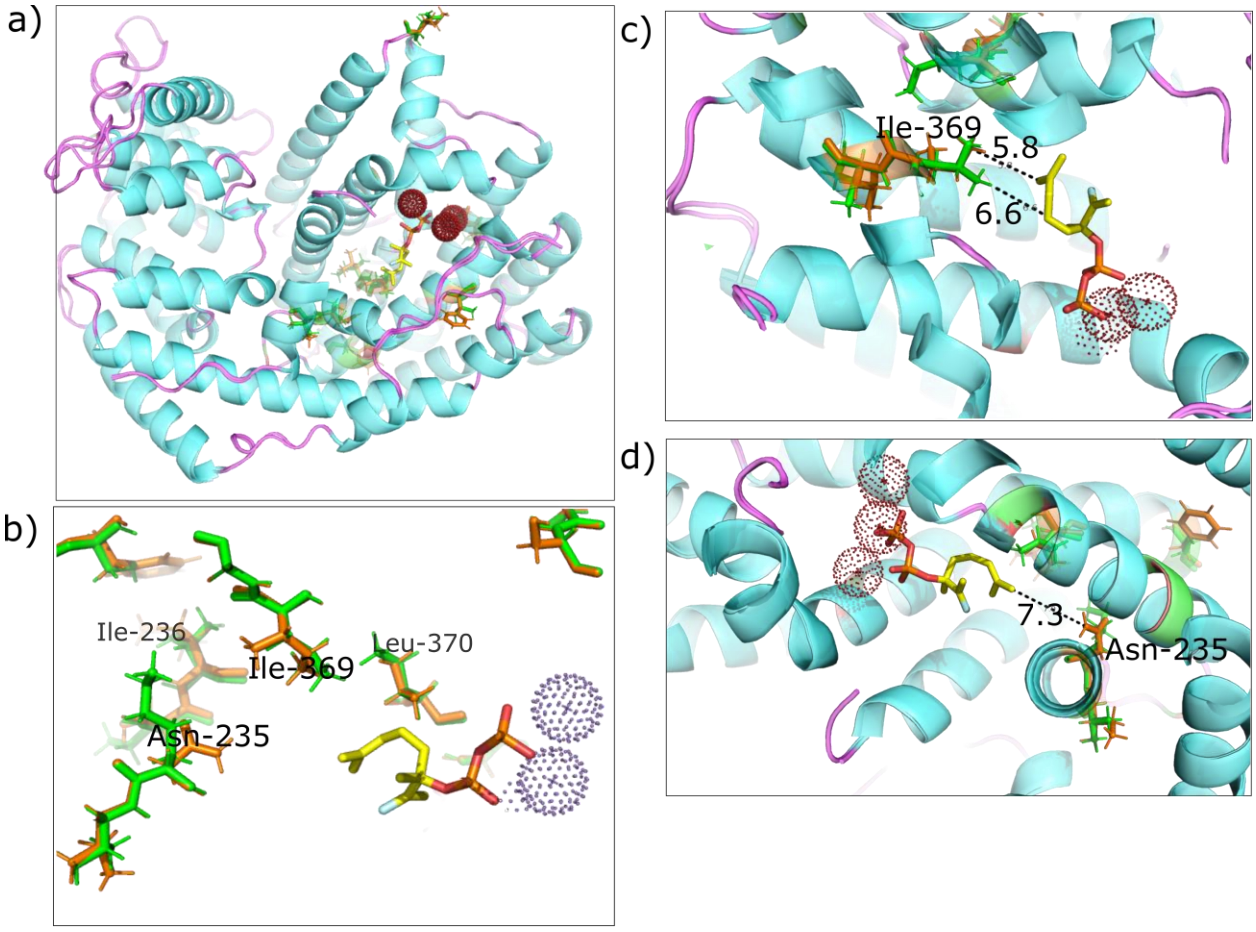


Figure 4.2 Homology modeling of TPS6 and TPS13 with substrate analog 2-fluorolinalyl diphosphate .
(a) Models of the complete proteins without transit peptides. (b) Amino acids of the active site cavity that differ between TPS6 and TPS13, and distance from the substrate analog. (c) Close-up of active site cavities highlighting isoleucine (I, Ile) 369. (d) Close-up of active site cavities highlighting asparagine (N, Asn) 235. Helices are shown in teal, loops in purple. Select amino acids are shown in orange (TPS6) or green (TPS13). The substrate analog 2-fluorolinalyl diphosphate is shown in yellow with the diphosphate group in red and orange. Mg2+ ions are dotted spheres. Distances in angstroms are shown as black dashed lines.

In total, 23 amino acids differed between CsTPS6 and CsTPS13, of which 21 are in the C-terminal domain (**Table 4.1**). Of these, nine amino acids are located on the three helices that

line the active site cavity, and five amino acids are located within 10 Å of the predicted position of the substrate. Nine of the amino acids pairs that differ between CsTPS and CsTPS13 represent different side-chain classes, e.g. nonpolar to polar or charged to neutral. Eleven residues or neighbouring pairs of residues were selected for site-directed mutagenesis in the CsTPS6 background, to test their role in affecting product stereochemistry (**Figure 4.3**). The selections were based on proximity to the active site and priority was given to residues that represent different types of amino acids between CsTPS6 and CsTPS13.

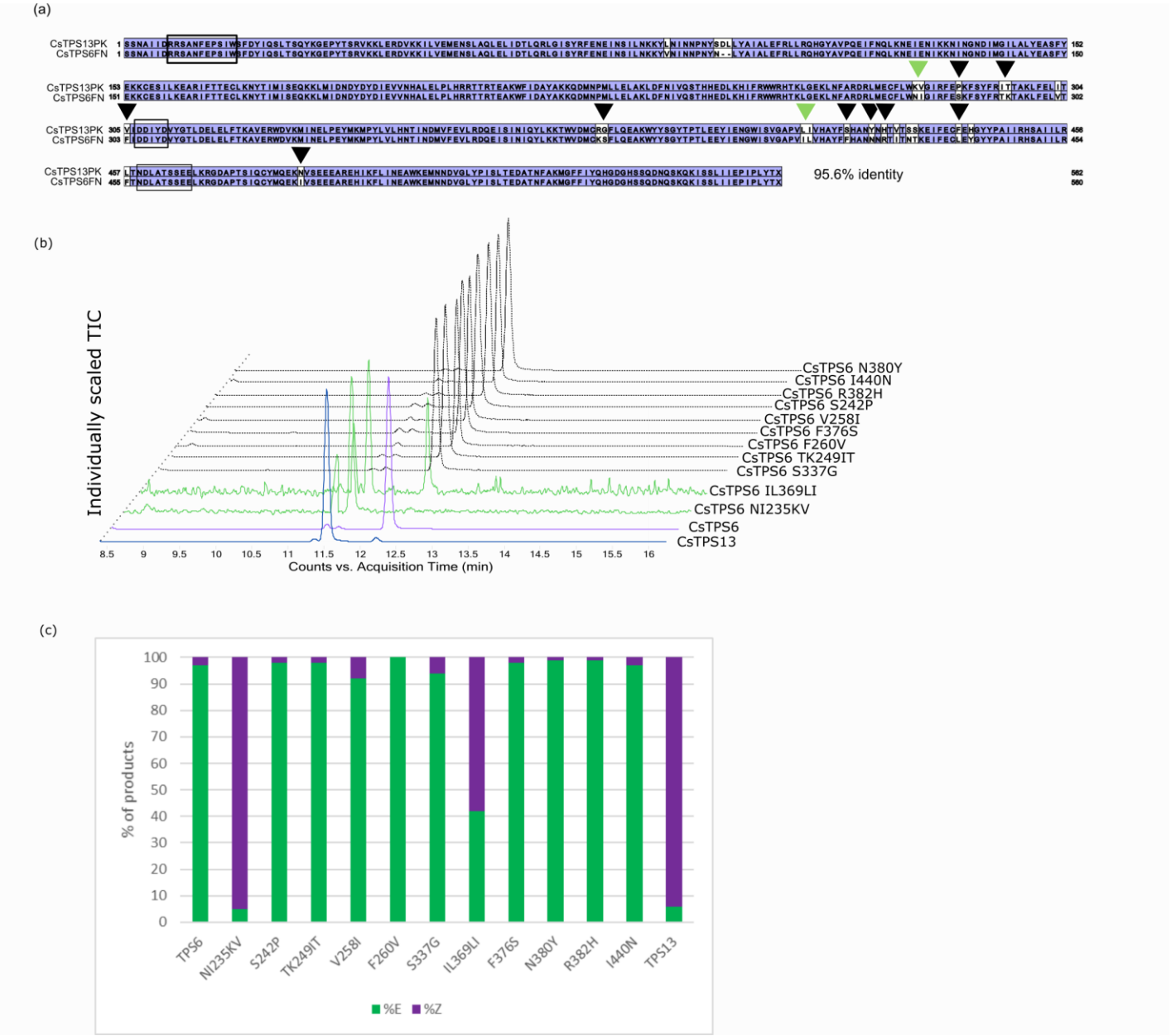


Figure 4.3 Ocimene synthase conversion.
a) Alignment of CsTPS6 and CsTPS13. Residues shown against a white background are different between CsTPS6 and CsTPS13. Black arrows indicate positions targeted for mutation in the CsTPS6 background to the corresponding amino acid in CsTPS13. Green arrows represent positions that altered product stereochemistry. Boxes indicate the conserved motifs RRX8W and DDXXD. b) Total ion chromatograms of products from original CsTPS6 and CsTPS13, and 11 TPS6 mutants. c) Mutants are based on the sequence of TPS6, numbered from the beginning of the RRX8W motif. Proportions are the percentof total ocimene products with trans (E) or cis (Z) stereochemistry.

Amino acid in TPS6	Amino acid in TPS13	Predicted distance from substrate (Å)	Position
Valine (V)	Leucine (L)	44.1	45
Asparagine (N)	Serine (S)	42.5	53
Asparagine (N)	Lysine (K)	7.3	235
Isoleucine (I)	Valine (V)	9.5	236
Serine (S)	Proline (P)	23.7	242
Threonine (T)	Isoleucine (I)	16.4	249
Lysine (K)	Threonine (T)	13.5	250
Valine (V)	Isoleucine (I)	3.5	258
Phenylalanine (F)	Valine (V)	20.3	260
Lysine (K)	Arginine (R)	12.7	336
Serine (S)	Glycine (G)	9.2	337
Isoleucine (I)	Leucine (L)	5.8	369
Leucine (L)	Isoleucine (I)	12.3	370
Phenylalanine (F)	Serine (S)	20.9	376
Asparagine (N)	Tyrosine (Y)	25.8	380
Arginine (R)	Histidine (H)	30.3	382
Isoleucine (I)	Valine (V)	32.7	384
Asparagine (N)	Serine (S)	23.8	386
Threonine (T)	Serine (S)	32.1	387
Leucine (L)	Phenylalanine (F)	20.1	394
Tyrosine (Y)	Histidine (H)	21.5	396
Phenylalanine (F)	Leucine (L)	15.3	412
Isoleucine (I)	Asparagine (N)	26.4	440

Table 4.1 Positions and distances of residues in CsTPS6 that are different in CsTPS13
Positions are numbered from the end of the predicted plastid target peptide. Distances are the nearest atom in the amino acid side chain to any atom in the predicted position of the linalyl cationic substrate.

Individual amino acid residues or two neighboring residues in CsTPS6 were mutated to the residues found in the corresponding position in CsTPS13. Mutated enzymes were heterologously expressed and tested for product ratios and specificity with regard to (*E*)- or (*Z*)-β-ocimene. The mutation CsTPS6$^{N235K/I236V}$ produced approximately 95% (*Z*)-β-ocimene and 5% (*E*)-β-ocimene. CsTPS6$^{I369L/L370I}$ also greatly increased the relative amount of (*Z*)-β-ocimene with a product ratio of 58% (*Z*) and 42% (*E*) β-ocimene (**Figure 4.3**). The activity of both

CsTPS6$^{N235K/I236V}$ and CsTPS6$^{I369L/L370I}$ was about 1000-fold lower compared to the activity of the original CsTPS6. 50 mg of purified recombinant CsTPS6 produced 1.8 ng ocimene during a 1 hr incubation. Products of both CsTPS6$^{N235K/I236V}$ and CsTPS6$^{I369L/L370I}$ were below the level of quantification under 1-hour standard assay conditions, and products were detectable only in assays with extended incubation times of at least 4 hours.

Nine of the individual amino acid mutations or mutations of sets of two neighboring amino acids retained the predominantly (*E*) product stereochemistry of the wild-type enzyme (**Figure 4.3**). CsTPS6^{F260V} produced no detectable *cis*-ocimene, making it the only mutant to produce 100% (*E*)-β-ocimene. Five other mutants, namely CsTPS6^{S242P}, CsTPS6$^{T249I/K250T}$, CsTPS6^{F376S}, CsTPS6^{N380Y}, CsTPS6^{R382H}, and CSTPS6^{I440N}, produced a higher ratio of *E* to *Z* ocimene than CSTPS6. Two mutants, CsTPS6^{V258I} and CsTPS6^{S337G}, produced slightly more cis-ocimene than CsTPS6.

4.5 Discussion

I used a targeted mutational approach to explore the effects of small sequence differences between CsTPS6 and CsTPS13 on their product profiles. CsTPS6 produced mostly (*E*)-β-ocimene, whereas CsTPS13 produced mostly (*Z*)-β-ocimene. Mutating sequential isoleucine-leucine residues to leucine-isoleucine in position 369-370, which is in the active site cavity of CsTPS6, caused a partial conversion of the CsTPS6 (*E*)-β-ocimene synthase to produce nearly equal amounts of the two ocimene stereoisomers. Conversion of asparagine-isoleucine (NI) to lysine-valine (KV) in position 235-236 caused nearly total reversal of product stereochemistry. Other mutations resulted in only minor changes to the product ratio of CsTPS6.

N_{235} in CsTPS6 is also an asparagine in the cannabis (*E*)-β-ocimene synthase CsTPS37 (Livingston et al., 2020, Appendix A). The two enzymes are 59.4% identical. This may indicate that N_{235} has a greater role in determining product stereochemistry than I_{236}, as it is conserved across less closely related enzymes with identical products.

In the *Mentha spicata* limonene synthase MsLimS, the proximity of any given amino acid residue to the position of the substrate is highly correlated with its impact on enzyme fidelity when mutated (Srividya et al., 2015). According to the model (**Table 4.1**), N_{325} in CsTPS6 is closer to the substrate than I_{326} and is located directly on the surface of the active site cavity. In CsTPS6 I_{326} is located on the face of the α-helix facing away from the substrate. This further supports the hypothesis that N_{325} is more important for product stereochemistry, as it can more directly affect the shape and size of the active site cavity than can I_{326}. I mutated the negatively charged N_{325} in CsTPS6 to a positively charged lysine (K). The change from a negatively- to a positively charged sidechain may alter the conformation of the surrounding residues, leading to a change in the shape of the active site cavity. By contrast, I_{326} was mutated to a valine (V), both nonpolar side-chains.

Similarly, I_{369} is closer to the substrate than L_{370}, which may suggest that I_{369} is the more important of the two for stereochemistry. While Ile and Leu are both nonpolar aliphatic amino acids, I_{369} in TPS6 projects further into the active site cavity than L_{369} in CsTPS13, limiting the available room for the substrate and intermediates to rotate. L_{370} faces the neighbouring α-helix, making it less likely than I_{369} to directly interact with the substrate.

Surprisingly, V_{258}, which is the closest residue to the substrate at 3.5Å (**Table 4.1**), had little effect when mutated to isoleucine, despite the aromatic structure and larger size of phenylalanine compared to valine. $TPS6^{V258I}$ did produce a slightly higher proportion (8%) of

(*Z*)-ocimene than TPS6 (3%). The difference was small compared to the two mutants discussed above. While enzyme efficiency was not tested for TPS6^{V258I}, there was no notable difference based on chromatogram peak area under the same assay conditions. This suggests that V_{258} is not directly involved in substrate binding or coordination of the substrate or intermediates.

Both CsTPS6$^{N235K/I236V}$ and CsTPS6$^{I369L/L370I}$ showed decreased enzyme activity compared to the wild-type enzyme. A variety of different mutations might be required to fully convert the enzymes, altering residues in the surrounding shell. CsTPS6^{V258I} or CsTPS6^{S337G}, which slightly increased the proportion of (*Z*)-ocimene, may be important for full and efficient conversion. Future studies including combinatorial mutations of CsTPS6 will clarify the roles of mutations in the outer shells in promoting the stereospecificity and catalytic efficiency of the enzyme.

The CsTPS family produce a wide array of different terpenes in cannabis (Allen et al., 2019; Booth and Bohlmann, 2019). The example of CsTPS6 and CsTPS13 highlight how a relatively small number of amino acid differences can result in new product stereochemistry. Similarly, CsTPS18 and CsTPS19 produce different isomers of linalool but differ only by one amino acid (Zager et al., 2019). Examples like these show how diverse profiles of different terpenes may have arisen in cannabis by expansion of the CsTPS gene family as a result of gene duplication and relatively minor sequence divergence.

For the model shown in **Figure 4.2**, I used an analogue of LPP to simulate binding of the phosphorylated early intermediate. In a model of tobacco (*Nicotiana tabacum*) 5-epi-aristolochene synthase, it was suggested that the carbocation slips deeper into the active site cavity from its initial binding position (Starks et al., 1997). This would bring the reactive intermediate closer to N_{325} and I_{369} than was shown in the model presented here. In this setting,

the positioning and orientation of the cation at the moment of deprotonation may be constrained differently from the model as presented in this study.

The stereochemistry of β-ocimene product is governed by the conformation of the bound substrate and the cationic intermediate, either geranyl or linalyl cations (**Figure 4.5**). The route through the geranyl cation is similar and more direct, but the formation of a linalyl cation cannot be ruled out. In an unconstrained context, there would be free rotation around bond 3-4. When the intermediate is constrained by the conformation of the active site cavity in a terpene synthase, it can be stabilized in a *cis* or *trans* conformation, leading to the formation of *cis* or *trans* stereochemistry after the double bond is formed by deprotonation.

In an recent study on *Mentha spicata* (-)-limonene synthase, the residue M^{458} was demonstrated to be a critical determinant of enantiospecificity (Srividya et al., 2020). This residue corresponds to I^{369} in CsTPS6. The authors of that study suggest M^{458} influences the product outcome by clashing with methyl groups of the substrate analog 2-Fluoro-LPP in a left-handed binding position, thus forcing the substrate into a right-handed binding position. Based on their modeling, their results support the hypothesis that this residue in the active site cavity by constraining the binding position of the substrate. This proposed mechanism is distinct from the one involved in the stereospecificity of two sabinene hydrate synthases in *Thymus vulgaris*, TvTPS6 and TvTPS7 (Krause et al., 2013). The conversion of an asparagine to an isoleucine in TvTPS6 produced inverted stereoisomers of the monoterpene sabinene hydrate by altering the stereochemistry of linalyl diphosphate, a chiral intermediate of many monoterpenes. A similar mechanism is proposed to explain four mutations that alter the stereochemistry of the suite of products from the maize enzymes TPS4 and TPS5 (Köllner et al., 2004). For cyclic terpenes, the stereochemistry of the linalyl intermediate determines the stereochemistry of the final product.

For β-ocimene, the absolute positioning of the intermediates seems to be more important than the stereochemistry. This may also be true for α-farnesene, the sesquiterpenoid structural equivalent of β-ocimene. The roles of the residues identified here in directing or stabilizing the cationic intermediate remain to be determined.

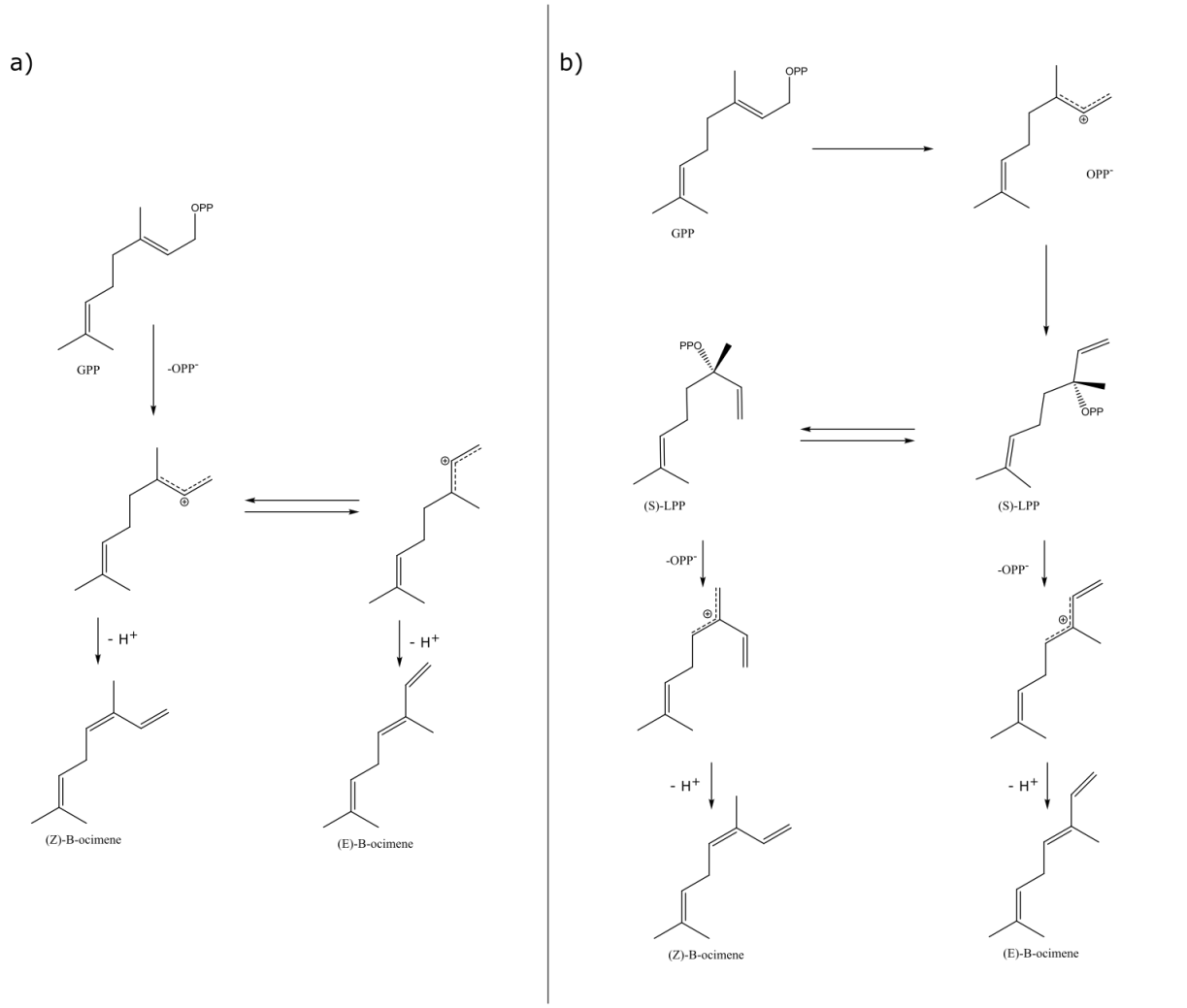


Figure 4.4 Two potential mechanisms leading to the formation of (*E*)- or (*Z*)-β-ocimene.
a) via rotation of geranyl diphosphate (GPP, second row) or the gerynyl cation (middle row), followed by deprotonation. b) via isomerization to linalyl diphosphate (LPP, second row) and rotation of the linalyl cation (third row), followed by deprotonation.

This study highlights the remarkable plasticity and complexity of TPS catalysis. I demonstrated the potentially large effects of small mutations on the products of a selected CsTPS, which may have direct implications for the terpene metabolite profiles of cannabis. Similarly, in the Sitka spruce (*Picea sitchensis*) mono-TPS, PsTPS-3car, a single mutation of leucine (L) to phenylalanine (F) altered the product profile from 3-carene to sabinene as the major product (Roach et al., 2014). Two amino acid mutations redirected the product profile of a sage (*Salvia fruticosa*) 1,8-cineole synthase towards sabinene (Kampranis et al., 2007). Similar effects have been described for sesquiterpene synthases. The grapevine (*Vitis vinifera*) selinane TPS VvTPS24 was fully converted to an α-guaiene synthase by two amino acid changes in the active site cavity. These changes provided the precursor for a distinctive wine aroma compound, rotundone (Drew et al., 2015). In maize (*Zea mays*), four amino acids are responsible for stereospecificity and product profile in two multi-product sesquiTPS involved in response to herbivory (Köllner et al., 2004). These large-effect mutations show unpredictable effects on plant fitness, as in the example of Sitka spruce and maize, or on plant products used by humans such as wine aroma and cannabis resin. These large-effect mutations are ultimately a source of the great diversity of terpenes found in Nature.

Chapter 5: Synthetic Biology of Cannabinoids and Cannabinoid Glycosides in *Nicotiana benthamiana* and *Saccharomyces cerevisiae*

5.1 Summary

Phytocannabinoids are a group of plant-derived metabolites that display a wide range of psychoactive as well as health-promoting effects. The regulated production of pharmaceutically relevant cannabinoids can rely on extraction and purification from cannabis (*Cannabis sativa*) plants yielding predominantly the major constituents Δ^9-tetrahydrocannabinol and cannabidiol. Heterologous biosynthesis of cannabinoids in *Nicotiana benthamiana* or *Saccharomyces cerevisiae* provides a cost-efficient and rapid production platform to acquire pure and high quantities of both the major and the rare cannabinoids as well as novel derivatives. Here, we used a meta-transcriptomic analysis of cannabis to identify genes for aromatic prenyltransferases of the UbiA superfamily and chalcone isomerase-like (CHIL) proteins. Among the aromatic prenyltransferases, CsPT4 showed CBGAS activity in both *N. benthamiana* and *S. cerevisiae*. Co-expression of selected CsPT pairs and co-expression of CHIL proteins with CsPT4 did not affect CBGAS catalytic efficiency. In a screen of different plant UGTs, we found that *Stevia rebaudiana* SrUGT71E1 and *Oryza sativa* OsUGT5 glucosylated olivetolic acid, cannabigerolic acid and Δ^9-tetrahydrocannabinolic acid. Metabolic engineering of *Nicotiana benthamiana* for production of cannabinoids revealed intrinsic glycosylation of olivetolic acid and cannabigerolic acid. *Saccharomyces cerevisiae* was engineered to produce olivetolic acid glycoside and cannabigerolic acid glycosides.

5.2 Introduction

Today, legal cannabis-derived products are a fast-growing global industry that is projected to reach $57 billion by 2027, with consumption of pharmaceutical cannabinoids making up one third of total profit (Pellechia, 2018). The production of cannabis-based pharmaceutical drug products can rely on extraction and purification of phytocannabinoids such as Δ^9-tetrahydrocannabinol and cannabidiol. For practical uses, the rare cannabinoids are present in too low abundance to be worthwhile extracting (Ahmed et al., 2015; Andre et al., 2016).

