Pathway Discovery and Characterization for Sesquiterpene and Higher-Molecular Weight Isoprenoids in Basidiomycetes

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Stephen Nicholas Michel

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Claudia Schmidt-Dannert, Advisor

September 2015

Acknowledgements

I would like to acknowledge all those who have helped me become a better scientist throughout the duration of this work. First and foremost, I would like to thank all current and former members of Dr. Schmidt-Dannert's laboratory for providing a fun and supportive place to work, but also for continually pushing me to hone my skills as a scientist for future endeavors. I would especially like to thank Dr. Maureen Quin, the postdoctoral researcher whom I worked under when I joined the lab, who always provided insightful advice, and offered great encouragement when it was needed. I would also like to thank Dr. Kathryn Bushley, Dr. Jon Menke, Dr. Jon Badalamenti, Dr. Kevin Silverstein, Karl Oles, Ying Zhang, Laura Nolden, Bruce Eckloff, LeeAnn Higgins, Todd Markowski, and Susan Van Riper for all of their help in better understanding genomic and proteomic analyses. Another thanks is in order to Dr. Peter Kennedy and Dr. Nhu Nguyen for helping me taxonomically identify *Lactarius* fruiting bodies in the field, and for the established *Lactarius* cultures we received. Furthermore, I would like to thank Ilana Aldor from Bridgestone Tires Inc. for being an excellent contact in our industrial collaboration. Finally, I would like to thank Claudia, for inspiring a truly innovative and exciting project, and providing excellent guidance throughout my time in her lab.

Dedication

This thesis is dedicated to my wife, Olivia Michel, for all of her love and support throughout our respective experience of going through graduate school together. We both know how life for us these past two years would be more difficult without the other around for support. I also dedicate this thesis to my parents, who always encourage me to challenge myself, and to take pride in all that I do. Finally, I also dedicate this thesis to my siblings, for always looking out for me growing up, and for making sure I never took myself too seriously.

Abstract

Isoprenoids represent the largest class of natural products, exhibiting a variety of physiological roles in organisms ranging from bacteria to plants. Mushroom-forming fungi (Basidiomycota) represent a largely diverse, but relatively uncharacterized natural product resource, and are prolific producers of bioactive isoprenoid compounds. In this work, two convergent lines of investigation are pursued involving the discovery and characterization of enzymes responsible for the biosynthesis of isoprenoids in Basidiomycota. The first investigation involved the structural characterization of a protoilludene synthase from the wood-rotting fungus, *Stereum hirsutum*. The second investigation followed the establishment of cultures for eight different *Lactarius* mushroom species, and their subsequent volatile headspace screening for novel isoprenoid compounds. Furthermore, three *Lactarius* species were selected for genome sequencing and mining to identify novel sesquiterpene synthases using an established predictive framework. This predictive framework was broadened in scope to include the search for enzymes responsible for the biosynthesis of higher molecular-weight isoprenoids, particularly those involved in natural rubber biosynthesis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1. Opening Remarks

In this thesis, two convergent lines of investigation are pursued. However, both studies have the same fundamental motivation behind them — discovery and characterization of natural product pathways in higher basidiomycetes (mushrooms). Coincidentally, both these analyses encompass work within the same initial natural product pathway— the synthesis of isoprenoids.

The first investigation follows the efforts of structurally characterizing one such isoprenoid biosynthetic gene from the mushroom *Stereum hirsutum*. The second narrative follows the establishment of cultures for eight different *Lactarius* mushroom species, and the subsequent genome mining for novel biosynthetic genes, particularly those involved in sesquiterpene and higher molecular weight isoprenoid biosynthesis.

The foundation of both inquiries is due in thanks to the previous work performed by current and former members of the Schmidt-Dannert laboratory. It is the goal of this thesis to build upon the basis of this former work to further expand the discovery and characterization of novel isoprenoid biosynthetic pathways in higher fungi to facilitate applications of this group of compounds in medicine and industry.

2. Thesis Overview

Chapter 1 provides a brief introduction on isoprenoid biosynthesis. Additional portions of this chapter pertain to the natural biosynthesis of isoprenoids in

basidiomycetes, and the comparison of isoprenoid content in basidiomycetes versus other living organisms. Finally, previous and ongoing work of the Schmidt-Dannert laboratory in the context of current genome mining and natural product biosynthetic discovery research is reviewed to provide the fundamental basis for the research conducted in this thesis.

Chapter 2 summarizes the previous work of the Schmidt-Dannert laboratory in the structural characterization of a terpene synthase from the mushroom *Omphalotus olearius*, and covers the extensive efforts in the structural characterization of a terpene synthase from the mushroom *Stereum hirsutum*.

Chapter 3 covers the establishment of eight different *Lactarius* mushroom species cultures. These cultivation efforts search for optimal growth media for fast growth, and production of isoprenoid compounds by these *Lactarius* strains. The latter part of this chapter focuses on the genomic mining of the mushroom *Lactarius uvidus* for terpene synthases. More specifically, this included the search for a drimane-like terpene synthase to clarify the mechanism behind drimane biosynthesis.

Chapter 4 covers the ongoing collaboration with Bridgestone Tires, Inc. (Nashville, TN). This chapter summarizes the general state of the natural rubber industry, and its reliance on a few plant species for its production. Furthermore, this chapter utilizes the same genomic fundamentals outlined in Chapter 3 along with complementary proteomics data towards the discovery of natural rubber producing genes from two *Lactarius* mushroom species. This chapter is not for public disclosure.

Chapter 5 summarizes the work of the previous four chapters and discusses potential future directions for the Schmidt-Dannert laboratory to take in natural products biosynthesis with Basidiomycetes.

3. Introduction

3.1. Isoprenoid biosynthesis

By definition, a secondary metabolite is a small molecule not generally associated with the general growth, development or reproduction of an organism [1]. With over 40,000 structures described in literature, isoprenoids encompass the largest group of secondary metabolites found in nature [1]. Be they from plants, animals, fungi or other microorganisms, isoprenoids are prevalent throughout a diverse set of species [2]. In addition to the near ubiquity in living organisms, isoprenoids also have diverse physiological functions. Toxins [3, 4], repellents, attractants [3], membrane components, hormones, pigments and antioxidants [2] are all prospective activities or roles isoprenoids elicit in nature.

Humans have also demonstrated a great interest in this class of molecules throughout history, whether consciously or not. Reports of use of essential oils have dated back to ancient Egyptian times for religious ceremonies [5], and the monoterpene camphor was introduced to Europe in the 11^{th} century and popularized as an ingredient for cooking [5] among other applications.

Modern day isoprenoids of human interest include the monoterpene menthol, often used as a topical analgesic [1], the sesquiterpene lactone artemisinin, a powerful

antimalarial drug [1], and illudin M and illudin S, anticancer therapeutic agents [6]. These isoprenoids are but a few of a plethora of examples with important human applications.

Given this large group of molecules, isoprenoids have come to be referred by various names in literature. Terpenes, terpenoids and isoprenoids are all names that have been used to describe these molecules. Terpenes themselves refer to the olefin carbon scaffold of this class of molecule, while isoprenoids and terpenoids are structurally modified terpenes where either oxygen atoms have been added to the scaffold, or the scaffolds have been modified in some other manner [1]. The terms terpenoids and isoprenoids are used synonymously. However some authors refer to terpenes, terpenoids and isoprenoids interchangeably. This work will use terpenes, terpenoids and isoprenoids interchangeably in that broad sense as well.

While naming conventions are inconsistent for this class of chemicals, the main building block for isoprenoids is not. All isoprenoids are fundamentally derived from the 5-carbon molecule isoprene, though more specifically isopentenyl pyrophosphate (IPP). IPP becomes available from central metabolism via two distinct biosynthetic pathways, depending upon the organism.

Animals, archaea, plants (cytoplasm), some bacterial lineages and most importantly fungi utilize the mevalonate pathway to generate IPP. The mevalonate pathway (or MVA, mevalonic acid pathway) in eukaryotes is realized by the conversion of acetyl-CoA to acetoacetyl-CoA by acetoacetyl-CoA thiolase. Another acetyl-CoA molecule is then added onto acetoacetyl-CoA to make 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) by HMG-CoA synthase. HMG-CoA is then reduced by HMG-CoA reductase to make

mevalonate. Mevalonate is then phosphorylated twice at the 5'-OH by mevalonate kinase and phosphomevalonate kinase, respectively, to make mevalonate pyrophosphate. Mevalonate pyrophosphate is then decarboxylated by mevalonate decarboxylase to finally yield IPP [1]. The full mevalonate pathway can be viewed in **Figure 1.1**.

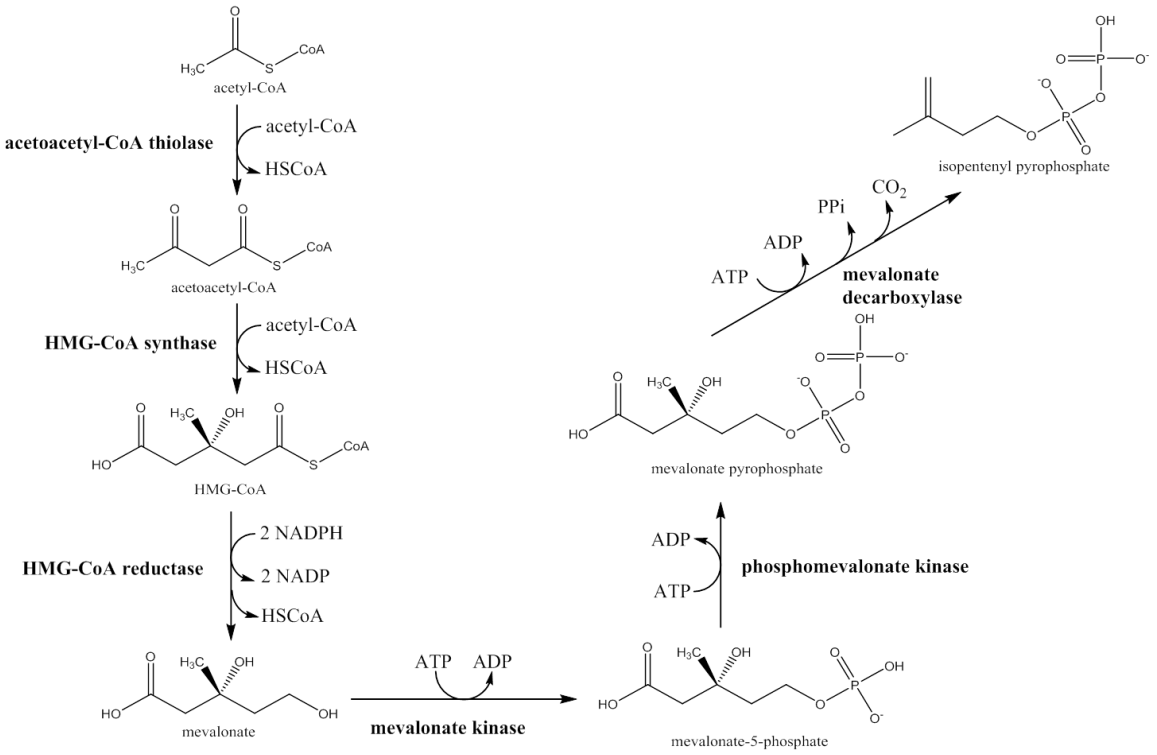


Figure 1.1 The mevalonate pathway. The mevalonate pathway in fungi constructs isopentenyl pyrophosphate (IPP) by first converting 3 acetyl-CoA molecules into HMG-CoA. HMG-CoA is then reduced to mevalonate, before mevalonate is phosphorylated twice to make IPP.

It was originally thought that IPP was solely produced through the mevalonate pathway, but it has since been discovered that plant plastids and most bacteria utilize an alternative pathway [7]. This pathway is called the DXP pathway named for the initial condensation product of pyruvate and glyceraldehyde-3-phosphate, which is 1-deoxy-D-xylulose-5-phosphate (DXP). DXP is further processed by six other enzymes (IspC, IspD,

IspE, IspF, IspG, and IspH in *E. coli*) to make IPP in a 5 to 1 ratio with its isomer dimethylallyl pyrophosphate (DMAPP) [8-12]. This pathway is sometimes also referred to as the MEP pathway, based on the second pathway intermediate 2C-methyl-D-erythrol-4-phosphate.

Isoprenoids are first formed when IPP is combined with its isomer DMAPP in a head-to-tail condensation by geranyl pyrophosphate synthase (GPPS) to make geranyl pyrophosphate (GPP). GPP is the skeleton building block for all C10 isoprenoids, further sub-classed as monoterpenoids. The "isoprenoid rule," first hypothesized by German chemist Otto Wallach in 1913 and modified by Croatian-Swiss Chemist Leopold Ruzicka [13], states that all terpenes are formed by the repeated joining of isoprene units. Indeed, isoprenoids are further sub-classified by the number of isoprene units present in the molecule. **Table 1.1** below represents the different classifications of isoprenoids.

Table 1.1 Parent molecules and classification of isoprenoids based on number of isoprene units.

Classification	Isoprene Units	Number of Carbon Atoms	Parent Molecule
Monoterpenoids	2	10	Geranyl Pyrophosphate (GPP)
Sesquiterpenoids	3	15	Farnesyl Pyrophosphate (FPP)
Diterpenoids	4	20	Geranylgeranyl Pyrophosphate (GGPP)
Sesterterpenoids	5	25	Geranylfarnesyl Pyrophosphate (GFPP)
Triterpenoids	6	30	Squalene
Carotenoids	8	40	Phytoene
Rubber	>100	>500	N/A

Like monoterpenoids, the skeleton building blocks for sesquiterpenoids and diterpenoids are originally derived from a linear prenyl molecule. The construction of all linear prenyl molecules in isoprenoid biosynthesis are made possible through prenyl chain synthases. For instance, sesquiterpenoids are originally formed through the head-to-tail addition of IPP to GPP by farnesyl pyrophosphate synthase (FPPS) to make (C15)

farnesyl pyrophosphate (FPP). Additionally, diterpenoids are originally formed through the head-to-tail addition of IPP to FPP to make (C20) geranylgeranyl pyrophosphate (GGPP) by geranylgeranyl pyrophosphate synthase (GGPPS). Finally, rare sesterterpenoids are formed through the addition of another IPP unit to GGPP by geranylfarnesyl pyrophosphate synthase (GFPPS) to make (C25) geranylfarnesyl pyrophosphate (GFPP). Triterpenoids and carotenoids, however, are formed through the tail-to-tail condensation of FPP to FPP and GGPP to GGPP respectively. Higher molecular weight isoprenoids like *cis*-polyisoprene (natural rubber) are formed from the repeated head-to-tail addition of IPP to *trans*-FPP by *cis*-prenyltransferases (CPTs) [14]. **Figure 1.2** illustrates the general isoprenoid pathways generated from IPP.

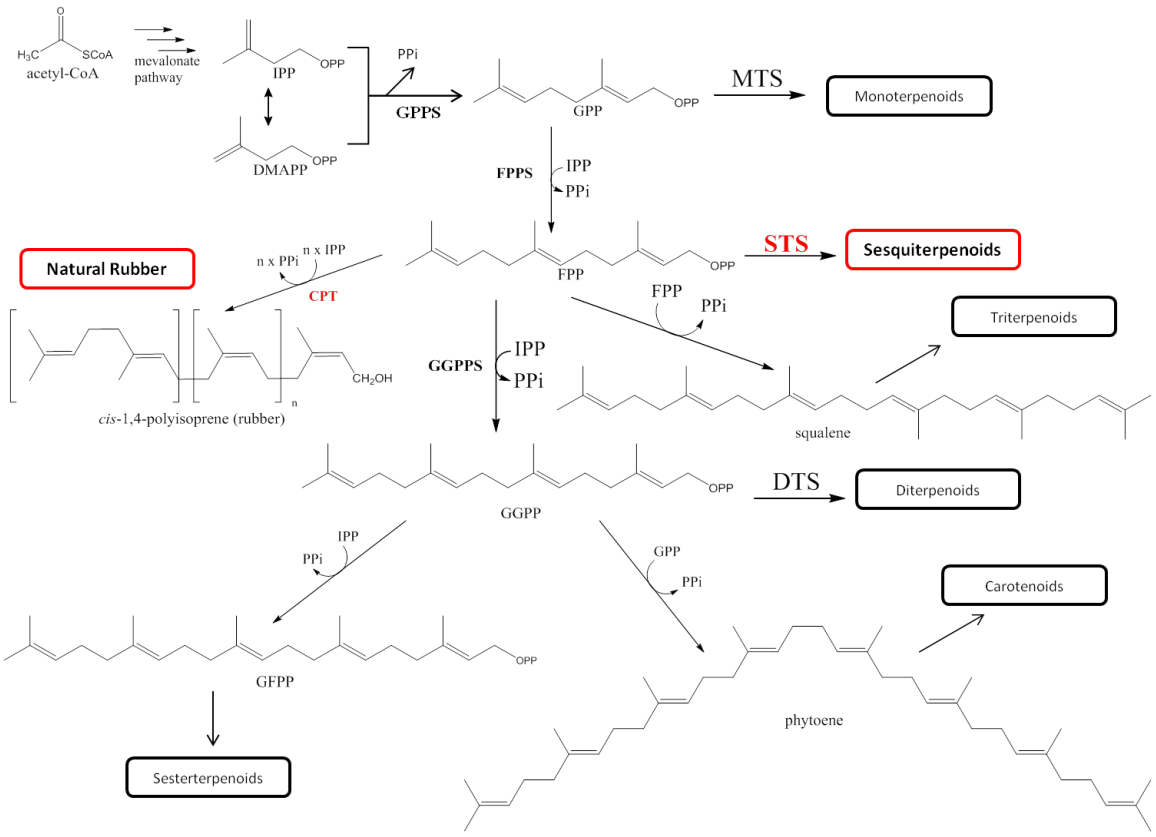


Figure 1.2 Potential isoprenoid pathways from repeated additions of isopentenyl pyrophosphate (IPP) to scaffold prenyl molecules. Addition of IPP to dimethylallylpyrophosphate (DMAPP) by geranyl pyrophosphate synthase (GPPS) yields geranyl pyrophosphate (GPP). GPP can then be converted by various monoterpene synthases (MTS) to monoterpenoids, or undergo a further IPP addition by farnesyl pyrophosphate synthase (FPPS) to yield farnesyl pyrophosphate (FPP). FPP has a number of possibilities for further processing. FPP can be cyclized by sesquiterpene synthases (STS) to sesquiterpenoids, combine with another FPP to make squalene (triterpenoid parent molecule), have another IPP addition by geranylgeranyl pyrophosphate synthase (GGPPS) to make geranylgeranyl pyrophosphate (GGPP), or have numerous additions of IPP by *cis*-prenyltransferases (CPT) to make natural rubber. GGPP can also be converted to diterpenoids by diterpene synthases (DTS), combine with another GGPP to make phytoene (carotenoid parent molecule), or undergo a further IPP addition by geranylfarnesyl pyrophosphate synthase (GFPP) to make geranylfarnesyl pyrophosphate (sesterterpenoid parent molecule).