The formation of cannabinoids requires precursors from two biosynthetic pathways, the polyketide and the methylerythritol phosphate (MEP) isoprenoid biosynthetic pathways (**Figure 5.1**). Hexanoic acid is converted to olivetolic acid (OA) by the actions of acyl-activating enzyme 1 (AAE1), olivetol synthase (OLS) and olivetolic acid cyclase (OAC) (Taura et al., 2009; Gagne et al., 2012; Stout et al., 2012). OA is then prenylated by a geranylpyrophosphate:olivetolate geranyltransferase (or cannabigerolic acid synthase, CBGAS) with geranyl diphosphate (GPP), provided by the plastidial MEP pathway, forming the first *bona fide* cannabinoid cannabigerolic acid (CBGA) (Fellermeier and Zenk, 1998; Fellermeier et al., 2001). CBGA serves then as substrate for the oxidocyclases tetrahydrocannabinolic acid synthase (THCAS), cannabidiolic acid synthase (CBDAS) and cannabichromenic acid synthase (CBCAS), forming tetrahydrocannabinolic acid, cannabidiolic acid and cannabichromenic acid, respectively (Taura et al., 1996; Morimoto et al., 1998; Sirikantaramas et al., 2005; Laverty et al., 2019). Beyond biosynthesis in cannabis, biotechnological production in heterologous hosts may provide access to both the main and rare cannabinoids in pure form, based on cost-efficient methods in comparison to extraction from the plant (Carvalho et al., 2017). Most enzymes of the cannabinoid biosynthetic pathway and their genes have been functionally described for several

years (Taura et al., 1996; Morimoto et al., 1998; Wada et al., 2004; Brenneisen, 2007; Taura et al., 2007; Taura et al., 2009; Shoyama et al., 2012; Zhao et al., 2016), with the notable exception of the aromatic prenyltransferase (aPT) CBGAS that catalyzes formation of the branch-point intermediate CBGA. Until recently, introduction of the prenyl group onto the aromatic ring posed a challenge in metabolic engineering.

In general, aPTs link metabolites of the shikimate or polyketide pathways with metabolites derived from the mevalonate or MEP pathways, providing the aromatic and prenyl (isoprenoid) components, respectively (Yazaki et al., 2009; Nagia et al., 2019). The addition of a prenyl moiety leads to increased lipophilicity and thereby stronger interaction with biological membranes (Botta et al., 2005; Alhassan et al., 2014; Chen et al., 2016). Aromatic PTs are ubiquitously found in plants (Chen et al., 2016), bacteria (Kuzuyama et al., 2005), fungi (Nagia et al., 2012) and animals (Fredericks et al., 2011) and contribute to the diversification of aromatic metabolites, such as phenylpropanoids, flavonoids, and coumarins (Yazaki et al., 2009). All known plant aPTs belong to the UbiA protein superfamily that consists of membrane-bound enzymes exhibiting absolute dependence on divalent cations (Tsurumaru et al., 2010; Fiesel et al., 2015). Enzymes of the UbiA superfamily catalyze biosynthetic steps in the production of ubiquinones, menaquinones, plastoquinones, hemes, chlorophylls, tocopherols, and structural lipids (Li, 2016). These proteins are characterized by multiple transmembrane α-helices and two aspartate-rich motifs that are essential for the catalytic activity and make up a common active site (Bräuer et al., 2008; Akashi et al., 2009; Tsurumaru et al., 2010). The first motif has the amino acid sequence N(D/Q) involved in complexing the isoprenoid pyrophosphate via divalent cations, such as Mg^{2+} (Brandt et al., 2009).

In *Humulus lupulus* trichomes (hop; Cannabaceae), the UbiA aPTs HlPT1L and HlPT2 localize to the plastid membranes, where they form heterodimeric complexes catalyzing three sequential prenylations in the biosynthesis of bitter acids (Li et al., 2015). The catalytic activity was influenced by non-catalytic chalcone isomerase-like proteins (CHIL proteins), which serve as chaperones and improve the production of chalcone-derived metabolites (Ban et al., 2018). Like hop, cannabis also expresses a range of UbiA aPTs. Eight UbiA aPTs (*Cs*aPT1-8) were identified by Rea et al (2019) as part of their investigation of the biosynthesis of cannflavin A and B (M.L. Banett, 1963; Werz et al., 2014; Rea et al., 2019). Cannflavins are prenylated flavonoids. CsPT3 prenylates chrysoeriol with either GPP (cannflavin A) or dimethylallyl diphosphate (cannflavin B) as prenyl donors (M.L. Banett, 1963; Rea et al., 2019). CBGAS activity was shown to be catalyzed by a Mg^{2+}-dependent membrane-bound aPT, prenylating OA with GPP (Fellermeier and Zenk, 1998; Li, 2016).

A gene for a putative cannabis CBGAS was initially described in a patent in 2012 (Page and Boubakir, 2012). A trichome-specific EST library for aPTs provided a sequence with high similarity to the homogentisate phytyltransferase VTE2-2 (Collakova and DellaPenna, 2003; Page and Boubakir, 2012). The gene displayed high and selective expression in cannabis trichomes. This gene, named CsPT1, when expressed in *Saccharomyces cerevisiae* or *Spodoptera frugiperda* 9 insect cells, encoded an enzyme catalyzing conversion of OA and GPP into CBGA (Page and Boubakir, 2012). CsPT1 also prenylated other substrates such as olivetol, phlorisovalerophenone, naringenin and resveratrol with GPP as the prenyl-donor. Functional characterization of CsPT1 expressed in *Escherichia coli* has also been reported (Kabiri et al., 2018). In eukaryotes, the promiscuous prenyltransferase *NphB* from *Streptomyces sp.* strain CL190 served to replace *Cannabis* CBGAS in pathway engineering (Zirpel et al., 2017).

Recently, Luo et al. (2019) described the cannabis aPT CsPT4 exhibiting CBGAS activity in *S. cerevisiae* and used CsPT4 to produce natural and non-natural cannabinoids in *S. cerevisiae* from sugar (Luo et al., 2019).

Here, we mined published transcriptomes from two cannabis cultivars used in previous genomic studies, 'Finola' and Purple Kush (van Bakel et al., 2011; Booth et al., 2017; Booth and Bohlmann, 2019), to investigate (i) different UbiA aPTs and their potential CBGAS activity, (ii) co-expression of selected CsPT encoding genes for possible enzyme hetero-dimerization and (iii) the effect of CsCHIL proteins on CsPT activity. We also explored *N. benthamiana* and *S. cerevisiae* as heterologous hosts for production of cannabinoids and glycosides of their intermediates. Glycosylation of small molecules enhances their aequeous solubility and and may alter their stability and bioactivity, thus increasing their value as as food additives, therapeutics or nutraceuticals (De Bruyn et al., 2015). Uridine diphosphate (UDP)-glycosyltransferases (UGTs) that transfer activated sugar moieties on acceptor molecules are widely abundant in the plant kingdom and can be applied biotechnologically for the glycosylation of small molecules both *in vivo* and *in vitro* (De Bruyn et al., 2015). Cannabinoids are highly hydrophobic, impeding their applicability as pharmaceutical agents and incentivizing the formation of cannabinoid glycoconjuagtes.

Our results substantiate findings by Luo et al. demonstrating functional CBGAS activity of CsPT4 in *Nicotiana benthamiana* (Luo et al., 2019).

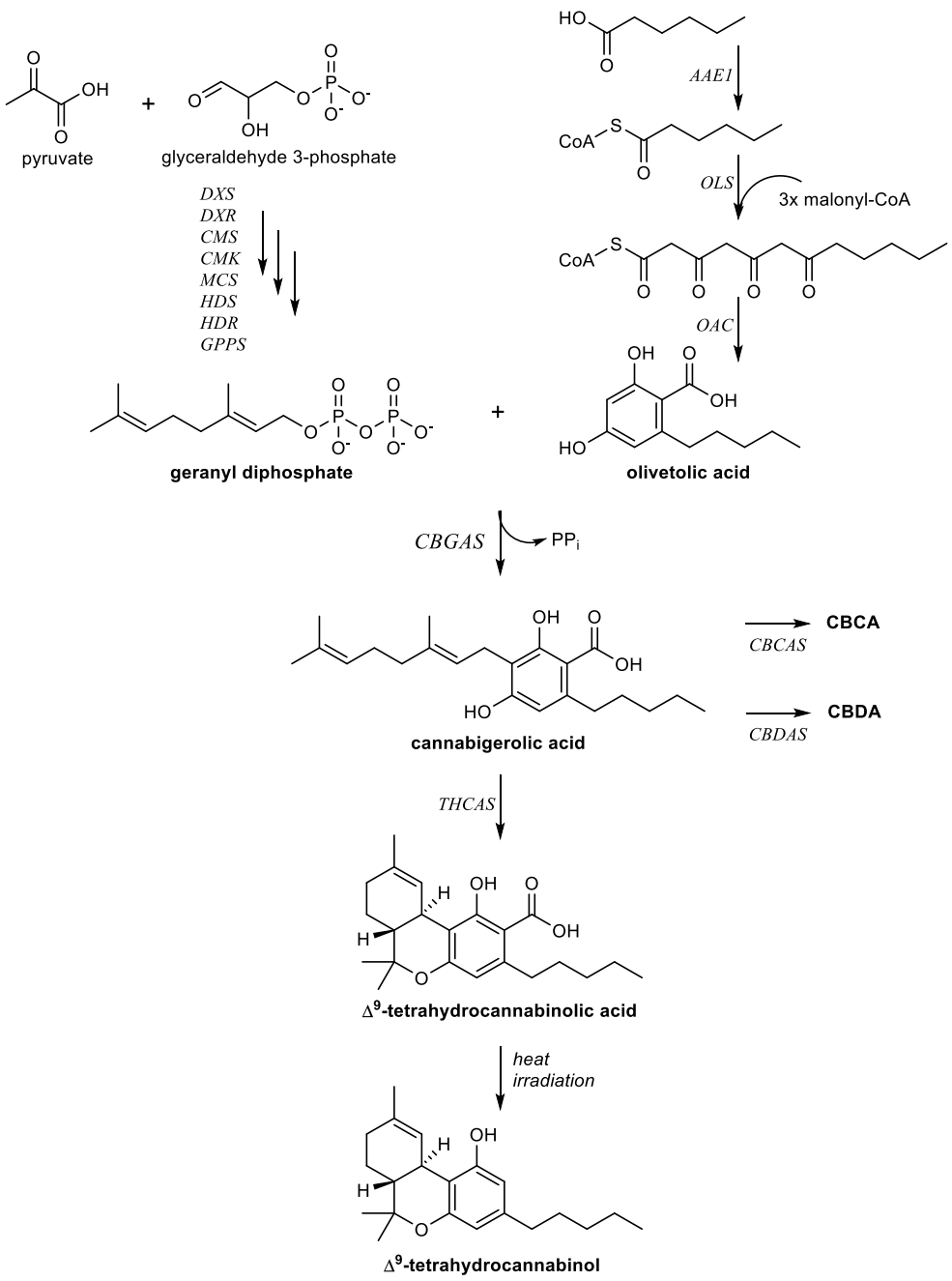


Figure 5.1 Cannabinoid biosynthetic pathway.
AAE1 = acyl-activating enzyme 1, CMK = 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, FAD = fatty acid desaturase, DXR = 1-deoxy-D-xylulose 5-phosphate reductoisomerase, DXS = deoxyxylulose-5-phosphate synthase, GPPS lsu = geranyl pyrophosphate_large subunit, GPPS ssu = geranyl pyrophosphate_small subunit, HDR = 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, HDS = 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, HPL = hydroperoxide lyase, IDI = isopentenyl diphosphate isomerase, LOX = lipoxygenase, MCT = 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, MDS = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, OAC = olivetolic acid cyclase, OLS = olivetol synthase, THCAS = Δ^9-tetrahydrocannabinolic acid synthase, aPT = UbiA aromatic prenyltransferase

5.3 Materials and Methods

5.3.1 Transcriptome meta-analysis:

A *de novo* transcriptome assembly (accession: PRJNA74271) and 30 Cannabis transcriptomes (with at least 10 million reads), were obtained from the National Center for Biotechnology Information (**Table 5.6**). The dataset was mined for aPTs using the target amino acid sequence CsPT1, obtained from Bureau, 2011(Bureau, 2011). RNA sequencing data were analyzed from 30 *Cannabis sativa* samples. Prior to further analysis, a quality check was performed on the raw sequencing data using FastQC and low quality portions of the reads were removed with BBDuk. The minimum length of the reads after trimming was set to 35 bp and the minimum base quality score to 25. Kallisto (v0.44.0) was used to calculate gene expression values as raw read counts across the samples and normalized transcripts-per-million values were calculated with edgeR. The nucleotide transcript sequences were translated using TransDecoder software (v-3.0.1). Interproscan (v5.28-67.0) was used to identify UbiA domains (PF01040) and ChloroP 1.1 Server-prediction was used to predict chloroplast transit peptides. In order to find groups of genes related to CBGAS and with high similarity to the target protein CsPT1, the OrthoMCL (Fischer et al., 2011) program was used. Co-expression analysis was performed using CEMiTool with the other known cannabinoid pathway genes as bait. All the transcripts with expression levels less than 17 cpm were removed using the R package NOISeq. Gene co-expression modules were obtained adding a filtering step based on the gene variance, P-value threshold (0.1) and the Pearson correlation coefficient (PCC). Only modules with at least 10 genes were considered. The genes in the pathway of cannabinoid were clustered in 5 expression modules.

5.3.2 Synthetic genes and cloning

Constructs for expression in *N. benthamiana* were generated with Gateway cloning (Katzen, 2007), constructs for expression in *E. coli* and *S. cerevisiae* were generated with USER-cloning (Geu-Flores et al., 2007). The polymerase chain reaction (PCR) for amplification of fragments for USER cloning was carried out with a mutant PfuX7 DNA polymerase designed for advanced uracil-excision DNA engineering (Nørholm, 2010). All other PCR reactions were performed using Phusion® High-Fidelity DNA polymerase (NEB). PCR products were purified using E.Z.N.A.® Gel Extraction Kit and SpinPrep™ PCR Clean Up Kit (VWR). Plasmid DNA was purified using E.Z.N.A.® Plasmid Mini Kit (VWR). *S. cerevisiae* genomic DNA was prepared using the protocol described by Lõoke (2011), for PCR based applications (Lõoke et al., 2011). DNA sequencing was performed by Macrogen, Inc. Synthetic genes for expression were ordered from Integrated DNA Technologies® or with attL1 and attL2 overhangs from GenScript®. Sequences for *CsPT1* and *CsPT4* were codon optimized for expression in *N. benthamiana, CsAAE1*, *CsOLS*, *CsOAC, CsPT4* and *CsTHCAS* were codon optimized for expression in *S. cerevisiae.* For *N. benthamiana*, gene expression constructs were generated in pEAQ-HT-DEST3[3] and GFP-tagged *aPTs* were generated in pMDC83 (University of Ghent), which contains a 35S promoter and a C-terminal GFP. For *S. cerevisiae,* genomic integration (**Table 5.1**) was chosen over expression via episomal plasmids to favor simultaneous and stable expression of up to 10 genes as well as to enable the use of selection marker recycling (EasyClone)(Mikkelsen et al., 2012; Stovicek et al., 2015) signal peptides were truncated according to the targetP prediction (Emanuelsson et al., 2000; Almagro Armenteros et al., 2019) before incorporation into *S. cerevisiae* genomic site XI-2 (Mikkelsen et al., 2012). *FAS1*/*FAS2* native promoter substitution by a galactose promoter (PGAL1) was performed using the

clustered regularly interspaced short palindromic repeats, associated endonuclease 9 (CRISPR/Cas9) system (Vanegas et al., 2017). Donor fragments were designed with sequences upstream and downstream of the open reading frame (ORF) to be deleted and ordered as duplex oligonucleotides (Integrated DNA Technologies®).

Plasmid number	**Assembler type**	**Genome integration site**	**Promoter 1**	**Gene 1**	**Terminator 1**	**Promoter 2**	**Gene 2**	**Terminator 2**
p*Cs*1	asb1	XI-2	pPGK1	*Cs*NbdaPT1	tCYC1	pTEF1	*AgGPPS*	tPGI1
p*Cs*2	asb1	XI-2	pPGK1	*Cs*NbdaPT4	tCYC1	pTEF1	*AgGPPS*	tPGI1
p*Cs*3	asb2A		pFBA1	*Cs*ScOAC	tFBA1	pSED1	*Cs*ScTHCAS	tENO2
p*Cs*6	asb3	XI-2	pTEF2	*Cs*ScAAE1	tADH1	pTDH3	*Cs*ScOLS	tTDH2
p*Cs*7	asb2		pFBA1	*Cs*ScOAC	tFBA1	pSED1	*Cs*ScTHCAS	tENO2
p*Cs*10	asb2B		-	-	-	pADH2	SrUGT71E1	tENO2
p*Cs*11	asb2C		-	-	-	pCCW12	OsUGT5	tFBA1
p*Cs*13	asb2B		pALD4	*Cs*ScdaPT4	tFBA1	pADH2	SrUGT71E1	tENO2

Table 5.1 List of integrative plasmids for stable integration into the *S. cerevisiae* genome.

5.3.3 RNA isolation and cDNA creation

Total RNA was isolated from ‘Finola’ and ‘Purple Kush’ tissues using the PureLink Plant RNA Reagent kit (Thermo Fisher) according to manufacturer’s instructions. RNA concentration and integrity were determined using a BioAnalyzer RNA Nanochip (Agilent). cDNA was produced

from total RNA using the Smarter RACE c7410DNA amplification kit (Clontech) or a Maxima First Strand cDNA Kit (Thermo Fisher).

5.3.4 Preparation of *Agrobacterium tumefaciens*

Transient expression in *N. benthamiana* leaves was performed as previously described (Wood et al., 2009), with some minor modifications. For *in planta* enzyme/pathway assays, 25 µL of *A. tumefaciens* AGL1 were transformed with 20-50 ng of the following expression plasmids: pEAQ3-HT-EXP_*Cs*ScAAE1 (*AEE1*), pEAQ3-HT-EXP_*Cs*ScOLS (*OLS*), pEAQ3-HT-EXP_*Cs*ScOAC (*OAC*), pEAQ3-HT-EXP_*Cs*NbPT01 (*CsPT01)*, pEAQ3-HT-EXP_*Cs*NbPT4 (*CsPT4*) and pEAQ3-HT-EXP_*Cs*ScTHCAS (Δ^9-*THCAS*) (**Table 6.2**). For i*n planta* localization studies, *CsPT*-GFP vectors were transformed into chemically competent *A. tumefaciens* GV3850 by heat-shock. Transformants were selected by antibiotic resistance and verified by colony PCR. Positive colonies were grown overnight at 28 °C, 200 rpm in 5 ml of LB containing appropriate antibiotic*s*. 15 mL of LB with appropriate antibiotic*s* were inoculated with 500 µL of the pre-culture and grown for 16 h at 28 °C, 200 rpm to an OD_{600} of about 0.8. AGL1 cell cultures were pelleted and resuspended in tap water supplemented with 100 µM acetosyringone at OD_{600}=0.4-0.5. This cell suspension was shaken at 28 °C, 200 rpm, 1-2 h before infiltration. Equal volumes of AGL1 cell cultures with different expression vectors were combined to obtain the desired gene combinations of the biosynthetic pathway. GV3850 cell cultures were pelleted and resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfonic acid pH5.6, 10 mM $MgCl_2$, 100 µM acetosyringone) at an OD_{600} of 0.1 and shaken at room temperature for 2-3 h.

Name with characteristics	Source
plasmids for *Nicotiana benthamiana* studies	
pEAQ-HT-DEST3	Thermo Fisher Scientific
pDONR207	Thermo Fisher Scientific
pENTR207_CsScOAC	This study
pUC_ENTR_CsNbPT1	GenScript®
pUC_ENTR_CsNbPT4	GenScript®
pEAQ3-HT-EXP_CsScAAE1	This study
pEAQ3-HT-EXP_CsScOLS	This study
pEAQ3-HT-EXP_CsScOAC	This study
pEAQ3-HT-EXP_CsNbPT1	This study
pEAQ3-HT-EXP_CsNbPT2	This study
pEAQ3-HT-EXP_CsNbPT3	This study
pEAQ3-HT-EXP_CsNbPT4	This study
pEAQ3-HT-EXP_CsNbPT5	This study
pEAQ3-HT-EXP_CsNbPT6	This study
pEAQ3-HT-EXP_CsNbPT7	This study
pEAQ3-HT-EXP_CsNbPT8	This study
pEAQ3-HT-EXP_CsNbPT9	This study
pEAQ3-HT-EXP_CsNbPT10	This study
pEAQ3-HT-EXP_CsNbPT11	This study
pEAQ3-HT-EXP_CsNbPT12	This study
pEAQ3-HT-EXP_CsScTHCAS	This study
pEAQ3-HT-EXP_CsCHIL1	This study
pEAQ3-HT-EXP_CsCHIL2	This study
pLIFE_AgGPPS	Kindly provided by Irini Pateraki
pLIFE_PbDXS	Kindly provided by Irini Pateraki

Plasmids	Genotype and characteristics	Resource
EPSC3911	pYU-URA3-3, Yeast Integrative plasmid (YPp), AmpR, URA3, XI-1, assembler 1, USER	[1]
EPSC2651	pU0002-YPp, AmpR, assembler 2, USER	[1]
EPSC3912	pU0002 YPp, AmpR, XI-1, assembler 3, USER	[1]
asb2A	pU0002-YPp, AmpR, assembler 2A, USER	Victor Forman
asb2B	pU0002-YPp, AmpR, assembler 2B, USER	Victor Forman
asb2C	pU0002-YPp, AmpR, assembler 2C, USER	Victor Forman
pSMG32	2 micron, AmpR, HygMX loxP,	Susanne Germann
pSMG33	2 micron, AmpR, BleMX loxP,	Susanne Germann
pRS315+FAS1(I306A)	pADH2-FAS1(I306A)	[2]
pRS313+FAS2(G1250S)	pADH2-FAS2(G1250S)	[2]
pCs1	EPSC3911{pPGK1-CsNbdaPT1-tCYC1/pTEF1-AgGPPS-tPGI1}	This study

pCs2	EPSC3911{pPGK1-CsNBdaPT4-tCYC1/pTEF1-AgGPPS-tPGI1}	This study
pCs3	asb2A{pFBA1-CsScOAC-tFBA1/pSED1-CsScTHCAS-tENO2}	This study
pCs6	EPSC3912{pTEF2-CsScAAE1-tADH1/pSED1-CSScTHCAS-tENO2}	This study
pCs7	EPSC2651{ pFBA1-CsScOAC-tFBA1/pSED1-CsScTHCAS-tENO2}	This study
pCs10	asb2B{pADH2-SrUGT71E1-tENO2}	This study
pCs11	asb2C{pCCW12-OsUGT5-tFBA1}	This study
pCs13	asb2B{pALD4-CsScdaPT4-tFBA1/pADH2-SrUGT71E1-tENO2}	This study
pYCA191	pSMG33-(pSNR52-FAS1_IS1_974-tCYC1) gRNA	This study
pYCA193	pSMG32-(pSNR52-FAS2_IS1_829-tCYC1) gRNA	This study

Table 5.2 List of plasmids used in this study

5.3.5 Nicotiana benthamiana growth conditions and infiltration

N. benthamiana plants were grown in greenhouse conditions (day temperature of 20°C, night temperature of 19°C, light period of 16 h light/8h dark, lighting conditions of 30 cm distance (plant-lamp, 230 µE) until 4 weeks old. On the day of infiltration, plants were brought to the laboratory, watered from the bottom and acclimatized for at least 3 h. *N. benthamiana* whole leaves were infiltrated on the abaxial side with AGL1 cell suspensions using a needleless 1 mL syringe. Per gene combination, 3 whole leaves on 3-4 plants were infiltrated. Leaf material was combined to provide 3-4 biological replicates. After infiltration, the plants were incubated for 1 night in the laboratory (low light conditions) and for 3-5 nights in the greenhouse (strong light conditions). *CsPT*-GFP localization was visualized using FV1000 Laser Scanning Microscope (Olympus) with a laser exciting at 390 nm. Tissues were mounted in deionized water. GFP fluorescence was detected between 450-550 nm, and chlorophyll at 610-710 nm. Image analysis was performed using Volocity software (Quorum technologies).

5.3.6 *Nicotiana benthamiana* enzyme assays

Enzyme assays in *N. benthamiana* were performed 4-5 days post infiltration. For *in planta* assays, infiltrated leaf areas were co-infiltrated with 1 mM hexanoic acid (SigmaAldrich) or 1mM olivetolic acid (Carbosynth) with 1 mM GPP (isoprenoids.com) in 5% ethanol using a needleless 1 mL syringe. After 24 h, ca. 400 mg of co-infiltrated tissue was macerated in 20% methanol and shaken at room temperature for 1 h, 200 rpm. The macerated leaf mixture was then filtered through a 0.22 µm mesh and the filtrate was analyzed by UPLC-Q-TOF-MS^2.

5.3.7 *Saccharomyces cerevisiae* transformation and fermentation

Yeast integration plasmids with cannabinoid biosynthetic genes were transformed into *S. cerevisiae* using the LiAc/SS carrier DNA/polyethylene glycol method[3]. All transformants were grown on synthetic complete medium lacking uracil (SC –URA) as auxotrophy marker to select positive transformants (**Table 5.3**). Using 2 mL 96-well plates, transformants were precultured in 500 µL of liquid SC -URA (24 h, 28 °C, 200 rpm). Then 50 µL preculture were used to inoculate 450 µL yeast peptone dextrose medium (YPD) at pH4, supplemented with 2% glucose 0.1 mM OA (72 h, 28 °C, 200 rpm). Cell cultures were extracted by addition of ethylacetate:formic acid 0.05% and metal bead-beating at 30 Hz for 3 min. Cell cultures were extracted three times. Organic phases were combined, freeze-dried and resuspended in 60 methanol for analysis. Cells cultures for UGT-activity tests were extracted by addition of 80% methanol 1:1. The mixture was incubated at room temperature with 100 rpm shaking for 60

minutes and subsequently centrifuged at max speed. Supernatant was analyzed by UPLC-ESI-tripleQuad-MS^2.

Strain name	Genotype or characteristic	Origin
Background strains		
IS1	MATa HO::KO Δhis3 Δleu2 Δura3	Evolva
S288C	MATalpha HO::KO Δura3 Δhis3 Δleu2	Evolva
Constructed strains		
yCs0.1	Mat alpha; ura3D0; his3D0; leu2D0; TRP1; lys2D0; MET15; fas1::uptag-kanMX4-downtag; fas2::uptag-kanMX4-;pFAS1 (I306A); pFAS2 (G1250S)	This study
yCs0.4	Mat alpha; ura3D0; his3D0; leu2D0; TRP1; lys2D0; MET15; fas1::uptag-kanMX4-downtag; fas2::uptag-kanMX4-;pRS34 (WT FAS1); pRS38 (WT FAS2)	This study
yCs1	Mat alpha; ura3D0; his3D0; leu2D0; TRP1; lys2D0; MET15; fas1::uptag-kanMX4-downtag; fas2::uptag-kanMX4-;pFAS1 (I306A); pFAS2 (G1250S);	This study
yCs2	yCs0.1{$AAE1^{Sc}$, OLS^{Sc}, OAC^{Sc}, $CsdaPT04^{Nb}$, $THCAS^{Sc}$, $AgGPPS^{WT}$}	This study
yCs3	yCs0.4{$AAE1^{Sc}$, OLS^{Sc}, OAC^{Sc}, $CsdaPT01^{Nb}$, $THCAS^{Sc}$, $AgGPPS^{WT}$}	This study
yCs4	yCs0.4{$AAE1^{Sc}$, OLS^{Sc}, OAC^{Sc}, $CsdaPT04^{Nb}$, $THCAS^{Sc}$, $AgGPPS^{WT}$}	This study
yCs8	yCs0.4{$AAE1^{Sc}$, OLS^{Sc}, OAC^{Sc}, $CsdaPT04^{Nb}$, $THCAS^{Sc}$, $AgGPPS^{WT}$, $SrUGT71E1^{WT}$; $OsUGT5^{WT}$}	This study
yCs16	yCs0.4{$AAE1^{Sc}$, OLS^{Sc}, OAC^{Sc}, $CsdaPT04^{Nb}$, $CsdaPT04^{Sc}$, $THCAS^{Sc}$, $AgGPPS^{WT}$, $SrUGT71E1^{WT}$; $OsUGT5^{WT}$ }	This study

Table 5.3 Yeast strains constructed

5.3.8 Glucosylation of OA, CBGA and Δ^9-THCA

Five plant UGTs were tested for *in vitro* glycosylation of OA, CBGA and Δ^9-THCA in *Escherichia coli* (**Table 5.4)**. The five UGT cDNAs described above were cloned into a pET30a+ expression vector (Novagen). The resulting plasmids and a pET30 control were transformed into *E. coli* Xjb-autolysis BL21 (DE3) (Zymo Research).