The first step in modifying linear isoprenoid precursors to bioactive products occurs through their cyclization by a terpene synthase. As such, monoterpene synthases (MTS) cyclize GPP into respective monoterpenes, while sesquiterpene synthases (STS) cyclize FPP into sesquiterpenes, and diterpene synthases (DTS) cyclize GGPP into diterpenes.

Terpene synthases are separated into two classes, Class I or Class II, which are characterized by which catalytic fold they contain.

Terpene synthases, prenyltransferases and prenyl chain synthases mostly share one of two structural folds (C1-fold or C2-fold) and associated reaction mechanisms [14-17]. The C1-fold of isoprenoid biosynthetic genes is also known as the α-domain fold, and catalyzes the initial cleavage of diphosphate off the respective substrate via an ionization event through the coordination of divalent Mg^{2+} ions. These C1-fold enzymes contain two metal binding motifs at the active site entrance (aspartate-rich regions, D(D/E)xx(D/E) or NSE/DTE motifs) [14]. C2-fold enzymes are characterized by a double α-helical barrel fold, with the active site situated between the two helices. Catalysis in C2-fold enzymes occur through a protonation cascade event from an aspartate residue starting at a C-C double bond. This catalytic aspartate region is characterized by a DxDD aspartate-rich motif [14].

Class I terpene synthases contain the C1-fold and catalyze the cyclization of the prenyl precursor through the initial ionization event described above, while Class II terpene synthases contain the C2-fold and depend upon a protonation cascade for catalysis. Examples of Class I terpene synthases include most STSs, while many DTSs and TTSs are Class II terpene synthases [14].

The focus of this thesis will largely be on fungal sesquiterpene synthases (STSs). Thus, more emphasis will be put on the reaction mechanism of Class I terpene synthases. Class I terpene synthases react by first cleaving the diphosphate moiety from the particular linear pyrophosphate molecule to create a reactive carbocation as previously

described. This carbocation then undergoes an intramolecular electrophilic attack to form a cyclized intermediate. Depending upon the terpene synthase, this intermediate can undergo further intramolecular electrophilic attacks, hydride and methyl shifts, or other modifications before being quenched through a protonation event. This protonation event is mediated by the cleaved inorganic phosphate remaining bound in the active site, or through nucleophilic attack by water. The nature of the amino acids lining the hydrophobic active site created by the C1-fold determines the promiscuity of the terpene synthase and their relative cyclization mechanism and whether further secondary and tertiary reactions occur.

Individual STSs catalyze one or several of four known possible initial cyclizations of FPP. (2*E*, 6*E*)-farnesyl pyrophosphate (FPP) has two initial outcomes depending upon entering the active site of a particular STS. FPP is either directly ionized from a divalent metal ion to cleave pyrophosphate, or it is isomerized at the C2-C3 bond to make (3R)-nerolidyl pyrophosphate (NPP) followed by metal dependent ionization and pyrophosphate cleavage. If FPP is ionized directly, the resulting reactive carbocation can either perform an intramolecular electrophilic attack on either C10 or C11 of the C10=C11 double bond to respectively make (*E*, *E*)-germacradienyl (1, 10 cyclization) or *trans*-humulyl carbocations (1, 11 cyclization). FPP that was isomerized to NPP before ionization will either perform an intramolecular electrophilic attack on C10 of the C10=C11 double bond to make a (*Z*, *E*)-germacradienyl carbocation (1, 10 cyclization), or attack C6 of the C6=C7 double bond to make a bisabolyl carbocation (1, 6 cyclization) [18-20]. Theoretically, intramolecular electrophilic attack on the C7 of the C6=C7 double

bond to make a cycloheptanyl carbocation (1, 7 cyclization) is also possible, but has not been mechanistically demonstrated yet [21]. **Figure 1.3** illustrates the possible sesquiterpene synthase initial cyclization mechanisms followed by examples of secondary or tertiary cyclization products derived from those initial reactive carbocations.

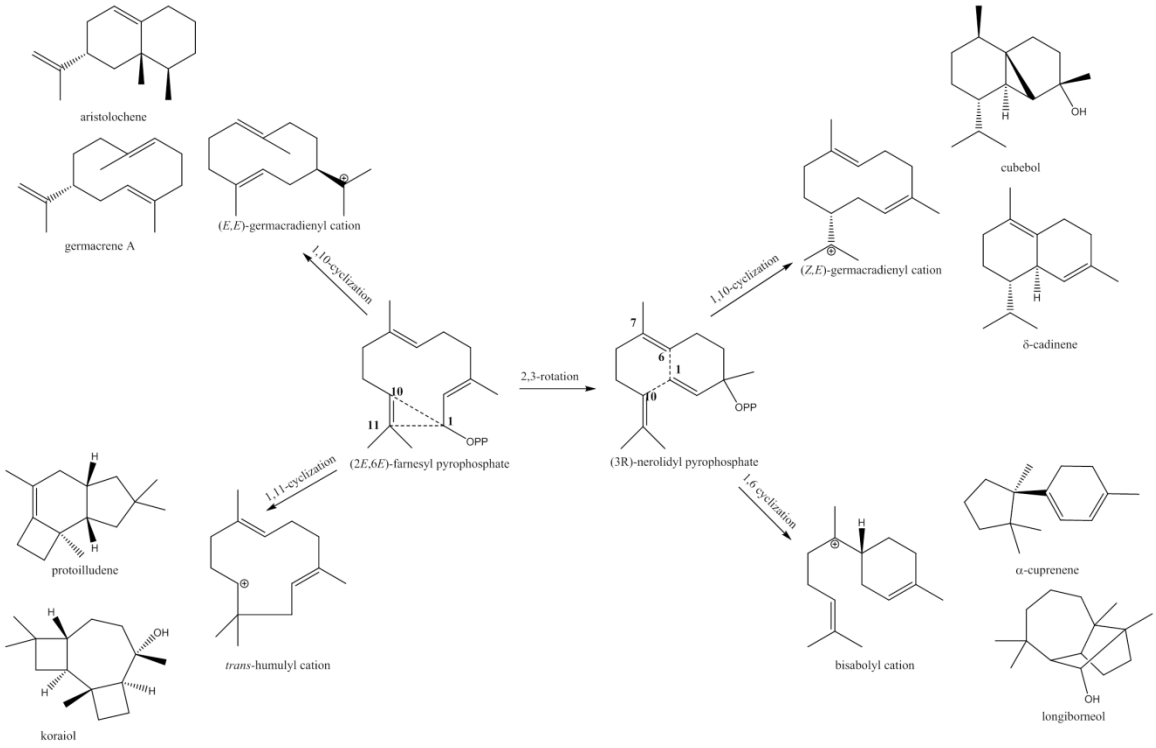


Figure 1.3 Known cyclization pathways of (2*E*,6*E*))-farnesyl pyrophosphate (FPP) within fungal sesquiterpene synthases. Upon direct pyrophosphate cleavage of FPP via ionization, the reactive carbocation can undergo a 1,10 or 1,11 cyclization via intramolecular electrophilic attack on the C10-C11 double bond. Conversely, FPP can first isomerize around the C2-C3 bond to make (3*R*)-nerolidyl pyrophosphate (NPP). After ionization and pyrophosphate cleavage of NPP, the reactive carbocation can undergo a 1,6 or 1,10 cyclization via the same intramolecular electrophilic attack on the C6-C7 or C10-C11 double bond. Further secondary or tertiary cyclizations can occur depending upon the hydrophobic cavity of the respective STS until the reaction is quenched with water or a proton.

Of the four initial cyclizations, the most bioactive and pharmaceutically relevant sesquiterpenoids are those derived from the 1,11-cyclizing *trans*-humulyl cation [22]. One *trans*-humulyl derivative of special note is the tertiary cyclization product

protoilludene. Protoilludene is a precursor of the anticancer compounds illudin M and illudin S [23]. This particular sesquiterpenoid scaffold is a major research interest of the Schmidt-Dannert laboratory, and will be described in a later section.

As previously mentioned, terpene synthases can have variable product promiscuity, largely dependent upon the hydrophobic residues lining the active site. Indeed, while some terpene synthases specifically make one final cyclization product, others may generate many different cyclization products. Some of these promiscuous enzymes may not even adhere to one particular initial cyclization mechanism [24].

3.2. Basidiomycota are an untapped natural products resource

Basidiomycota, or mushrooms, present a diverse but relatively uncharacterized natural product resource. The use of mushrooms for medicinal purposes has been well recorded throughout history [25-28], and an immense amount of credit for their medical applicability can be attributed towards their secondary metabolic profiles. Additionally, the vast diversity of distinct fungi outnumber that of plants by an order of magnitude [29]. However, given the complex life cycles, cultivation difficulties and general genetic intractability, very little attention has been given to the Basidiomycota phylum as a source for new enzymes and biosynthetic pathways [30].

With the advent of the decreasing genome sequencing costs, and the Department of Energy's 1000 Fungal Genomes project [31], the number of sequenced Basidiomycota genomes has exploded within the past decade. Currently, 174 Basidiomycota genomes are listed on the Joint Genome Institute (JGI) website, which is an increase in two orders of magnitude since 2005.

Another particular advantage with Basidiomycota is that their secondary metabolism profile is distinct from other organisms' secondary metabolome. Filamentous fungi (Ascomycota) and bacteria more often utilize polyketides (PKS) and non-ribosomal peptides (NRPS) within their secondary metabolite profiles, [32-35] while preliminary surveys of sequenced Basidiomycota genomes show more putative isoprenoid biosynthetic genes than those for PKSs or NRPSs. The following **Table 1.2** describes the breakdown of sequenced genomes and putative sesquiterpene synthase (STS) genes between each class of organism.

Table 1.2 Comparison of sesquiterpenoid content between classes of organisms.

Phylum Surveyed	Number of Genomes Surveyed	Sesquiterpene Synthases
Ascomycota (filamentous fungi)	80	250
Basidiomycota (mushroom forming fungi)	100	1200
Bacteria	2000	179 terpene synthases

Almost 1200 putative STSs were estimated from 100 Basidiomycota genomes, while only about 250 STSs were estimated from about 80 Ascomycota genomes. Additionally, only 179 putative terpene synthase genes in total were identified from almost 2000 sequenced bacterial genomes.

Further advantages in mining Basidiomycota genomes include the enriched amount of *trans*-humulyl derived sesquiterpenoids present almost exclusively in Basidiomycota, while also maintaining relatively small genomes. **Table 1.3** represents relative haploid genome sizes between various forms of life. Fungal genomes typically fall on the smaller end ranging from 9Mb to 1.5 GB[36, 37].

Table 1.3 Relative haploid genome sizes in nature.

Organism	Relative Genome Size Per Haploid Genome
Viruses	4.5-200 kb
Bacteria	0.3-10 Mb
Fungi	9 Mb-1.5 Gb
Protists	25 Mb-800 Gb
Plants	0.1-200 Gb
Insects	0.12-6 Gb
Mollusks	0.3-4.5 Gb
Bony Fish	0.4-5 Gb
Amphibians	0.6-120 Gb
Birds	0.7-1.5 Gb
Reptiles	1.5-4 Gb
Mammals	2.5-8 Gb
Cartilagenous Fish	3-10 Gb

However, the most beneficial feature for the rational mining of Basidiomycota genomes is their ability to physically cluster related biosynthetic genes together. This close tethering of genes provides an elegant means towards potentially elucidating entire biosynthetic pathways. It was for this reason, and for the other reasons explained in section 3.2 that the Schmidt-Dannert laboratory took interest in performing rational mining of Basidiomycota for STSs. **Figure 1.4** illustrates an example biosynthetic gene cluster surrounding a protoilludene synthase, Stehi7, from the wood-rotting fungus *Stereum hirsutum*.

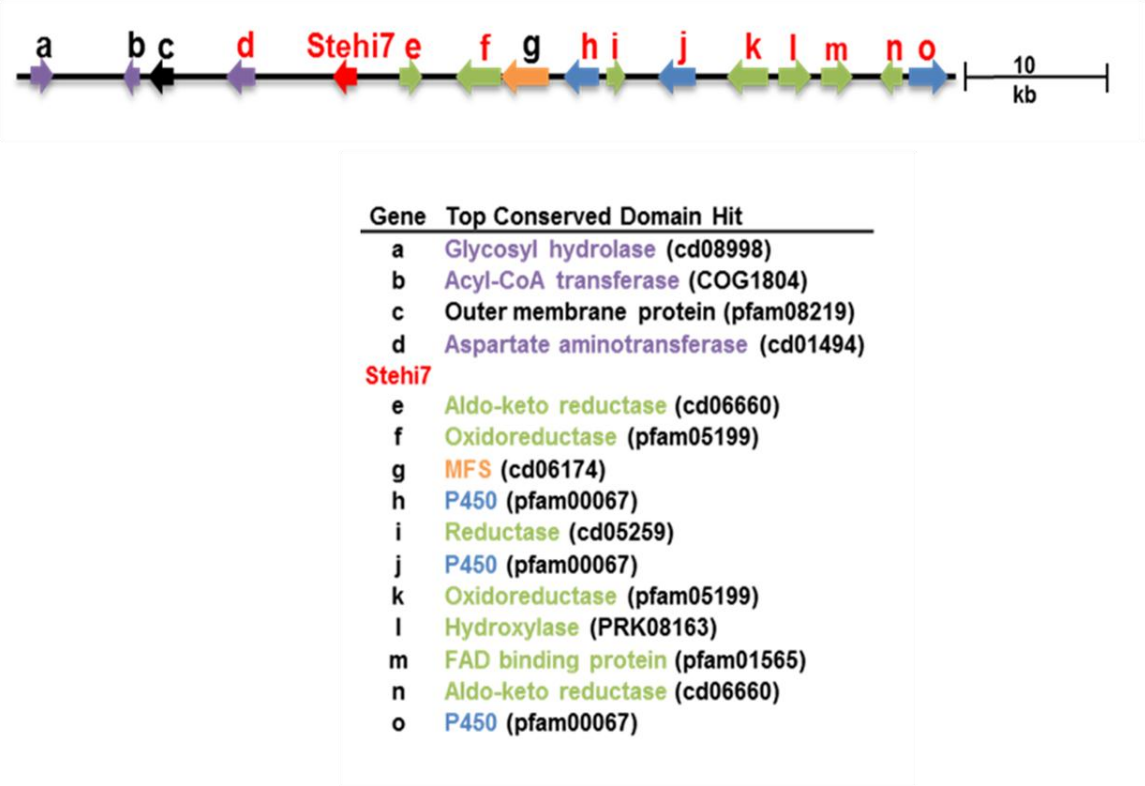


Figure 1.4 Biosynthetic gene cluster of sesquiterpene synthase Stehi7 from the wood-rotting Basidiomycete, Stereum hirsutum. Image taken from Quin et al. 2013 [38]**.**

3.3. Mining of Basidiomycota genomes for novel sesquiterpene natural products pathways

The Schmidt-Dannert laboratory initially began functional characterization of fungal STSs in 2005, where the only available mushroom genome sequence at the time was from the mushroom *Coprinus cinereus*, where considerable time was spent characterizing six of its STSs [39-41]. However, it was also known since 1950 that anticancer illudins were identified from the Jack O'Lantern mushroom, *Omphalotus olearius* [42]. Thus, one of the initial goals for the Schmidt-Dannert laboratory was to identify the genes responsible for the illudin biosynthesis. Two years were spent unsuccessfully attempting

to clone out putative *O. olearius* STSs with degenerate primers. It was not until genome sequencing became cheaper that the *O. olearius* genome could be affordably sequenced.

Upon sequencing the *O. olearius* genome, a diverse network of STSs and a number of biosynthetic gene clusters were surveyed [24]. By biochemically analyzing the putative STSs from *O. olearius*, two Δ6-protoilludene synthases were identified, the enzymes necessary for the initial step in illudin biosynthesis [24]. The broad sweeping biochemical analysis also developed a predictive framework to forecast an STS reaction mechanism based upon sequence conversion in the hydrophobic active site cavity. As previously mentioned, the nature of a terpene synthase reaction mechanism is largely predicated upon by the amino acids lining the active site cavity of the enzyme [14].

With a far more comprehensive predictive framework in place, the Schmidt-Dannert laboratory was able to perform large scale phylogenetic analysis of all predicted fungal STSs available from JGI. Large scale phylogenetic analyses of all available fungal STSs illustrate four distinct clades that coincide with their particular initial cyclization mechanisms [43]. **Figure 1.5** is a neighbor-joining dendrogram of previously characterized and putative STSs from Basidiomycota. This *in silico* genome mining method was utilized for directed discovery to further identify, and characterize novel STSs within the wood-rotting fungus *Stereum hirsutum*.

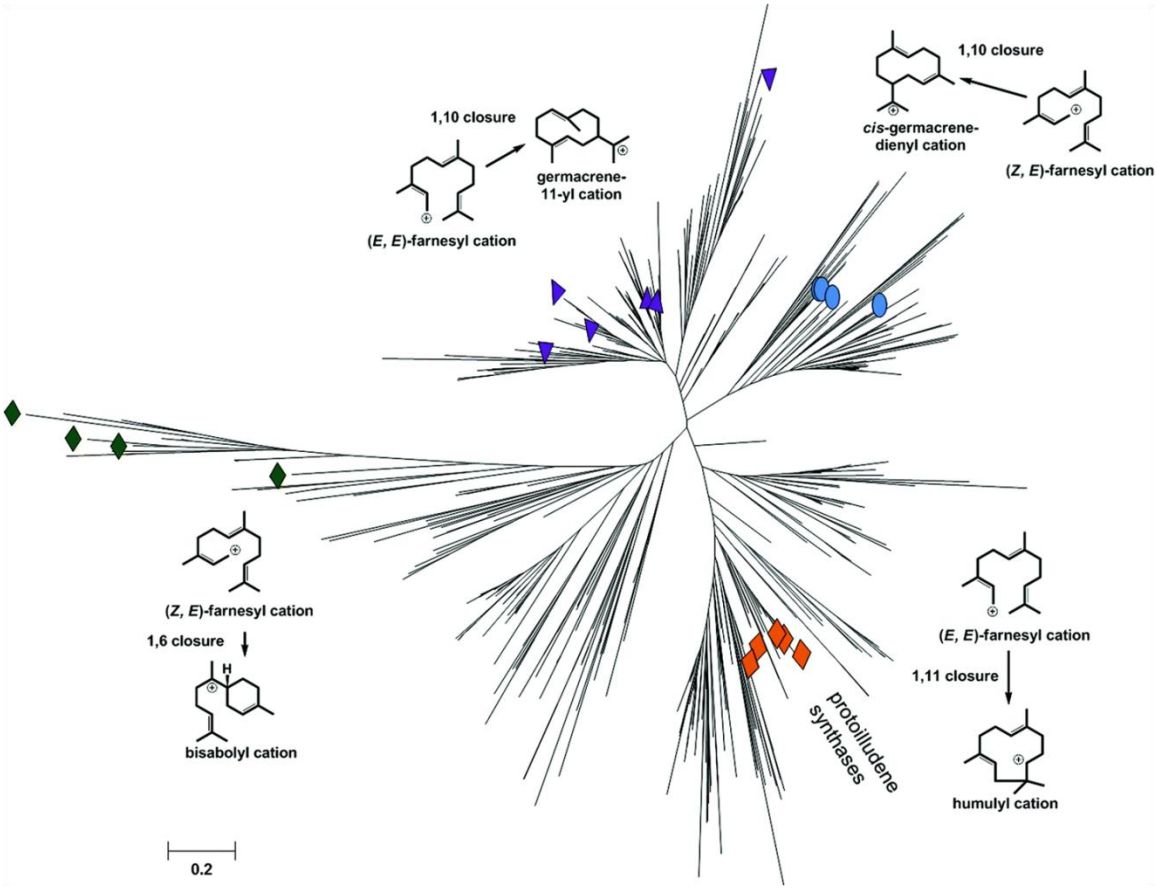


Figure 1.5 Neighbor joining dendrogram of predicted and characterized fungal sesquiterpene synthases (STS). Four distinct clades are present according to the cyclization mechanism. Additionally labeled are the protoilludene synthases, STSs important for the creation of the protoilludene scaffold, which is a precursor to important anticancer therapeutics illudin M and illudin S. Image taken from Quin et al. 2014 [43].