For *in vitro* studies, crude lysate preparation was performed with the following method. Colonies of *E. coli* strains constructed to express a UGT enzyme were placed into sterile 96

deep well plates with 1 mL of NZCYM bacterial culture broth (SigmaAldrich®) with ampicillin or kanamycin. Samples were incubated overnight (37°C, 200 rpm). The following day, 50 μL of each culture were transferred to a new sterile 96 deep well plate with 1 mL of NZCYM bacterial culture broth with ampicillin and polypeptide expression inducers. Samples were incubated (20°C, 200 rpm, 20 h). On day 3, the plate was centrifuged (4000 rpm, 10 min, 4°C). After decanting the supernatant, 50 μL of buffer comprising Tris-HCI, $MgCI_2$, $CaCI_2$, and protease inhibitors were added to each well and cells were resuspended by shaking (200 rpm, 5 min, 4°C). The contents of each well (i.e., cell slurries) were transferred to a PCR plate and frozen at -80°C overnight. Frozen cell slurries were thawed at room temperature for up to 30 min. If the thawing mix was not viscous due to cell lysing, samples were frozen and thawed again. When samples were nearly thawed, 25 μL of binding buffer containing DNase and $MgCI_2$ were added to each well. The PCR plate was incubated at room temperature (5 min, 500 rpm), until samples became less viscous. Finally, samples were centrifuged at 4000 rpm for 5 min, and supernatants were used to measure UGT activity. UGT enzyme samples were screened for *in vitro* for activity on substrates OA, CBGA and Δ^9-THCA by preparing a reaction mixture according to **Table 5.5**. The reaction mixture was incubated overnight at 30°C. The reaction was stopped by adding 30 μl of 100% DMSO. The resultant mixture was diluted further with 90 μI 50% DMSO for LC-MS analysis.

In vivo expression of Sr*UGT71E1* and *UGT5* in *S. cerevisiae* was performed with integrative plasmids (EasyClone system) and each UGT was flanked by a unique yeast promoter and terminator. *UGT71E1* and *UGT5* were tested individually in *S. cerevisiae* by feeding with CBGA or Δ^9-THCA. Feeding experiments with CBGA were performed in

synthetic glucose media at pH 4.0 supplemented with 200 µM of CBGA. Feeding experiments with Δ^9-THCA were performed with synthetic glucose media at pH 4.0 with 4% DMSO supplemented with 55 µM of Δ^9-THCA.

Name	Organism	Genbank no.
UGT73B5	*Arabidopsis thaliana*	NM_127108.4
UGT76C5	*Arabidopsis thaliana*	NM_120671.4
UGT73B3	*Arabidopsis thaliana*	NM_119574.3
UGT71E1	*Stevia rebaudiana*	AY345976
UGT5	*Oryza sativa*	XP_015622068.1

Table 5.4 List of genes encoding plant UDP-glycosyltransferases tested for *in vitro* glycosylation of OA, CBGA and Δ9-THCA.

Component	Volume (µL)
H_20	4.2
Alkaline phosphatase	0.3
4X Buffer (10 mM Tris-HCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$)	7.5
UDP-Glucose (1mM)	9
Substrate	3
UGT sample	6

Table 5.5 UGT Activity Assay Reaction Mixture.

5.3.9 Chemical synthesis of olivetolic acid glycoside [4-β-D-glucopyranosyloxy-2-hydroxy-6-pentylbenzoic acid (4)]

Chemical synthesis of olivetolic acid glycoside is shown in **Figure 5.2**.

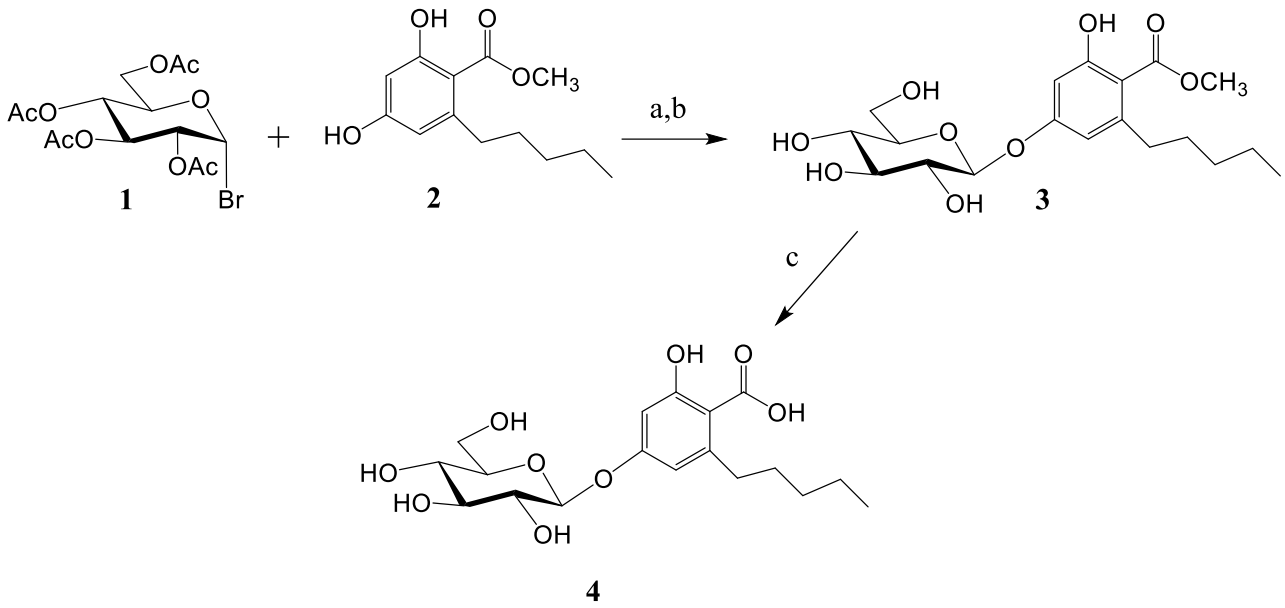


Figure 5.2 Reaction conditions.
a tris[2-(2-methoxyethoxy)ethyl]amine (TMEA), 0.1 M NaHCO3-0.1 M KCl (1:1 v/v), relux, 48 h. b. MeONa cat / MeOH, stirring at room temperature, 4 h. c.1 M NaOH /H_2O/MeOH (1:2:3), stirring at room temperature, 8 h.

Commercially available 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide (**1**) and methyl 2,4-dihydroxy-6-pentylbenzoate (methyl olivetolate) (**2**) were subjected to glucosylation under phase-transfer catalysis (PTC) conditions according to procedure previously used for the synthesis of flavonoid and hydroxycinnamic acid glycosides using tris[2-(2-methoxyethoxy)ethyl]amine (TMEA) as the catalyst (Alluis and Dangles, 1999; Galland et al., 2007). The synthesized target molecules were purified by chromatography and their structures confirmed by ^{1}H- and ^{13}C-NMR spectroscopy.

5.3.10 Metabolomics on cannabinoids by LC-MS/Q-TOF

LC-MS/MS was performed on a Dionex UltiMate 3000 Quaternary Rapid Separation UHPLC+ focused system (Thermo Fisher Scientific, Germering, Germany). Separation was achieved on a Kinetex 1.7 µm XB-C18 column (150 × 2.1 mm, 1.7 µm, 100 Å, Phenomenex). For eluting,

0.05% (v/v) formic acid in H2O and MeCN [supplied with 0.05% (v/v) formic acid] were employed as mobile phases A and B, respectively. Gradient conditions were as follows: 0.0−0.5 min 5% B; 0.5−5.0 min 5−33% B; 5.0−11.0 min 33−35% B, 11.0−20.0 min 35−100% B, 20.0−21.0 min 100% B, 21.0−21.2 min 100−5% B, and 21.2−23.0 min 5% B. The flow rate of the mobile phase was 300 μL/min. The column temperature was maintained at 55°C. UV spectra for each sample were acquired at 210, 275, 310, and 343 nm. The UHPLC was coupled to a Compact micrOTOF-Q mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ion source (ESI) operated in positive ionization mode. The ion spray voltage was maintained at 4500 V. The dry temperature was set to 250°C, and the dry gas flow was set to 8 L/min. Nitrogen was used as the dry gas, nebulizing gas, and collision gas. The nebulizing gas was set to 2.5 bar and collision energy to 10 eV. HRESIMS and MS/MS spectra were acquired in an m/z range from 50 to 1000 amu at a sampling rate of 2 Hz. Na-formate clusters were used for calibration. All files were automatically calibrated by post-processing.

5.3.11 t-quad analysis of *in vitro* and *in vivo* glucosylation assays and quantification

Sample analysis was performed by UPLC coupled to a triple-quadrupole mass spectrometer interfaced with an electrospray ion source (ESI) (Waters, Milford, MA). 1uL of the extracted sample was injected into the LC-MS system and separation was achieved in reversed phase using a C18 BEH (1.7μm, 2.1x50mm) column equipped with a C18 BEH (1.7μm) pre-column (Waters, Milford, MA) and mobile phases consisted of 0.1% formic acid (Sigma-Aldrich) in Milli-Q© grade water (A) and 0.1% formic acid in MS grade acetonitrile (B) with a flow rate of 0.6 mL/min. Masslynx software (version 1.6) was used for instrument control, while Markerlynx for data integration.

Different mobile phase compositions and MS experiments were set for the analysis of the different compounds. For OA analysis, the separation began with a linear gradient from 20% B to 80% B in 1.6 min, reaching 100% B in 0.4 min and maintained for 0.2 min, then the column was re-equilibrated at 20% B for 0.6 min before the next injection. The total run time for the method was 2.8 min. The mass spectrometer was operated in positive and negative ion mode using Single Ion Monitoring (SIM) or untargeted mode. The m/z value was set at 223.09 (in negative mode) for olivetolic acid. The capillary voltage was set at 2.2 kV and 3.0 kV in negative and positive mode, respectively. The sampling cone was set at 35V. The source and the desolvation gas (nitrogen) temperature were set at 150 and 350 ˚C, respectively, while cone and desolvation gas flow rates at 30 and 600 l/hr.

For CBGA and Δ^9-THCA analysis, the separation was achieved using a linear gradient from 50% B to 100% B in 1.0 min, and maintained for 0.5 min, then the column was re-equilibrated at 50% B for 0.7 min before the next injection. The total run time for the method was 2.2 min. SIM mode was used for THCA detection. The m/z value selected for THCA was 357.2, while 359.2 was set for CBGA. The cone voltage was set at 55V. For all the different MS analyses, the capillary voltage was set at 2.2 kV. Bruker DataAnaysis was used to analyze Q-TOF data. Extracted ion chromatograms were generated with ±0.01 tolerance.

5.4 Results

5.4.1 Meta-analysis of cannabis transcriptomes

To obtain a comprehensive set of candidate aPTs of the UbiA superfamily, we analysed 30 different transcriptomes (**Table 5.6**) of female cannabis plants including root, primary stem, shoot, stem-petioles, young leaf, mature leaf, pre-flower, early-stage flower, mid-stage flower,

mature flower, flower buds and trichome (**Figure 5.3**). This provided a total of 14 aPT unigenes, two of which were redundant (PK11068.2 redundant with PK11068.1; PK151949.2 redundant with PK15949.1), resulting in 12 cannabis aPTs (*Cs*PT1-12), including the previously reported *Cs*PT1-8 (Rea et al., 2019). Co-expression analysis between these 12 CsPTs and other known cannabinoid pathway genes was used to prioritize CBGAS candidates (Tai et al., 2018; Zheng et al., 2018), by common expression pattern in trichomes. Most cannabinoid pathway genes and CsPTs grouped into six co-expression modules (**Figure 5.3**, **Table 5.7**). The co-expression modules showed that CsPT10 and 7 co-express with genes of the MEP-pathway and the early cannabinoid pathway (AAE, OLS). *Cs*PT1, 3, 4, 11 and 12 co-expressed with each other and later cannabinoid pathway genes (OAC, THCAS). A fatty acid desaturase (FAD), a putative upstream enzyme in OA biosynthesis, also grouped with OAC and THCAS. CsPT8 and 9 grouped with the MEP pathway enzyme DXR, and CsPT2 and CsPT5 did not group with cannabinoid pathway genes.

Sample Name	Experiment Title	Tissue	Publication
SRR352196	Cannabis sativa Purple Kush Early-stage Flower polyA RNA-Seq library PK-EFLW	Early-stage Flower	Allen et al., 2019
SRR306884	Cannabis sativa flower buds RNA-Seq (CSA_AP)	Flower Buds	N/A
SRR306867	Cannabis sativa flower-buds RNA-Seq (CSA_AG)	Flower Buds	N/A
SRR306866	Cannabis sativa flower- buds RNA-Seq (CSA_AF)	Flower Buds	N/A
SRR192369	Cannabis sativa Flower Buds PE RNA-Seq (CSA_RE)	Flower Buds	N/A
SRR192371	Cannabis sativa Mature and Immature Leaf PE RNA-Seq (CSA_RG)	Leaf	N/A
SRR306870	Cannabis sativa mature flower (fully expanded) RNA-Seq (CSA_AJ)	Mature Flower	N/A
SRR306869	Cannabis sativa mature flower (fully expanded) RNA-Seq (CSA_AI)	Mature Flower	N/A
SRR306868	Cannabis sativa mature flower (fully expanded) RNA-Seq (CSA_AH)	Mature Flower	N/A
SRR192370	Cannabis sativa Mature Flower PE RNA-Seq (CSA_RF)	Mature Flower	N/A
SRR306886	Cannabis sativa mature leaf (fully expanded) RNA-Seq (CSA_AR)	Mature Leaf	N/A
SRR306885	Cannabis sativa mature leaf (fully expanded) RNA-Seq (CSA_AQ)	Mature Leaf	N/A
SRR306875	Cannabis sativa mature leaf (fully expanded) RNA-Seq (CSA_AN)	Mature Leaf	N/A
SRR352195	Cannabis sativa Purple Kush Mid-stage Flower polyA RNA-Seq library PK-MFLW	Mid-stage Flower	Allen et al., 2019
SRR352198	Cannabis sativa Purple Kush Pre-Flower polyA RNA-Seq library PK-PFLW	Pre-Flower	Allen et al., 2019
SRR306877	Cannabis sativa primary stem (entire) RNA-Seq (CSA_AO)	Primary Stem	N/A
SRR192373	Cannabis sativa Primary Stem PE RNA-Seq (CSA_RI)	Primary Stem	N/A
SRR352202	Cannabis sativa Purple Kush Root polyA RNA-Seq library PK-RT	Root	Allen et al., 2019
SRR306863	Cannabis sativa entire root RNA-Seq (CSA_AC)	Root	N/A
SRR306862	Cannabis sativa entire root RNA-Seq (CSA_AB)	Root	N/A

SRR306861	Cannabis sativa entire root RNA-Seq (CSA_AA)	Root	N/A
SRR192372	Cannabis sativa Entire Root PE RNA-Seq (CSA_RH)	Root	N/A
SRR352200	Cannabis sativa Purple Kush Shoot polyA RNA-Seq library PK-SHT	Shoot	Allen et al., 2019
SRR306865	Cannabis sativa stem- petioles (entire) RNA-Seq (CSA_AE)	Stem-Petioles	N/A
SRR306864	Cannabis sativa stem- petioles (entire) RNA-Seq (CSA_AD)	Stem-Petioles	N/A
SRR684087	Transcriptome analysis of the Cannabis sativa trichome.	Trichome	N/A
SRR292255	Transcriptome analysis of the Cannabis sativa trichome.	Trichome	N/A
SRR306874	Cannabis sativa young leaf (<25% expanded) RNA-Seq (CSA_AM)	Young Leaf	N/A
SRR306872	Cannabis sativa young leaf (<25% expanded) RNA-Seq (CSA_AL)	Young Leaf	N/A
SRR306871	Cannabis sativa young leaf (<25% expanded) RNA-Seq (CSA_AK)	Young Leaf	N/A

5.6 Transcriptomes analyzed in this study

Focusing on trichome specific expression, *CsPT1* and *CsPT4* were the most highly expressed, followed by *CsPT3, CsPT7* and *CsPT10*, which displayed moderate expression in trichomes and *CsPT2, CsPT5, CsPT6, CsPT8, CsPT9, CsPT11 and CsPT12*, which displayed low expression in trichomes. Most cannabinoid biosynthetic genes as well as genes of the MEP isoprenoid biosynthetic pathway were highly expressed in trichomes. Transcripts of cannabinoid biosynthetic genes were in general most highly expressed in trichomes, moderately expressed in flower tissues, shoot and young leaves and low or not detectable in roots and primary stem. qRT-PCR confirmed highest expression levels of all cannabinoid biosynthetic genes as well as *CsPT1* and *CsPT4* in trichomes compared to other tissue (**Figure 5.10**). Shotgun proteomic analysis confirmed the presence of CsPT1 and CsPT4 enzymes in female flowers but no other tissue (**Table 5.8**).

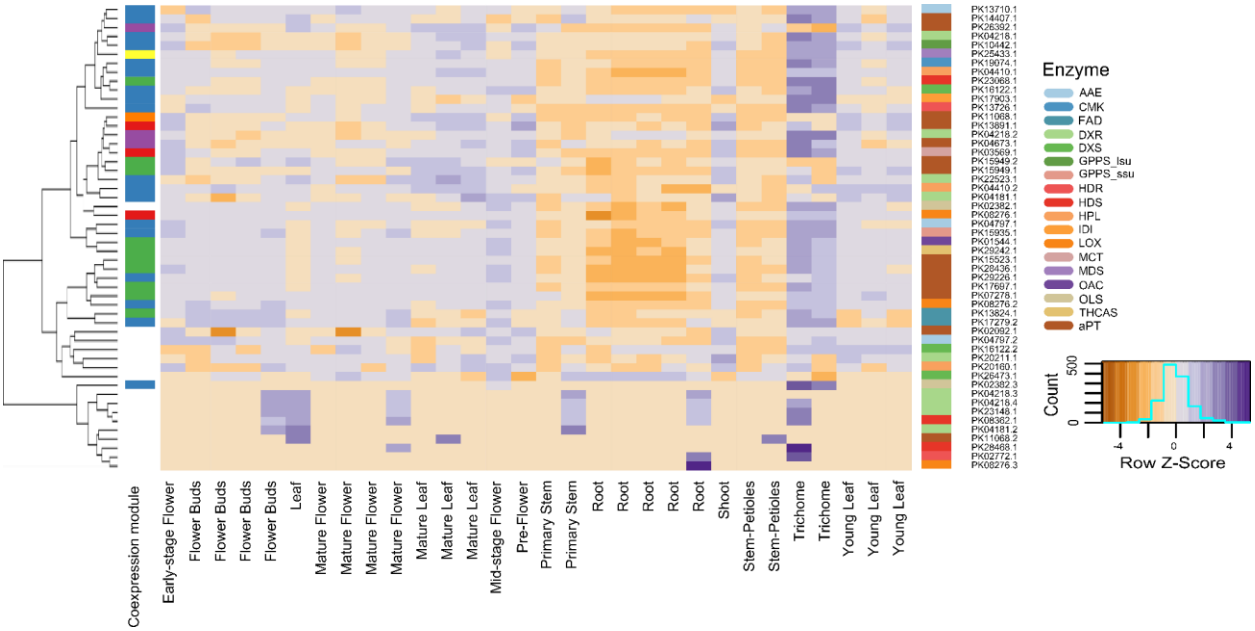


Figure 5.3 Heatmap representation of the expression of cannabinoid pathway genes and *aPT* candidate genes. Each transcript is associated to a color label on the right-hand side, indicating the putative enzyme identity. Tissues of transcriptome origin are indicated in the bottom row. Co-expression modules are indicated by colors on the left-hand side of the plot, and were determined using CEMiTool (Russo et al., 2018). White indicates that no co-expression module was defined. Clustering is unsupervised hierarchical clustering analysis using the R package gplots, clustering method hclust. Relative expression strength indicated by Row Z-score, where the teal line indicates the frequency of each Z-score bin. *AAE1* = acyl-activating enzyme 1, *CMK* = 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, *FAD* = fatty acid desaturase, *DXR* = 1-deoxy-D-xylulose 5-phosphate reductoisomerase, *DXS* = deoxyxylulose-5-phosphate synthase, *GPPS_lsu* = geranyl pyrophosphate_large subunit, *GPPS_ssu* = geranyl pyrophosphate_small subunit, *HDR* = 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase, *HDS* = 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, *HPL* = hydroperoxide lyase, *IDI* = isopentenyl pyrophosphate isomerase, *LOX* = lipoxygenase, MCT = 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, *MDS* = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *OAC* = olivetolic acid cyclase, *OLS* = olivetol synthase, *THCAS* = Δ^9-tetrahydrocannabinolic acid synthase, *aPT* = UbiA aromatic prenyltransferase.

Accession	IDI	Co-expression module	Accession	IDI	Co-expression module
PK13710.1	*AAE*	Blue	PK13891.1	*CsPT6*	Red
PK14407.1	*CsPT10*	Blue	PK03569.1	*GPPS_ssu*	Red
PK04218.1	*DXR*	Blue	PK08276.1	*LOX1*	Red
PK10442.1	*GPPS_lsu*	Blue	PK26392.1	*CsPT9*	Purple
PK19074.1	*CMK*	Blue	PK04218.2	*DXR*	Purple
PK04410.1	*HPL*	Blue	PK04673.1	*CsPT8*	Purple
PK23068.1	*HDS*	Blue	PK11068.1	*CsPT5*	Orange
PK16122.1	*DXS*	Blue	PK25433.1	*MDS*	Yellow
PK17903.1	*IPP*	Blue	PK02382.1	*OLS*	unassigned
PK13726.1	*HDR*	Blue	PK02092.1	*CsPT2*	unassigned
PK22523.1	*DXR*	Blue	PK04797.2	*AAE*	unassigned
PK04410.2	*IDI*	Blue	PK16122.2	*DXS*	unassigned
PK04181.1	*DXR*	Blue	PK20211.1	*DXR*	unassigned
PK04797.1	*AAE1*	Blue	PK20160.1	*HPL*	unassigned
PK15935.1	*GPP_ssu*	Blue	PK26473.1	*DXS*	unassigned
PK29226.1	*CsPT7*	Blue	PK04218.3	*DXR*	unassigned
PK08276.2	*LOX1*	Blue	PK04218.4	*DXR*	unassigned
PK17279.2	*FAD*	Blue	PK23148.1	*DXR*	unassigned
PK02382.3	*OLS*	Blue	PK08362.1	*HDS*	unassigned
PK15949.2	*(CsPT11)*	Green	PK04181.2	*DXR*	unassigned
PK15949.1	*CsPT11*	Green	PK11068.2	*(CsPT5)*	unassigned
PK01544.1	*OAC*	Green	PK28468.1	*HDS*	unassigned
PK29242.1	*THCAS*	Green	PK02772.1	*HDR*	unassigned
PK15523.1	*CsPT4*	Green	PK08276.3	*LOX*	unassigned
PK28436.1	*CsPT1*	Green			
PK17697.1	*CsPT3*	Green			
PK07278.1	*CsPT12*	Green			
PK13824.1	*FAD*	Green			

Table 5.7 Co-expression analysis of cannabinoid pathway genes

. Unigenes with matching expression patterns were grouped using the R package 'CEMItool' (Russo et al., 2018). *AAE1* = acyl-activating enzyme 1, *CMK* = 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, *FAD* = fatty acid desaturase, *DXR* = 1-deoxy-D-xylulose 5-phosphate reductoisomerase, *DXS* = deoxyxylulose-5-phosphate synthase, *GPPS_lsu* = geranyl pyrophosphate_large subunit, *GPPS_ssu* = geranyl pyrophosphate_small subunit, *HDR* = 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase, *HDS* = 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, *HPL* = hydroperoxide lyase, *IDI* = isopentenyl pyrophosphate isomerase, *LOX* = lipoxygenase, *MCT* = 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, *MDS* = 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, *OAC* = olivetolic acid cyclase, *OLS* = olivetol synthase, *THCAS* = Δ^9-tetrahydrocannabinolic acid synthase, *aPT* = UbiA aromatic prenyltransferase. Parentheses indicate redundant transcripts.

Protein ID	Transcript number	Female flower	Leaf	Root	Stem
AAE1	PK04797.1:140-2305(+)	++	-	-	-
PKS	PK02382.1:1-1227(+)	++	+	-	++
OAC	PK01544.1:96-401(+)	++	++	-	++
THCAS	PK29242.1:43-1680(+)	++	+	-	++
aPT1	PK28436.1:1-1254(+)	+	-	-	-
aPT4	PK15523.1:2-1213(+)	+	-	-	-

Table 5.8 Proteomic analysis of CsPTs and cannabinoid pathway enzymes in Purple Kush tissues. Three technical replicates from three clonal individuals were used for shotgun proteomics. ++: detected in all 3 samples, +: detected in 1 or 2 samples, -: not detected. Transcripts are translated from van Bakel et al., 2011.

5.4.2 Phylogenetic analysis of CsPTs

In a maximum likelihood phylogeny of protein sequences of cannabis UbiA aPTs and UbiA aPTs from other plants (**Figure 5.4**), the 12 CsPTs fell into two clades generally associated with specialized metabolism and two clades associated with primary metabolism. CsPT1 clustered with CsPT4 and CsPT7. These CsPTs formed a phylogenetic group with HlPT1L and HlPT2 from hop. CsPT3, which is involved in the biosynthesis of cannflavins (Rea et al., 2019), clusters with CsPT12 in the same clade. CsPT5 clusters with two isoliquiritigenin dimethylallyltransferases involved in the biosynthesis of prenylated flavonoids and chalcones in mulberry (Wang et al., 2014). Together, CsPT1, CsPT3, CsPT4, CsPT5, CsPT7 and CsPT12 form a clade known or putatively involved in specialized metabolism. CsPT2 and CsPT6 group with aPTs involved in prenylated isoflavonoid biosynthesis in both primary and specialized metabolism (Lin et al., 1999; Sasaki et al., 2011; Shen et al., 2012; Munakata et al., 2014; Yoneyama et al., 2016). This clade includes Arabidopsis (*Arabidopsis thaliana*) homogentisate phytyltransferase (AtHPT), which is involved in tocopherol biosynthesis. It also includes enzymes involved in specialized metabolism, such as the *Citrus limon* coumarin

geranyltransferase ClPT1, and the naringenin dimethylallyl prenyltransferase SfN8DT. The range of different biological functions associated with this clade make the likely functions of CsPT2 and CsPT6 difficult to predict. CsPT8 forms a group with genes involved in ubiquinone biosynthesis across multiple angiosperm lineages (Okada et al., 2004; Ohara et al., 2006; Ohara et al., 2009; Ohara et al., 2013). CsPT10 is most closely related to Arabidopsis homogentisate solanesyltransferase involved in plastoquinone biosynthesis. Lastly, CsPT9 and 11 do not cluster with any of the other aPTs and prenyltransferases, but are most closely related to CsPT8 and 10, and are likewise probably associated with primary metabolism.

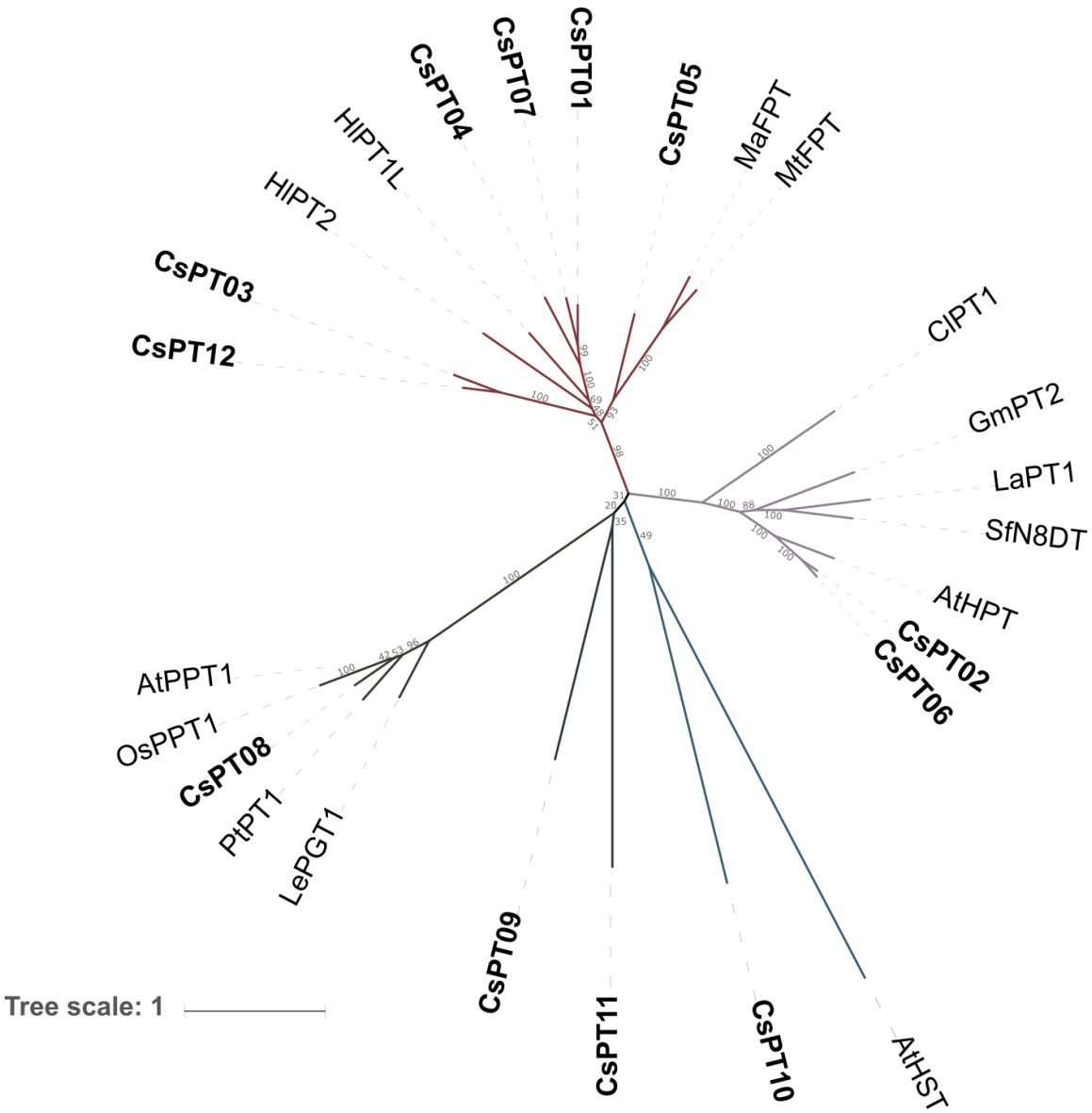


Figure 5.4 Maximum likelihood phylogeny of aPT amino acid sequences
Cannabis sativa (*Cs*) sequences are in bold. Grey numbers indicate bootstrap values (100 replicates, min BP = 29.9). *Hl*PT1 and *Hl*PT2 = *Humulus lupulus* acylphloroglucinol dimethylallyltransferases (AB543053.1 and A0A0B4ZTQ2.1)(Tsurumaru et al., 2010); *Ma*FPT = *Morus alba* isoliquiritigenin dimethylallyltransferase (KM262659.1)(Wang et al., 2014); *Ct*FPT = *Cudrania tricuspidata* isoliquiritigenin dimethylallyltransferase (KM262660.1)(Wang et al., 2014); *Le*PGT1 = *Lithospermum erythrorhizon* p-hydroxybenzoate polyprenyltransferase (MK585559.1)(Ohara et al., 2013); *Pt*PPT1 = *Populus trichocarpa* p-hydroxybenzoate polyprenyltransferase (PNT35788.1)(Tuskan et al., 2006); *Os*PPT1 = *Oryza sativa* p-hydroxybenzoate polyprenyltransferase (BAE96574.1)(Ohara et al., 2006); *At*PPT1 = *Arabidopsis thaliana p*-hydroxybenzoate polyprenyltransferase (NM_001203886.1)(Mayer et al., 1999); *At*HST = *Arabidopsis thaliana* homogentisate solanesyltransferase (AF324344.1)(Tanaka et al., 1999); *Cl*PT1 = *Citrus limon* coumarin geranyltransferase (A0A077K8G3.1)(Munakata et al., 2014); *Gm*PT2 = *Glycine max* glycinol dimethylallyltransferase (NP_001235990.1)(Akashi et al., 2009); *La*PT1 = *Lupinus albus* genistein dimethylallyltransferase (JN228254)(Shen et al., 2012); *Sf*N8DT = *Sophora flavescens* naringenin 8-dimethylallyltransferase (AB325579)(Suzuki et al., 2002); *At*HPT = *Arabidopsis thaliana* homogentisate phytyltransferase (NM_179653.4)(Suzuki et al., 2002).

5.4.3 Identification of CsCHIL1 and CsCHIL2

The transcriptome analysis also identified two cannabis chalcone isomerase-like proteins (*Cs*CHIL1 and *Cs*CHIL2) that are orthologous to HlCHIL1-L and HlCHIL2 from hop (Ban et al., 2018). CsCHIL1 shows high and specific expression in trichomes. In addition to trichomes, CsCHIL2 is expressed in the shoot, flower tissues and young leaves and is overall less strongly expressed than CsCHIL1. Neither CsCHIL1 nor CsCHIL2 is expressed in root tissues (**Figure 6.5**). CsCHIL1 shares 81% sequence identity at the nucleotide level and 68% identity on the amino acid level with HlCHIL1-L. CsCHIL2 shares 91% sequence identity on nucleotide level and 91% identity on the amino acid level with *Hl*CHIL2.

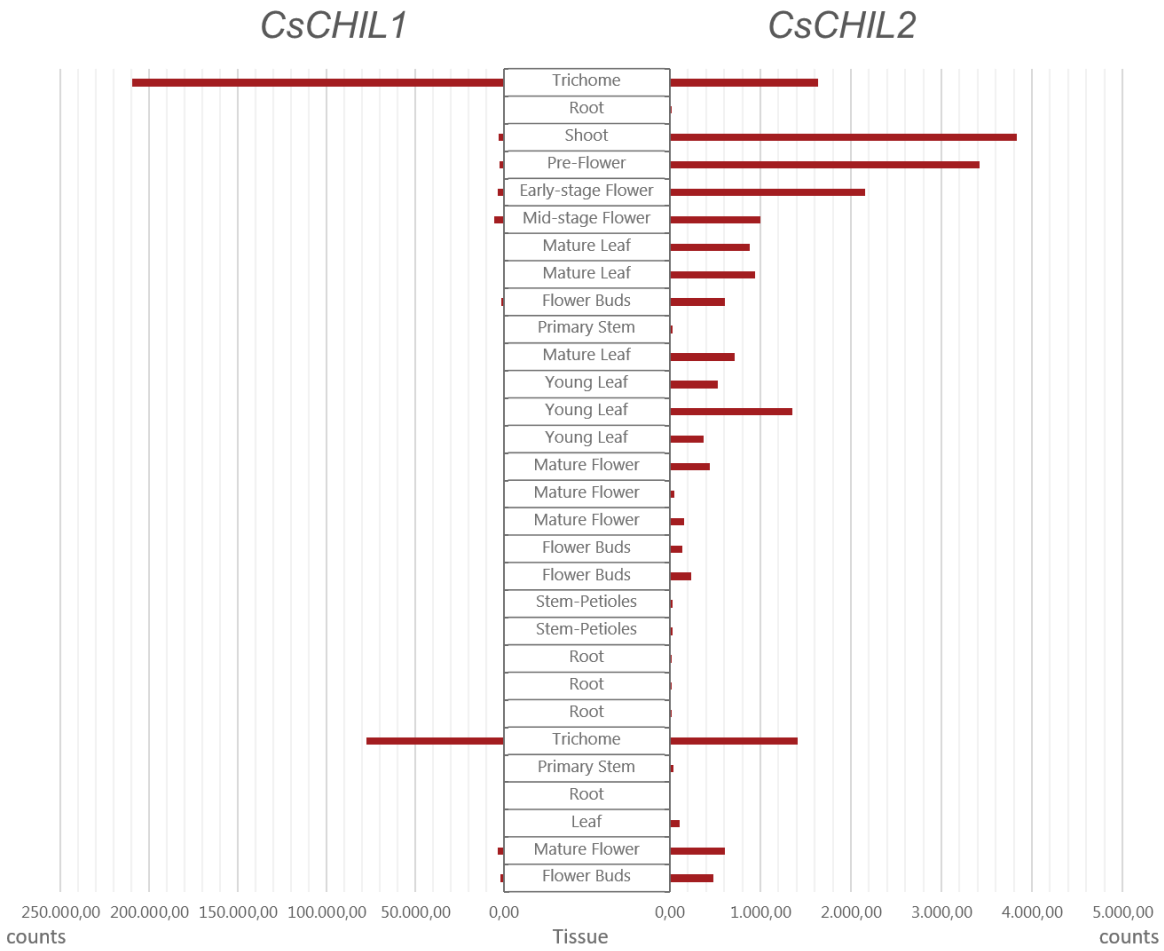


Figure 5.5 Expression of *CsCHIL1* and *CsCHIL2* across analyzed *Cannabis* tissues.

5.4.4 Intrinsic glucosylation of intermediates

N. benthamiana is known to metabolize hydrophobic non-endogenous compounds, glucosylating OA to OA-glucoside (OA-glc) and approximately 90% of CBGA into CBGA-glc and CBGA-di-glc (**Figure 5.6**). THCA-glycosides were not observed. Glucosidic constituents were determined present in the LC-MS-chromatograms (liquid chromatography-mass spectrometry) according (1) a shift towards shorter retention time (RT), (2) presence of the aglucon m/z I and (3) spectra with neutral loss transitions corresponding to the glucose moiety (difference of 162.0528). The OA analytical standard eluted at RT 9.6 min with m/z [225.11] and m/z [207.10] (corresponding to loss of water: 225.11->207.10), and a fragment with m/z [123.04]. To unequivocally document the formation of OA-glc, the compound was chemically synthesized and its structure verified by NMR. The OA-glc standard eluted at RT 6.2-6.3 min with glucosidic neutral loss transitions [387.15 -> 225.11] and [369.15 -> 207.10]. The mass spectrum of the authentic standard matched the spectrum of the compound produced in *N. benthamiana.* The CBGA analytical standard eluted at RT 18.3 min with m/z [361.23], but predominantly the loss-of-water ion with m/z [$M\text{-}H_2O+H^+$ = 343.22], and a fragment with m/z [219.10]. CBGA-glc was identified at RT 15.3 min with glucosidic neutral loss transitions [523.29 -> 361.23 and 505.28 -> 343.22]. CBGA-diglucoside was identified at RT 10.3 min with glucosidic neutral loss transitions [667.33 -> 505.28 -> 343.22]. The THCA analytical standard eluted at RT 19.9 min with m/z [359.2187], but predominantly the loss-of-water ion with m/z [341.20], and a fragment with m/z [219.10]. The formation of two novel constituents with m/z-transitions resembling the analytical standard of THCA [341.20->219.10] was observed but their origin remained unclear.

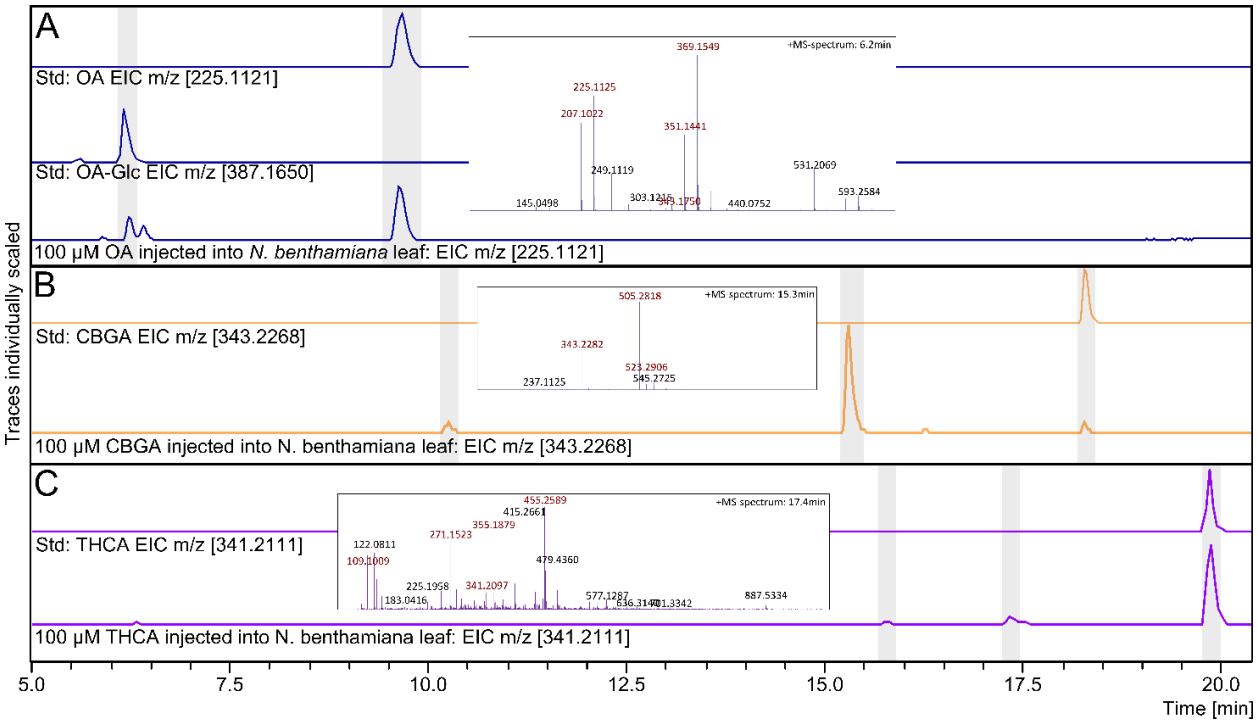


Figure 5.6 Enzyme and pathway assays in *Nicotiana benthamiana*.
A: Conversion of OA to OA-glucoside. First lane: extracted ion chromatogram (EIC) of OA analytical standard, m/z [225.1121] at RT=9.6 min; second lane: EIC of OA-glucoside analytical standard, m/z [387.1650] at RT=6.2-6.3 min; third lane: tobacco leaf was injected with 100 µM OA and incubated for 24 h, EIC of OA m/z [225.1121]. Inlet box: mass spectrum at 6.2 min. **B**: Conversion of CBGA to CBGA-glucoside. First lane: EIC of CBGA analytical standard, m/z [343.2268] at RT=18.3 min; second lane: tobacco leaf was injected with 100 µM CBGA and incubated for 24 h, EIC of CBGA m/z [343.2268]. Inlet box: mass spectrum at 15.3 min. **C**: Application of THCA to tobacco leaves. First lane: EIC of THCA analytical standard, m/z [341.2111] at RT=19.9 min. Second lane: tobacco leaf was injected with 100 µM THCA and incubated for 24 h, EIC of THCA m/z [341.2268]. Inlet box: MS-spectrum at 17.4 min.

5.4.5 Testing CBGAS activity of CsPT1-12

Agrobacterium tumefaciens mediated transient expression in *N. benthamiana* was used to individually test CsPT1-12 in combination with the genes encoding GPP synthase from *Abies grandis* (*AgGPPS;* AAN01134.1), DXS from *Plectranthus barbatus* (*PbDXS;* KP889115.1) and the *p19* silencing suppressor (Voinnet et al., 2003). In all experiments, *AgGPPS* and *PbDXS* were co-expressed with *CsPTs* with the intend to boost product formation based on availability of GPP, while *p19* generally increases accumulation of heterologously expressed proteins (Voinnet et al., 2003). OA and GPP were provided as substrates by leaf infiltration. Leaves

expressing CsPT4 accumulated CBGA, CBGA-glycoside and CBGA-diglycoside (**Figure 5.7**). None of the other CsPTs displayed CBGAS activity in *N. benthamiana*.

Expressing CsCHIL1 and CsCHIL2 in combination with CsPT1-12 did not result in CBGAS activity from candidate aPT*s* other than CsPT4. Co-expression of CsCHIL1 and CsCHIL2 with CsPT4 did not result in increased product formation. Experiments involving co-expression of CsPT1, CsPT4 and CsPT7 in any combination in *N. benthamiana* and in each of these combinations in the presence or absence of CsCHIL1/2 were also carried out. In none of these experiments was an increase of product formation achieved compared to individual expression of *CsPT4*. Aside from OA, we tested CsPT1-12's potential to prenylate phloroglucinol, resveratrol, naringenin and homogentisic acid using the prenyl-donors dimethylallyl pyrophosphate (C5), GPP (C10) and farnesyl pyrophosphate (C15) in *N. benthamiana.* No such activities were detected.

5.4.6 Engineering native cannabinoid biosynthesis in Nicotiana benthamiana

Following identification of CsPT4 as a functional CBGAS in *N. benthamiana*, approaches towards engineering of the cannabinoid pathway into *N. benthamiana* (**Figure 5.7**) were undertaken. Simultaneous expression of the genes encoding AAE1, OLS and OAC and concomitant infiltration of hexanoic acid resulted in efficient production of OA and OA-glc, as indicated by compounds formed possessing the same RT and m/z patterns as the reference compounds. Expression of THCAS in *N. benthamiana* with co-infiltration of CBGA led to efficient production of THCA. Combining expression of AAE1, OLS, OAC and *Cs*PT4 in *N. benthamiana* with infiltration of hexanoic acid resulted in formation OA and OA-glc, but not CBGA. Combining expression of CsPT4 and THCAS with administration of OA manifested

formation of low amounts of CBGA and CBGA-glc, but not THCA. Co-expression of AAE1, OLS, OAC, CsPT4 and THCAS lead to formation of OA and OA-glc, but neither CBGA nor THCA (**Figure 5.7**).

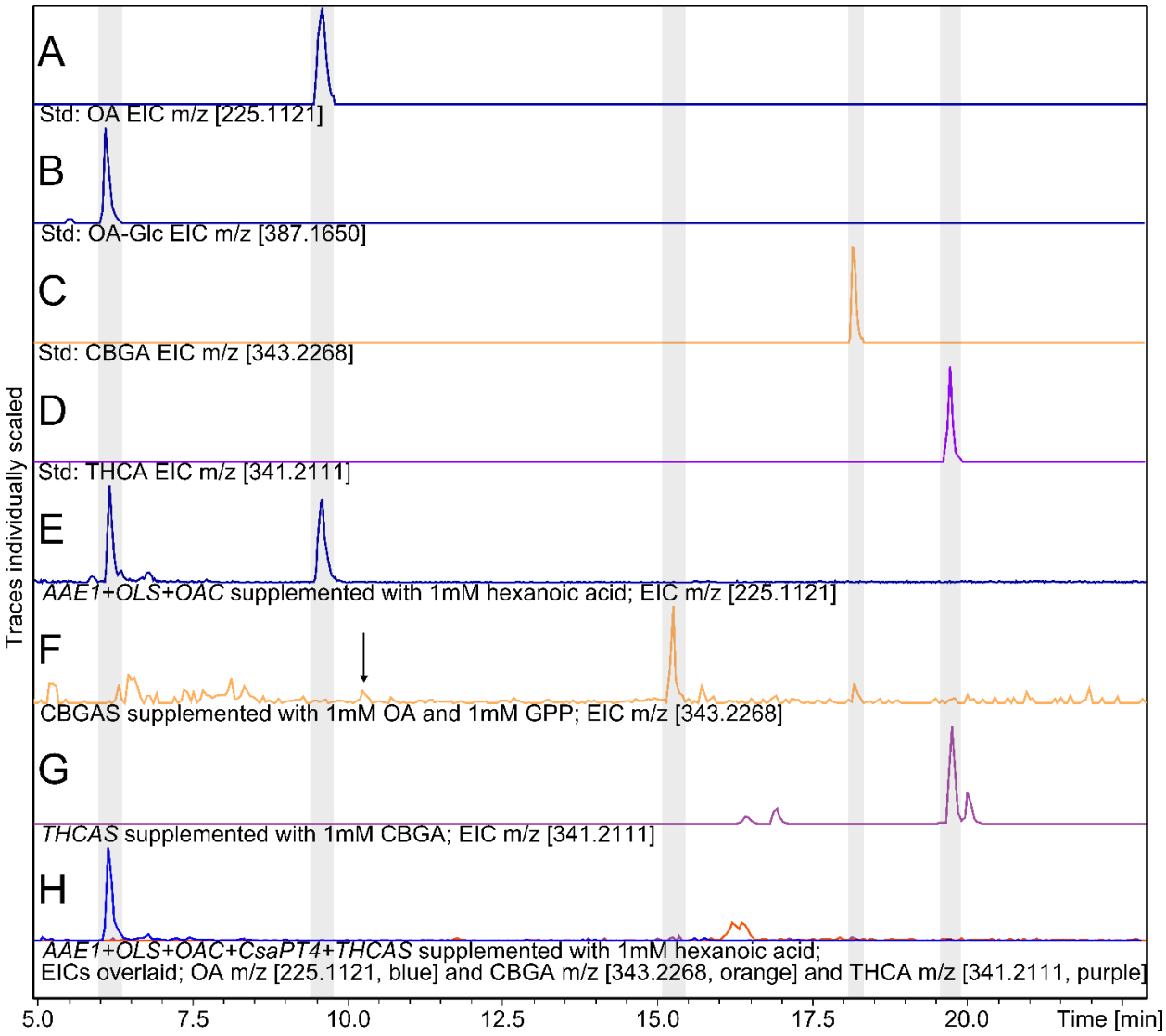


Figure 5.7 **Enzyme and pathway assays in *Nicotiana benthamiana***

A: OA standard (RT=9.6 min, EIC m/z [225.1121]); **B:** OA-glycoside standard (RT=6.2-6.3 min; EIC m/z [387.1650]); **C:** CBGA standard (RT=18.3 min, EIC m/z [343.2268]); **D:** THCA standard (RT=19.9 min, EIC m/z [341.2111]) **E**: tobacco leaf expressing *AAE1*, *OLS* and *OAC*, supplemented with 1 mM hexanoic acid; **F**: tobacco leaf expressing *CsPT04*, supplemented with 1 mM OA, and 0.1 mM GPP. Arrow indicates CBGA-di-glc peak at 10.3 min; **G**: tobacco leaf expressing *THCAS* supplemented with 1mM CBGA; **H**: tobacco leaf expressing *AAE1*, *OLS*, *OAC*, *CsPT4* and *THCAS* supplemented with 5 mM hexanoic acid.

5.4.7 Subcellular localization of CsPT1 and CsPT4 in transformed Nicotiana benthamiana

CsPT1 and CsPT4 both carry chloroplast targeting signal peptides according to the target prediction server (Emanuelsson et al., 2007). Their subcellular localization in *N. benthamiana* mesophyll cells was monitored by transient expression of C-terminal *eGFP*-fusions of full-length *CsPT1* and *CsPT4* (**Table 5.2**). Comparing the localization of the fluorescence from eGFP with the auto-fluorescence of chlorophyll shows co-localization in the chloroplast (**Figure 5.8**).

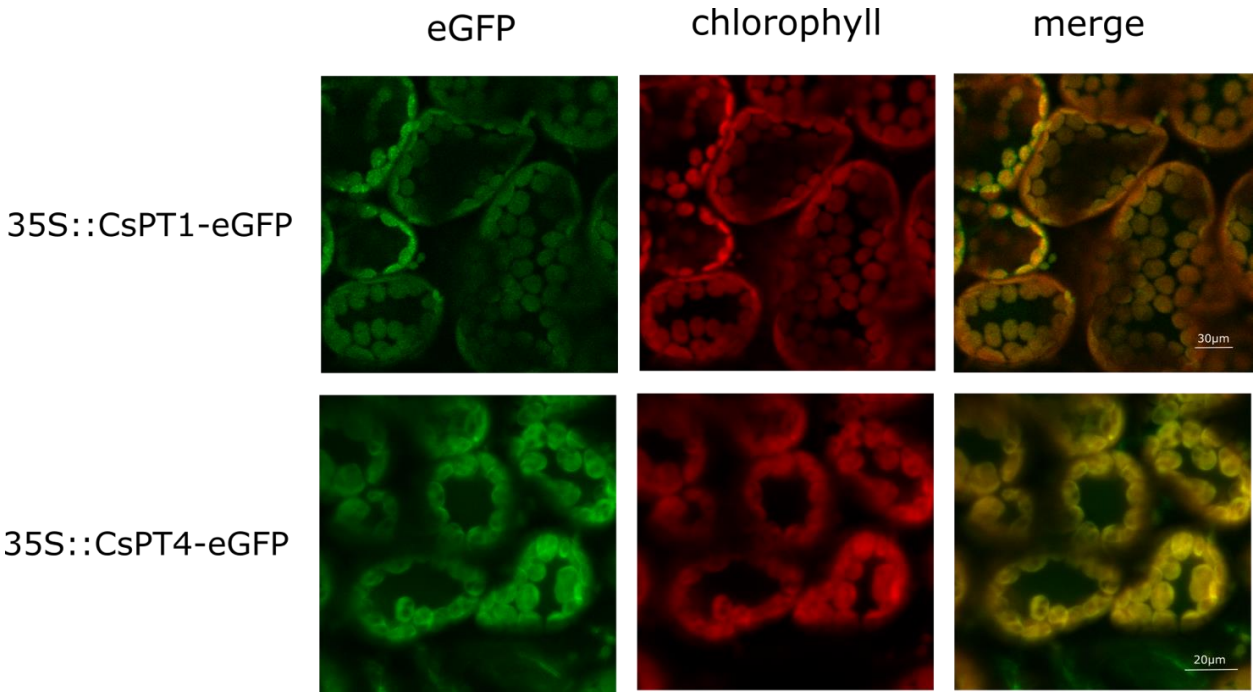


Figure 5.8 Subcellular localization of *CaPT1-GFP* and *CsPT4-GFP* fusion proteins in *N. benthamiana* mesophyll cells
Top row: *CsPT1-GFP*. Bottom row: *CsPT4-GFP*. First column: *eGFP*-channel detected between 450-550 nm. Second column: chlorophyll auto-fluorescence detected between 610-710 nm. Third column: Merged image of *eGFP* and chlorophyll fluorescence emission. Image analysis was performed using Volocity software (Quorum technologies).