The *S. hirsutum* genome was mined particularly because preliminary bioinformatics analysis predicted an abundance of 1,11-cyclizing *trans*-humulyl derived STSs [38]. Indeed, biochemical analysis of expressed genes confirmed three Δ6-protoilludene synthases [38], along with representative STSs for 1,6 and *Z,E*-1,10 cyclizations. **Table 1.4** illustrates all biochemically characterized STSs from *C. cinereus*, *O. olearius* and *S. hirsutum* by the Schmidt-Dannert laboratory [24, 38, 40, 41, 43-46].

Table 1.4 Previously identified products of sesquiterpene synthases of *C. cinereus, O. olearius and S. hirsutum*

Enzyme	Identified Products	Predominant Product Cyclization	Symbol on Dendrogram
Cop1	germacrene A (β-elemene)	E,E-1,10	Purple Triangles
Cop2	germacrene A (β-elemene)	E,E-1,10	Purple Triangles
Cop3	γ-muurolene	E,E-1,10	Purple Triangles
Cop4	δ-cadinene	Z,E-1,10	Blue Circles
Cop5	putative pentalanene synthase	1,11	Not Shown
Cop6	α-cuprenene	1,6	Green Diamonds
Omp1	α-muurolene	E,E-1,10	Purple Triangles
Omp3	germacrene A (β-elemene), selina-4,7-diene, α-muurolene, δ-cadinene	E,E-1,10	Purple Triangles
Omp4	δ-cadinene	Z,E-1,10	Blue Circles
Omp5a	γ-cadinene, epi-zonarene, germacrene A (β-elemene)	Z,E-1,10	Blue Circles
Omp5b	γ-cadinene, germacrene A (β-elemene)	Z,E-1,10	Blue Circles
Omp6	Δ6-protoilludene	1,11	Orange Squares
Omp7	Δ6-protoilludene, pentalanene	1,11	Orange Squares
Omp9	α-barbatene, β-barbatene	1,6	Green Diamonds
Omp10	daucene, trans-dauca-4(11),8-diene	1,6	Green Diamonds
Stehi1|159379	β-barbetene, α-barbetene, α-cuprenene	1,6	Green Diamonds
Stehi1|128017	δ-cadinene, β-copaene, sativene, γ-muurolene, β-cubebene, germacrene D, sesqui	Z,E-1,10	Blue Circles
Stehi1|25180	Δ6-protoilludene	1,11	Orange Squares
Stehi1|64702	Δ6-protoilludene, germacrene A (β-elemene)	1,11	Orange Squares
Stehi1|73029	Δ6-protoilludene, germacrene A (β-elemene)	1,11	Orange Squares

As shown in **Table 1.4**, therein lays numerous and diverse types of STSs present within 3 fungal genomes alone, which further exemplify the untapped resource that Basidiomycetes represent. Furthermore, while some of the STSs and their associated biosynthetic gene clusters encode pathways for previously isolated bioactive compounds from these fungi, the function of other STSs and/or STS pathway products remain to be characterized.

3.4. *Lactarius* mushrooms produce polyisoprene

Natural rubber, or *cis*-1,4-polyisoprene is a vastly important resource for the modern industrial world, particularly because of its superior qualities to synthetic rubber [47]. While polyisoprene was known to be produced in the latex of over 2500 species of plants, the production of polyisoprene was not proven in Basidiomycota until 1994, when Tanaka et al. demonstrated polyisoprene synthesis in 5 different *Lactarius* (also known as "Milk-Cap" mushrooms) species [48]. A more comprehensive review on this subject can be found in Chapter 4, detailing our collaboration with Bridgestone Tires Inc. (Nashville,

TN) towards the characterization of natural rubber biosynthetic genes in *Lactarius* mushrooms.

3.5. Summary of thesis goals

It was the aim of this thesis to fully understand the cyclization mechanism and variable product specificity of bioactive Δ6-protoilludene synthases. Better understanding of the product specificity, and the 1,11-cyclization mechanism was to be achieved through solving the three-dimensional structure of at least one of the *S. hirsutum* Δ6-protoilludene synthases via X-ray diffraction analysis. Chapter 2 covers the ongoing progress towards the crystallization of these three Δ6-protoilludene synthases.

The other major aim of this thesis was to build upon the foundation set forth by the previous established *in silico* genome mining analysis by the Schmidt-Dannert laboratory to identify novel STSs within 3 *Lactarius* mushroom genomes. Chapter 3 covers the terpene screening of 8 *Lactarius* cultures and the subsequent genome assembly and mining of *Lactarius uvidus* for novel STSs. Chapter 4 is an extension of this process to identify *cis*-prenyl transferases (CPTs) and farnesyl pyrophosphate synthases (FPPSs) involved with natural rubber biosynthesis in the genome assemblies of *L. rufus* and *L. vinaceorufescens.* STSs were also identified in these genome assemblies. This project was an industrial collaboration with Bridgestone Tires Inc. (Nashville, TN).

Chapter 2: Towards the crystallization of the protoilludene synthase, Stehi7

1. Introduction

1.1 Cyclization mechanisms of sesquiterpene synthases

As previously discussed in chapter 1, terpene synthases generally convert linear prenyl compounds into cyclic products of varying complexity. Sesquiterpene synthases (STSs) in particular are capable of 4 different initial cyclization mechanisms [14]. The possible confirmed cyclizations include a 1,11-cyclization of E,E-FPP, a 1,10-cyclization of E,E-FPP, a 1,10-cyclization of NPP or a 1,6-cyclization of NPP [24, 38, 43].

Additional electrophilic attacks, carbocation rearrangements, methyl and hydride shifts, and secondary/tertiary cyclizations are all possible after these initial cyclizations, but the additional modifications to the scaffold are quenched with a deprotonation event mediated by residual inorganic phosphate in the active site, or nucleophilic attack with water [14].

Again, these processes are mediated by the nature of the amino acids found lining the hydrophobic enzyme active site cavity. The Schmidt-Dannert lab has been able to demonstrate that the sequence conservation of these amino acids dictate which type of initial cyclization will occur for the major product of a given STS [38].

1.2 Previously crystallized STSs and active site motifs

There has been much interest in better understanding the conservation of amino acid sequence and predicted cyclization mechanisms of sesquiterpene synthases through

structural analysis. This is evidenced by the solved sesquiterpene structures from several plant, fungal and microbial hosts via X-ray diffraction analyses [49-60]. **Table 2.1** illustrates the unique solved crystal structures of all sesquiterpene synthases to date [61].

Table 2.1 Solved crystal structures of sesquiterpene synthases. Plant source organisms are green and have a (-) symbol, bacterial source organisms are purple and have a (+) symbol, and fungal source organisms are orange and have a (*) symbol.

Year	Enzyme	Source Organism	Cyclization Mechanism (Confirmed or Inferred)	PDBID
1997	5-epi-aristolochene synthase	*Nicotania tabacum (-)*	E,E-1,10	5EAS
1998	Pentalanene synthase	*Streptomyces exfoliatus (+)*	1,11	1PS1
2000	Aristolochene synthase	*Penicillium roqueforti (*)*	E,E-1,10	1DI1
2001	Trichodiene synthase	*Fuserium sporotrichoides (*)*	1,6	1JFA
2007	Aristolochene synthase	*Aspergillus terreus (*)*	E,E-1,10	2E4O
2010	Epi-isozizanene synthase	*Streptomyces coelicolor (+)*	1,6	3KB9
2011	α-bisabolene synthase	*Abies grandis (-)*	1,6	3SAE
2013	α-bisabolene synthase	*Artemisia annua (-)*	1,6	4FJQ
2014	Hedycaryol synthase	*Kitasatospora setae (+)*	Z,E-1,10	4MCO
2014	Selinadiene synthase	*Streptomyces pristinaespiralis (+)*	E,E-1,10	4OKM

As seen in **Table 2.1,** trichodiene synthase from *Fusarium sporotrichoides* and aristolochene synthases from *Penicillium roqueforti* and *Aspergillus terreus* are the only solved crystal structures of fungal sesquiterpene synthases. Also noteworthy is the lack of 1,11-cyclizing STSs, with pentalanene synthase from *Streptomyces exfoliatus* being the only representative. As previously stated, most medicinally relevant sesquiterpenes isolated from Basidiomycota are of the 1,11 cyclized *trans*-humulyl cation-derived variety[43]. Of special interest are the *trans*-humulane-protoilludane type sesquiterpenes, as their derivatives hold promise as anticancer, antifungal and antibiotic agents [24].

1.3 Previous work for crystallization of Omp6

The Schmidt-Dannert lab previously set out to obtain a crystal structure of the highly active and specific protoilludene synthase, Omp6, from the mushroom *Omphalotus olearius* [62]. This was done to improve the representation of *trans*-humulyl cation synthesizing known STS structures, and in addition facilitate the understanding of the

cyclization mechanism and product-specificity of protoilludene synthases. **Figure 2.1** below is a structural model of Omp6 using pentalanene synthase from *Streptomyces exfoliates* as the template [63].

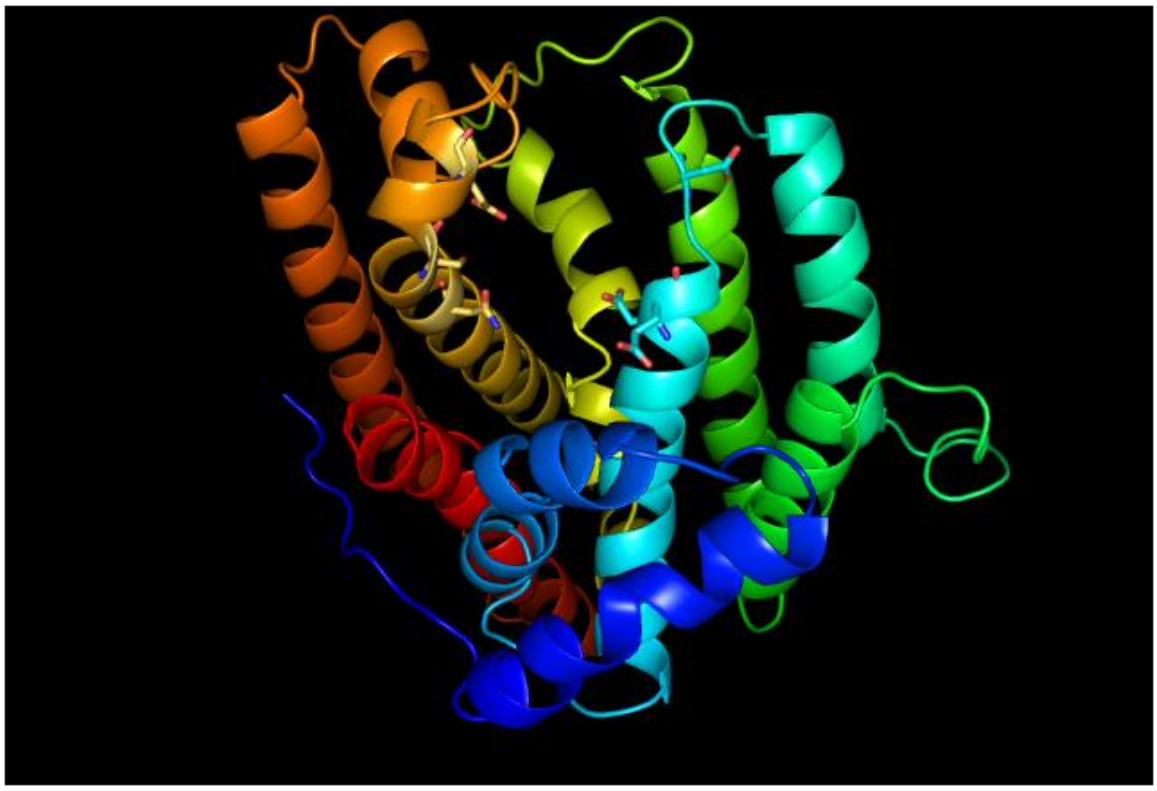


Figure 2.1 Structural model of Omp6 from *Omphalotus olearius* with conserved D(D/E)xx(D/E) and NSE/DTE motifs shown as sticks. Omp6 was modeled from N219L pentalanene synthase from *Streptomyces exfoliatus* (PDB 1HM7). Pentalanene synthase is the only such 1,11-cyclizing sesquiterpene synthase with a solved structure.

The Schmidt-Dannert laboratory was able to successfully obtain 100 x 50 x 20μm crystals of Omp6 in 0.01M $ZnCl_2$, 20% (w/v) PEG 6000, 0.1M MES pH 6.0 from manual hanging-drop vapor-diffusion trays. Diffraction data of these crystals could be collected down to a resolution of 2.9 Å to allow for a native data set to be collected. However, this resolution was not sufficient to solve the crystal structure by molecular replacement. Additionally, manual hanging drop crystals in the above listed conditions were unable to

be reproduced to allow for suitable crystals to solve via single-wavelength anomalous dispersion (SAD) phasing, despite exhaustive attempts [62].

An added complexity of this study was the prevalence of false positive salt crystals caused by the presence of extraneous metal ions in purified protein samples and reagents. This led to the tedious requirement of washing and rinsing all glassware involved in the purification of Omp6 with 10% nitric acid to strip away extraneous metal ions. Additionally, purified Omp6 was dialyzed against 5mM EDTA and then reagent grade water to ensure no metal ions were present within the purified protein sample itself. Furthermore, all crystallization reagents were required to be prepared in reagent grade water.

When the genome of *Stereum hirsutum* was mined as mentioned in Chapter 1, three homologues to Omp6 with high sequence identity were identified [38]. The three homologues were Stehi1|25180 (Stehi2), Stehi1|64702 (Stehi6), and Stehi1|73029 (Stehi7). With Omp6 strictly producing Δ6-protoilludene, it was initially hypothesized that the three *Stereum* homologues would elicit the same behavior. Surprisingly, both Stehi6 and Stehi7 were shown to be capable of dual cyclization mechanisms, when low levels of β-elemene (a Cope rearrangement of the 1,10 cyclization product germacrene A) were present in addition to Δ6-protoilludene in heterologous expression assays.

It was my goal in this study to obtain a solved crystal structure of Stehi2, Stehi6, or Stehi7 in order to investigate the subtleties of protoilludene synthases fidelity, and potentially succeed where the Omp6 crystallization was unsuccessful. Furthermore, obtaining a solved structure of a protoilludene synthase structure would be the first

solved protoilludene synthase structure, and only the second 1,11-cyclizing STS representative solved so far [64].

2. Materials and Methods

2.1 Stehi1|25180, Stehi1|64702 and Stehi1|73029 Expression and Purification

Respective pET28a constructs of N-terminally His-tagged Stehi1|25180 (Stehi2), Stehi1|64702 (Stehi6) and Stehi1|73029 (Stehi7) with 19 amino acid linker gene inserts were designed and constructed by Dr. Maureen Quin of the Schmidt-Dannert laboratory. *Escherichia coli* Rosetta™ (DE3) pLysS cells (Novagen, Madison, WI) were transformed with the respective pET28a constructs. Single transformant colonies were picked and inoculated into 50mL LB liquid starter cultures supplemented with 50μg mL^{-1} chloramphenicol and 30μg mL^{-1} kanamycin, and incubated overnight at 37^{o}C with shaking. 500mL LB liquid cultures, supplemented with 50μg mL^{-1} chloramphenicol and 30μg mL^{-1} kanamycin were used for large scale protein production. Isopropyl β-D-thiogalactopyranoside (IPTG) was added to each culture to a concentration of 0.5mM for induction after growing cultures with shaking at 30^{o}C to an OD_{600} of 0.6. Cultures were allowed to incubate 3 hours longer at 30^{o}C before each culture headspace was sampled for GC-MS analysis to verify protoilludene synthase activity.

Culture headspace was sampled for 10 minutes via solid phase micro extraction (SPME) using a 100μm polydimethylsiloxane fiber. Headspace sampling was achieved by inserting the SPME fiber through an aluminum foil seal on top of each culture flask for adsorption of the culture gas phase. Post adsorption, the SPME fiber was inserted into

the injection port of the GC-MS for thermal desorption. Further GC-MS details will be described in the section below.

Cells were centrifuged at 4,000rpm for 20min at 4^{o}C upon verification of Δ6-protoilludene synthase activity. All glassware involved with purification were thoroughly washed and rinsed with 10% nitric acid and reagent grade water (ThermoFisher Scientific, Waltham, MA) prior to cell lysis. Additionally, all purification buffers were made in reagent grade water. Resulting cell pellets were resuspended in 10mL of Buffer "A" (50mM tris pH 7.5, 50mM NaCl, 5mM imidazole), and sonicated for a total of 4 minutes in 1s on/2s off pulses at 30% amplitude. The resulting soluble protein fraction was separated from the cell slurry via centrifugation for 30min at 12,000 rpm at 4^{o}C. Soluble protein fractions were added to 2mL Ni-NTA resin pre-equilibrated with 3x4mL washes of Buffer A, and allowed to mix for 1hr at 4^{o}C. After mixing, remaining unbound protein was collected in fraction X1, while fractions X2 and X3 respectively corresponded to collections of a 10mL wash of the resin with buffer A and a 5mL wash with buffer "B" (50mM Tris pH 7.5, 50mM NaCl, 250mM imidazole) . Fractions were analyzed via 15% SDS-PAGE gels to determine expression levels and purity.

2.2 GC-MS Analysis

GC-MS analysis was performed on an HP GC 7890A instrument coupled to an anion-trap mass spectrometer HP MSD triple axis detector. Volatile compounds were allowed 15min for desorption at the injection port. After desorption, volatile compounds were carried via helium gas through a HP 5MS Capillary Column (30m x 0.25mm x 1.0μm). Mass spectra were recorded in electron impact ionization mode with scans in the range of

5-300amu at 1 second intervals. The oven temperature began at 100^{o}C and was increased to a final temperature of 250^{o}C at a rate of 10^{o}C per minute.