5.4.8 Engineering of cannabinoid biosynthesis in *Saccharomyces cerevisiae*

CsPT1 was included in *S. cerevisiae* studies to test for possible CBGAS activity. The hexanoic acid producing *S. cerevisiae* strain y*Cs*0.1 and the control strain y*Cs* 0.4 were used as background strains to incorporate the cannabinoid pathway with either CsPT1 and CsPT4 (both

with the chloroplast targeting signals truncated: CsdPT1 and CsdPT4, respectively) generating the strains y*Cs*1 and y*Cs*4 (**Table 5.3**). The strains were supplemented with glucose, glucose+hexanoic acid (0.3mM) or glucose+OA (0.1mM) as substrates. Strain y*Cs*1 (expressing *CsdPT1*) produced OA from sugar alone or from hexanoic acid, but no formation of CBGA or THCA was observed (**Figure 5.9**). Strain y*Cs*2 (expressing *CsdPT4*) produced OA from sugar or hexanoic acid, detectable amounts of CBGA from sugar and 1.0 mg CBGA from 0.1 mM OA, confirming CsPT4 functionality as a CBGAS. Strain y*Cs*3 produced OA from hexanoic acid, but not from sugar, confirming the control strains inability to generate sufficient intrinsic hexanoic acid. Strain y*Cs*4 produced OA from hexanoic acid and CBGA from OA. None of the strains produced detectable amounts THCA.

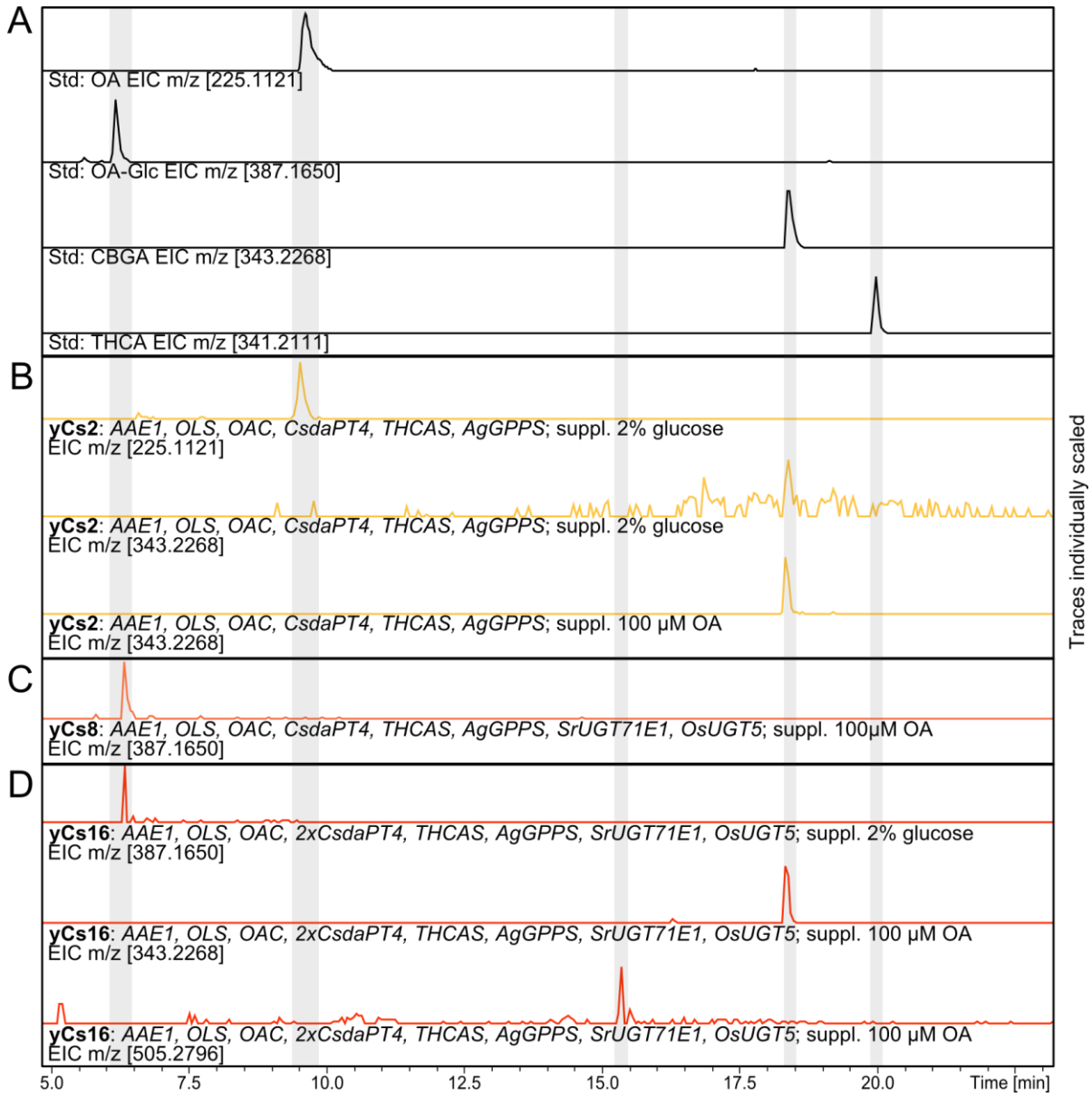


Figure 5.9 Engineering of the cannabinoid pathway in *Saccharomyces cerevisiae* strains
A: analytical standards of OA, OA-glycoside, CBGA and THCA); **B**: Biosynthetic capacity of y*Cs*2 strain. **C**: Biosynthetic capacity of the y*Cs*8 strainexpressing two UGT encoding genes. **D**: Biosynthetic capacity of the y*Cs*16 expressing two copies of *CsdaPT4* and two UGT encoding genes.

5.4.9 Glucosylation of OA and CBGA in *Saccharomyces cerevisiae*

Administration of OA or CBGA to *N. benthamiana* leaves led to the production of glycosides of the respective compounds. We attempted to target production of these glycosides by introducing specific plant UGTs into the yeast genome. We tested five different plant UGTs, of which *Stevia*

rebaudiana UGT71E1 (*Sr*UGT71E1, AY345976.1) and *Oryza sativa* UGT5 (OsUGT5, XP_015622068.1) glucosylated CBGA (**Figure 5.11**) and Δ^9-THCA *in vitro* (**Figure 5.12** and **Figure 5.13**). The genes encoding these UGTs were co-expressed with those expressing the cannabinoid pathway in *S. cerevisiae* strains, generating the strain y*Cs*8 (THCA pathway+*SrUGT71E1+OsUGT5*). In parallel, y*Cs*16 with an additional second copy of *CsPT4* codon optimized for expression in *S. cerevisiae* (THCA pathway + *SrUGT71E1 + OsUGT5 + CsScdPT4*) was engineered. The y*Cs*8 strain was demonstrated to form OA-glucoside and CBGA from OA. y*Cs*16 led to the formation of OA-glucoside from incubation with just glucose, and CBGA-glucoside from incubation with OA and glucose. Neither THCA nor THCA-glycosides were detected in the y*Cs*8 and y*Cs*16 strains. Interestingly, OA-glc formation was observed in the yeast strains, though *SrUGT71E1* and *OsUGT5* showed only activity on CBGA and THCA *in vitro*.

5.5 Discussion

5.5.1 Functions of CsPTs

Cannabis produces a variety of terpenophenolics known as cannabinoids, including CBGA and Δ^9-THCA. Here, we demonstrate that CsPT4 is active as CBGAS in *N. benthamiana* and confirm recent work describing the function of CsPT4 in *S. cerevisiae*. Furthermore, we engineer cannabinoid and cannabinoid glucoside biosynthesis in *S. cerevisiae* using CsPT4 together with the other cannabinoid pathway genes and UGTs.

We tested if activity of CsPTs are enhanced in the presence of other CsPTs. In hop, HlPT1-L and HlPT2 require formation of a heterodimeric complex to perform three sequential prenylations in the biosynthesis of lupulone (Li et al., 2015). The aPTs HcPT8px and HcPTpat

from *Hypericum calycinum* show coordinated expression upon elicitation in *H. calycinum* cell cultures, and their co-expression in *S. cerevisiae* microsomes strongly enhanced production of the prenylated xanthone patulone (Nagia et al., 2019). In contrast, CsPT1, CsPT4 and CsPT7, when co-expressed in *N. benthamiana* in various combinations did not increase product level in comparison to CsPT4 alone.

Gene families of enzymes with similar or identical functions are a hallmark of plant specialized metabolism (Lee Chae, Taehyong Kim, Ricardo Nilo-Poyanco, 2014). Among the CsPTs, CsPT1 and CsPT4 are highly and selectively expressed in cannabis glandular trichomes, and both enzymes have CBGAS activity. Testing whether CsPT1 and CsPT4 are strictly redundant or support cannabinoid biosynthesis under different conditions *in planta* will require future work by selectively down-regulating these two aPTs in cannabis.

CsPT1 and CsPT4 localize to chloroplasts when expressed in *N. benthamiana* leaves, demonstrating the functionality of their plastidial targeting signals. It remains to be tested if the CsPTs reside in the inner or the outer membrane of the chloroplast envelope and whether the Mg^{2+}-binding aspartate-rich motifs of the active site are orienting to the inner or the outer face. Proper location and orientation in the envelope membrane system is likely to affect access of aPTs to substrates. Auxiliary proteins may also differentially influence catalytic activity of CsPT1 and CsPT4. This may offer a potential explanation for different reports of functional activity observed for CsPT1.

5.5.2 Chalcone isomerase like proteins in Cannabis

We identified CsCHIL1 and CsCHIL2, which represent homologues of non-catalytic HlCHIL involved in the biosynthesis of prenylated aromatics in hops (Ban et al., 2018). CHILs are

members of a recently defined protein class of polyketide binding proteins (PBPs) (Ban et al., 2018). CsCHIL1 was highly and selectively expressed in cannabis trichomes, while CsCHIL2 transcripts were present in different tissues, but not in roots. A partial sequence of *CsCHIL1* was previously reported by Gagne et al. (2012) (Gagne et al., 2012). The full-length CsCHIL1 described here contains additional 24 amino acids at the N-terminus. We co-expressed CsCHIL1 and CsCHIL2 with all 12 individual CsPTs but did not observe effects on the activities of those prenyltransferases.

5.5.3 Cannabinoid engineering in *N. benthamiana*

Identification of the functional CBGAS CsPT4 enabled metabolic engineering of *N. benthamiana* for cannabinoid production. *N. benthamiana* is a well-established host system for metabolic engineering and production of a diverse array of complex specialized plant metabolites, including cyanogenic glycosides (Hansen et al., 2018; Thodberg et al. 2018; Knoch et al. 2016), glucosinolates (Geu-Flores et al., 2009; Crocoll et al., 2016), alkaloids (Miettinen et al., 2014), flavonoids (Irmisch et al., 2018; Irmisch et al., 2019b; Irmisch et al., 2019a), ketides (Andersen-Ranberg et al., 2017), MEV-derived terpenoids (Reed et al., 2017) and MEP-derived terpenoids (Andersen-Ranberg et al., 2016; Heskes et al., 2018). A common challenge in *N. benthamiana* engineering is intrinsic glycosylation of pathway intermediates and end products of heterologous biosynthetic pathways (Bártíková et al., 2015; Tiwari et al., 2016). We assessed the potential conversion of cannabinoid pathway intermediates in *N. benthamiana* by leaf infiltration of OA, CBGA and THCA and observed formation of OA- and CBGA-glycosides. These compounds have, to our knowledge, not been described before and represent novel-to-nature cannabinoids.

Attempts to produce cannabinoids in *N. benthamiana* from hexanoic acid resulted in formation of OA and OA-glycoside but not CBGA or THCA, despite expression of *CsPT4* and *THCAS*, which are functional in *N. benthamiana* when supplied with their respective substrates. Lack of CBGA and THCA formation may be due to glycosylation of OA impeding biosynthesis of downstream cannabinoid products. Production of cannabinoids in *N. benthamiana* may require down-regulation of the UGTs that glycosylate OA and CBGA. Another key challenge for engineering the cannabinoid biosynthetic pathway in tobacco is the toxicity of pathway intermediates. CBGA and THCA induce cell-death via induction of apoptosis that lead to necrosis of leaf tissue (Sirikantaramas et al., 2005). For cannabinoid production in *N. benthamiana,* efficient pathway channeling and targeted glycosylation of final products may be essential to avoid auto-toxicity and provide a metabolic sink that could alleviate feedback inhibition. If desired, glycosylated products can be de-glycosylated during downstream-processing, affording the pure aglycone (van Herpen et al., 2010).

5.5.4 Cannabinoid production in *Saccharomyces cerevisiae*

S. cerevisiae was chosen as an alternative host to *N. benthamiana* to confirm CsPT4 CBGAS activity and for the production of cannabinoids and cannabinoid glucosides from sugar or hexanoic acid. Following the work by Gajewski et al. (2017), we first reprogrammed fatty acid biosynthesis to optimize production of the cannabinoid pathway precursor hexanoic acid (Gajewski et al., 2017a; Gajewski et al., 2017b), as confirmed by production of OA. Carboxylic acids such as hexanoic, octanoic and decanoic acid have been reported to cause cell membrane leakage that leads to the inhibition of cell growth and ultimately, cell death (Legras et al., 2010; Borrull et al., 2015). Weak carboxylic acids such as hexanoic acid partially dissociate in aqueous

systems, establishing an equilibrium between undissociated, uncharged molecules and their anionic forms, according to their pK_a and to the pH of the solution. When the pH is below the pK_a of the acid, the protonated undissociated form predominates and is able to cross biological membranes by passive diffusion, cause lowering of the intracellular pH and accumulation of toxic levels of anions (Casal et al., 2008) To prevent the acidification of the cytosol, membrane transporters such as Pdr12p export these carboxylic acids to the extracellular environment (Cabral et al., 2001; Legras et al., 2010; Liu et al., 2013; Borrull et al., 2015). To minimize the active export of hexanoic acid, *PDR12* may need to be deleted in future work to enhance channeling of hexanoic acid towards cannabinoid production. Next, we incorporated cannabinoid pathway genes and GPPS into the *S. cerevisiae* genome, each gene flanked by unique yeast promoters and terminators. This led to production of up to 1.0 mg L^{-1} OA and detectable amounts of CBGA in *S. cerevisiae* growth media supplemented with 2% glucose, but not production of Δ^9-THCA. Supplementation of 0.1 mM OA led to the production of 5.3 µg L^{-1} CBGA. OA levels were measured either in whole cell cultures, i.e. with combined pellet and supernatant fractions, or separately in pellets or supernatant. In growth assays without OA supplementation, approximately 50% of OA was exported to the medium. Furthermore, supplementation of OA to the medium increased CBGA production, indicating either influx of substrate or lower efflux of intrinsically produced OA due to a weaker diffusion gradient. Both observations suggest that OA was not limiting. One possible explanation may be incorrect membrane orientation of *CsPT4* in the engineered *S. cerevisiae* affecting access of substrate at the catalytic site.

Stable integration of the pathway genes as single copies allows greater control over gene dosage effects, enabling fine-tuning of gene expression on the genomic level (Da Silva and

Srikrishnan, 2012). A typical unwanted side product in cannabinoid biosynthesis is olivetol, which in the absence of OAC is a product of OLS (Taura et al., 2009; Gagne et al., 2012). Olivetol cannot serve as a substrate for CBGAS (Fellermeier and Zenk, 1998). Accumulation of olivetol despite the presence of OAC caused by an imbalance between OAC and OLS may be compensated by introducing additional copies of *OAC* (Luo et al., 2019). Likewise, the lack of Δ^9-THCA production can probably be alleviated by introducing more copies of *CsPT4* and/or *THCAS*. However, CBGA and THCA (and the cannabinoid synthase side product H_2O_2) are toxic also to *S. cerevisiae* and require adjustments to maintain high productivity.

5.5.5 Glucosylation of cannabinoid pathway intermediates

The OA- and CBGA-glucosides described here are novel-to-nature compounds expanding the chemical space of known cannabinoids. Enzymatic glycosylation provides advantages in natural products biotechnology due to enzyme stereospecificity, making chemical methods including blocking and deblocking unnecessary (Tiwari et al., 2016). Formation of hydrophilic conjugates can alleviate feedback inhibition and thereby increase metabolic pathway flux by altering the product. Subsequently, O-glycosylated compounds can easily be recovered by the use of commercially available glucosidases (Singh et al., 2016). We found that SrUGT71E1 and OsUGT5 glycosylate CBGA and THCA *in vitro* and OA and CBGA in *S. cerevisiae*. The addition of UGTs to the pathway-expressing strains was intended reduce toxicity of CBGA and THCA and to provide a metabolic sink to increase pathway flux. However, expression of the *UGTs* did not result in THCA production. Production of CBGA may have been insufficient, indicating a problem on the push-side and not on the pull-side of the pathway. It should be noted that CBGA-glc was only produced in the strain yCs16, which carries a second copy of *CsPT4*,

substantiating the importance of adjusting the gene expression of pathway components that create a bottleneck.

5.5.6 Conclusions

Cannabinoid production in *N. benthamiana* suffers from unwanted modification of pathway intermediates and toxicity effects of pathway products. Establishment as a competitive production host requires extensive further engineering efforts addressing these issues. In addition to functional expression of the cannabinoid biosynthetic genes, the pathway flux must maximized using transcription factors, bottleneck alleviation and optimal supply of co-factors. The fungal host *Saccharomyces cerevisiae* emerged as a production platform for cannabinoids and cannabinoid derivatives with altered side-chains (Luo et al., 2019) or glucose conjugates.

The role of CsPT1 in *Cannabis* remains elusive. Together with CsPT4, it shares high and selective expression in *Cannabis* trichomes and both enzymes carry a plastid localization signal. On the genome, both genes' gene loci are in close proximity and show tight linkage to a known marker for total cannabinoid content (ANU*CS*501)(Laverty et al., 2019). It could be reasoned that CsPT1 and 4 are functionally redundant in *Cannabis sativa*. To learn their biological roles, we require the development of transformation protocols and knock-out studies for *Cannabis*.

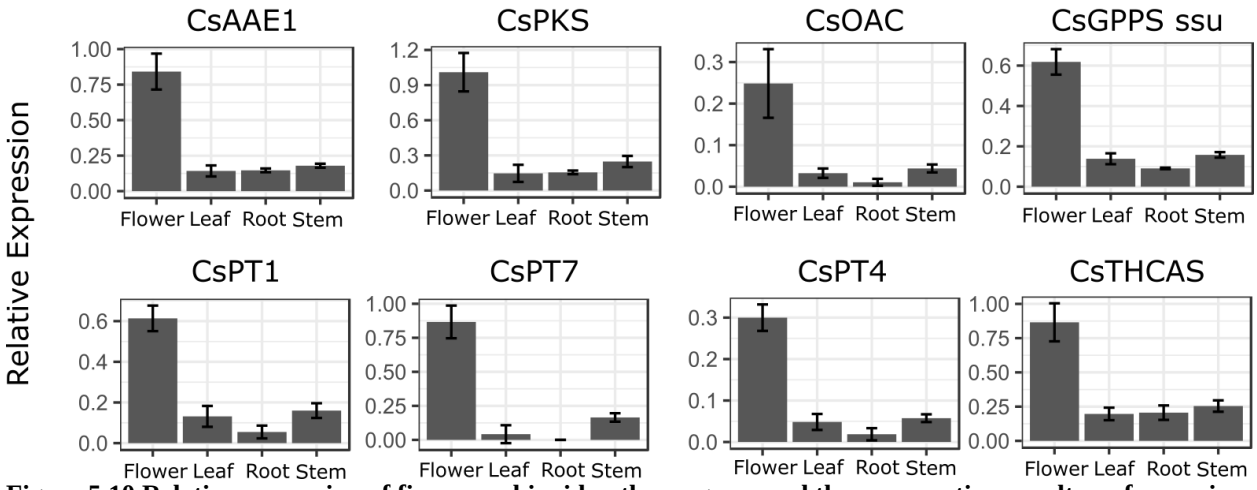


Figure 5.10 Relative expression of five cannabinoid pathway genes and three aromatic prenyltransferases in four cannabis tissues from the cultivar Purple Kush

Whole flower, mature fan leaf, root, and stem were used. CsAAE1: acyl activating enzyme; CsPKS: polyketide synthase; CsOAC: olivetolic acid cyclase; CsGPPS ssu: geranyl diphosphate synthase small subunit; CsTHCAS: tetrahydrocannabinolic acid synthase. Error bars represent standard deviation of three biological and two technical replicates.

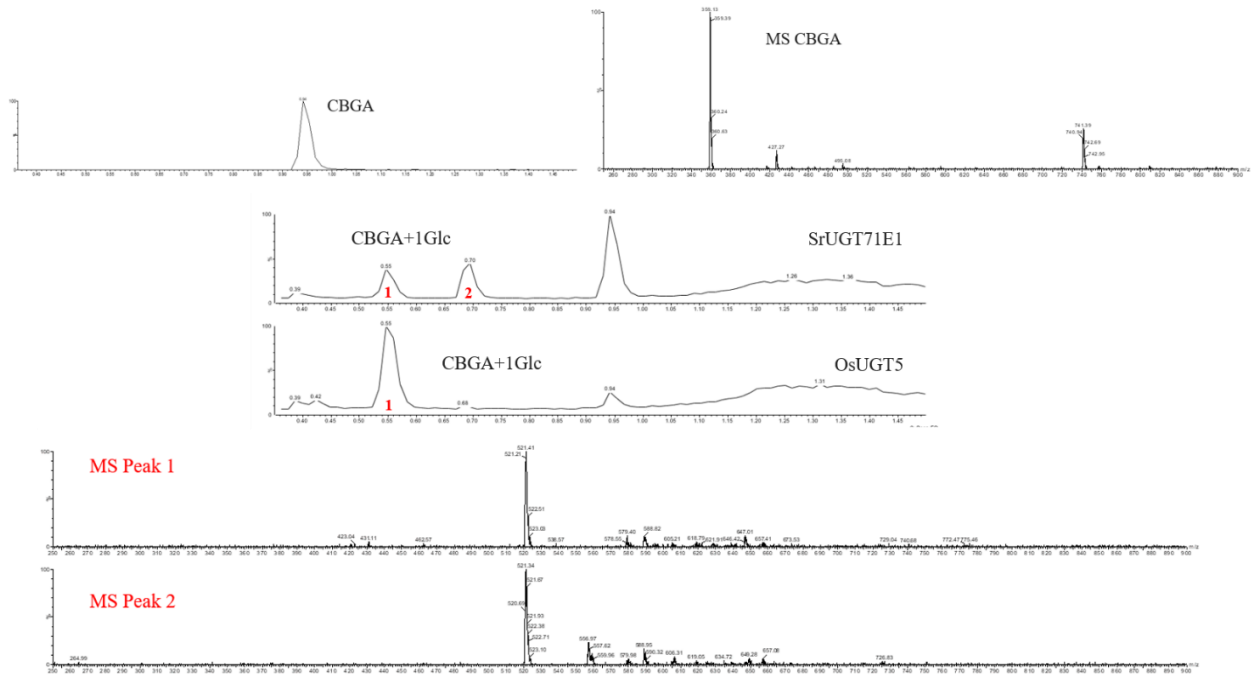


Figure 5.11 *in vitro* glucosylation of CBGA by *Sr*UGT71E1

The LC-triple-quad MS analysis for CBGA+1Glucose. Above, the chromatogram showing CBGA+1Glc detection. Below, the corresponding MS spectra of two CBGA+1Glc.

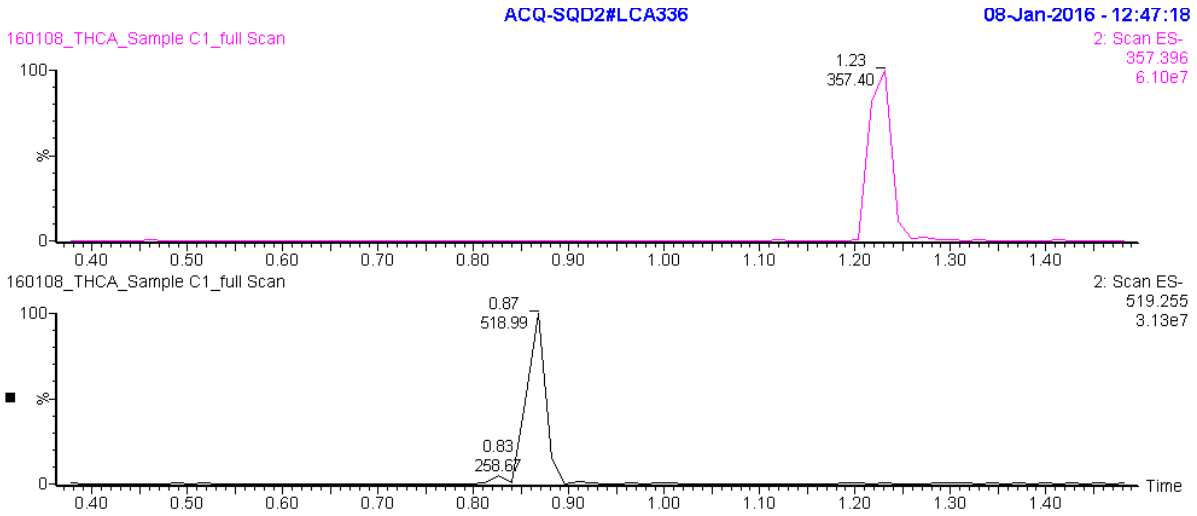


Figure 5.12 *in vitro* glucosylation of THCA by *Sr*UGT71E1

The LC-tQMS analysis for Δ^9-THCA+1Glc. Above, the chromatogram showing Δ^9-THCA+1Glc detection. Below, the corresponding MS spectrum of Δ^9-THCA+1Glc and of Δ^9-THCA.

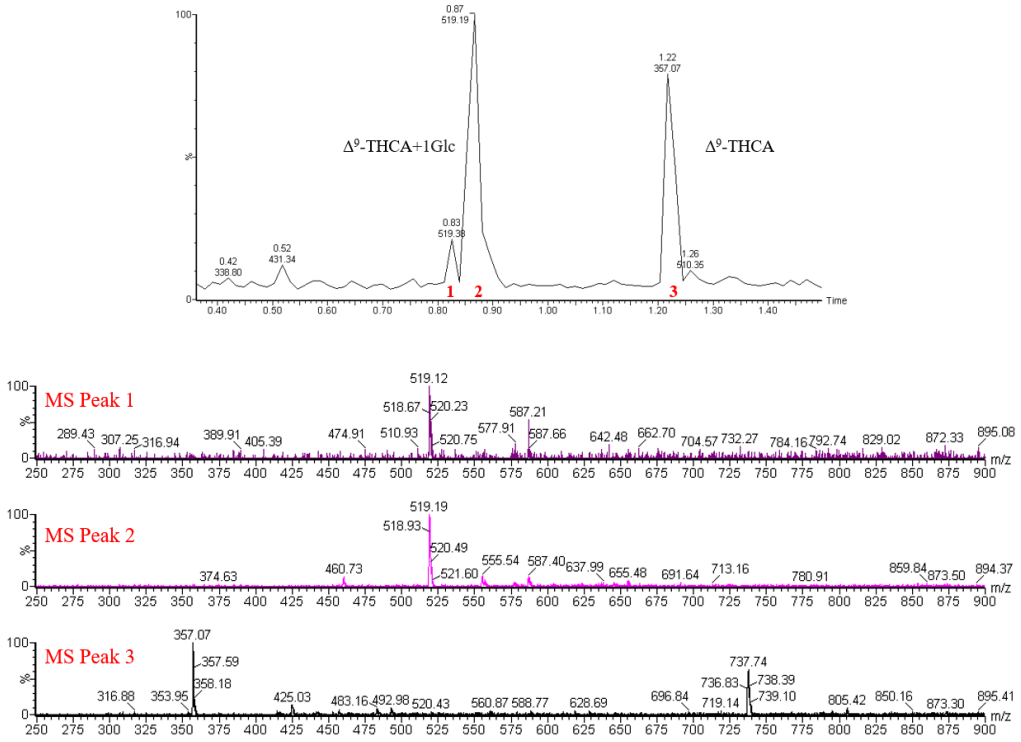


Figure 5.13 *in vitro* glucosylation of THCA by OsUGT5

A: Chromatogram and spectra obtained in positive ionization mode. B: Chromatogram and spectra obtained in negative ionization mode.

Chapter 6: Conclusion

6.1 Terpene and isoprenoid biosynthesis in *Cannabis sativa*

Cannabis sativa produces a variety of isoprenoid compounds in its glandular trichomes. The most abundant of these are cannabinoids and terpenes, especially mono- and sesquiterpenes. These compounds form a resin that is most abundant in trichomes on female flowers. In this thesis, I explored the biosynthesis of these compounds in glandular trichomes, and how their abundance changes based on physiological, developmental, and genetic variables.

Terpenes are made by terpene synthases (TPS). The cannabis genome encodes at least 35 TPS genes associated with specialized metabolism (Gunnewich et al., 2007; Booth et al., 2017; Allen et al., 2019; Livingston et al., 2019; Zager et al., 2019). Some of these are expressed in the trichomes of all cultivars sequenced to date, whereas others are expressed only in a subset of cultivars. Collectively, these TPS enzymes are responsible for the diversity of terpenes found within and between cultivars. The activities of the TPS responsible for most terpenes found in cannabis resin are elucidated in Chapters 2, and 3, as well as appendix A.

Cannabis displays a plethora of terpene metabolite profiles, which vary between cultivars. Compounds range from apparently universal in cannabis (e.g. myrcene, β-caryophyllene) to highly discriminant, (e.g. terpinolene, bisabolol). Chapters 2 and 3 explored the genetic basis of that variability, focusing on variation between cultivars. When all these factors are controlled, terpene profiles may provide a toolkit to verify the identity of cannabis cultivars. This toolkit will include minor terpenes that are often excluded from commercial terpene analyses (Booth and Bohlmann, 2019).

The diversity of terpenes across cannabis cultivars represents a combination of variation in enzyme activities between cultivars, enzyme expression levels, and the presence of a given

gene in the genome of any cultivar. In Chapter 4, I explored the basis of that diversity by examining the mechanisms of two ocimene synthases – CsTPS6, an (*E*)-β-ocimene synthase from 'Finola', and CsTPS13, a (*Z*)-β-ocimene synthase from Purple Kush.

Collectively, the TPS and aromatic prenyltransferase (aPT) enzymes discussed in this thesis are responsible for the direct formation of the dominant components of cannabis resin. TPS share a similar mechanism with aPTs (**Figure 1.3)**. At least two aPTs have been implicated in cannabinoid biosynthesis via the formation of cannabigerolic acid (CBGA). CBGA is formed by the prenylation of olivetolic acid (OA) by geranyl diphosphate (GPP). The isoprenoid portion of the molecule is then cyclized to form downstream cannabinoids. CsPT1 and CsPT4 have both been shown possess CBGA synthase (CBGAS) activity, and both are highly expressed in glandular trichomes. The CBGAS activity of CsPT4 is covered in Chapter 6.

This thesis describes many of the enzymes directly involved in producing isoprenoid specialized metabolites in cannabis resin. By studying transcript expression and metabolite abundance of cannabinoids and especially terpenes in different cannabis cultivars and tissues, I was able to demonstrate the individual and overlapping contributions of TPS and aPTs to the overall specialized metabolite profiles of cannabis resin.

6.2 Future directions

In this section, I highlight conclusions related to each major goal of this thesis. For each topic, I propose future work that builds on the findings of my thesis.

6.2.1 Terpene synthases in *Cannabis sativa*

One goal of my thesis was to characterize the enzymes responsible for isoprenoid biosynthesis in cannabis resin, with a focus on TPS enzymes.

In Chapters 2 and 3 and Appendix A, I aimed to characterize the TPS that produce the terpenes found in cannabis flowers, and to explain the genetic and biochemical basis of terpene variation between cultivars. In Chapter 2, I focused on terpene biosynthesis in 'Finola'. In Chapter 3, I expanded the search into Purple Kush and five additional cultivars for which no previous data had been collected: Lemon Skunk, Chocolope, Afghan Kush, CBD Skunk Haze, and Blue Cheese. The cultivars were selected to encompass the widest diversity of terpene metabolites available.

The results presented in this thesis, as well as several studies published while this thesis was in progress (Allen et al., 2019; Braich et al., 2019; Zager et al., 2019), combine to describe a total of 29 characterized TPS and a further 35 that are named but not yet characterized. Most of these are involved in specialized metabolism, and mostly in trichomes. As I stated in Chapter 3, the diversity of the cannabis genome makes a true estimate of the size of the cannabis TPS family difficult to obtain. As more cultivars are sequenced, new enzymes will be revealed. Progress over the last several years, including the data presented in this thesis, has led to the characterization of most of the enzymes responsible for most of the major terpene metabolites in common North American cannabis cultivars.

In Chapters 2 and 3, I showed that the expression of TPS transcripts in the trichomes does not predict the abundance of terpene metabolites. A similar lack of correlation has been shown in the literature for cannabis (Zager et al., 2019; Livingston et al., 2020). This highlights an unexplored facet of terpene biosynthesis – the stability and turnover of TPS transcript and

proteins. To date, no data has been published examining the regulation of transcript or protein stability for any cannabis enzymes including TPS. A preliminary cannabis proteome revealed that TPS may be acetylated (Jenkins and Orsburn, 2020). Future work may explore the role of modifications such as farnesylation, ubiquitination, glycosylation, and other modes of regulating protein stability and turnover. This will hopefully help illuminate the route from gene expression to metabolite accumulation for all aspects of cannabis resin biosynthesis.

Improved genome sequencing techniques will speed the discovery of further TPS genes. The trajectory is evident in this thesis. When this work started, the genome assembly that was available was too fragmented to resolve any complete TPS genes (van Bakel et al., 2011). A later, unpublished version of the PK genome using PacBio technology (**Figure 2.7**) allowed for an early estimate of the TPS family size in PK. The most recent version of the PK genome (Laverty et al., 2019) was able to fully resolve all known TPS enzymes in PK (**Figure 3.3**). The improvement from fragmented draft genome to chromosome-scale assembly provided new insights into the genomics of terpene biosynthesis in cannabis, including gene copy number, identification of gene arrays, and annotation of full- and pseudogenes associated with resin biosynthesis. Continued sequencing of more diverse cultivars would rapidly uncover any novel TPS enzymes, and may also help clarify the roles of heterozygosity and gene copy number in determining terpene metabolite profiles.

In order to fine tune the terpene profile of a cultivar, the impact of each TPS gene on the terpene phenotype should be determined. One way to address this would be to use genomic markers associated with terpene traits. A QTL associated with terpene profile located near a known TPS gene could indicate a crucial role for that enzyme.

6.2.2 Variation of terpene biosynthesis gene expression in flower trichomes and roots

Genes for terpene, as well as cannabinoid biosynthesis, were highly expressed in trichomes in all five cultivars investigated here (**Figure 3.6b**), in agreement with previous reports on the cannabis MEP and MEV pathways and selected *CsTPS* genes preferentially expressed in trichomes (Booth et al., 2017; Braich et al., 2019). Many are also highly expressed in root tissue (Allen et al., 2019). TPS expressed in roots include members of the TPS-a2 clade (**Figure 3.7**), TPS12, two enzymes not identified in the transcriptomic resources developed here (CsTPS11-Like and CsTPS49), and in some cases TPS9 (Allen et al., 2019; Braich et al., 2019). We have been unable to clone CsTPS12 from floral or leaf tissue of any cultivar. Root-specific transcriptome analysis and enzyme cloning may lead to the characterization of CsTPS12 and other previously unidentified TPS in cannabis.

While cannabis roots are known to contain specialized metabolites of interest, including cannflavins (M.L. Banett, 1963; Ryz et al., 2017; Rea et al., 2019), their volatile terpene profiles have not been explored. Given the high, and sometimes specific, expression of TPS enzymes in cannabis roots, terpene metabolite profiling of cannabis roots under different conditions and developmental stages should be undertaken at the first opportunity.

6.2.3 Terpene catalysis

The third major goal of my thesis was to investigate terpene catalysis in cannabis. In Chapter 4, I addressed this goal by focusing on two similar TPS. Both CsTPS6 and CsTPS13 are β-ocimene synthases, but CsTPS6 produces mostly (*E*)-β-ocimene while CsTPS13 is a (*Z*)-β-ocimene synthase. In this chapter, I originally intended to use protein modeling and site-directed mutagenesis to determine the mechanism by which the stereochemistry of these enzymes'

products is determined. While I was able to identify some important residues, the research was cut short due to data loss and software license issues after my computer was stolen in 2016. I identified 11 residues or neighbouring pairs of residues in CsTPS6 to target for mutagenesis. Using site-directed mutagenesis, I converted the candidate residues in CsTPS6 to their corresponding residues in CsTPS13. Most had little to no effect. Two residue pairs lining the active site cavity, when mutated to the corresponding residues in CsTPS13, caused some conversion of the product to (*Z*)-β-ocimene. Both also came with a loss of enzyme efficiency, although K_ms were not measured.

This work is relevant in light of the broader question of predicting TPS products in cannabis. Both CsTPS6 and CsTPS13 are β-ocimene synthases. They are also closely related to another linear monoTPS, the geraniol synthase CsTPS17 (**Figure 3.9 and Figure 3.10**). This suggests that CsTPS36, which I was not able to express, may also be a linear monoTPS.

Sequence comparisons between these linear monoTPS and related TPS with cyclic products like CsTPS2 could help identify residues involved in the emergence of cyclization in cannabis monoTPS. This approach has been used successfully in species including *Thymus vulgaris*, *Artemisia annua*, and *Mentha spicata* (Krause et al., 2013; Salmon et al., 2015; Srividya et al., 2020) Conceivably, a similar approach to the one I used in Chapter 5 could be used to convert CsTPS2 to a linear TPS, or to develop a CsTPS6 mutant capable of producing cyclic terpenes.

I addressed the subject of sesquiterpene catalysis more broadly in Chapter 3, where I demonstrated that highly related sesqui-TPS form similar products via similar catalytic routes (**Figure 3.10**). These enzymes descended from a recent common ancestor, and likely retain some of the same mechanisms of product formation. While this method cannot perfectly predict

products, or even product classes, it was able to provide a reasonable estimate of likely product classes for many cannabis sesqui-TPS. Expanding this approach to more sesqui-TPS may yield more precise product predictions for as yet uncharacterized CsTPS.

6.2.4 Distribution of isoprenoid biosynthesis

One major goal of my thesis was to explore the distribution of terpene biosynthesis across tissues and developmental stages of the plant. In Chapter 2, I showed that many transcripts related to isoprenoid biosynthesis are highly expressed in trichomes relative to other tissues (**Figure 2.2**). In Chapter 3, I demonstrated that while terpenes and cannabinoids are present in both flowers and leaves, they are much more abundant in flowers. This confirms the previous consensus that trichomes on female flowers and the primary site of resin biosynthesis and storage.

I identified transcripts corresponding to every known step in the MEP and MEV pathways, as well as GPPS and FPPS candidates. Several of these genes are represented by multiple isoforms, whose expression patterns differ by tissue. For example, in 'Finola', DXS, DXR, and FPPS are each represented by two isoforms whose expression is preferentially in trichomes or roots.

The ratio of cannabinoids and monoterpenes to sesquiterpenes is higher in trichomes on flowers than on leaves (Livingston et al., 2020, Appendix A). Monoterpenes and cannabinoids both stem from the MEP pathway, whereas sesquiterpenes generally arise from the MEV pathway. While this is an overly simplistic view of isoprenoid biosynthesis, it may explain the higher proportions of monoterpenes and cannabinoids in floral trichomes as they mature. Trichomes on flowers increase their production of monoterpenes and cannabinoids, both products of the MEP pathway, as the flower matures. This may suggest that sessile and stalked

trichomes regulate metabolite production differently, with the first primarily committed to sesquiterpene production and the second increasingly committed to monoterpenes and cannabinoids. In other words, sessile trichomes probably commit more flux through the MEV than the MEP pathway, compared to stalked trichomes.

Leaves have generally not been selected for cannabinoid content, so a higher proportion of MEV pathway-derived sesquiterpenes are observed there. This same phenomenon explains the increasing proportion of monoterpenes to sesquiterpenes over floral development, as cannabinoid biosynthesis is upregulated.

6.3 Conclusion

I was able to address all three major goals of my thesis. I identified the TPS enzymes responsible for the most abundant terpenes found in cannabis resin. Using transcriptome analysis and metabolite profiling, I was able to associate expression of genes related to terpene biosynthesis to the sites where terpenes accumulate in cannabis plants. I also performed some initial investigations into the mechanisms of catalysis that give rise to terpene diversity in cannabis. Through collaborative work I helped establish the spatial and developmental patterns of terpene accumulation in glandular trichomes. Further collaborative work established the role of an aPT in early cannabinoid biosynthesis. With these results, this thesis lays a groundwork for many aspects of specialized metabolism in cannabis, a field which may see significant advancement in the coming years.

References

Aaron JA, Christianson DW (2010) Trinuclear metal clusters in catalysis by terpenoid synthases. Pure Appl Chem **82**: 1585–1597

Aharoni A, Giri AP, Verstappen FWA, Bertea CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ (2004) Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell **16**: 3110–31

Ahmed SA, Ross SA, Slade D, Radwan MM, Khan IA, ElSohly MA (2015) Minor oxygenated cannabinoids from high potency Cannabis sativa L. Phytochemistry **117**: 194–199

Aizpurua-Olaizola O, Soydaner U, Öztürk E, Schibano D, Simsir Y, Navarro P, Etxebarria N, Usobiaga A (2016) Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes. J Nat Prod **97**: 324–331

Akashi T, Sasaki K, Aoki T, Ayabe S, Yazaki K (2009) Molecular cloning and characterization of a cDNA for pterocarpan 4-dimethylallyltransferase catalyzing the key prenylation step in the biosynthesis of glyceollin, a soybean phytoalexin. Plant Physiol **149**: 683–93

Aldington S, Harwood M, Cox B, Weatherall M, Beckert L, Hansell A, Pritchard A, Robinson G, Beasley R, Cannabis and Respiratory Disease Research Group (2008) Cannabis use and risk of lung cancer: a case–control study. Eur Respir J **31**: 280–286

Alhassan A, Abdullahi M, Uba A, Umar A (2014) Prenylation of Aromatic Secondary Metabolites: A New Frontier for Development of Novel Drugs. Trop J Pharm Res **13**: 307

Ali M, Li P, She G, Chen D, Wan X, Zhao J (2017) Transcriptome and metabolite analyses reveal the complex metabolic genes involved in volatile terpenoid biosynthesis in garden sage (Salvia officinalis). Sci Rep **7**: 16074

Allen KD, McKernan K, Pauli C, Roe J, Torres A, Gaudino R (2019) Genomic characterization of the complete terpene synthase gene family from Cannabis sativa. PLoS One **14**: e0222363

Alluis B, Dangles O (1999) Acylated flavone glucosides: Synthesis, conformational investigation, and complexation properties. Helv Chim Acta **82**: 2201–2212

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S,

von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol **37**: 420–423

Andersen-Ranberg J, Kongstad KT, Nafisi M, Staerk D, Okkels FT, Mortensen UH, Lindberg Møller B, Frandsen RJN, Kannangara R (2017) Synthesis of C-Glucosylated Octaketide Anthraquinones in Nicotiana benthamiana by Using a Multispecies-Based Biosynthetic Pathway. ChemBioChem **18**: 1893–1897

Andersen-Ranberg J, Kongstad KT, Nielsen MT, Jensen NB, Pateraki I, Bach SS, Hamberger B, Zerbe P, Staerk D, Bohlmann J, et al (2016) Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. Angew Chemie - Int Ed **55**: 2142–2146

Andre CM, Hausman J-F, Guerriero G (2016) Cannabis sativa: The Plant of the Thousand and One Molecules. Front Plant Sci **7**: 19

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of Cannabis sativa. Genome Biol **12**: R102

Ban Z, Qin H, Mitchell AJ, Liu B, Zhang F, Weng J-K, Dixon RA, Wang G (2018) Noncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. Proc Natl Acad Sci **115**: E5223–E5232

Bártíková H, Skálová L, StuchÍková L, Vokřál I, Vaněk T, Podlipná R (2015) Xenobiotic-metabolizing enzymes in plants and their role in uptake and biotransformation of veterinary drugs in the environment. Drug Metab Rev **47**: 374–387

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T (2017) The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res **45**: D313–D319

BLAZQUEZ CRIS, CASANOVA ML, PLANAS ANNA, DEL PULGAR TG, VILLANUEVA CONC, FERNANDEZ-ACENERO MJ, ARAGONES JULI, HUFFMAN JW, JORCANO JL, GUZMAN MANU (2003) Inhibition of tumor angiogenesis by cannabinoids. FASEB J **17**: 529–531

Bohlmann J, Meyer-Gauen G, Croteau R (1998) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc Natl Acad Sci **95**: 4126–4133

Booth JK, Bohlmann J (2019) Terpenes in Cannabis sativa – From plant genome to humans.

Plant Sci. doi: 10.1016/J.PLANTSCI.2019.03.022

Booth JK, Page JE, Bohlmann J (2017) Terpene synthases from Cannabis sativa. PLoS One **12**: e0173911

Borrull A, López-Martínez G, Poblet M, Cordero-Otero R, Rozès N (2015) New insights into the toxicity mechanism of octanoic and decanoic acids on saccharomyces cerevisiae. Yeast **32**: 451–460

Botta B, Vitali A, Menendez P, Misiti D, Monache G (2005) Prenylated Flavonoids: Pharmacology and Biotechnology. Curr Med Chem **12**: 713–739

Braich S, Baillie RC, Jewell LS, Spangenberg GC, Cogan NOI (2019) Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis. Sci Rep **9**: 16583

Brandt W, Bräuer L, Günnewich N, Kufka J, Rausch F, Schulze D, Schulze E, Weber R, Zakharova S, Wessjohann L (2009) Molecular and structural basis of metabolic diversity mediated by prenyldiphosphate converting enzymes. Phytochemistry **70**: 1758–1775

Bräuer L, Brandt W, Schulze D, Zakharova S, Wessjohann L (2008) A Structural Model of the Membrane-Bound Aromatic Prenyltransferase UbiA fromE. coli. ChemBioChem **9**: 982–992

Brenneisen R (2007) Chemistry and Analysis of Phytocannabinoids and Other Cannabis Constituents. Forensic Sci. Med. Marijuana Cannabinoids. pp 17–49

Brenneisen R, elSohly M a (1988) Chromatographic and spectroscopic profiles of Cannabis of different origins: Part I. J Forensic Sci **33**: 1385–404

De Bruyn F, Maertens J, Beauprez J, Soetaert W, De Mey M (2015) Biotechnological advances in UDP-sugar based glycosylation of small molecules. Biotechnol Adv **33**: 288–302

Bureau I (2011) AROMATIC PRENYLTRANSFERASE FROM CANNABIS.

Burke C, Croteau R (2002) Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. J Biol Chem **277**: 3141–9

Burke YD, Stark MJ, Roach SL, Sen SE, Crowell PL (1997) Inhibition of pancreatic cancer growth by the dietary isoprenoids farnesol and geraniol. Lipids **32**: 151–156

Cabral MG, Viegas CA, Sá-Correia I (2001) Mechanisms underlying the acquisition of resistance to octanoic-acid-induced-death following exposure of Saccharomyces cerevisiae to mild stress imposed by octanoic acid or ethanol. Arch Microbiol **175**: 301–7

Carretero-Paulet L, Cairó A, Talavera D, Saura A, Imperial S, Rodríguez-Concepción M, Campos N, Boronat A (2013) Functional and evolutionary analysis of DXL1, a non-essential gene encoding a 1-deoxy-D-xylulose 5-phosphate synthase like protein in Arabidopsis thaliana. Gene **524**: 40–53

Carvalho Â, Hansen EH, Kayser O, Carlsen S, Stehle F (2017) Designing microorganisms for heterologous biosynthesis of cannabinoids. FEMS Yeast Res **17**: 1–11

Casal M, Paiva S, Queirós O, Soares-Silva I (2008) Transport of carboxylic acids in yeasts. FEMS Microbiol Rev **32**: 974–994

Casano S, Grassi G, Martini V, Michelozzi M (2011) Variations in terpene profiles of different strains of Cannabis sativa L. Acta Hortic **925**: 115–122

Celedon JM, Chiang A, Yuen MMS, Diaz-Chavez ML, Madilao LL, Finnegan PM, Barbour EL, Bohlmann J (2016) Heartwood-specific transcriptome and metabolite signatures of tropical sandalwood (*Santalum album*) reveal the final step of (*Z*)-santalol fragrance biosynthesis. Plant J **86**: 289–299

Chappell J, Wolf F, Proulx J, Cuellar R, Saunders C (1995) Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants? Plant Physiol **109**: 1337–1343

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J **66**: 212–29

Chen H, Köllner TG, Li G, Wei G, Chen X, Zeng D, Qian Q, Chen F (2020) Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in oryza. Plant Physiol **182**: 480–492

Chen R, Gao B, Liu X, Ruan F, Zhang Y, Lou J, Feng K, Wunsch C, Li S-M, Dai J, et al (2016) Molecular insights into the enzyme promiscuity of an aromatic prenyltransferase. Nat Chem Biol **13**: 226–234

Cheng W, Li W (2014) Structural insights into ubiquinone biosynthesis in membranes. Science

343: 878–81

Christianson DW (2017) Structural and Chemical Biology of Terpenoid Cyclases. Chem Rev acs.chemrev.7b00287

Christianson DW (2006) Structural biology and chemistry of the terpenoid cyclases. Chem Rev **106**: 3412–42

Collakova E, DellaPenna D (2003) Homogentisate Phytyltransferase Activity Is Limiting for Tocopherol Biosynthesis in Arabidopsis. Plant Physiol **131**: 632–642

Crocoll C, Mirza N, Reichelt M, Gershenzon J, Halkier BA (2016) Optimization of Engineered Production of the Glucoraphanin Precursor Dihomomethionine in Nicotiana benthamiana. Front Bioeng Biotechnol. doi: 10.3389/FBIOE.2016.00014

Croteau R (1987) Biosynthesis and catabolism of monoterpenoids. Chem Rev **87**: 929–954

Cunillera N, Arro M, Delourme D, Karst F, Boronat A, Ferrer A (1996) Arabidopsis thaliana Contains Two Differentially Expressed Farnesyl-Diphosphate Synthase Genes. J Biol Chem **271**: 7774–7780

Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E, Searle SMJ, Clamp M (2004) The Ensembl automatic gene annotation system. Genome Res **14**: 942–50

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

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res **36**: W465-9

Drew DP, Andersen TB, Sweetman C, Møller BL, Ford C, Simonsen HT (2015) Two key polymorphisms in a newly discovered allele of the Vitis vinifera TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. J Exp Bot erv491

Dudareva N (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

Durairaj J, Di Girolamo A, Bouwmeester HJ, de Ridder D, Beekwilder J, van Dijk AD (2019) An analysis of characterized plant sesquiterpene synthases. Phytochemistry **158**: 157–165

ElSohly MA, ed (2007) Marijuana and the Cannabinoids. doi: 10.1007/978-1-59259-947-9

Elzinga S, Fischedick JT, Podkolinski R, Raber J (2015) Cannabinoids and Terpenes as Chemotaxonomic Markers in Cannabis. Nat Prod Chem Res. doi: 10.4172/2329-6836.1000181

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc **2**: 953–971

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300**: 1005–1016

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, et al (2011) The tomato terpene synthase gene family. Plant Physiol **157**: 770–89

Fellermeier M, Eisenreich W, Bacher A, Zenk MH (2001a) Biosynthesis of cannabinoids. Eur J Biochem **268**: 1596–1604

Fellermeier M, Eisenreich W, Bacher A, Zenk MH (2001b) Biosynthesis of cannabinoids Incorporation experiments with 13C-labeled glucoses. Eur J Biochem **268**: 1596–1604

Fellermeier M, Zenk MH (1998a) Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett **427**: 283–285

Fellermeier M, Zenk MH (1998b) Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett **427**: 283–285

Fiesel T, Gaid M, Müller A, Bartels J, El-Awaad I, Beuerle T, Ernst L, Behrends S, Beerhues L (2015) Molecular Cloning and Characterization of a Xanthone Prenyltransferase from Hypericum calycinum Cell Cultures. Molecules **20**: 15616–15630

Fischedick JT (2017) Identification of Terpenoid Chemotypes Among High (−)- trans -Δ 9 -Tetrahydrocannabinol-Producing Cannabis sativa L. Cultivars. Cannabis Cannabinoid Res **2**: 34–47

Fischedick JT, Hazekamp A, Erkelens T, Choi YH, Verpoorte R (2010) Metabolic fingerprinting of Cannabis sativa L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. Phytochemistry **71**: 2058–73

Fischer S, Brunk BP, Chen F, Gao X, Omar S, Iodice JB, Shanmugam D, Roos DS, Jr CJS (2011) Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster

Proteomes Into New Ortholog Groups. 1–19

Fredericks WJ, McGarvey T, Wang H, Lal P, Puthiyaveettil R, Tomaszewski J, Sepulveda J, Labelle E, Weiss JS, Nickerson ML, et al (2011) The bladder tumor suppressor protein TERE1 (UBIAD1)modulates cell cholesterol: Implications for tumor progression. DNA Cell Biol **30**: 851–864

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

Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE (2012) Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. Proc Natl Acad Sci U S A **109**: 12811–6

Gajewski J, Buelens F, Serdjukow S, Janßen M, Cortina N, Grubmüller H, Grininger M (2017a) Engineering fatty acid synthases for directed polyketide production. Nat Chem Biol **13**: 363–365

Gajewski J, Pavlovic R, Fischer M, Boles E, Grininger M (2017b) Engineering fungal de novo fatty acid synthesis for short chain fatty acid production. Nat Commun **8**: 1–8

Galland S, Mora N, Abert-Vian M, Rakotomanomana N, Dangles O (2007) Chemical synthesis of hydroxycinnamic acid glucosides and evaluation of their ability to stabilize natural colors via anthocyanin copigmentation. J Agric Food Chem **55**: 7573–7579

Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structures: a so far incomplete view of complex catalysis. Nat Prod Rep **29**: 1153

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol **3**: 408–14

Gertsch J (2018) The Intricate Influence of the Placebo Effect on Medical Cannabis and Cannabinoids. 60–64

Gertsch J, Leonti M, Raduner S, Racz I, Chen J-Z, Xie X-Q, Altmann K-H, Karsak M, Zimmer A (2008) Beta-caryophyllene is a dietary cannabinoid. Proc Natl Acad Sci U S A **105**: 9099–104

Geu-Flores F, Nielsen MT, Nafisi M, Møldrup ME, Olsen CE, Motawia MS, Halkier BA (2009) Glucosinolate engineering identifies a γ-glutamyl peptidase. Nat Chem Biol **5**: 575–577

Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res **35**: e55

Gietz RD, Schiestl RH (2007) Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc **2**: 35–37

Gilbert AN, DiVerdi JA (2018) Consumer perceptions of strain differences in Cannabis aroma. PLoS One **13**: e0192247