2.3 Stehi7 Purification

Large scale expression of *E. coli* Rosetta (DE3) pLysS, transformed with pET28a-Stehi7, was performed as previously described. Again, all glassware involved in Stehi7 purification was thoroughly washed and rinsed with 10% nitric acid and reagent grade water, and all reagents were prepared in reagent grade water. Purification of Stehi7 diverged from the above stated purification method after the soluble protein fraction was obtained from sonication and centrifugation. Soluble protein found in the supernatant of cell lysate was bound to a HisTrap FF Column (GE Healthcare, Cleveland, OH) in the presence of buffer A, and eluted in buffer B on an Akta FPLC (GE Healthcare, Cleveland, OH). Combined protein fractions were concentrated to a volume of 500uL via centrifugal filtration, and concentrated protein was passed over an S200 size exclusion purification column (GE Healthcare, Cleveland, OH) pre-equilibrated with buffer C (20mM Tris pH 7.5, 50mM NaCl, 1mM β-mercaptoethanol, 10% (v/v) glycerol). Combined protein fractions were dialyzed overnight into buffer D (20mM Tris pH 7.5). Post-dialysis, protein was bound to a MonoQ column (GE Healthcare, Cleveland, OH) in the presence of buffer D, and eluted with buffer E (20mM Tris pH 7.5, 1M NaCl). Combined protein fractions were dialyzed into 50mM EDTA at RT for 1hr before being dialyzed into reagent grade water overnight at 4^{o}C and stored at 4^{o}C indefinitely.

2.4 *In vitro* Stehi7 activity assay

A 100µL reaction containing 100µg of Stehi7, 10mM $MgCl_2$ and 2µM (2E,6E)-farnesyl pyrophosphate was incubated in a sealed glass GC-MS vial for 1 hour at RT. The headspace of the vials was sampled with an SPME fiber for 10min prior to GC-MS analysis as described above.

2.5 Crystallization Trials

Purified Stehi7 was concentrated down to 10mg/ml, and a 150µL sample was submitted to the University of Minnesota Nanoliter Crystallization Facility. At the facility, crystallization trials were performed using the sitting-drop vapor-diffusion technique on a Rigaku CrystalMation automated system. Stehi7 was subjected to the following high throughput crystallization screens: Crystal Screen HT, Index, PEG/Ion HT and SaltRx surveying a total of 384 different crystallization conditions. Trials were performed in 100µL wells with 100nL of protein combined with 100nL crystallant. Plates were sealed with clear sealing film, and incubated at 4^oC. Results from the automated screens were published on a web interface using in house CrystalTrak software.

A subset of the four most promising crystallization conditions was selected for manual crystallization screens using the hanging-drop vapor-diffusion crystallization technique in 24-well trays. Each well contained 1mL of crystallant solution. On a microscope cover slide, 1µL of crystallant was combined with 1µL of purified Stehi7, and 1µL crystallant was combined with 1µL reagent grade water as a negative control. The microscope slide was placed inverted above the well containing the crystallant,

creating an airtight seal with high vacuum grease. Completed trays were incubated at 4^{o}C or 18^{o}C in perpetuity upon completion (see **Supplemental Table S2.1** for complete list of manual crystallization assays and conditions).

2.6 Microseeding Crystallization Trials

A subset of manual hanging-drop trays were established using microseed stocks of crystalline precipitate from previously completed manual hanging-drop trays using a Seed Bead Kit (Hampton Research, Mission Viejo, CA) with an adaptation of the recommended protocol. Wells containing quality crystalline precipitate were selected for seed stocks and the microscope slides containing the crystallization drops were removed from the top of the well. 50µL of crystallant was removed from the well and placed in a 1.5mL eppendorf tube containing a seed bead. A glass probe was then used to crush crystal material found within the crystallization drop. 5µL of crystallant from the seed bead tube was used to resuspend crushed crystal material, and returned to the seed eppendorf tube. The resuspension process was repeated two more times. The eppendorf tube containing the suspension and seed bead was vortexed at the highest setting for a total of 2 minutes in 30s pulses with 1minute breaks on ice. 4-fold and 10-fold dilutions were performed from the vortexed seed stock. New crystallization drops were formed using 1µL crystallant, 1µL purified Stehi7, and 0.5µL seed stock (or one of its dilutions) on a microscope slide and sealed inverted over its respective well. Trays were either incubated at 4^{o}C or 18^{o}C indefinitely upon completion.

2.7 Co-crystallization Trials

Another subset of manual crystallization assays were created involving co-crystallization with 2-12mM NaPP, 1-100mM $MgCl_2$, 10mM $MnCl_2$, 10mM $CaCl_2$ and 10mM $ZnCl_2$ in addition to the established crystallization conditions. Trays again were incubated at either 4^{o}C or 18^{o}C indefinitely upon completion (see **Supplemental Table S2.1** for more details).

2.8 Macroseeding Crystallization Trials

A further subset of hanging-drop crystallization trays were established using macroseeding techniques from human and feline hosts. For human host macroseeding, 0.16g of human facial hair was ground in liquid nitrogen into a crude powder, and resuspended in reagent grade water. The crude powder resuspension was further sonicated at 30% amplitude in 1s pulses with a 2s break for a total pulse time of 30s. The sonicated powder resuspension was then diluted by a factor of five. Crystallization drops for the hanging drop trays included 1µL crystallant, 1µL protein and 0.5µL ground beard suspension.

For feline host macroseeding, one feline whisker was cut into ~0.5mm pieces using a sterile scalpel. Whisker pieces were dipped in a 5µL drop of freshly purified Stehi7 and placed in a 2µL drop already containing 1µL protein and 1µL crystallant. Trays were incubated at 4^{o}C indefinitely upon completion.

2.9 Circular Dichroism

Purified Stehi7 was dialyzed into HEPES buffer (20mM HEPES, 50mM NaCl, 5% glycerol pH 7.5) prior to CD measurements. Post-dialysis, Stehi7 (15μM) was separated into 200μL aliquots and pre-incubated with $CaCl_2$ or $MgCl_2$ (1mM), or reagent grade water as a control. Protein was loaded into a 1mm pathlength cuvette. CD spectra were obtained using a JASCO J-815 spectropolarimeter, and data was collected in 1nm increments from 200-260nm.

2.10 Truncation Cloning and Expression

The amino acid sequence of Stehi7 was systematically truncated down by one amino acid at the N-terminus *in silico*, with the resulting truncations used as input onto the Xtal-Pred web server [65], a protein crystallization prediction program. It was predicted that truncating the first 41 amino acids off of Stehi7 would allow for higher crystallization likelihood than full length Stehi7. Stehi7 (42-369) was sub-cloned from pET28a-Stehi7, and into empty pET28a vector using NdeI and EcoRI restriction sites. Transformation, expression, and purification of *E. coli* Rosetta (DE3) pLysS cells with pET28a-Stehi7(42-369) were performed as previously described for pET28a-Stehi7.

3 Results and Discussion

3.1 Stehi7 has better soluble heterologous expression than Stehi2 or Stehi6 and exhibits higher Δ6-protoilludene synthase activity

In order to fully understand the reaction cyclization mechanism and product specificity of the predicted Δ6-protoilludene synthases, pET28a constructs of N-terminal His-tagged Stehi2, Stehi6, and Stehi7 were heterologously expressed in *E. coli* Rosetta (DE3) pLysS cells. This large scale protein production was performed to investigate potential downstream crystallization viability. The first experiment to perform on these heterologously expressed proteins was to verify the predicted heterologous Δ6-protoilludene activity post-induction in recombinant *E. coli*. The gas phase of the recombinant culture headspaces were sampled using solid phase micro extraction (SPME) and the resultant volatile compounds were identified using GC-MS analysis. As shown in **Figure 2.2**, all three heterologously expressed cultures were confirmed to make Δ6-protoilludene, with Stehi7 as the most active sample. The starred peak in each chromatogram is indole, which is naturally produced in *E. coli* as a breakdown of tryptophan, and was used as an internal standard. Interestingly, low level peaks of β-elemene were not observed as previously found with Stehi6 and Stehi7.

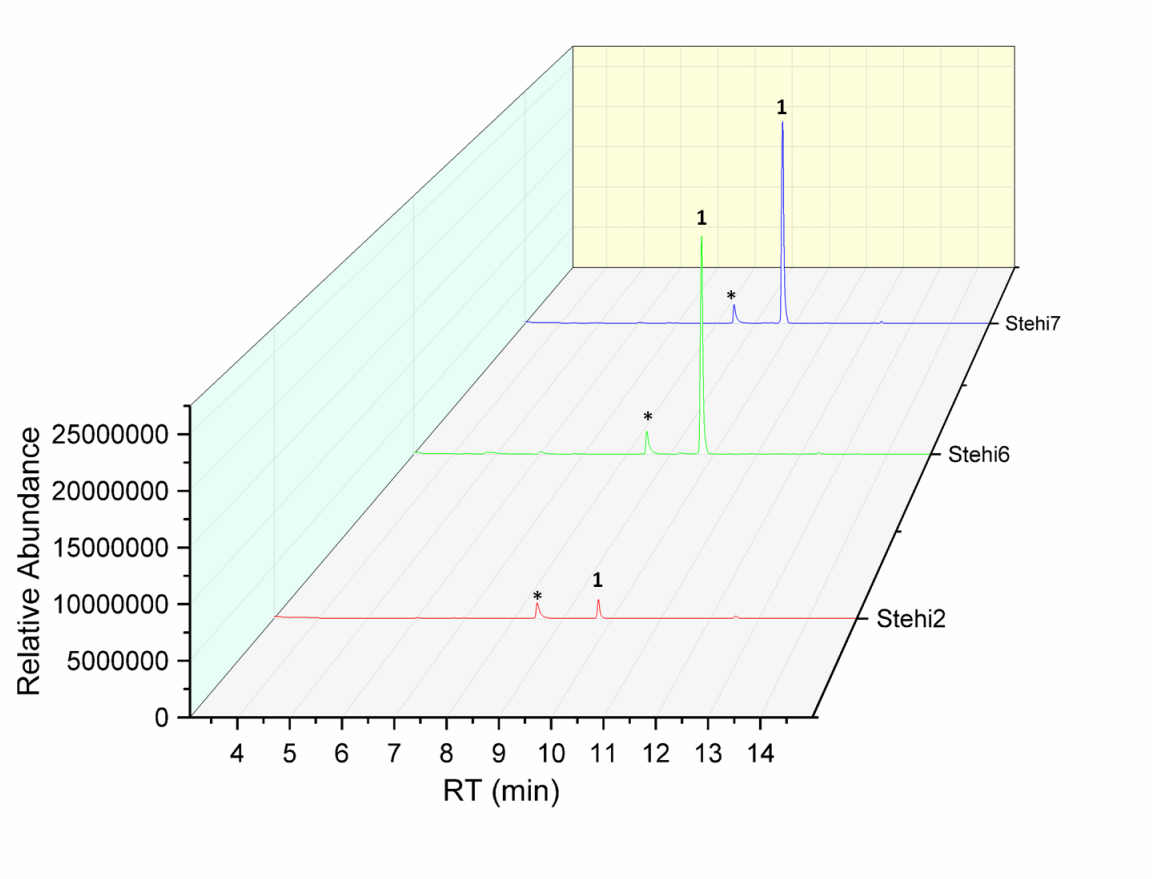


Figure 2.2 GC-MS chromatogram of heterologously expressed Stehi2, Stehi6 and Stehi7. Starred (*) peaks are identified as indole, a natural breakdown product of tryptophan in *E. coli*. Peaks labeled as (1) are identified as Δ6-protoilludene.

Upon verification of Δ6-protoilludene synthase activity, cultures were harvested for protein purification. Fractions of the His-tagged purification of these heterologously expressed proteins via Ni-NTA resin were viewed on a 15% SDS-Gel shown in **Figure 2.3**.

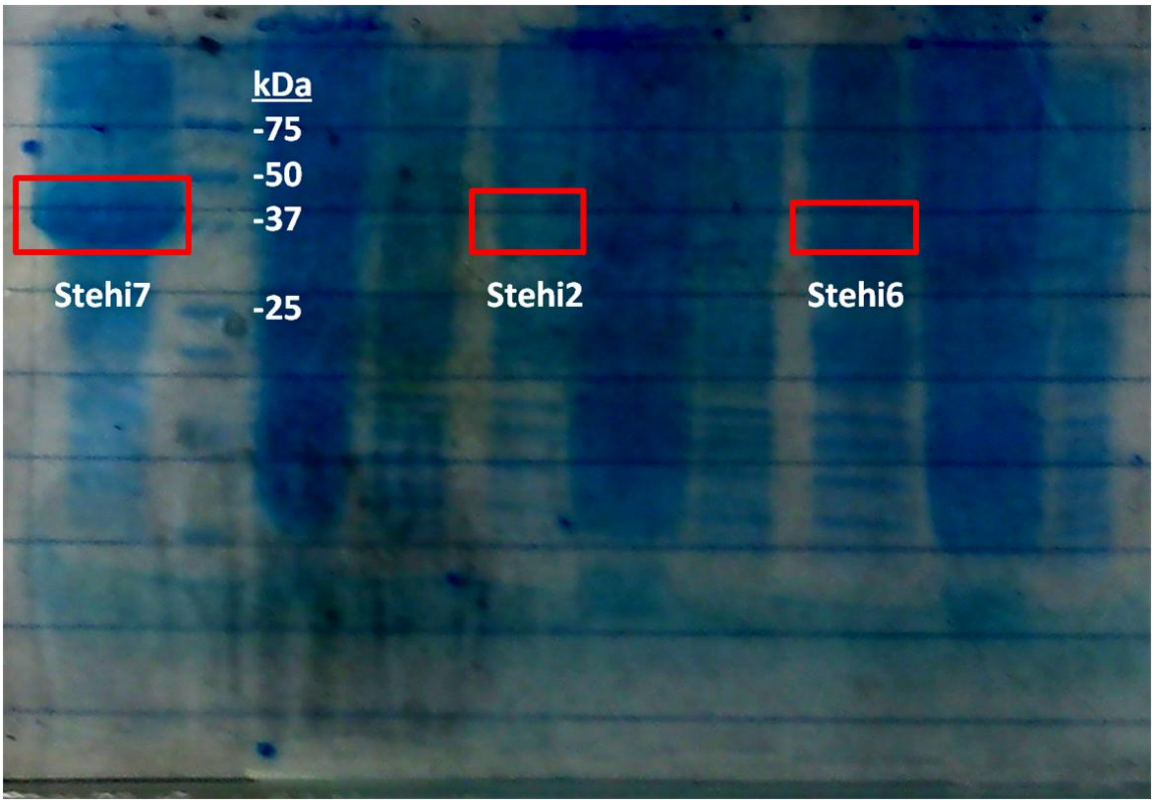


Figure 2.3 15% SDS-PAGE gel of His purified fractions of Stehi2, Stehi6, and Stehi7. Fractions were not particularly pure, and Stehi2 and Stehi6 yields were difficult to distinguish.

Though Stehi2 and Stehi6 protein bands were present at the expected sizes, the immense quantity of expressed and purified Stehi7 dwarfed the quantity of purified Stehi2 and Stehi6. Given proteins generally crystallize at concentrations of 10mg/mL [66] or more, it became abundantly clear that Stehi7 was the best candidate protein of the three to use for downstream crystallization trials.

3.2 Extensively purified Stehi7 maintains Δ6-protoilludene synthase activity *in vitro,* and is in sufficient quantities for crystallization trials

Upon selecting Stehi7 as the best candidate protein to crystallize, further purity of Stehi7 would need to be obtained. Thus, large scale expression of Stehi7 in *E. coli* Rosetta (DE3) pLysS cells was repeated. Extensive large scale purification of Stehi7 was

performed through three different rounds of purification on an Akta FPLC. The first round of Stehi7 was through His purification on a HisTrap FF column. The second round of purification was through size-exclusion chromatography on an S200 column, and the final round of purification was anion-exchange purification through a MonoQ column. **Figure 2.4** illustrates the respective fractions of Stehi7 through the multiple rounds of purification on an SDS-gel, highlighting the high enough yield for crystallization. The three rounds of purification yielded 1.5mg purified protein from 500mL culture.

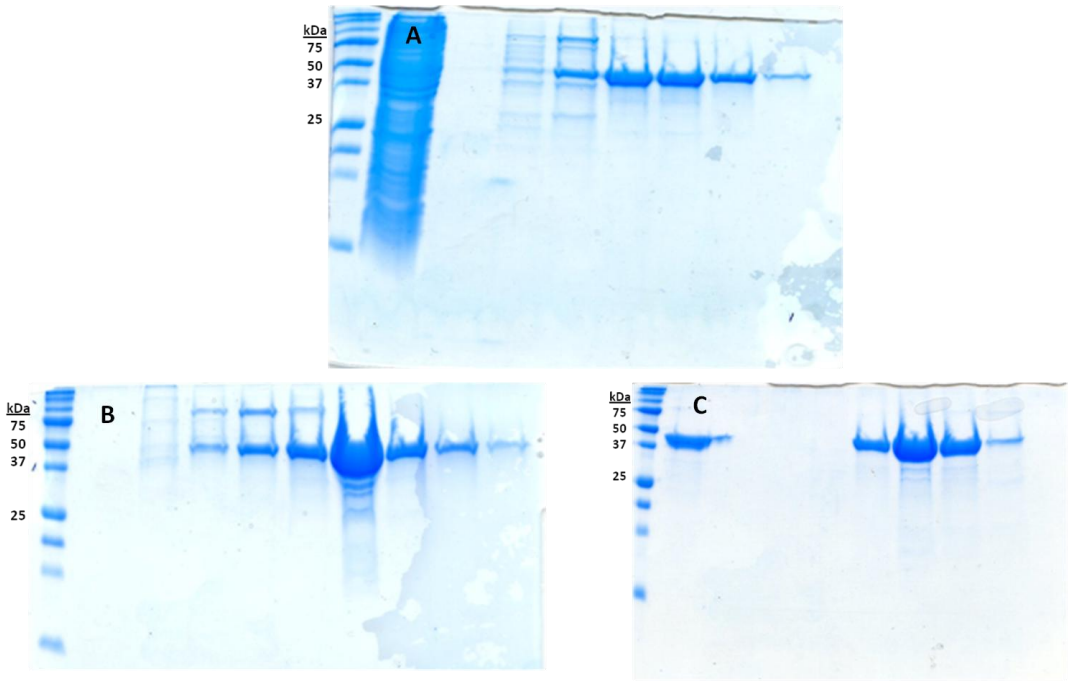


Figure 2.4 Fractions from the purifications of Stehi7 from *E. coli* Rosetta (DE3) pLysS cells. Corresponding gels are (**A**) eluted fractions from His affinity purification, (**B**) eluted fractions from S200 size exclusion purification, and (**C**) eluted fractions from MonoQ anion exchange purification. Bands at 40kDa correspond to Stehi7.

It was also important to verify the activity of Stehi7 after purification for two reasons. The first reason was to completely verify that the purified protein was in fact Stehi7. The second reason was to verify all metal ions had been sufficiently stripped from purified

Stehi7, as previous attempts to crystallize Omp6 had been made difficult by the presence of co-purified metal ions (unpublished data). Activity of purified Stehi7 was confirmed with its incubation with (E,E)-FPP and $MgCl_2$ and subsequent GC-MS analysis as seen in **Figure 2.5**.