Gilmore S, Peakall R, Robertson J (2003) Short tandem repeat (STR) DNA markers are hypervariable and informative in Cannabis sativa: implications for forensic investigations. Forensic Sci Int **131**: 65–74

Gould MN (1997) Cancer chemoprevention and therapy by monoterpenes. Environ Health Perspect **105**: 977–979

Greenhagen BT, O'Maille PE, Noel JP, Chappell J (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proc Natl Acad Sci U S A **103**: 9826–31

Gunnewich N, Page JE, Koellner T, Degenhardt J, Kutchan TM (2007) Functional expression and characterization of trichome- specific (-)-limonene synthase and (+)- α -pinene synthase from Cannabis sativa. Nat Prod Commun **0**: 1–3

Hall DE, Robert J a, Keeling CI, Domanski D, Quesada AL, Jancsik S, Kuzyk M a, Hamberger B, Borchers CH, Bohlmann J (2011) An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (Picea sitchensis) genotypes that are resistant or susceptible to white pine weevil. Plant J **65**: 936–948

Hammond CT, Mahlberg PG (1973) MORPHOLOGY OF GLANDULAR HAIRS OF CANNABIS SATIVA FROM SCANNING ELECTRON MICROSCOPY. Am J Bot **60**: 524–528

Hanuš LO, Meyer SM, Muñoz E, Taglialatela-Scafati O, Appendino G (2016) Phytocannabinoids: A unified critical inventory. Nat Prod Rep. doi: 10.1039/c6np00074f

Hattan J, Shindo K, Ito T, Shibuya Y, Watanabe A, Tagaki C, Ohno F, Sasaki T, Ishii J, Kondo A, et al (2016) Identification of a novel hedycaryol synthase gene isolated from Camellia brevistyla flowers and floral scent of Camellia cultivars. Planta **243**: 959–972

Hazekamp A, Fischedick JT (2012) Cannabis - from cultivar to chemovar. Drug Test Anal **4**: 660–7

van Herpen TWJM, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J (2010) Nicotiana benthamiana as a Production Platform for Artemisinin Precursors. PLoS One **5**: e14222

Heskes AM, Sundram TCM, Boughton BA, Jensen NB, Hansen NL, Crocoll C, Cozzi F, Rasmussen S, Hamberger B, Hamberger B, et al (2018) Biosynthesis of bioactive diterpenoids in the medicinal plant Vitex agnus-castus. Plant J **93**: 943–958

Hillig KW (2004) A chemotaxonomic analysis of terpenoid variation in Cannabis. Biochem Syst Ecol **32**: 875–891

Irmisch S, Jancsik S, Yuen MMS, Madilao LL, Bohlmann J (2019a) Biosynthesis of the Anti-Diabetic Metabolite Montbretin A: Glucosylation of the Central Intermediate mini-MbA. Plant J 1–13

Irmisch S, Jo S, Roach CR, Jancsik S, Yuen MM Saint, Madilao LL, O'neil-Johnson M, Williams R, Withers SG, Bohlmann J (2018) Discovery of UDP-glycosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite montbretin A. Plant Cell **30**: 1864–1886

Irmisch S, Ruebsam H, Jancsik S, Man Saint Yuen M, Madilao LL, Bohlmann J (2019b) Flavonol Biosynthesis Genes and Their Use in Engineering the Plant Antidiabetic Metabolite Montbretin A. Plant Physiol **180**: 1277–1290

Jenkins C, Orsburn B (2020) The Cannabis Proteome Draft Map Project. Int J Mol Sci **21**: 965

Jones CG, Moniodis J, Zulak KG, Scaffidi A, Plummer JA, Ghisalberti EL, Barbour EL, Bohlmann J (2011) Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b subfamilies, including santalene synthases. J Biol Chem **286**: 17445–54

Kabiri M, Kamal SH, Pawar S V., Roy PR, Derakhshandeh M, Kumar U, Hatzikiriakos SG, Hossain S, Yadav VG (2018) A stimulus-responsive, in situ-forming, nanoparticle-laden hydrogel for ocular drug delivery. Drug Deliv Transl Res **8**: 484–495

Kampranis SC, Ioannidis D, Purvis A, Mahrez W, Ninga E, Katerelos NA, Anssour S, Dunwell JM, Degenhardt J, Makris AM, et al (2007) Rational Conversion of Substrate

and Product Specificity in a Salvia Monoterpene Synthase: Structural Insights into the Evolution of Terpene Synthase Function. Plant Cell **19**: 1994–2005

Karlson J, Borg-Karlson A-K, Unelius R, Shoshan M, Wilking N, Ringborg U, Linder S (1996) Inhibition of tumor cell growth by monoterpenes in vitro: evidence of a Ras-independent mechanism of action. Anticancer Drugs **7**: 422–429

Katzen F (2007) Gateway ® recombinational cloning: a biological operating system . Expert Opin Drug Discov **2**: 571–589

Kaufman L, Rousseeuw PJ (1990) Finding groups in data : an introduction to cluster analysis. Wiley-Interscience

Kessler A, Baldwin IT (2001) Defensive function of herbivore-induced plant volatile emissions in nature. Science **291**: 2141–4

Kim E-S, Mahlberg PG (1991) Secretory Cavity Development in Glandular Trichomes of Cannabis sativa L. (Cannabaceae). Am J Bot **78**: 220–229

Köllner TG, Held M, Lenk C, Hiltpold I, Turlings TCJ, Gershenzon J, Degenhardt J (2008) A maize (E)-beta-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. Plant Cell **20**: 482–94

Köllner TG, Schnee C, Gershenzon J, Degenhardt J (2004) The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell **16**: 1115–31

Krause ST, Köllner TG, Asbach J, Degenhardt J (2013) Stereochemical mechanism of two sabinene hydrate synthases forming antipodal monoterpenes in thyme (Thymus vulgaris). Arch Biochem Biophys **529**: 112–121

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins Struct Funct Bioinforma **77**: 114–122

Külheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg AA, Degenhardt J, Foley WJ (2015) The Eucalyptus terpene synthase gene family. BMC Genomics **16**: 450

Kuzuyama T, Noel JP, Richard SB (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. Nature **435**: 983–7

Lange BM, Turner GW (2013) Terpenoid biosynthesis in trichomes--current status and future opportunities. Plant Biotechnol J **11**: 2–22

Lange BM, Wildung MR, Stauber EJ, Sanchez C, Pouchnik D, Croteau R (2000) Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. Proc Natl Acad Sci U S A **97**: 2934–9

Lange I, Poirier BC, Herron BK, Lange BM (2015) Comprehensive Assessment of Transcriptional Regulation Facilitates Metabolic Engineering of Isoprenoid Accumulation in Arabidopsis. Plant Physiol **169**: 1595–606

Laverty KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, Deikus G, Sebra R, Hughes TR, Page JE, et al (2019) A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Res **29**: 146–156

Lê S, Josse J, Husson F (2008) **FactoMineR** : An *R* Package for Multivariate Analysis. J Stat Softw **25**: 1–18

Lee Chae, Taehyong Kim, Ricardo Nilo-Poyanco SYR (2014) Genomic Signatures of Specialized Metabolism in Plants Lee. Science (80-) **344**: 510–513

Legras JL, Erny C, Le Jeune C, Lollier M, Adolphe Y, Demuyter C, Delobel P, Blondin B, Karst F (2010) Activation of two different resistance mechanisms in Saccharomyces cerevisiae upon exposure to octanoic and decanoic acids. Appl Environ Microbiol **76**: 7526–35

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res **47**: W256–W259

Li H-L (1973) An archaeological and historical account of cannabis in China. Econ Bot **28**: 437–448

Li H, Ban Z, Qin H, Ma L, King AJ, Wang G (2015a) A Heteromeric Membrane-Bound Prenyltransferase Complex from Hop Catalyzes Three Sequential Aromatic Prenylations in the Bitter Acid Pathway. Plant Physiol **167**: 650–659

Li H, Ban Z, Qin H, Ma L, King AJ, Wang G (2015b) A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. Plant Physiol **167**: 650–9

Li W (2016a) Bringing Bioactive Compounds into Membranes: The UbiA Superfamily of Intramembrane Aromatic Prenyltransferases. Trends Biochem Sci **41**: 356–370

Li W (2016b) Bringing Bioactive Compounds into Membranes: The UbiA Superfamily of Intramembrane Aromatic Prenyltransferases. Trends Biochem Sci **41**: 356–370

Lichtenthaler HK, Rohmer M, Schwender J (1997) Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. Physiol Plant **101**: 643–652

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, et al (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature **402**: 761–8

Liu P, Chernyshov A, Najdi T, Fu Y, Dickerson J, Sandmeyer S, Jarboe L (2013) Membrane stress caused by octanoic acid in Saccharomyces cerevisiae. Appl Microbiol Biotechnol **97**: 3239–3251

Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, Castellarin SD, Bohlmann J, Page JE, Samuels AL (2019) Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. Plant J tpj.14516

Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, Castellarin SD, Bohlmann J, Page JE, Samuels AL (2020) Cannabis glandular trichomes alter morphology and metabolite content during flower maturation. Plant J **101**: 37–56

Lõoke M, Kristjuhan K, Kristjuhan A (2011) Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques **50**: 325–8

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550

Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W, et al (2019) Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nat 2019 1

Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, Ruthenburg TC, DeCesare K, Land DP, Kane NC (2016) Genomic and Chemical Diversity in Cannabis. CRC Crit Rev Plant Sci **35**: 349–363

M.L. Banett AMSAFJEL (1963) Cannflavin A and B, prenylated flavones from Cannabis

sativa L. Archs Insect Biochem PhysioL **8**: 243

Ma Y, Yuan L, Wu B, Li X, Chen S, Lu S (2012) Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in Salvia miltiorrhiza. J Exp Bot **63**: 2809–23

Marchini M, Charvoz C, Dujourdy L, Baldovini N, Filippi J-J (2014) Multidimensional analysis of Cannabis volatile constituents: identification of 5,5-Dimethyl-1-vinylbicyclo[2.1.1]hexane as a volatile marker of Hashish, the resin of Cannabis sativa L. J Chromatogr A **1370**: 200–215

Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, He J, Gang DR, Weiblen GD, Dixon RA (2009) Identification of candidate genes affecting Delta9-tetrahydrocannabinol biosynthesis in Cannabis sativa. J Exp Bot **60**: 3715–26

Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J (2010) Functional annotation, genome organization and phylogeny of the grapevine (Vitis vinifera) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol **10**: 226

Martin DM, Toub O, Chiang A, Lo BC, Ohse S, Lund ST, Bohlmann J (2009) The bouquet of grapevine (Vitis vinifera L. cv. Cabernet Sauvignon) flowers arises from the biosynthesis of sesquiterpene volatiles in pollen grains. Proc Natl Acad Sci U S A **106**: 7245–50

Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Düsterhöft A, Stiekema W, Entian K-D, Terryn N, et al (1999) Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature **402**: 769–777

McDowell ET, Kapteyn J, Schmidt A, Li C, Kang JH, Descour A, Shi F, Larson M, Schilmiller A, An L, et al (2011) Comparative functional genomic analysis of solanum glandular trichome types. Plant Physiol **155**: 524–539

Meehan-Atrash J, Luo W, McWhirter KJ, Strongin RM (2019) Aerosol Gas-Phase Components from Cannabis E-Cigarettes and Dabbing: Mechanistic Insight and Quantitative Risk Analysis. ACS Omega **4**: 16111–16120

Meehan-Atrash J, Luo W, Strongin RM (2017) Toxicant Formation in Dabbing: The Terpene Story. ACS Omega **2**: 6112–6117

de Meijer EPM, Bagatta M, Carboni A, Crucitti P, Moliterni VMC, Ranalli P, Mandolino

G (2003) The Inheritance of Chemical Phenotype in Cannabis sativa L. Genetics **163**: 335–346

de Meijer EPM, Hammond KM, Sutton A (2009) The inheritance of chemical phenotype in Cannabis sativa L. (IV): cannabinoid-free plants. Euphytica **168**: 95–112

Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lugan R, Ilc T, et al (2014) The seco-iridoid pathway from Catharanthus roseus. Nat Commun **5**: 3606

Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, Halkier BA (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng **14**: 104–111

Morehouse BR, Kumar RP, Matos JO, Olsen SN, Entova S, Oprian DD (2017) Functional and Structural Characterization of a (+)-Limonene Synthase from *Citrus sinensis*. Biochemistry **56**: 1706–1715

Morimoto S, Komatsu K, Taura F, Shoyama Y (1998) Purification and characterization of cannabichromenic acid synthase from Cannabis sativa. Phytochemistry **49**: 1525–1529

Mugoni V, Postel R, Catanzaro V, De Luca E, Turco E, Digilio G, Silengo L, Murphy MP, Medana C, Stainier DYR, et al (2013) Ubiad1 is an antioxidant enzyme that regulates eNOS activity by CoQ10 synthesis. Cell **152**: 504–518

Munakata R, Inoue T, Koeduka T, Karamat F, Olry A, Sugiyama A, Takanashi K, Dugrand A, Froelicher Y, Tanaka R, et al (2014) Molecular cloning and characterization of a geranyl diphosphate-specific aromatic prenyltransferase from lemon. Plant Physiol **166**: 80–90

Nagia M, Gaid M, Biedermann E, Fiesel T, El-Awaad I, Hänsch R, Wittstock U, Beerhues L (2019) Sequential regiospecific gem-diprenylation of tetrahydroxyxanthone by prenyltransferases from Hypericum sp. New Phytol **222**: 318–334

Nagia MM, El-Metwally MM, Shaaban M, El-Zalabani SM, Hanna AG (2012) Four butyrolactones and diverse bioactive secondary metabolites from terrestrial Aspergillus flavipes MM2: isolation and structure determination. Org Med Chem Lett **2**: 9

Nip KM, Chiu R, Yang C, Chu J, Mohamadi H, Warren RL, Birol I (2019) RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. bioRxiv 701607

Nørholm MHH (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. BMC Biotechnol **10**: 21

Nuutinen T (2018) Medicinal properties of terpenes found in Cannabis sativa and Humulus lupulus. Eur J Med Chem **157**: 198–228

O'Maille PE, Chappell J, Noel JP (2004) A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases. Anal Biochem **335**: 210–7

Ohara K, Mito K, Yazaki K (2013) Homogeneous purification and characterization of LePGT1 - a membrane-bound aromatic substrate prenyltransferase involved in secondary metabolism of *Lithospermum erythrorhizon*. FEBS J **280**: 2572–2580

Ohara K, Muroya A, Fukushima N, Yazaki K (2009) Functional characterization of LePGT1, a membrane-bound prenyltransferase involved in the geranylation of p-hydroxybenzoic acid. Biochem J **421**: 231–41

Ohara K, Yamamoto K, Hamamoto M, Sasaki K, Yazaki K (2006) Functional characterization of OsPPT1, which encodes p-hydroxybenzoate polyprenyltransferase involved in ubiquinone biosynthesis in Oryza sativa. Plant Cell Physiol **47**: 581–90

Okada K, Ohara K, Yazaki K, Nozaki K, Uchida N, Kawamukai M, Nojiri H, Yamane H (2004) The AtPPT1 gene encoding 4-hydroxybenzoate polyprenyl diphosphate transferase in ubiquinone biosynthesis is required for embryo development in Arabidopsis thaliana. Plant Mol Biol **55**: 567–577

Orlova I, Nagegowda DA, Kish CM, Gutensohn M, Maeda H, Varbanova M, Fridman E, Yamaguchi S, Hanada A, Kamiya Y, et al (2009) The small subunit of snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. Plant Cell **21**: 4002–17

Paetzold H, Garms S, Bartram S, Wieczorek J, Urós-Gracia E-M, Rodríguez-Concepción M, Boland W, Strack D, Hause B, Walter MH (2010) The isogene 1-deoxy-D-xylulose 5-phosphate synthase 2 controls isoprenoid profiles, precursor pathway allocation, and density of tomato trichomes. Mol Plant **3**: 904–16

Page JE, Boubakir Z (2012) Aromatic Prenyltransferase from Cannabis.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods **14**: 417–419

Pellechia T (2018) Legal Cannabis Industry Poised For Big Growth, In North America And Around The World. Forbes,

Radwan MM, Elsohly MA, Slade D, Ahmed SA, Wilson L, El-Alfy AT, Khan IA, Ross SA (2008) Non-cannabinoid constituents from a high potency Cannabis sativa variety. Phytochemistry **69**: 2627–33

Rea KA, Casaretto JA, Al-Abdul-Wahid MS, Sukumaran A, Geddes-McAlister J, Rothstein SJ, Akhtar TA (2019) Biosynthesis of cannflavins A and B from Cannabis sativa L. Phytochemistry **164**: 162–171

Reed J, Stephenson MJ, Miettinen K, Brouwer B, Leveau A, Brett P, Goss RJM, Goossens A, O'Connell MA, Osbourn A (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. Metab Eng **42**: 185–193

Reimann-Philipp U, Speck M, Orser C, Johnson S, Hilyard A, Turner H, Stokes AJ, Small-Howard AL (2019) Cannabis Chemovar Nomenclature Misrepresents Chemical and Genetic Diversity; Survey of Variations in Chemical Profiles and Genetic Markers in Nevada Medical Cannabis Samples. Cannabis Cannabinoid Res can.2018.0063

Richins RD, Rodriguez-Uribe L, Lowe K, Ferral R, O'Connell MA (2018) Accumulation of bioactive metabolites in cultivated medical Cannabis. PLoS One **13**: e0201119

Roach CR, Hall DE, Zerbe P, Bohlmann J (2014) Plasticity and Evolution of (+)-3-Carene Synthase and (-)-Sabinene Synthase Functions of a Sitka Spruce Monoterpene Synthase Gene Family Associated with Weevil Resistance. J Biol Chem **289**: 23859–23869

Rodziewicz P, Loroch S, Marczak Ł, Sickmann A, Kayser O (2019) Cannabinoid synthases and osmoprotective metabolites accumulate in the exudates of Cannabis sativa L. glandular trichomes. Plant Sci **284**: 108–116

Ross SA, ElSohly MA (1996) The volatile oil composition of fresh and air-dried buds of Cannabis sativa. J Nat Prod **59**: 49–51

Rousseau MA, Sabol K (2018) THE ROLE OF CANNABINOIDS AND TERPENES IN CANNABIS MEDIATED ANALGESIA IN RATS.

Russo EB (2011) Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br J Pharmacol **163**: 1344–64

Russo EB, Jiang H-E, Li X, Sutton A, Carboni A, del Bianco F, Mandolino G, Potter DJ,

Zhao Y-X, Bera S, et al (2008) Phytochemical and genetic analyses of ancient cannabis from Central Asia. J Exp Bot **59**: 4171–4182

Russo PST, Ferreira GR, Cardozo LE, Bürger MC, Arias-Carrasco R, Maruyama SR, Hirata TDC, Lima DS, Passos FM, Fukutani KF, et al (2018) CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics **19**: 56

Ryz NR, Remillard DJ, Russo EB (2017) Cannabis Roots: A Traditional Therapy with Future Potential for Treating Inflammation and Pain. Cannabis Cannabinoid Res **2**: 210–216

Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) PEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J **7**: 682–693

Salmon M, Laurendon C, Vardakou M, Cheema J, Defernez M, Green S, Faraldos JA, O'Maille PE (2015) Emergence of terpene cyclization in Artemisia annua. Nat Commun **6**: 6143

Sasaki K, Tsurumaru Y, Yamamoto H, Yazaki K (2011) Molecular characterization of a membrane-bound prenyltransferase specific for isoflavone from Sophora flavescens. J Biol Chem **286**: 24125–34

Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S (2015) The Genetic Structure of Marijuana and Hemp. PLoS One **10**: e0133292

Seeliger D, de Groot BL (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J Comput Aided Mol Des **24**: 417–422

Setzer WN (2008) Ab initio analysis of the Cope rearrangement of germacrane sesquiterpenoids. J Mol Model **14**: 335–342

Shen G, Huhman D, Lei Z, Snyder J, Sumner LW, Dixon RA (2012a) Characterization of an Isoflavonoid-Specific Prenyltransferase from *Lupinus albus*. Plant Physiol **159**: 70–80

Shen G, Huhman D, Lei Z, Snyder J, Sumner LW, Dixon RA (2012b) Characterization of an isoflavonoid-specific prenyltransferase from Lupinus albus. Plant Physiol **159**: 70–80

Shoyama Y, Tamada T, Kurihara K, Takeuchi A, Taura F, Arai S, Blaber M, Shoyama Y, Morimoto S, Kuroki R (2012) Structure and function of Δ1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of Cannabis sativa. J Mol Biol

423: 96–105

Da Silva NA, Srikrishnan S (2012) Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. FEMS Yeast Res **12**: 197–214

Singh G, Verma AK, Kumar V (2016) Catalytic properties, functional attributes and industrial applications of β-glucosidases. 3 Biotech **6**: 3

Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F (2004) The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. J Biol Chem **279**: 39767–74

Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, Shoyama Y (2005) Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. Plant Cell Physiol **46**: 1578–1582

Small E (2015) Evolution and Classification of Cannabis sativa (Marijuana, Hemp) in Relation to Human Utilization. Bot Rev **81**: 189–294

Smit SJ, Vivier MA, Young PR (2019) Linking Terpene Synthases to Sesquiterpene Metabolism in Grapevine Flowers. Front Plant Sci **10**: 177

Sparkes IA, Runions J, Kearns A, Hawes C (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc **1**: 2019–2025

Sperschneider J, Catanzariti A-M, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM (2017) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep **7**: 44598

Srividya N, Davis EM, Croteau RB, Lange BM (2015) Functional analysis of (4S)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. Proc Natl Acad Sci U S A 1501203112-

Srividya N, Lange I, Lange BM (2020) Determinants of Enantiospecificity in Limonene Synthases. Biochemistry. doi: 10.1021/acs.biochem.0c00206

Starks CM, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science **277**: 1815–20

Stevens JF, Page JE (2004) Xanthohumol and related prenylflavonoids from hops and beer: to your good health! Phytochemistry **65**: 1317–30

Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE (2012) The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. Plant J **71**: 353–65

Stovicek V, Borja GM, Forster J, Borodina I (2015) EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial Saccharomyces cerevisiae strains. J Ind Microbiol Biotechnol **42**: 1519–1531

Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Yamaguchi M, Tanaka Y, Kusumi T, Nishino T (2002) Cloning and Characterization of Naringenin 8-Prenyltransferase, a Flavonoid-Specific Prenyltransferase of Sophora flavescens. Plant Physiol **130**: 2142–2151

Tai Y, Liu C, Yu S, Yang H, Sun J, Guo C, Huang B, Liu Z, Yuan Y, Xia E, et al (2018) Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (Camellia sinensis). BMC Genomics **19**: 616

Tanaka R, Oster U, Kruse E, Rüdiger W, Grimm B (1999) Isolation and Functional Analysis of Homogentisate Phytyltransferase from Synechocystis sp. PCC 6803 and Arabidopsis. Plant Physiol **120**: 695–704

Tantillo DJ (2011) Biosynthesis via carbocations: theoretical studies on terpene formation. Nat Prod Rep **28**: 1035–53

Taura F, Morimoto S, Shoyama Y (1996a) Purification and Characterization of Cannabidiolic-acid Synthase from Cannabis sativa L.: BIOCHEMICAL ANALYSIS OF A NOVEL ENZYME THAT CATALYZES THE OXIDOCYCLIZATION OF CANNABIGEROLIC ACID TO CANNABIDIOLIC ACID. J Biol Chem **271**: 17411–17416

Taura F, Morimoto S, Shoyama Y (1996b) Purification and Characterization of Cannabidiolic-acid Synthase from Cannabis sativa L . J Biol Chem **271**: 17411–17416

Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S (2007) Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. FEBS Lett **581**: 2929–2934

Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y, Morimoto S (2009)

Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Lett **583**: 2061–6

Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS (2006) Artemisia annua L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. FEBS Lett **580**: 1411–1416

Tholl D (2015) Biosynthesis and Biological Functions of Terpenoids in Plants. Adv Biochem Eng Biotechnol. doi: 10.1007/10_2014_295

Tholl D, Kish CM, Orlova I, Sherman D, Gershenzon J, Pichersky E, Dudareva N (2004) Formation of Monoterpenes in Antirrhinum majus and Clarkia breweri Flowers Involves Heterodimeric Geranyl Diphosphate Synthases. Plant Cell **16**: 977–992

Tian W, Chen C, Lei X, Zhao J, Liang J (2018) CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res **46**: W363–W367

Tiwari P, Sangwan RS, Sangwan NS (2016) Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. Biotechnol Adv **34**: 714–739

Tsurumaru Y, Sasaki K, Miyawaki T, Momma T, Umemoto N, Yazaki K (2010) An aromatic prenyltransferase-like gene HlPT-1 preferentially expressed in lupulin glands of hop. Plant Biotechnol **27**: 199–204

Tsurumaru Y, Sasaki K, Miyawaki T, Uto Y, Momma T, Umemoto N, Momose M, Yazaki K (2012) HlPT-1, a membrane-bound prenyltransferase responsible for the biosynthesis of bitter acids in hops. Biochem Biophys Res Commun **417**: 393–8

Turner GW, Croteau R (2004) Organization of monoterpene biosynthesis in Mentha. Immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. Plant Physiol **136**: 4215–27

Turner JC, Hemphill JK, Mahlberg PG (1978) QUANTITATIVE DETERMINATION OF CANNABINOIDS IN INDIVIDUAL GLANDULAR TRICHOMES OF CANNABIS SATIVA L. (CANNABACEAE). Am J Bot **65**: 1103–1106

Turner JC, Hemphill JK, Mahlberg PG (1980) Trichomes and Cannabinoid Content of

Developing Leaves and Bracts of Cannabis sativa L. (Cannabaceae). Am J Bot **67**: 1397–1406

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al (2006) The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science (80-) **313**: 1596–1604

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3--new capabilities and interfaces. Nucleic Acids Res **40**: e115

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3**: RESEARCH0034

Vanegas KG, Lehka BJ, Mortensen UH (2017) SWITCH: A dynamic CRISPR tool for genome engineering and metabolic pathway control for cell factory construction in Saccharomyces cerevisiae. Microb Cell Fact. doi: 10.1186/s12934-017-0632-x

Vergara D, Baker H, Clancy K, Keepers KG, Mendieta JP, Pauli CS, Tittes SB, White KH, Kane NC (2016) Genetic and Genomic Tools for Cannabis sativa. CRC Crit Rev Plant Sci **35**: 364–377

Vergara D, Huscher EL, Keepers KG, Givens RM, Cizek CG, Torres A, Gaudino R, Kane NC (2019) Gene copy number is associated with phytochemistry in Cannabis sativa. bioRxiv 736181

Vickery CR, La Clair JJ, Burkart MD, Noel JP (2016) Harvesting the biosynthetic machineries that cultivate a variety of indispensable plant natural products. Curr Opin Chem Biol **31**: 66–73

Voinnet O, Rivas S, Mestre P, Baulcombe D (2003) An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant J **33**: 949–956

Wada Y, Sirikantaramas S, Morimoto S, Ishikawa Y, Shoyama Y, Shoyama Y, Taura F (2004) The Gene Controlling Marijuana Psychoactivity. J Biol Chem **279**: 39767–39774

Wagner H, Ulrich-Merzenich G (2009) Synergy research: approaching a new generation of phytopharmaceuticals. Phytomedicine **16**: 97–110

Walter MH, Hans J, Strack D (2002) Two distantly related genes encoding 1-deoxy-d-xylulose

5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. Plant J **31**: 243–254

Wang G, Dixon RA (2009) Heterodimeric geranyl(geranyl)diphosphate synthase from hop (Humulus lupulus) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci U S A **106**: 9914–9

Wang R, Chen R, Li J, Liu X, Xie K, Chen D, Yin Y, Tao X, Xie D, Zou J, et al (2014) Molecular characterization and phylogenetic analysis of two novel regio-specific flavonoid prenyltransferases from Morus alba and Cudrania tricuspidata. J Biol Chem **289**: 35815–25

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res **46**: W296–W303

Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, Marks MD (2015) Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytol **208**: 1241–50

Werz O, Seegers J, Schaible AM, Weinigel C, Barz D, Koeberle A, Allegrone G, Pollastro F, Zampieri L, Grassi G, et al (2014) Cannflavins from hemp sprouts, a novel cannabinoid-free hemp food product, target microsomal prostaglandin E2 synthase-1 and 5-lipoxygenase. PharmaNutrition **2**: 53–60

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci U S A **99**: 15375–80

Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of limonene synthase preprotein provides a fully active “pseudomature” form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. Biochemistry **37**: 12213–20

Wood CC, Petrie JR, Shrestha P, Mansour MP, Nichols PD, Green AG, Singh SP (2009) A leaf-based assay using interchangeable design principles to rapidly assemble multistep recombinant pathways. Plant Biotechnol J **7**: 914–924

Yazaki K, Sasaki K, Tsurumaru Y (2009) Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. Phytochemistry **70**: 1739–45

Yoneyama K, Akashi T, Aoki T (2016) Molecular Characterization of Soybean Pterocarpan 2-Dimethylallyltransferase in Glyceollin Biosynthesis: Local Gene and Whole-Genome Duplications of Prenyltransferase Genes Led to the Structural Diversity of Soybean Prenylated Isoflavonoids. Plant Cell Physiol **57**: 2497–2509

Zager JJ, Lange I, Srividya N, Smith A, Lange BM (2019) Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis. Plant Physiol pp.01506.2018

Zerbe P, Chiang A, Yuen M, Hamberger B, Hamberger B, Draper J a, Britton R, Bohlmann J (2012) Bifunctional cis-abienol synthase from Abies balsamea discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. J Biol Chem **287**: 12121–31

Zerbe P, Hamberger B, Yuen MMS, Chiang A, Sandhu HK, Madilao LL, Nguyen A, Hamberger B, Bach SS, Bohlmann J (2013) Gene discovery of modular diterpene metabolism in nonmodel systems. Plant Physiol **162**: 1073–91

Zhao J, Bao X, Li C, Shen Y, Hou J (2016) Improving monoterpene geraniol production through geranyl diphosphate synthesis regulation in Saccharomyces cerevisiae. Appl Microbiol Biotechnol **100**: 4561–4571

Zheng J, He C, Qin Y, Lin G, Park WD, Sun M, Li J, Lu X, Zhang C, Yeh C, et al (2018) Co-expression analysis aids in the identification of genes in the cuticular wax pathway in maize. Plant J **97**: tpj.14140

Zirpel B, Degenhardt F, Martin C, Kayser O, Stehle F (2017) Engineering yeasts as platform organisms for cannabinoid biosynthesis. J Biotechnol **259**: 204–212

Appendix: Excerpts from "Cannabis glandular trichomes alter morphology and metabolite content during flower maturation"

Materials and Methods

Plant growth.

Female pistillate Cannabis sativa L. plants of the auto-flowering hemp-type variety 'Finola' were soil-grown from seed at Anandia Laboratories in Vancouver, British Columbia, in a Health Canada-permitted research lab. Seeds were planted into a soil mixture containing 1.5 tablespoons of FlorikoteTM (15-5-15) per 1L scoop of soil (Sungro, Sunshine Mix #4). Plants were grown under T5 linear fluorescent lamps (Plusrite, FL54T5/865/HO), using an 18/6 cycle (hours of light/hours of dark). Plants were watered with tap water for the first two weeks after planting and subsequently watered with Peter's Excel® (15-5-15) at a concentration of 0.05 tbsp/L (tablespoons per litre) for the duration of vegetative growth. Sterile bamboo stakes were used to support plant weight to best maintain an upright position when necessary. Biological controls were applied to the plants on a weekly basis. Female Cannabis sativa L. plants of the marijuana-type varieties 'Purple Kush' and 'Hindu Kush' were grown via clonal propagation. Clones were rooted in rockwool, then transferred directly into soil in a growth chamber (BC Northern Lights) under LED lights (BC Northern lights, 3000K 80 CRI spectrum). The plants were subjected to vegetative growth for 2-3 weeks using an 18/6 light cycle and watered with Peter's Excel® (15-5-15). To induce flower development, the light cycle was switched to 12/12, and plants were watered with Maxibloom (5-15-14).

Solid-phase microextraction (SPME) GC-MS

Following UV-light assisted microextraction of trichome metabolites, terpene content was analyzed by the Solid Phase Microextraction technique on an Agilent 7890A/5975C Inert XL MSD Triple Axis gas chromatograph mass spectrometer equipped with an Agilent GC Sampler 80 autosampler. Supelco SPME 50/30 µm DVB/CAR/PDMS fiber and Agilent DB-WAX Column (30 m, 0.25 mm internal diameter, 0.25 µm film thickness) were used. The SPME GC-MS cycle included a 300 s pre-incubation, 40°C incubation on SPME-GC agitator unit, 600 s extraction (fiber exposure), and 300 s desorption, with an oven program of 35°C for 4 min, then increase 4° per min to 110°C, then increase 3° per min to 150°C, then increase 25° per min to 230°C and hold for 3 min. The Helium flow rate was 0.9 mL/min and injection was split with a ratio of 4:1. The MS acquisition was performed in electron ionization mode with a mass range from 33.0-400.0. Selected ion monitoring was performed simultaneously for masses 93, 121, 136, 161, 189, and 204 with a dwell of 35. Data processing was performed with the Enhanced Chem Station and mass spectra were matched against NIST/WILEY library spectra (W9N08).

TPS cloning and characterization

Total RNA was isolated from 'Finola' flowers using Invitrogen PureLink Plant RNA reagent (www.thermofisher.com). RNA quality and concentration was measured with a Bioanalyzer 2100 RNA Nanochip assay (www.agilent.ca). Full-length sequences of FN15171.1 and FN20433.1 were obtained using 5` Rapid Amplification of cDNA ends (www.clontech.com). 5` plastid target peptides were predicted using the TargetP algorithm (Emanuelsson et al., 2007). Sequences were truncated to the RRX8W motif (Bohlmann et al., 1998) and cloned into the pASK IBA37 vector (www.iba-lifesciences.org) with a 5` 6X-HIS tag.

Vectors were transformed into E. coli strain BL21DE3 for heterologous protein expression. Bacterial cultures were grown in 50 mL of Luria Broth containing ampicillin (100 mg/ml).

Cultures were grown at 37°C at 200 rpm until they reached OD600 0.6, then at 18°C for two more hr. Protein production was then induced using 200 µg/l anhydrous tetracycline, and the cultures were shaken for a further 18 hr before harvesting by centrifugation.

Recombinant protein was purified using the GE healthcare His SpinTrap kit (www.gelifesciences.com) according to the manufacturer's instructions. Binding buffer for purification was 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) (pH 7.5), 500 mM NaCl, 25 mM imidazole, and 5% glycerol. Cells were lysed in binding buffer supplemented with Roche complete protease inhibitor tablets (lifescience.roche.com) and 0.1 mg/mL lysozyme. Elution buffer was 20 mM HEPES (pH 7.5), 500 mM NaCl, 350 mM imidazole, and 5% glycerol. Purified protein was desalted through Sephadex into TPS assay buffer.

In vitro assays were performed at 500 µL volume by incubating purified protein with isoprenoid diphosphate substrates as previously described (O'Maille et al., 2004). TPS assay buffer was 25 mM HEPES (pH 7.5), 100 mM KCl, 10 mM MgCl2, 5% glycerol, and 5 mM DTT. GPP was dissolved in 50% methanol and used at 32 µM. Assays were overlaid with 500 µL pentane containing 31 g/l isobutyl benzene as internal standard.

Product identification was performed using an Agilent 7890A gas chromatographer with a 7683B series autosampler and a 7000A TripleQuad mass spectrometer. Ionization was at 70 eV electrospray with a flow rate of 1 ml per minute of He. The column was an Agilent VP-5MS (30 m, 250 µm internal diameter, 0.25 µm film). The temperature program was: 50°C for 1 minute, then increase 10°C per minute to 280°C, then hold for 5 min. Injection was 1 µl pulsed splitless. Products were identified by comparison to authentic standards.

Whole organ metabolite analysis

Metabolite analysis was performed on whole organs from 5 ‘Finola’ plants harvested at week 7 post-seed germination (immature calyces) and week 13 post-seed germination (mature calyces and mature vegetative leaves). Three floral clusters were harvested for each time point, one from an axillary bud of a lateral branch from the first node (oldest flowers), one from an axillary bud near the shoot apex, and one from the tip of the main cola (youngest flowers). Three replicate calyces were dissected from each cluster and used for metabolite analysis. At least three vegetative leaves were harvested from the shoot base at the oldest branching nodes. Calyces were immature if there were no conspicuous stalked trichomes and the calyx was less than 4 mm in length, while mature calyces contained >70% stalked trichomes and greater than 4mm in length, as assessed on an Olympus SZX10 stereomicroscope. Vegetative leaves were devoid of stalked glandular trichomes.

Whole calyces were immersed in 500 μL of pentane with 31 g/L isobutyl benzene as internal standard. Resin was extracted by vigorous vortexing with glass beads followed by shaking for 4 hours at room temperature. Tissue was dried at 60°C for 16 hr, then weighed. The same procedure was followed for leaves, but using 4 mL of solvent.

Compounds were identified using the same GC-MS equipment as above, but with the following temperature program: 50°C for 3 min, then increase 10°C per minute to 150°C, then increase 15°C per minute to 320°C, hold for 5 min. Identifications were made by comparison to authentic standards and NIST/WILEY library spectra. Monoterpenes were quantified using standard curves of α-pinene, myrcene, limonene, and linalool. Sesquiterpenes were quantified using β-caryophyllene, (E)-β-farnesene, and bisabolol standard curves. No standards were available for cannabinoids, so THC and CBD were identified by retention index and quantified by relative peak area.

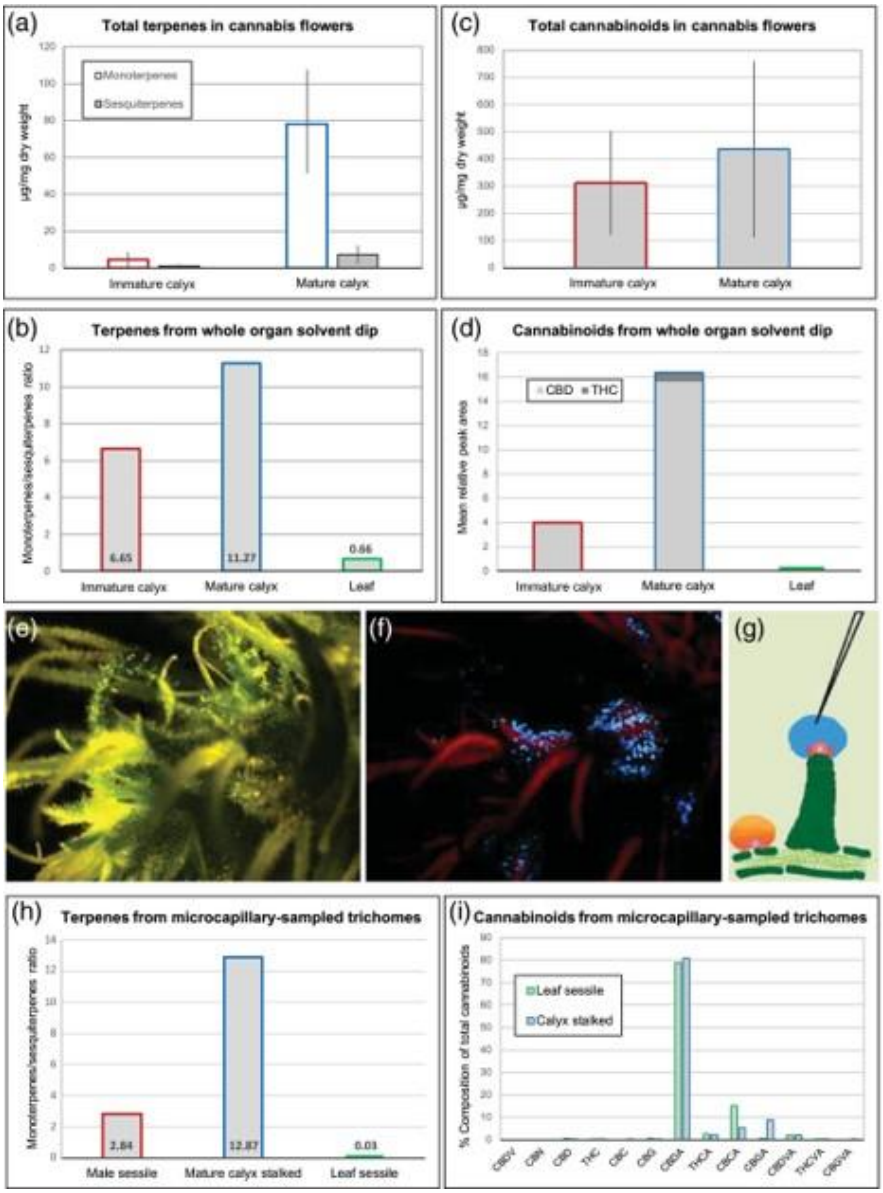


Figure A1. Mature calyces and microcapillary-sampled stalked glandular trichomes have monoterpene-dominant terpene profiles.

(a) Absolute levels of monoterpenes and sesquiterpenes from developing whole calyces dipped in solvent showing increases in monoterpenes as calyces mature. (b) The ratio of monoterpenes:sesquiterpenes is high in both immature calyces, with predominantly premature stalked trichomes, and mature calyces, with stalked trichomes. Leaves, which produce exclusively sessile trichomes, have more sesquiterpenes. (c) Absolute levels of cannabidiolic acid, mean+/-SE, n = 5 plants with 3 technical replicates per plant. (d) Relative amounts of total cannabinoids in immature calyx and mature calyx compared to low levels in leaf; n = 5 plants with 3 technical replicates per plant. (e-g) A floral cluster (e), illuminated with ultraviolet light (f) to produce blue fluorescence from stalked glandular trichomes to assist microcapillary sampling of resin, as shown in (g). (h) The ratio of monoterpenes:sesquiterpenes from microcapillary sampling of storage cavities from anther sessile trichomes, floral stalked trichomes, or leaf sessile trichomes. (i) The cannabinoid profiles of microcapillary-sampled storage cavities of floral stalked trichomes (100 cavities each) and leaf sessile trichomes (200 cavities each), n = 3 plants with 3 technical replicates per plant.

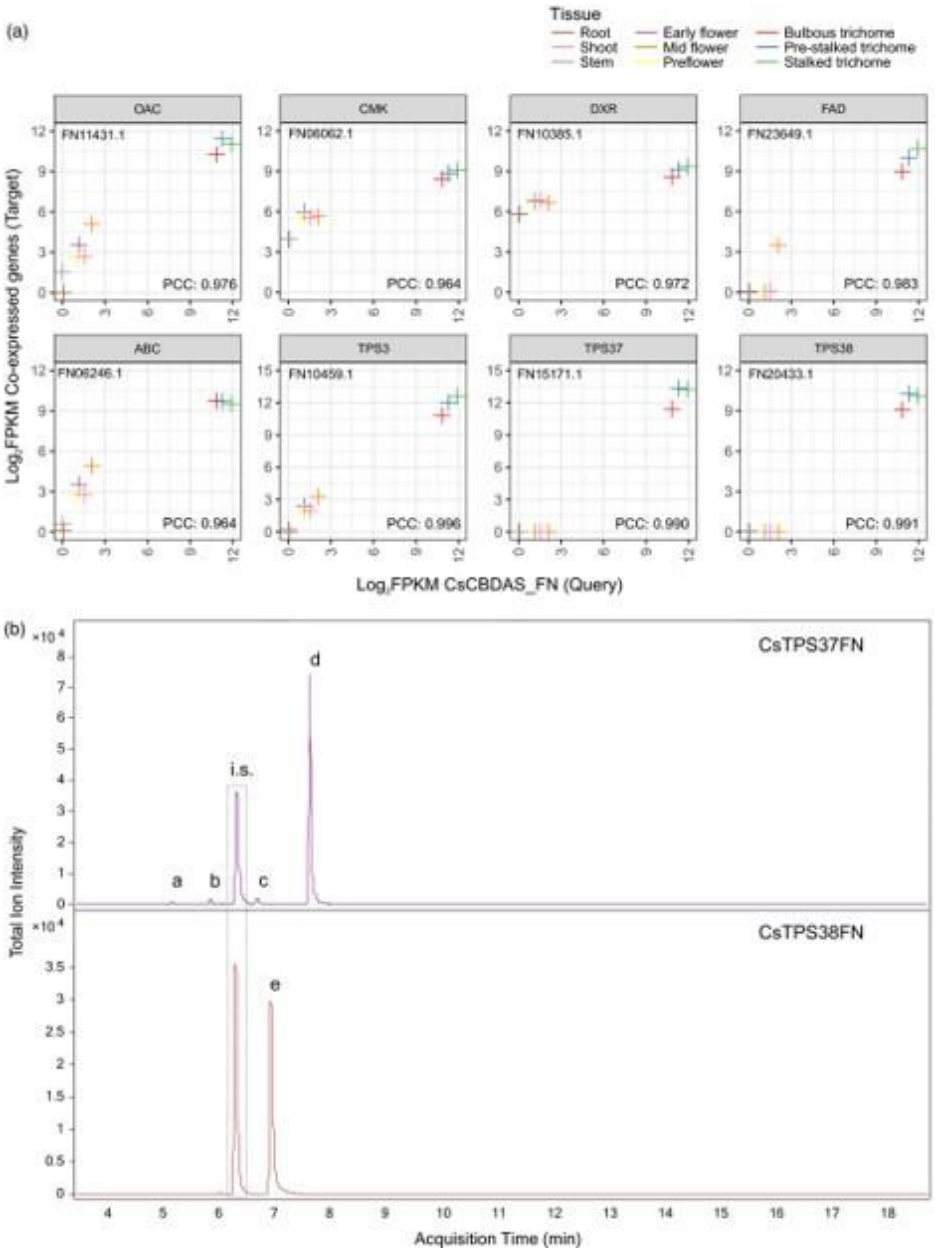


Figure A2. Co-expression analysis using CBDAS as a query reveals numerous genes putatively involved in metabolite biosynthesis, transport, and storage.

(a) Co-expression relationship of CBDAS with selected highly co-expressed genes in various tissues and organs. The Pearson's product-moment correlation coefficient (PCC) depicts the co-expression strength of each interaction. (b) Products of CsTPS37 (top panel) and CsTPS38 (bottom panel) when incubated with GPP. a: α-pinene, b: β-pinene, c: limonene, d: terpinolene, e: (E)-β-ocimene. i.s. = internal standard, isobutylbenzene.

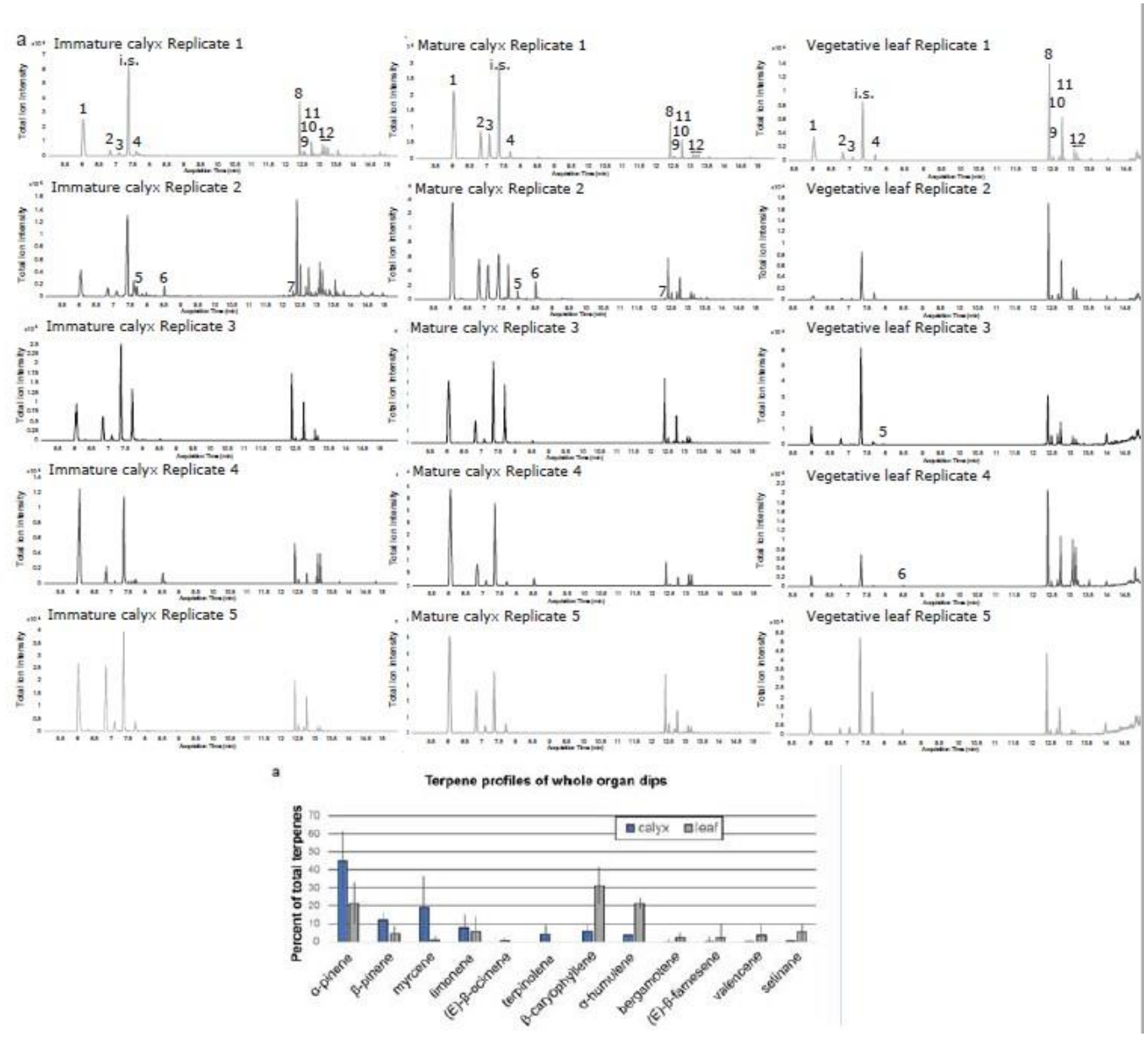


Figure A3. Terpene profiles and chromatograms from whole organ solvent extractions
(a) Total ion chromatograms from five 'Finola' solvent dips of immature flowers, mature flowers, and vegetative leaves. Each chromatogram represents three pooled flowers. 1: α-pinene, 2: β-pinene, 3: myrcene, 4: limonene, 5: (*E*)- β-ocimene, 6: terpinolene, 7: bergamotene, 8: β-caryophyllene, 9: (*E*)- β-farnesene 10: valencene, 11: α-humulene, 12: 3-5 selinane-type sesquiterpenes. i.s. = internal standard, isobutyl benzene. (b) Total terpene profiles of whole-organ solvent extractions of calyces (pooled immature and mature) and leaves.

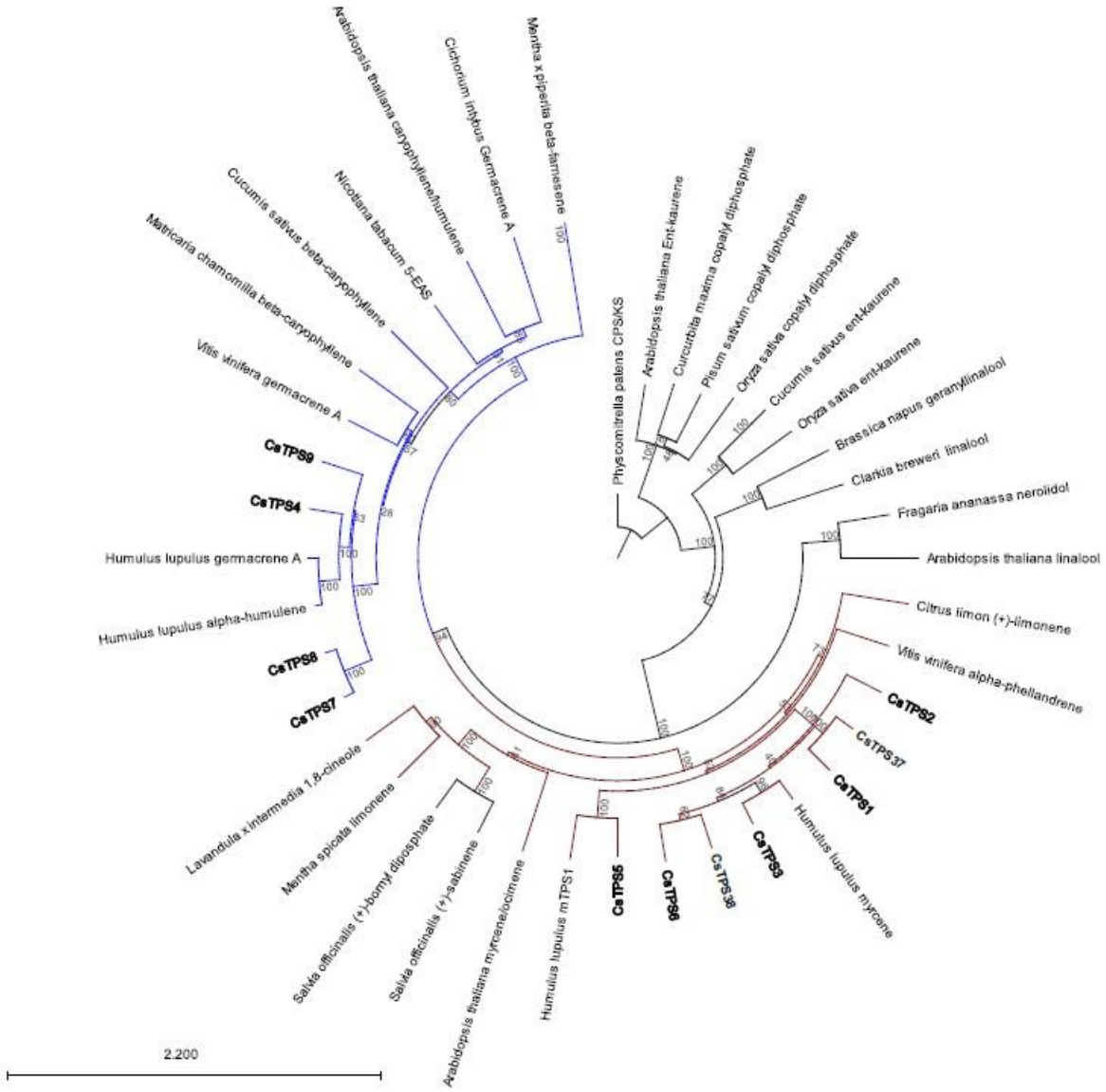


Figure A4. Maximum likelihood phylogeny of terpene synthase (TPS) amino acid sequences
TPS-a (sesquiterpene synthase) clade is shown in blue, and TPS-b (monoterpene synthase) clade in red. *Cannabis sativa* sequences are in bold. Grey numbers represent bootstrap values from 100 replicates.