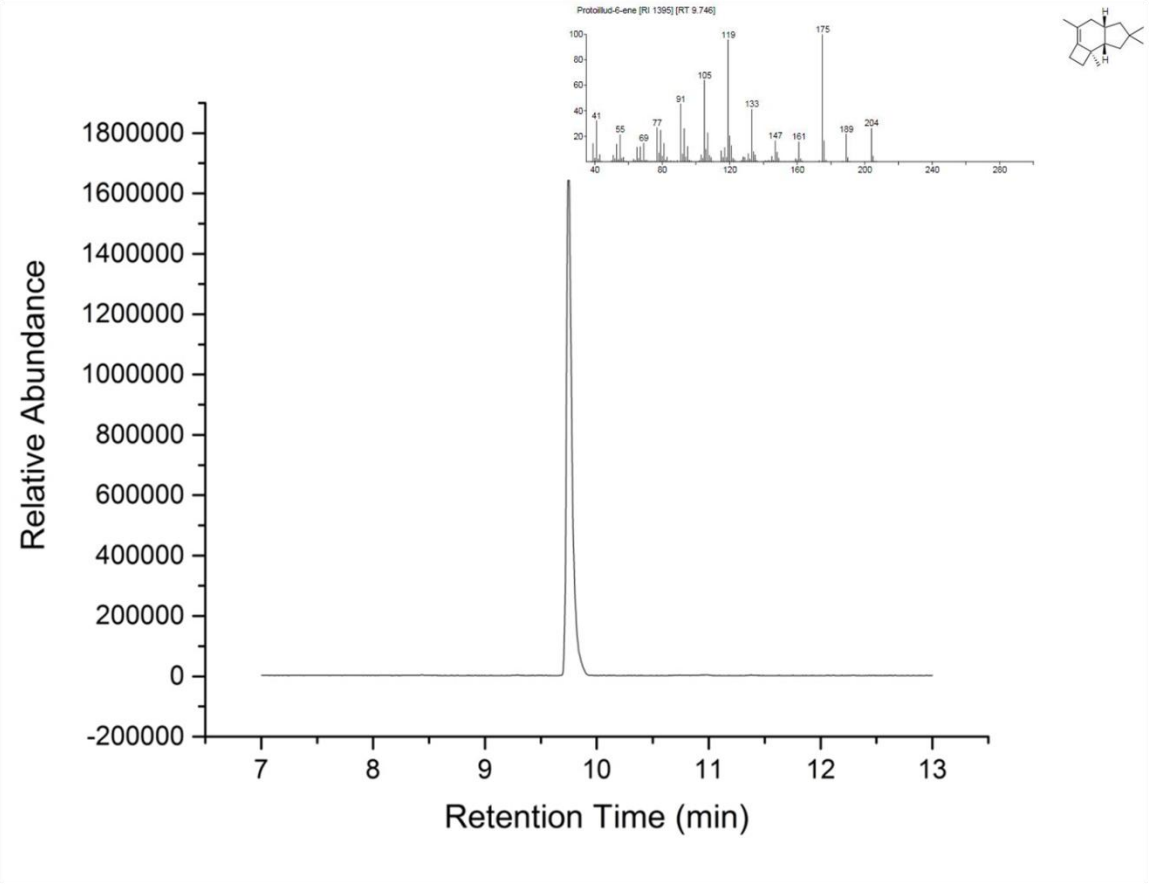


Figure 2.5 GC-MS chromatogram of purified Stehi7 incubated with 10mM $MgCl_2$ and 2μM FPP. Sole peak was identified as Δ6-protoilludene.

Purified Stehi7 was confirmed to be free of extraneous metal ions through its incubation with (E,E)-FPP without $MgCl_2$, as no GC-MS peak was shown. Upon this second level of verification, purified Stehi7 was concentrated and submitted for crystallization trials.

3.3 Exhaustive attempts to crystallize Stehi7

Sitting-drop vapor-diffusion assays were performed on Stehi7 at the University of Minnesota Nanoliter Crystallization Facility in the form of four different high throughput screens. Of the 96 different conditions between each high throughput screen, the SaltRx high throughput screen had the most hits with nucleation events occurring within 34 different conditions. Crystal Screen HT and Index high throughput screens each had 18 such conditions, while the PEG/Ion HT high throughput screens only had 9 such conditions. Though the SaltRx screen did have the most hits, 2 of the most promising crystallization conditions came from Crystal Screen HT, with 1 from Index, and 1 from SaltRx. The most meaningful crystallization hits are defined as the conditions that provided the most abundant and largest crystals (~20μm in diameter at the largest). A summary of the most promising crystallization conditions are discussed in **Table 2.2**. **Figure 2.6** illustrates the drops of these most promising crystallization conditions.

Table 2.2 The most promising 4 of the 384 different crystallization conditions screened for Stehi7.

High Throughput Screen	Well ID	Condition Name	Crystallization Conditions
Crystal Screen HT	A4	A	100mM tris pH 8.5, 2.0M ammonium sulfate
Crystal Screen HT	A9	B	200mM ammonium acetate, 100mM sodium citrate tribasic dihydrate pH 5.6, 30% w/v PEG 4000
Index	A4	C	100mM BIS-TRIS pH 6.5, 2M ammonium sulfate
SaltRx	F1	D	1.5M ammonium Sulfate, 100mM sodium acetate pH 4.6

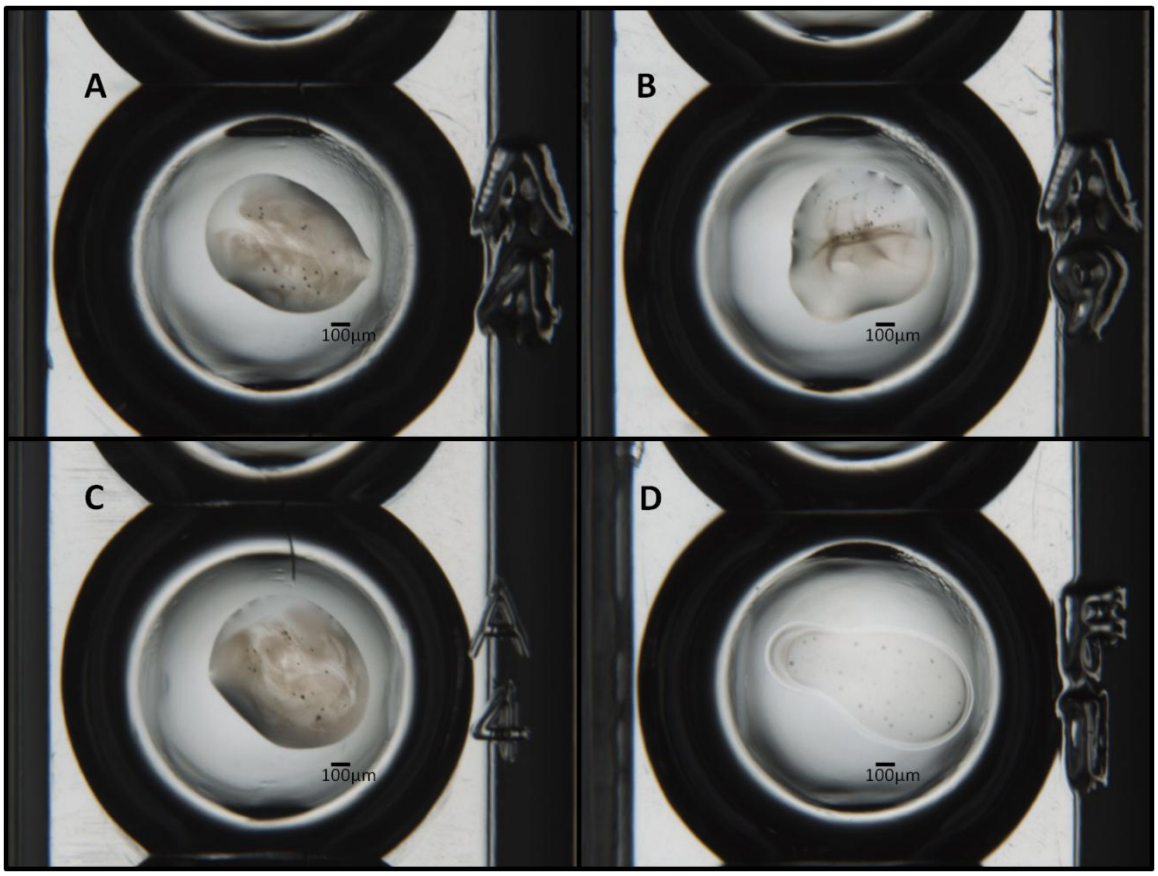


Figure 2.6 Most promising initial sitting drop crystallization conditions of Stehi7. (**A**) Crystal Screen HT Well A4: 100mM Tris pH 8.5, 2.0M ammonium sulfate. (**B**) Crystal Screen HT Well A9: 200mM ammonium acetate, 100mM sodium citrate tribasic dihydrate pH 5.6, 30% w/v PEG 4000. (**C**) Index Well A4: 100mM BIS-Tris pH 6.5, 2.0M ammonium sulfate. (**D**) SaltRx Well F1: 1.5M ammonium sulfate, 100mM sodium acetate pH 4.6. Specks within the drop indicate nucleation events.

The previously mentioned subset of four crystallization conditions was then selected for manual optimization screens via the hanging-drop vapor-diffusion technique. No crystals were found in the initial hanging drop trials, with most crystallization drops containing heavy amorphous precipitate and skin. However, for conditions A and D, the optimization screens for conditions showed lighter non-amorphous precipitate at lower ammonium sulfate concentrations (1.2-1.8M). Thus, all attempts at crystallization with conditions B and C were suspended, while the focus centered towards conditions A and D.

New hanging drop trays were optimized around the lower ammonium sulfate concentrations for conditions A and D, and light amorphous precipitate was reproducibly observed around the 1.4-1.6M ammonium sulfate range for both conditions. This was regardless of pH of sodium acetate or tris in the respective mother liquors.

It was hypothesized that Stehi7 was entering the supersaturation phase too quickly, and precipitating into unusable protein in these hanging-drop vapor-diffusion assays. Previous literature has shown that providing a seed stock in addition to crystallant can allow crystals to grow in the metastable solubility zone, where they otherwise would have precipitated [67-71]. In order test this concept for Stehi7, the previously mentioned hanging-drop trays optimized around 1.4M ammonium sulfate were repeated with the addition of a microseed stock of Stehi7. Microseed stocks of Stehi7 were made possible with a Seed Bead Kit from Hampton Research (Aliso Viejo, CA). Examination of these new hanging drop trays showed no detectable change in crystallization conditions for Condition D, leading to the suspension of further crystallization attempts with this condition. However, adding seed stock to trays optimizing condition A, allowed for the formation of crystalline precipitate at 1.2M ammonium acetate regardless the pH of tris.

Additional hanging-drop trays testing condition A with seed stocks could be reproduced, but the microcrystals in the crystalline precipitate never grew in size. Previous literature has also shown that the addition of macrocontaminants can also be used as stable seeds for nucleation [72]. However, ground human beard and chopped cat whisker coated in Stehi7 were also unable to introduce stable nucleation points for Stehi7 crystals to grow.

Final attempts at crystallizing Stehi7 was through the co-crystallization screening with sodium pyrophosphate (NaPP), and various metal ions. Co-crystallants could provide rigidity throughout the protein molecule as the pyrophosphate and metals would occupy the respective active and metal binding sites. However, addition of NaPP caused no significant change to the crystalline precipitate that was already observed, while the addition of metal ions introduced metal salt crystals to the solution. Through all hanging drop screens, the size of microcrystals formed in the crystalline precipitate were again on the order of 15-20μm in diameter as seen in **Figure 2.7**.

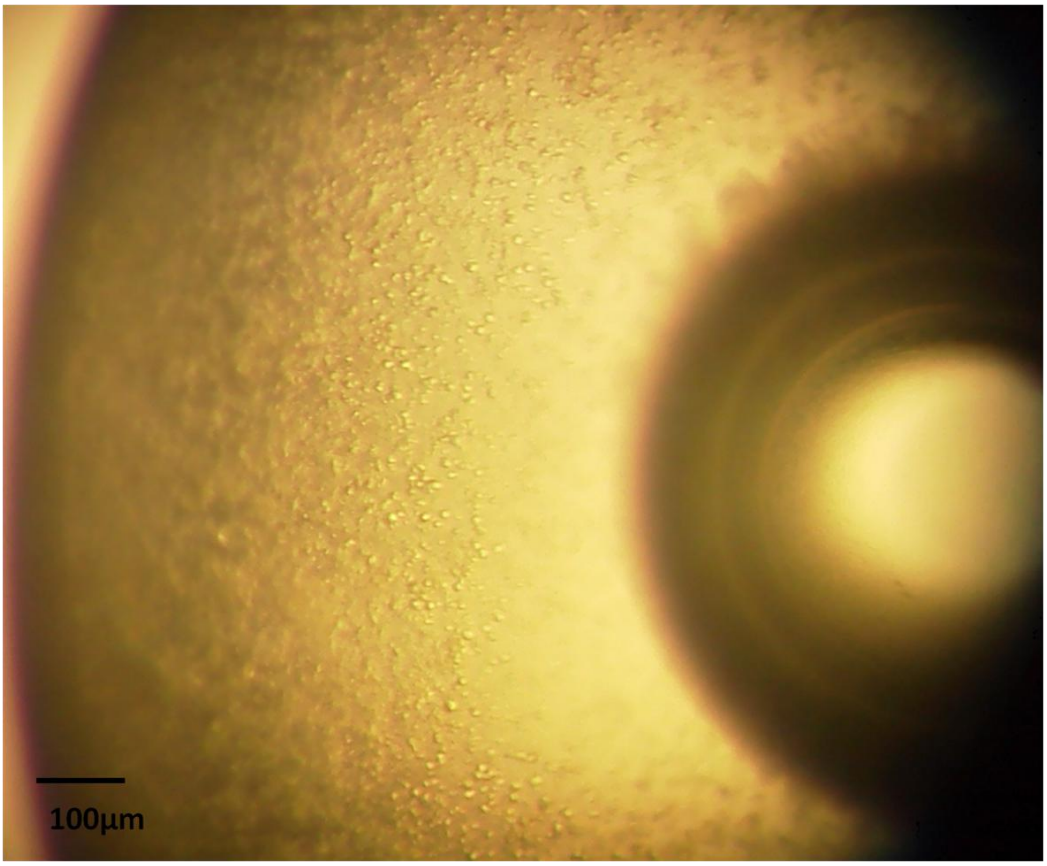


Figure 2.7 Hanging drop crystallization assay of Stehi7 in 1.2M ammonium sulfate, 100mM Tris pH 7.5, 4mM NaPP, additionally prepared with a microseed stock of previous Stehi7 crystallization drops.

3.4 Stehi7 does not undergo large scale structural rearrangements in the presence of different metal ions

Stehi7 was subjected to circular dichroism analysis in order to verify that Stehi7 did not undergo large structural rearrangements in the presence of metal ions. Additionally, circular dichroism experiments would provide some secondary structural information on Stehi7 as the crystallization attempts became increasingly tedious. In testing Stehi7 in the presence of $MgCl_2$ and $CaCl_2$, it was found that no large scale secondary structural rearrangements were occurring in comparison to native Stehi7. Additionally, the characteristic CD spectra in **Figure 2.8** illustrate that Stehi7 is largely α-helical, consistent with the literature of general sesquiterpene synthase structures [73].

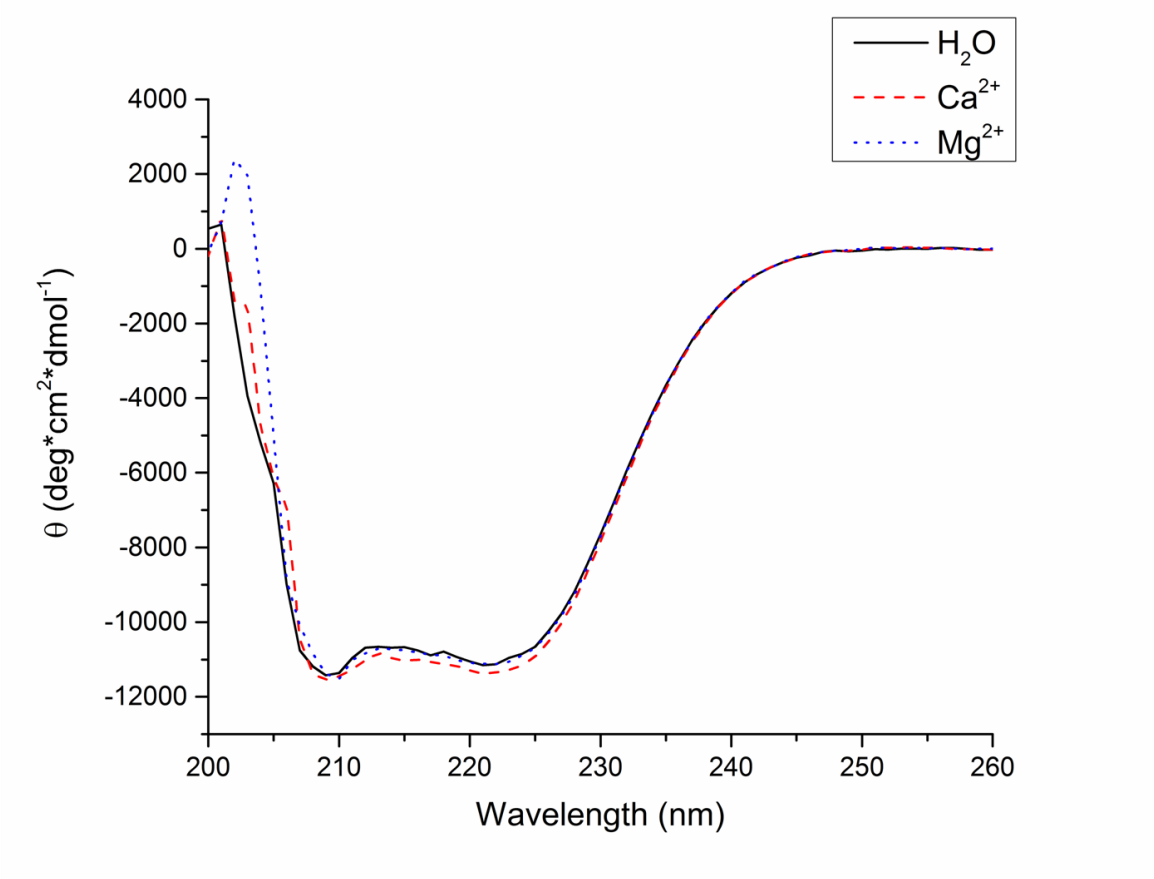


Figure 2.8 Circular dichroism reveals that addition of divalent metal ions does not affect the secondary structure of Stehi7. In the presence or absence of 1mM divalent metal ions Mg2+ and Ca2+ CD spectra of Stehi7 displays double minima at 208 and 222nm, characteristic of α-helical secondary structure.

3.5 Stehi7 (42-369) truncation does not express as well solubly as full length Stehi7

Many solved STS structures from Dave Christianson's laboratory were constructs with N-terminal truncations [74-77]. This is due to the relatively unstable nature of the N-terminus of STSs. Using this information, along with the guidance of the Xtal-Pred web server [65] crystallization prediction program, N-terminally truncated Stehi7(42-369) was cloned and heterologously expressed in *E. coli* Rosetta (DE3) pLysS strains. While the predicted crystallization of Stehi7(42-369) scored higher than wildtype Stehi7 on Xtal-Pred, the soluble expression lacked considerably in comparison. **Figure 2.9** illustrates the

stark contrast in soluble expression from identical culture conditions between Stehi7 and Stehi7(42-369) on an SDS gel after expression and His purification on Ni-NTA resin.

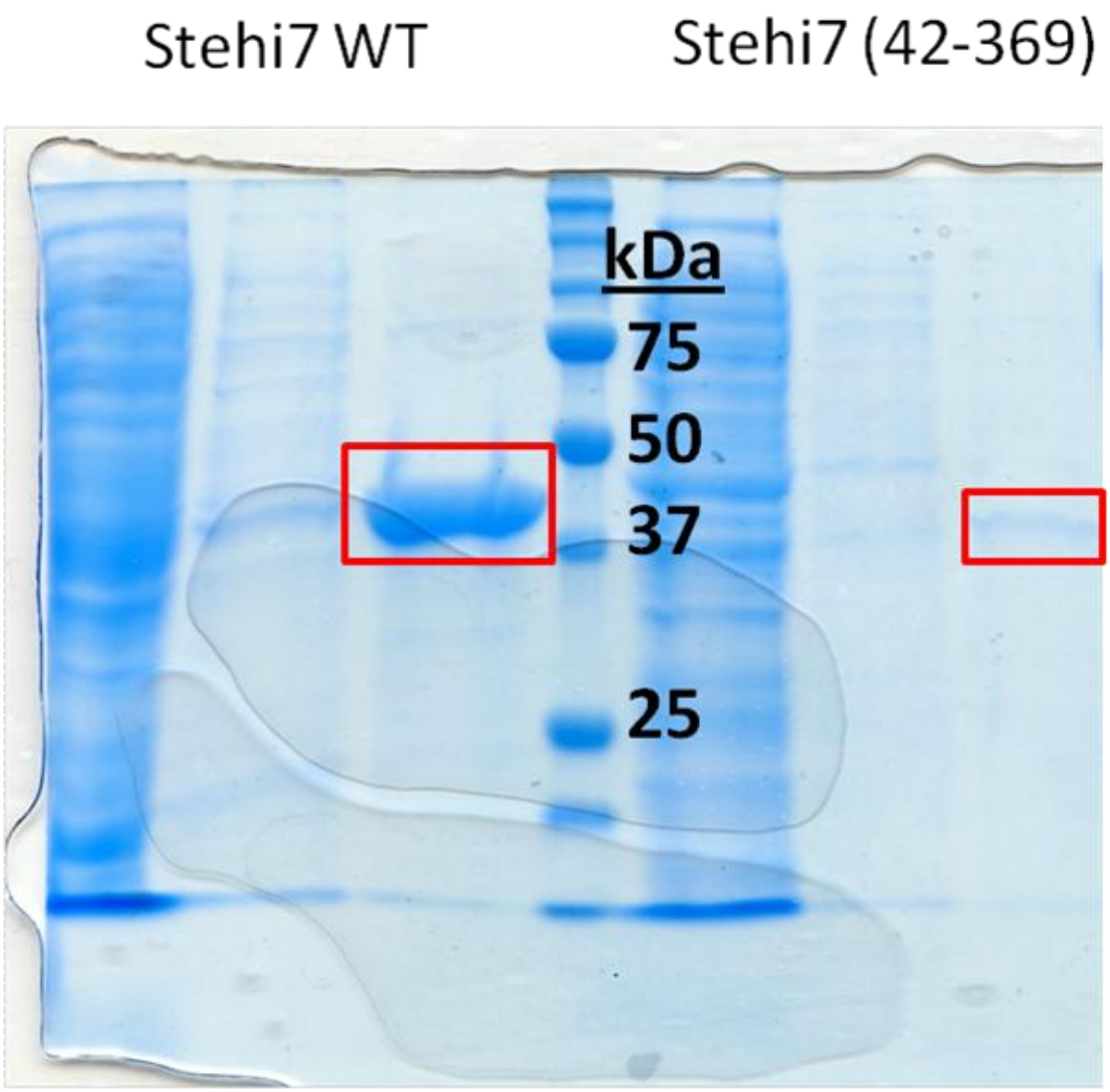


Figure 2.9 SDS-PAGE gel of Ni-NTA purified fractions of full length Stehi7 and Stehi7 (42-369) from same expression conditions. The discrepancy between expression of Stehi7 and its truncation lies in that most of Stehi7(42-369) was not expressed solubly.

4 Conclusion

Given the tedious nature of having to acid wash all glassware with 10% nitric acid to ensure Stehi7 purifications were free of metal ions, in addition to the tedious nature of protein crystallization in general, obtaining a solved X-ray crystal structure of a difficult to crystallize enzyme type was always going to be an ambitious endeavor within the scope of a Master's project timeline. However, this project was not without merit, as heterologous expression of fungal STSs was demonstrated, with Δ6-protoilludene synthase activity additionally conserved throughout purifications.

It should also be said that obtaining the crystal structure of Stehi7 is still plausible. 79 positive hits out of 384 crystallization conditions from the initial screen support this assertion. Many more conditions could be tested. Indeed, there were still many different parameters to optimize on this project, before work on this project was suspended for the promise of more tangible results with *Lactarius* mushrooms to be outlined in Chapters 3 and 4. For instance, all of the attempted manual scale up crystallization trays attempted utilized the hanging-drop vapor-diffusion method. In contrast, all initial crystal condition hits involved the sitting-drop vapor-diffusion method. Manual scaled-up crystallization trays could have been performed utilizing the sitting-drop vapor diffusion method to be more consistent with the initial automated crystallization screens. Additionally, no attempts were made at cleaving the N-terminal His tag from Stehi7 after His purification, which may have had a negative impact on the crystallization of Stehi7 as well, though no discernible difference was predicted on XtalPred.

Finally, if the N-terminal truncated construct were to be completely necessary for the crystallization of Stehi7, expression could have been performed at lower induction temperatures, in the presence of GroEL and GroES molecular chaperones, or optimized in a different *E. coli* strain to allow for better soluble expression of protein. The above strategies or large scale parallel cultures could have been set up for the Stehi7 N-terminal truncation construct. However, future work on this aim would require an extensive time commitment regardless of crystallization strategy.

5 Supplemental Material

Table S2.1 All hanging drop crystallization conditions screened for Stehi7.

Tray #	Crystallization Condtions	Protein Notes	Additional Notes
20131205A	1.4-2.4M Ammonium Sulfate, 100mM Tris HCl pH 7.5-9.0		
20131205B	24-34% w/v PEG 4000, 200mM Ammonium Acetate, 100mM Sodium citrate pH 5.2-5.8		
20131205C	1.4-2.4M Ammonium Sulfate, 100mM BIS-TRIS pH 5.5-7.0		
20131210A	1.2-2.2M Ammonium Sulfate, 100mM Tris pH 7.5-9.0		
20131210B	1.2-1.7M Ammonium Sulfate, 100mM Sodium Acetate pH 4.2-4.8		
20140224	1.4-2.0M Ammonium Sulfate, 100mM Sodium Acetate pH 4.2-4.7		
20140310	1.4-2.0M Ammonium Sulfate, 100mM Sodium Acetate pH 4.2-4.7	Purified protein underwent extra dialysis into RG H2O to remove NaCl	
20140331A	1.4-2.0M Ammonium Sulfate, 100mM Sodium Acetate pH 4.2-4.7	Protein samples seeded with well A6 from tray 20140310 using Hampton Research Seed Bead Kit	
20140331B	1.4-2.4M Ammonium Sulfate, 100mM Tris HCl pH 7.5-9.0	Protein samples seeded with well A6 from tray 20140310 using Hampton Research Seed Bead Kit	
20140331C	1.4-2.0M Ammonium Sulfate, 100mM Sodium Acetate pH 4.4-4.9	Protein samples seeded with well A6 from tray 20140310 using Hampton Research Seed Bead Kit	
04092014A	0.6-1.6M Ammonium Sulfate, 100mM TrisHCl pH 7.5-9.0		4oC
04092014B	1.1 M Ammonium Sulfate, 100mM TrisHCl pH 8.0 with 2-12mM MgCl2/BTAC/NaPP/AllThree		
04152014A	0.6-1.6M (NH4)2SO4, 100mM TrisHCl pH 7.5-9.0, 4mM MgCl2, 4mM BTAC, 4mM NaPP		
05122014A	0.6-1.6M (NH4)2SO4, 100mM Tris pH 7.0-8.5, 4mM NaPP, 4mM MgCl2	Protein Diluted to ~5.5mg/mL	
05122014B	0.6-1.6M (NH4)2SO4, 100mM Tris pH 7.0-8.5, 4mM NaPP, 4mM MgCl2	Protein Seed Stock Diluted 1:4	
5132014	1.2M Ammonium Sulfate, 100mM Tris pH 7.5, 4mM CoSubstrates	Proteins various freezing/glycerol/dialysis conditions	
20140519A	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 4mM NaPP, 4mM MgCl2		Seeded with Human Beard Micro seeds 1:5 Dilution (0.5uL added to drop) 4oC
20140519B	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 4mM NaPP, 4mM MgCl2		Seeded with Cat Whisker Macro seeds *** (Chopped Whisker Coated in Fresh Protein and added to protein drop) 4oC
20140520A	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 4mM NaPP, 1-100mM MgCl2		4oC
20140520B	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 4mM NaPP, 10mM Zn/Mg/Ca/MnCl2		4oC
20140527	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 4mM NaPP, 10mM Zn/Mg/Ca/MnCl2		20oC
20140529	0.6-1.6M Ammonium Sulfate, 100mM Tris pH 7-8.5, 10mM Zn/Mg/Ca/MnCl2		20oC
20140711	0.4-1.4M LiSO4, 0.5M Ammonium Sulfate, 100mM Citrate pH 4.8-6.0		

Chapter 3: Towards the sesquiterpenome screening of *Lactarius* and the genome assembly of *L. uvidus*

1. Introduction

1.1 The genus *Lactarius*

Lactarius mushroom species provide an ideal model system for further sesquiterpenoid characterization in Basidiomycota. This genus of ectomycorrhizal fungi is distributed worldwide, but mostly within the Northern hemisphere. There are conservative estimates of at least 450 different known species of *Lactarius*, with at least 150 different species described in Europe alone [78]. Perhaps the most well-known *Lactarius* species is the choice edible *Lactarius deliciosus, w*hich is also known as the saffron-milk cap [79]. However, other *Lactarius* species, like *Lactarius turpis* are known to be poisonous [80]. This variability between species can be explained by differences in the constituents of the most defining characteristic of *Lactarius* fruiting bodies— their latex.

Lactarius species are easily identified and differentiated from other fungal genera in the field by the presence of a milky or aqueous juice (latex) that is exuded from wounded fruiting bodies. This exuded latex is the cytoplasm of specialized cells within the *Lactarius* hyphae known as laticifers, which can be viewed and differentiated from neighboring cell types microscopically through the use of stains including sulpho-vanillin [81]. Further differentiation within the genus *Lactarius* can be made through determining the color and taste of the particular *Lactarius* species latex.

Depending upon the species of *Lactarius*, the exuded latex may be white or colored, or become colored over time. Additionally, the taste of the latex may be mild, hot or acrid, or become progressively hotter or more acrid [78]. This latex provides an additional means of chemical defense to protect the fruiting bodies from parasites and other predators [82]. Indeed, the color and taste variety of the *Lactarius* latex can be explained by the presence of a multitude of secondary metabolites found within. It is hypothesized that the change in color and taste in the *Lactarius* latex can be described by the conversion of inactive fatty acid esters of sesquiterpene alcohols from intact fruiting bodies, to biologically active sesquiterpenoid derivatives upon injury [78, 83]. In fact, a plethora of highly bioactive sesquiterpenoid scaffolds alone have been isolated from *Lactarius* fruiting bodies and latex [78], which will be described in greater detail in the following section 1.2.

1.2 *Lactarius sp.* produce diverse *trans*-humulyl derived sesquiterpenoids

As previously stated throughout this thesis, most of the medicinally and industrially relevant sesquiterpenoids isolated from Basidiomycota come from 1,11-cyclization *trans*-humulyl derived cations [43]. *Lactarius* mushrooms are known for producing a wide variety of sesquiterpenoid scaffolds derived from the *trans*-humulyl cation, making them an appealing target for further characterization of sesquiterpenoid natural products in Basidiomycota with the goal of accessing additional bioactive sesquiterpenoid scaffolds. Of the particular *trans*-humulyl derived scaffolds, the majority of *Lactarius* sesquiterpenoids isolated from fruiting bodies are of the marasmane, lactarane and secolactarane varieties [78]. These three scaffold classes are all putatively derived from a

protoilludene scaffold, and have demonstrated antimicrobial, cytotoxic, antifeedant and mutagenic activities [78, 84-86].

The biosynthesis of almost all of these sesquiterpenoid scaffolds made by *Lactarius* species is uncharacterized. More specifically, it is unknown whether the cyclization pathways to these different scaffolds are derived from a sesquiterpene synthase (STS) alone, or whether these are secondary cyclizations reactions from additional enzymes. **Figure 3.1** below illustrates the putative biosynthetic routes from FPP to bioactive sesquiterpenoid scaffolds in the *Lactarius* genus, including the marasmanes, lactaranes and secolactaranes.

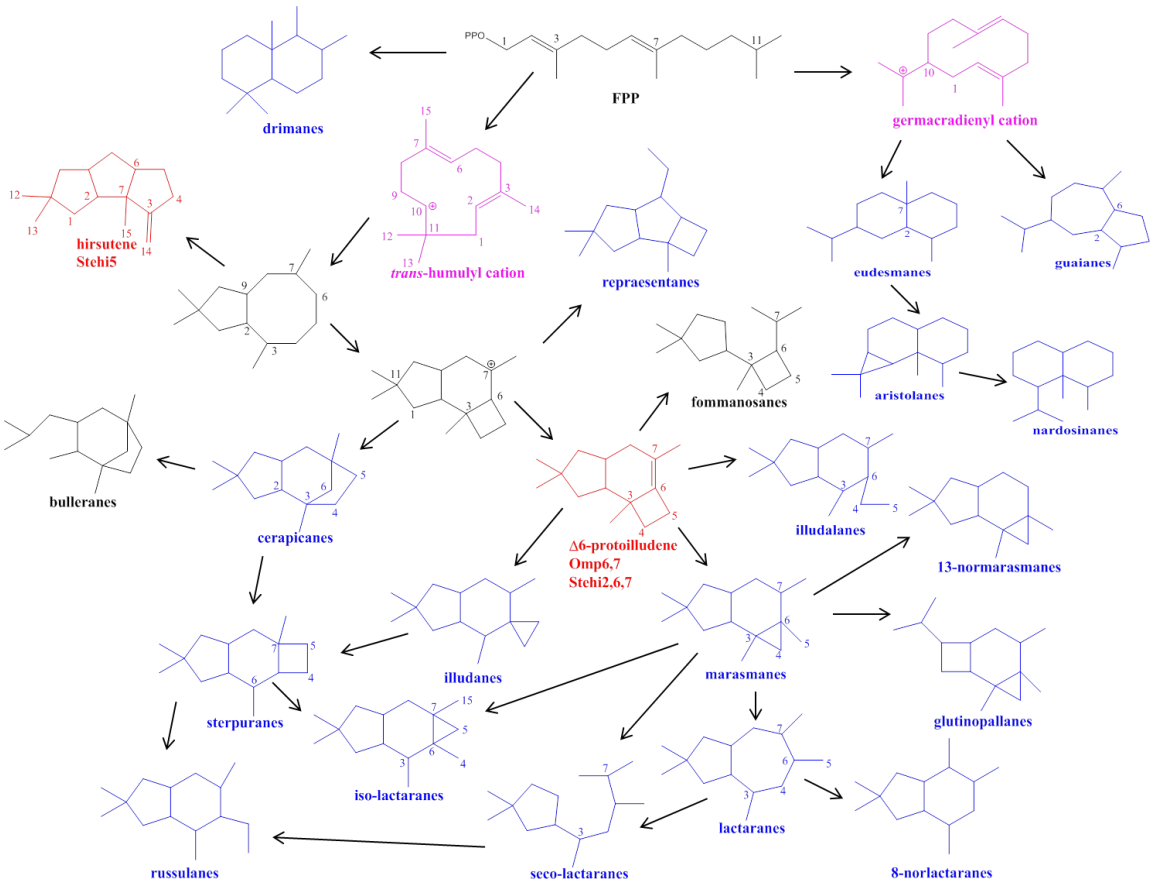


Figure 3.1 Possible biosynthetic pathways for *Lactarius* sesquiterpenoid scaffolds. Blue scaffolds have bioactive representative sesquiterpenoids. Red scaffolds have characterized STSs. Pink scaffolds are the initial carbocation.

In addition to the *trans*-humulyl cation derived scaffolds isolated from *Lactarius* species shown in **Figure 3.1,** these mushrooms are known to produce sesquiterpenoids derived from the drimane scaffold. Drimanes have been isolated from a number of different plants and fungi [87], and have been found to exhibit antifungal [88], cytotoxic [89, 90], antiviral [91], mTOR inhibiting [92] and biofilm disrupting [93] activities. What further sets drimanes apart from other *Lactarius* sesquiterpenoid scaffolds is the nature of their biosynthesis, which have not been well characterized. It is postulated that drimane sesquiterpene synthases may be the missing link between higher and lower terpenoids, as

their biosynthesis may require a protonation cascade event akin to Class II terpene synthases [87]. Class II mechanisms are less common in fungal STSs and more common in plant and fungal diterpene synthases [14]. As previously mentioned in Chapters 1 and 2, Class I and Class II terpene synthases can be distinguished by their conserved motifs (D(D/E)xx(D/E) and NSE/DTE for Class I, DxDD for Class II).

Characterization of drimane biosynthesis is difficult because drimane itself is rarely isolated from nature. Thus, mechanistic studies are typically performed on the related drimane-type sesquiterpenoid, drimenol. A 2014 study further added to the mystery of drimane biosynthesis through the cloning and characterization of a drimenol synthase from the valerian plant *Valeriana officinalis*, in which the cloned drimenol synthase appeared to exhibit both Class I ionization and Class II protonation mechanism attributes [94].

1.3 Genome mining of *Lactarius sp.* for novel natural product biosynthetic pathways

Given the diverse chemical space afforded by the diversity of sesquiterpenoid scaffolds found in *Lactarius sp.* chemical defense systems, it is of great interest to discover and characterize novel 1,11-cyclizing STSs from various *Lactarius* species. Additional emphasis is placed on the search for a novel drimane-type sesquiterpene synthase to help determine the nature of drimane biosynthesis in fungi. Furthermore, it is hypothesized that since many sesquiterpenoid scaffolds found from *Lactarius* species are putatively derived from a protoilludane scaffold, that these protoilludene synthases should be readily discovered through subsequent genome mining using the predictive framework developed in the Schmidt-Dannert laboratory.

Therefore, the experiments performed in this study first aimed at establishing mycelial cultivation protocols from field collected fruiting bodies for various *Lactarius* species with the goal of analyzing their volatile sesquiterpenoid profiles and growth phenotypes. Suitably fast growing strains that produce protoilludane- and/or drimane-type sesquiterpenoids will be selected for subsequent genome sequencing and mining. Mycelial growth cultures of strains will be complemented by chemical analysis of their latex collected from fruiting bodies.

One potential challenge in the genome sequencing of *Lactarius* is that growing homokaryotic strains from *Lactarius* spores has not been accomplished. Thus, all *Lactarius* in subsequent analyses exists as dikaryotic strains. Dikaryotic growth occurs in Basidiomycota when compatible cell types of mycelium fuse together in a process called plasmogeny. After the initial fusing event, the presence of nuclei from both original cells is maintained throughout growth and cell divisions through clamp connections [95]. While this is a novel growth feature, this does complicate genomic studies as genome assembly software is often unable to resolve allelic differences between two nuclei [96].

2. Materials and Methods

2.1 *Lactarius* species collection and division

Mature fruiting bodies of *Lactarius turpis* (Cedar Creek Ecosystem Science Reserve, East Bethel, MN), *L. vinaceorufescens* (Reservoir Woods Park, Roseville, MN), and *L. uvidus* (Cedar Creek Ecosystem Science Reserve, East Bethel, MN) were collected in the field. The lab was also graciously donated *L. rufus* and *L. xanthogalactus* fruiting bodies

from Alija Mujic of Oregon State University, and established cultures of *L. deterrimus*, *L. indigo, L. psammicola* and *L. rufus* from Dr. Peter Kennedy's lab at the University of Minnesota. *Lactarius* species were differentiated taxonomically via the following phenotypic characteristics: KOH staining, cap diameter, cap color, stem color, stem length, gill morphology, latex color, latex abundance, gill-wounding color, and tree symbiont species. Further species identification was confirmed through ITS sequencing and phylogenetic analysis by the Kennedy lab from genomic DNA preps to be described in Section 2.2. Fruiting bodies that were either collected or received as gifts were placed in 3 groups based on their respective downstream processing goals. Those three groups were: (1) cultivation, (2) tissue prep for RNA extraction and (3) latex extraction.

2.2 *Lactarius* species cultivation

A slight incision was performed across the convex diameter of the fruiting body cap with a scalpel, and the fruiting body was split in half by hand. Interior mycelium from the cap was cut out with a sterile scalpel and used to inoculate Modified Melin-Norkrans (MMN) agar plates (5 g/L glucose, 2 g/L malt extract, 1 g/L Yeast Extract, 0.5 g/L KH_2PO_4, 0.15 g/L $MgSO_4$, 0.25 g/L $(NH_4)_2HPO_4$, 0.05 g/L $CaCl_2$, 0.025 g/L NaCl, 0.012 g/L $FeCl_3$) whose glucose concentration was increased to 10g/L (designation "MMN10" plates). MMN10 inoculations were allowed to grow at room temperature (RT) in the dark for 3-4 weeks. Using a sterile scalpel, agar plugs containing mycelium cut from the outer radius of resultant growth were used for subsequent serial inoculations of cultures onto regular MMN and other solid growth media.

2.3 Tissue prep for mRNA extraction

Large fruiting bodies were placed in 1L glass beakers containing 500mL sterile water and sonicated for 20 minutes in a Branson 3510 Ultrasonic Cleaner. Fruiting bodies were then split in half by hand after a slight incision with a scalpel as previously described. Material from the interior of the fruiting body was cut out using a sterile scalpel and separated into three tissue groups (1) cap, (2) gills and (3) stem. Each tissue sample was diced into ~2x2x2mm pieces. Diced material was dried between filter paper, placed in 1.5mL eppendorf tubes in 100mg portions, and stored at -80^{o}C until mRNA extractions were performed.

2.4 Fruiting body latex extractions

Fruiting bodies for latex extractions were placed in 1L glass beakers containing 500mL sterile water, and sonicated for 20 minutes in a Branson 3510 Ultrasonic Cleaner. Fruiting body caps were cut from the stem and laid on UV-treated aluminum foil with gills facing upward. Grid-like incisions were performed with a scalpel across the gills to bleed the latex. A Pasteur pipette containing small aliquots of sterile water were continuously aspirated and ejected onto the latex to prevent coagulation, and the latex/water mixture was placed in a 1.5mL eppendorf tube. The latex/water mixture was then combined with storage buffer (10mM Tris pH 8.0, 1mM dithiothreitol, 10% w/v glycerol, 1X Roche protease inhibitor cocktail) and stored in 250µL aliquots at -80^{o}C.

2.5 Growth comparison of *Lactarius* cultures

A ~2 x 4mm agar plug was cut from MMN solid cultures of *L. deterrimus, L. indigo, L. psammicola, L. rufus, L. turpis, L. uvidus, L. vinaceorufescens,* and *L. xanthogalactus* and inoculated on triplicate agar plates of biotin aneurin folic acid (BAF) media (50 g/L glucose, 2 g/L peptone, 1 g/L yeast extract, 0.5 g/L KH_2PO_4, 0.5 g/L $MgSO_4*7H_2O$, 0.1 g/L $CaCl_2$, 0.005 g/L $FeCl_3$, 0.005 g/L $MNSO_4$, 0.001 g/L $ZnSO_4*7H_2O$) , MMN media (recipe previously mentioned), potato dextrose agar (PDA) media (24 g/L Difco potato dextrose broth) and Rich media (20 g/L malt extract, 20 g/L glucose, 5 g/L peptone). Major and minor radii of outward mycelium growth from the agar plug were measured in mm at time points of 6, 20, 25, 38 and 54 days. Mycelium growth area was calculated using the area formula for an ellipse, correcting for the initial agar plug area.

$$A = (\pi * a * b) - (c * d)$$

As such, a and b are the major and minor axes of respective mycelium growth, and c and d are the length and width of initial agar plug for inoculation. The average mycelium growth area was found for each triplicate culture. Average mycelium growth rate was determined by taking the average change in mycelium growth area between each time point for each respective culture. The average mycelium growth rate was ranked between species for each growth medium, and also between growth medium for each species. Each rank was averaged to determine fastest general growing *Lactarius* species and most suitable growth media for fastest mycelium growth for all *Lactarius* species.

2.6 Mycelial culture head-space and latex analysis by GC-MS

Agar plugs containing mycelium of *L. deterrimus, L. indigo, L. psammicola, L. rufus, L. turpis, L. uvidus, L. vinaceorufescens,* and *L. xanthogalactus* were placed on 30mL agar slants containing MMN or Rich media in 50mL glass culture tubes. Agar plugs containing mycelium from the previously mentioned *Lactarius* species were also used to inoculate 50mL liquid cultures of Rich media and 50mL liquid cultures of MMN. Additional 50mL liquid cultures of "A" media (20g/L soybean flour, 5 g/L bactopeptone) [97], "C" media (30g/L glucose, 30g/L soybean flour) [98], and M media (12 g/L soybean peptone, 3 g/L yeast extract, 28 g/L glucose) [99] were inoculated with agar plugs containing *Lactarius uvidus* mycelium.

After 35-118 days in culture, solid and liquid medium growth headspace was sampled by solid phase microextraction (SPME). SPME was achieved by inserting a 100μm polydimethylsiloxane fiber into the culture headspace after piercing aluminum foil seals on the tops of the cultures. The polydimethylsiloxane fiber sampled the culture headspace for 30min. For latex GC-MS analysis, 100μL reactions containing 10mM $MgCl_2$, 2μM (2E,6E)-farnesyl pyrophosphate and thawed latex samples were incubated at RT overnight in glass GC-MS vials. Vial headspace was sampled with the polydimethylsiloxane fiber for 30min. GC-MS analysis was performed on the same machine as mentioned in chapter 2. However, the oven in the mycelial culture headspace assays was ramped from 60^{o}C to 250^{o}C at 3.5^{o}C/min for 54min, and ramped from 60^{o}C to 250^{o}C at 5^{o}C/min for 38min for the *in vitro* latex assays.

Volatile chemical products produced from the mycelial headspace and latex analyses were identified by comparing retention indices and mass spectra of the compound peaks with published reference spectra in the terpene library of MassFinder and in the National Institute of Standards and Technology (NIST) standard reference database 08 (refer to **Supplemental Table S3.1** for MS spectra of identified peaks).

2.7 *Lactarius uvidus* genomic DNA extraction

An agar plug containing *L. uvidus* mycelium was placed onto a nitrocellulose membrane covering an MMN agar plate. After 3 weeks of growth, mycelium tissue was harvested from the plate and frozen at -80^{o}C, and lyophilized overnight. Lyophilized tissue was ground into a fine powder in liquid nitrogen and genomic DNA (gDNA) was extracted with the OmniPrep for Fungus commercial kit (G Biosciences, St. Louis, MO) by adapting the manufacturer's gDNA extraction protocol.

Briefly, 500µL of "Genomic Lysis Buffer" and 50µg/mL proteinase K was added to 20mg of finely ground mycelium powder and incubated at 60^{o}C for two hours. The resultant mixture was extracted with 0.4 volumes of chloroform. Further nucleotide-protein interactions were disrupted via the addition of 50µL "DNA Stripping Solution" and incubated at 60^{o}C for 10min. Stripped protein was precipitated out via addition of 100µL "Precipitation Solution" and centrifugation at 14,000xg for 5 min to form a protein pellet. The genomic DNA was precipitated out from the resultant supernatant by addition of 500µL of isopropanol and centrifugation at 14,000xg for 5 min. Resultant DNA pellets were washed with 700µL 70% ethanol and dissolved in 50-100µL TE buffer

(10mM Tris pH 8, 1mM EDTA) overnight at 4^{o}C. Dissolved gDNA was treated with 100µg/mL RNase A at 37^{o}C for 2 hours.

Further gDNA purification was performed via an additional round of isopropanol precipitation by adjusting the sample salt concentration to 0.3M sodium acetate and adding 0.65 volumes of isopropanol at RT, and centrifuging at 14,000xg for 10min. The resultant gDNA pellet was washed with 70% ethanol and resuspended in 50-100µL TE buffer for resolubilization overnight at 4^{o}C.

The quality of gDNA sample was determined by confirming double stranded DNA concentration of the sample with a Qubit ® fluorometer (Thermo Fisher Scientific, Waltham, MA) and the sample purity with a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Additional quality control measures included running 2.5µL of gDNA sample on a 1% agarose gel to determine whether sample gDNA degradation occurred. Lastly, upon passing the quality control measures, the purified *L. uvidus* gDNA sample was shipped overnight on dry ice to the Molecular Biology Core at the Mayo Clinic's Medical Genome Facility (Rochester, MN) for Illumina library preparation and sequencing.

2.8 Illumina sequencing of *Lactarius uvidus*

At the Mayo Clinic Next Generation Sequencing Center, a 101bp paired-end Illumina library was prepared from purified *Lactarius uvidus* gDNA using the NEBNext® Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) according to the manufacturer's suggested protocol. The paired-end *L. uvidus* library was run on an Illumina HiSeq 2500 platform in one lane. Raw sequencing reads were packaged as

forward and reverse read .fastq files and transferred to the University of Minnesota's Supercomputing Institute (MSI) via file transfer protocol (FTP).

2.9 Quality control of Illumina reads

Raw .fastq files of *L. uvidus* were processed for quality control via the Galaxy [100-102] platform provided by the Minnesota Supercomputing Institute (MSI). Raw initial reads with adapter contamination were first filtered out using Galaxy's Trimmomatic tool [103]. Trimmomatic was also used to crop away 5 bases at the 5' end of each read, and to crop the 3' end of the reads down to a final length of 80bp. Finally, Trimmomatic filtered out any remaining reads that were not exactly 80bp in length. The output of the Trimmomatic tool was four .fastq files (forward paired, forward unpaired, reverse paired and reverse unpaired reads).

The four outputted Trimmomatic .fastq files underwent further quality control using the Filter by Quality tool on Galaxy. Reads that contained at least one base pair with a quality score of less than 20 were removed. Filtered forward paired and filtered reverse paired .fastq files were then interleaved using the FASTQ interlacer tool in Galaxy, which generated an output of two files (Paired Interleaved and Paired Singles .fastq files). Statistical analysis was performed on the Paired Interleaved, Paired Singles, Filtered Forward Unpaired and Filtered Reverse Unpaired .fastq files via Galaxy's FASTQC tool [104]. Files containing greater than 1 million reads passing the majority of quality statistical tests in the FASTQC output were exported for genome assembly.

2.10 Genome assembly and functional annotation

Paired Interleaved and Paired Singles .fastq files for *L. uvidus* reads were assembled into genomic sequences with the Velvet [37] assembler using a k-mer size of 53. Assembly statistics were computed using the python script summarizeAssembly.py available from MSI's PBsuite [105]. Large scale repeats in the genome were masked using the RepeatMasker [106] program. Gene predictions were performed on the masked genome files using Augustus [107] with gene models derived from the agaricomycete *Laccaria bicolor* supplied with the Augustus package. An in house perl computer program called AugustusParserv3.0.pl was created, and used to parse and extract the predicted amino acid sequences from the Augustus .gff3 file. AugustusParserv3.0.pl also wrote all predicted amino acid sequences into an output .fasta file. BLAST [108] (blastp specifically) was used with anchor genes of interest against the amino acids .fasta file to identify gene candidates for functional analysis.

2.11 Bioinformatics analysis of candidate genes

New sesquiterpene synthases (STSs) were searched for in the predicted amino acids sequences of *L. uvidus* using a library of previously functionally characterized STSs as query sequences in a blastp search. The library included STSs identified in *Coprinus cinereus*, *Omphalotus olearius*, and *Stereum hirsutum* as described previously by the Schmidt-Dannert laboratory [24, 38, 40, 43, 44, 46]. A multiple sequence alignment of the STS library and candidate genes were made using ClustalW [109] on the MEGA [110] version 6 bioinformatics suite. The alignment was manually adjusted to remove

gene predictions that were not the characteristic length of STSs (see **Supplemental Table S3.2** for detailed gene information, including possession of conserved active site motifs).

A new multiple sequence alignment was created using MAFFT [111], and the best protein evolutionary model to use for maximum likelihood phylogenetic tree construction was determined using ProtTest3.2 [112]. In ProtTest3.2, a consensus between the Akaike Information Criterion (AIC) [113], Bayesian Information Criterion (BIC) [114], corrected Akaike Information Criterion (AICc) [115] and performance-based Decision Theory (DT) [116] determined the Le Gascuel substitution model [117, 118] with an additional gamma distribution parameter (LG+G) as the best substitution model to use. A maximum likelihood tree was then constructed based on the multiple sequence alignment using the LG+G substitution model using the program PhyML3.1 [119].

2.12 Cloning of *L. uvidus* Sesquiterpene Synthases

Tissues prepped for RNA extraction from *L. uvidus* were ground into a fine powder in liquid nitrogen. Extraction of RNA was performed using TRIzol® (Invitrogen, Carlsbad, CA) according to the suggested protocol. Extracted RNA was treated with DNaseI (New England Biolabs, Ipswich, MA) according to the manufacturer's protocol to further get rid of residual gDNA found in the prep. The synthesis of cDNA from the extracted mRNA was performed using the Superscript® III First-Strand Synthesis System for RT-PCR kit (Invitrogen, Carlsbad, CA) according to the suggested protocol using oligo(dT) primers to enrich cDNA from mRNA. Newly synthesized cDNA was then treated with RNaseH for 20min at 37^{o}C and then stored at -20^{o}C indefinitely.

PCR scans using Phusion High Fidelity Polymerase (New England Biolabs, Ipswich, MA) with gene specific primers and conserved D(D/E/N)xx(D/E) and NSE motif specific primers were performed on the cDNA template for putative *Lactarius uvidus* sesquiterpene synthases. Gel extracted positive amplicons were blunt cloned into pCR™-BluntII-TOPO® plasmids using the Zero Blunt® TOPO® PCR Cloning Kit according to the manufacturer's instructions (Invitrogen, Carlsbad, CA), and submitted to the University of Minnesota Genomics Center for sequencing verification.

3 Results and Discussion

3.1 MMN media best for cultivating relative fast growth of *Lactarius sp.* and *L. rufus* is fastest growing *Lactarius* species

To determine the effect of different culture media on mycelial growth rates, *L. deterrimus, L. indigo, L. psammicola, L. rufus, L. turpis, L. uvidus, L. vinaceorufescens* and *L. xanthogalactus* were cultivated on biotin aneurin folic acid agar (BAF), modified Melin-Norkrans agar (MMN), potato dextrose agar (PDA), and rich agar (Rich) (see Methods for media composition). This experiment was modeled similarly to a study involving the growth of *Lactarius controversus* and *Lactarius pyrogalus* [120]. In this current study, the resulting mycelium areas of growth were calculated from the measured major and minor axes of growth measured to the nearest millimeter. Furthermore, average mycelium area growth rate was determined by averaging the change in mycelium growth area between each random measurement time point.

There was a large amount of variation in growth rate between each species for each type of media. By ranking the average mycelium growth rate between species for each growth medium, and ranking the average mycelium growth rate between growth medium for each species, allowed for the intuitive determination of which was the most suitable medium for the fastest *Lactarius* growth, and which *Lactarius* was the fastest growing species. In terms of best mycelium growth, MMN provided the most suitable medium for culturing *Lactarius* for faster relative growth. Under these conditions, *L. rufus* was generally the fastest growing species in this survey. Conversely, PDA was the least suitable medium for *Lactarius* growth and *L. xanthogalactus* was the slowest growing species in this survey. **Figure 3.2** illustrates the average mycelium growth rate between each species for each growth medium.

A

Culture Medium	Average Ranking
MMN	1.75
BAF	2.25
Rich	2.75
PDA	3.25

B

Species	Average Ranking
L. rufus	2
L. uvidus	3
L. psammicola	3.5
L. indigo	4
L. deterrimus	4.25
L. vinaceorufescens	4.5
L. turpis	7.25
L. xanthogalactus	7.5

C

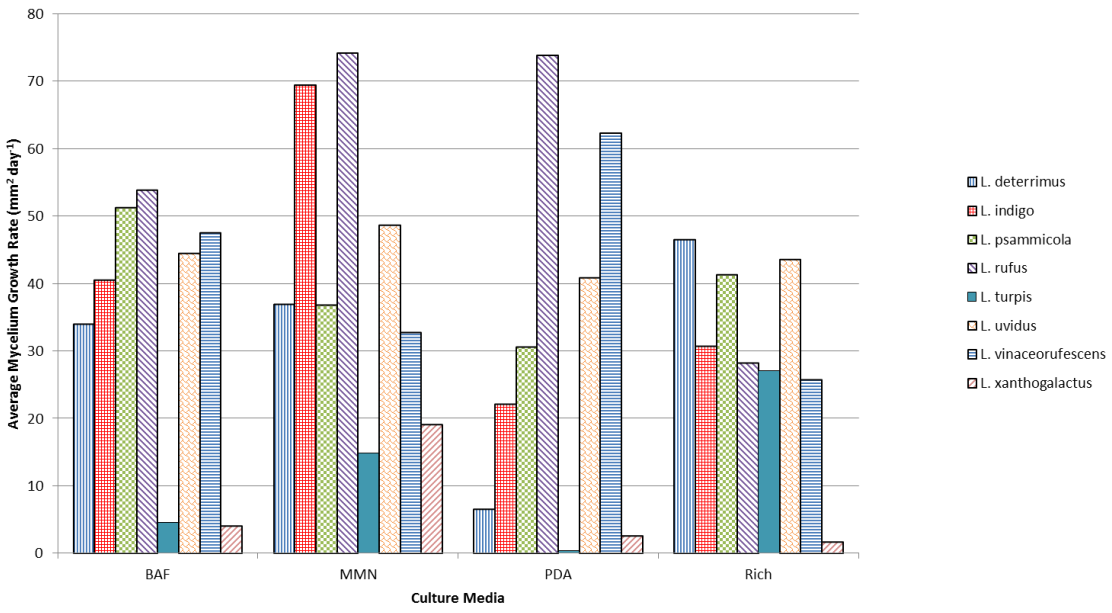


Figure 3.2 Average mycelium area growth rates for *L. deterrimus, L. indigo, L. psammicola, L. rufus, L. turpis, L. uvidus, L. vinaceorufescens*, and *L. xanthogalactus*. (A) Ranked average of the average mycelium area growth rate for each growth medium between each *Lactarius* species. **(B)** Ranked average of the average mycelium area growth rate for each species between each growth medium. **(C)** Average mycelium area growth rate for each *Lactarius* species for each growth medium.

Of general note, the overall mycelium area growth rates were not very high for any of the *Lactarius* species surveyed. For instance, *L. rufus* on MMN had the highest average mycelium area growth rate of any species in any condition at 74.2 mm^2 day^{-1}. At this rate, it would take roughly 3.5 months to cover the surface of a typical 100mm diameter Petri dish. This slow growth rate forced several considerations to be put in place for large scale culturing of *Lactarius* species for genomic DNA extraction, to be explained in a later section.

3.2 *Lactarius sp.* mycelium and latex produce diverse sesquiterpenes

The Schmidt-Dannert laboratory had previously successfully screened the headspace of *Omphalotus olearius* and *Stereum hirsutum* liquid cultures in Rich medium [24, 38]. The same culture conditions were therefore tested with the first *Lactarius* strain that the Schmidt-Dannert laboratory received in culture from the Kennedy Laboratory at the University of Minnesota, *L. rufus*. Initial attempts at growing this fungus in Rich cultures showed that it grew fairly poorly in this medium. Mycelial cultures never reached a sufficient density to produce volatile compound levels in the culture headspace for confident GC-MS identification. Assuming from these growth experiments that other *Lactarius* would likely grow poorly in Rich medium (which was later found to be incorrect, see experiments shown in **Figure 3.2**), an alternative approach was developed where fungal strains were grown on Rich (for comparison to *O. olearius* and *s. hirsutum* profiles) and MMN media agar slants (because of fastest *Lactarius* growth rate on this media). Previous growth experiments showed that *Lactarius* could be more readily cultured on solid media. Furthermore, by using agar slants, a smaller headspace would yield a higher concentration of volatile compounds for subsequent GC-MS analysis.

Growth experiments with additional *Lactarius* species later showed that several of them grew well in liquid Rich medium. Liquid mycelial cultures were revisited for volatile compound analysis by GC-MS. Liquid cultures in Rich and MMN media were inoculated with fungal strains and cultivated for extended times with periodic headspace sampling by SPME followed by GC-MS analysis of the absorbed volatiles.

Complementary to the chemical analysis of the cultures, *in vitro* assays for sesquiterpene synthase activity were performed with extracted latex (note latex is cytoplasmic material) from fresh fruiting bodies. Specifically, extracted latex from *L. rufus, L. uvidus, L. vinaceorufescens and L. xanthogalactus* was incubated with FPP in GC-MS vials, and their respective headspaces were sampled for volatile compounds. **Figure 3.3** summarizes the sesquiterpene compounds identified by GC-MS analysis of headspaces from slant and liquid mycelial cultures, and *in vitro* transformation of FPP by latex samples of the various *Lactarius* strains tested. **Supplemental Figures S3.1 and S3.2** show the MS spectra of all identified compounds for these analyses.

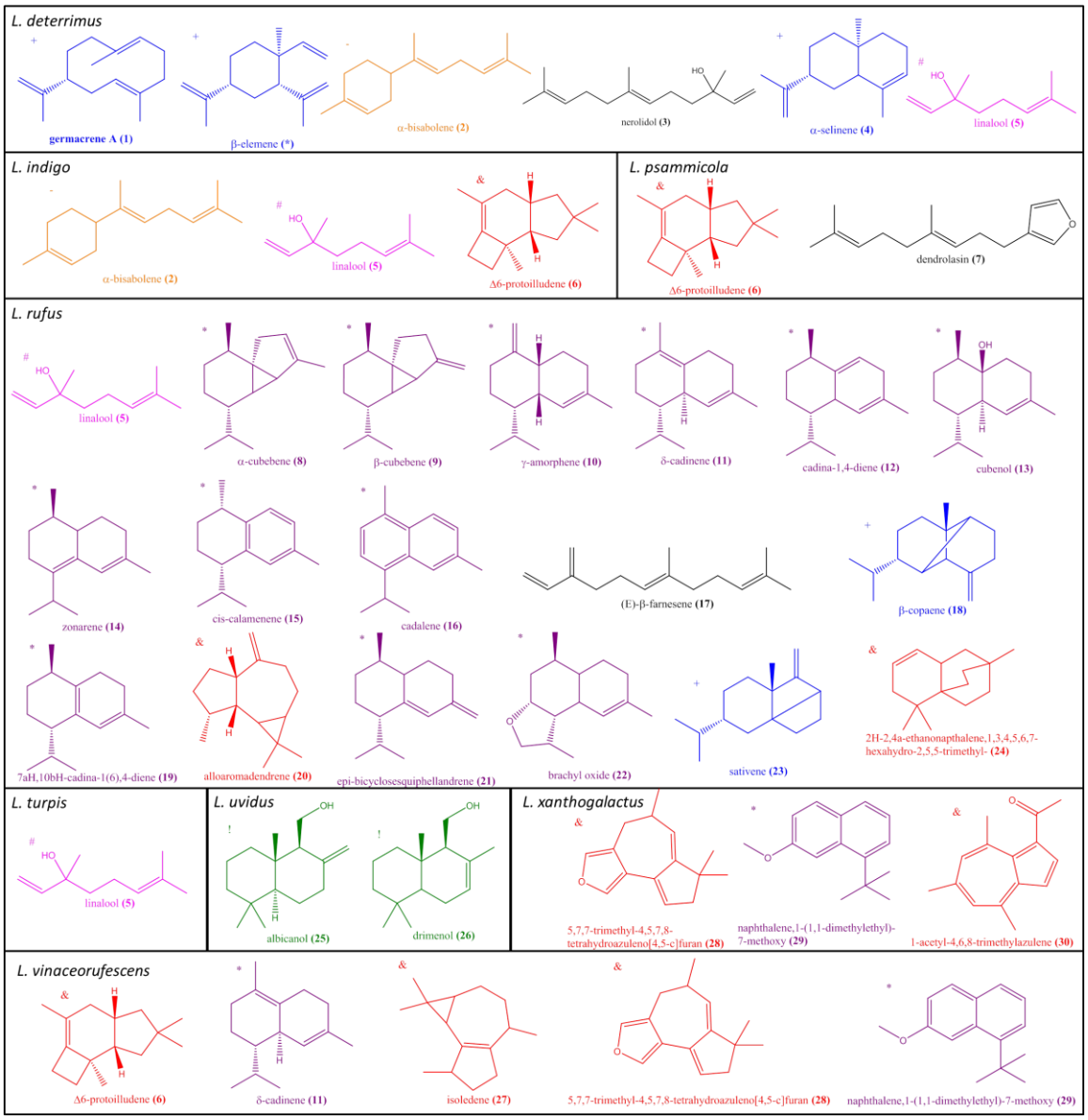


Figure 3.3 Identified mono- and sesquiterpenoids from *Lactarius* mycelium cultures and latex. Compounds are colored according to their proposed cyclization mechanism. Orange (-) scaffolds represent bisabolyl cation derived sesquiterpenoid. Purple (*) scaffolds represent *Z,E*-germacradienyl cation derived sesquiterpenoids. Blue (+) scaffolds represent *E,E*-germacradienyl derived sesquiterpenoids. Red (&) scaffolds represent *trans*-humulyl derived sesquiterpenoids. Green (!) scaffolds represent drimane sesquiterpenoids and pink (#) scaffolds represent monoterpenes.

For *L. deterrimus* only β-elemene**(*),** the heat based Cope rearrangement of germacrene A**(1)** (the major product of the 1,10-cyclized *E,E*-germacradienyl compound), was detected after 19 days as an MMN slant culture, while only α-

bisabolene**(2)** was detected at day 115 in the same culture. On Rich media slants, β-elemene and nerolidol**(3)** were detected at 35 days and 75 days, and α-selinene**(4)** only after 75days of cultivation. No volatile sesquiterpenes were detected in any liquid cultures for *L. deterrimus*, except for the monoterpene linalool**(5)** and β-elemene after prolonged cultivation of 118 days in MMN.

As previously stated, β-elemene is a heat induced Cope rearrangement of germacrene A, the major product of the 1,10-cyclized *E,E*-germacradienyl cation. α-selinene is also derived from germacrene A. However, the presence of nerolidol and α-bisabolene indicate the presence of STSs utilizing (3R)-nerolidyl pyrophosphate, in this case to make the 1,6-cyclized bisabolyl cation. Linalool is a desired monoterpene used in many commercial fragrances due to its flowery yet spicy odor, and has been found in many plants and fungi [121].

Much of the literature for *L. deterrimus* focuses more on the alkaloid, phenol and flavonoid content of *L. deterrimus* [122-126] but guaiane sesquiterpenoids have been isolated from *L. deterrimus* before [79]. No volatile guaiane sesquiterpenoids were found in the headspace, indicating they are either not produced under these conditions or are so heavily modified that they are no longer volatile. However, β-elemene was produced, indicating the presence of the guaiane precursor *E,E*-germacradienyl cation.

For *L. indigo*, no identifiable peaks were present at 19 days in the MMN slant, but α-bisabolene**(2)** was found at 115 days in MMN. Additionally, Δ6-protoilludene**(6)** was identified at 34 and 75 days in the Rich culture slant, while small amounts of the

monoterpene linalool**(5)** were detected only after 75 days. No volatile sesquiterpenes were detected in any liquid culture of *L. indigo*.

The literature of *L. indigo* biochemical characterization included the identification of linoleic acid amongst the methyl esters of free fatty acid esters [127], which supports the identification of linalool in the Rich culture headspace. Additionally, the blue color of *L. indigo* is explained by the presence of (7-isopropyl-4-methylazulen-1-yl)methyl stearate, which contains a guaiazulene sesquiterpene component [128]. While it was expected to see a 1,11-cyclization product in Δ6-protoilludene, and even a bisabolyl derived product, the lack of an *E,E*-germacradienyl cation derived product was surprising, as it would indicate the presence of guaiane precursors.

Production of volatile compounds by *L. psammicola* was consistent across all headspace analyses with the sesquiterpenes Δ6-protoilludene**(6)** and dendrolasin**(7)** identified throughout cultivation in MMN and Rich slants. Sesquiterpene identification was less consistent with liquid cultures, with Δ6-protoilludene and dendrolasin only present at 54 days in the Rich liquid culture. Dendrolasin is an interesting sesquiterpenoid, as it has previously been isolated from diverse sources including the ant, *Lasius fullginosus,* sweet potato, the wood oil of *Torreya nucifera*, and the sponges *Oligoceras hemorrhages* and *Spongia mycofijiensis* [129]. Literature of *L. psammicola* is quite limited involving its biochemical characterization, though a major distinguishing feature of *L. psammicola* is its strongly acrid taste [130], indicating a possible presence of marasmane, lactarane or secolactarane dialydehydes found in other acrid tasting species [78].

The most diverse volatile sesquiterpene product profile was seen from *L. rufus* cultures. An abundance of sesquiterpenoids derived from the 1,10-cyclized *Z,E*-germacradienyl cation were identified from many of the *L. rufus* cultures. Within the solid MMN slant cultures, α-cubebene**(8)**, β-cubebene**(9)**, γ-amorphene**(10)**, δ-cadinene**(11)**, cadina-1,4-diene**(12)**, cubenol**(13)**, zonarene**(14)**, *cis*-calamenene**(15)**, cadalene**(16)** and linalool**(5)** were all identified. With the exception of the monoterpene linalool, all were products derived from the *Z,E*-germacradienyl cation.

A similar array of volatile sesquiterpenoids was made by *L. rufus* grown on Rich slants. In the Rich slants, α-cubebene, β-cubebene, δ-cadinene, zonarene, cadina-1,4-diene and cubenol were identified as previously, with additional identifications of (*E*)-β-farnesene**(17)**, β-copaene**(18)**, 7aH,10bH-cadina-1(6),4-diene**(19)**, alloaromadendrene**(20)**, epibicyclosesquiphellandrene**(21)** and brachyl oxide**(22)**. 7aH,10bH-cadina-1(6),4-diene, epibicyclosesquiphellandrene and brachyl oxide are further *Z,E*-germacradienyl cation derived sesquiterpenoids, but β-copaene is thought to be derived from the 1,10-cyclized *E,E*-germacradienyl cation and alloaromadendrene is derived from the 1,11-cyclized *trans*-humulyl cation. (*E*)-β-farnesene is the quenched product of hydrolyzed FPP.

Similar sesquiterpenoids were identified in the *L. rufus* liquid MMN cultures. Linalool, α-cubebene, β-cubebene, β-copaene, (*E*)-β-farnesene, 7aH,10bH-cadina-1(6),4-diene, alloaromadendrene, epibicyclosesquiphellandrene, δ-cadinene, epizonarene, cadina-1,4-diene and cubebol were all identified as previously, though an additional

presence of sativene**(23)** was detected. Sativene is thought to be derived from the 1,10-cyclized *E,E*-germacradienyl cation.

Analysis of *L. rufus* latex indicated the presence of 2H-2,4a-ethanonaphthalene,1,3,4,5,6,7-hexahydro-2,5,5-trimethyl**(24)**, which is thought to be derived from the 1,11-cyclized *trans*-humulyl cation.

The literature on *L. rufus* is arguably the most substantial of the eight *Lactarius* species screened in this thesis work. A number of modified sesquiterpenes have been isolated from *L. rufus*, namely rufuslactone, lactarorufins and isolactarorufins, all *trans*-humulyl derived sesquiterpenes of the lactarane scaffold [131, 132]. While it was encouraging to find representatives of three different cyclization pathways amongst the volatile headspace of *L. rufus,* it was surprising to see the abundance of *Z,E*-germacradienyl cation derived sesquiterpenes, and lack of *trans*-humulyl derived sesquiterpenes. The reason for this discrepancy is likely for the same reasons for the *L. deterrimus* sesquiterpenoid discrepancies; either the pathways are not induced under these conditions, or the hydrocarbon scaffolds are further modified into non-volatile derivatives, or perhaps because some products are heat labile compounds.

No sesquiterpenes peaks were found in any culture of *L. turpis*, though abundant peaks of the monoterpene linalool**(5)** were consistently present throughout cultivation in MMN and Rich slants. Additionally, abundant peaks of linalool were also present after prolonged cultivation of 117 days in MMN liquid culture and 54 days in Rich liquid culture.

Much of the biochemical characterization of *L. turpis* in the literature has been performed on *L. necator*, a synonymous name for the species of mushroom. The most famous natural product isolated from *L. necator* is the alkaloid necatorone [133], but the lactarane sesquiterpenoids *epi*-piperalol [134] and 8-epi-pipertriol [135] have also been isolated.

Unlike the other *Lactarius* strains, no sesquiterpenoid peaks were detected in any of the analyzed solid and liquid headspaces of *L. uvidus* cultures. However, albicanol**(25)** and drimenol**(26)** were found in the headspace of *L. uvidus* latex incubated with FPP. Albicanol and drimenol are important sesquiterpenoids that represent the drimane scaffold. Indeed, drimenol, albicanol, and uvidins A and B from *Lactarius uvidus* have all been isolated and described previously in the literature [136-138]. The confirmed presence of drimane sesquiterpenes in the *L. uvidus* latex made *L. uvidus* relevant for later study.

Because no sesquiterpenoids were produced by *L. uvidus* under the tested cultivation conditions, 3 additional types of complex plant-based media, rich in carbon and nitrogen, were tested for inducing the production of sesquiterpenes in subsequent *L. uvidus* cultures. The use of these media types was based on studies detailing induction of fungal unspecific peroxygenases in respective strains *Agrocybe aegerita* [97], *Coprinus radians* [98], and *Marasmus rotula* [99] (see Methods for media recipes) when grown with complex plant-based media. *L. uvidus* was therefore grown in media “A” [97], “C” [98] and “M” [99]. However, no visible mycelium growth of *L. uvidus* was observed in any of the solid or liquid cultures of these media.

The volatile headspace of *L. vinaceorufescens* of all solid slant cultures consistently contained Δ6-protoilludene**(6)**, found at 115 days in MMN and 35 and 76 days in Rich media. Only the headspace of Rich cultures of *L. vinaceorufescens* contained Δ6-protoilludene, in addition to δ-cadinene**(11)** and a further 1,11-cyclized *trans*-humulyl derived sesquiterpene, isoledene**(27)**.

In vitro assays with the latex of *L. vinaceorufescens,* however, yielded two sesquiterpenoids identified as: 5,7,7-trimethyl-4,5,7,8-tetrahydroazuleno[4,5-c]furan**(28)** (abbreviated as 5TTF) and 1-(1,1-dimethylethyl)-7-methoxy-napthalene**(29)** (abbreviated as 1-7MN). 5TTF also represents a *trans*-humulyl derived sesquiterpenoid, while 1-7MN appears to be derived from a *Z,E*-germacradienyl derived scaffold. These compounds were also identified in the closely related *L. xanthogalactus* latex assays, though no sesquiterpene peaks were found in any of the solid or liquid culture headspaces of *L. xanthogalactus*. In addition to 5TTF and 1-7MN, 1-acetyl-4,6,8-trimethylazulene**(30)**, another *trans*-humulyl cation derivative, was identified in the *L. xanthogalactus* latex assays.

Literature involving the biochemical characterization of *L. vinaceorufescens* and *L. xanthogalactus* is scarce. However, furan sesquiterpenoids have genuinely been proven to be present in the closely related species *L. chrysorrheus*, including an abundance of lactaranes and secolactaranes [139]. 5TTF found in both *L. vinaceorufescens* and *L. xanthogalactus* latex was very similar to the identified furanolactaranes from *L. chrysorrheus*.

3.3 Genome sequencing and analysis of *Lactarius uvidus*

As the presence of drimenol was confirmed in the *L. uvidus* latex, a genomics driven approach (similar to the approach that led to the identification of the protoilludene synthases in the anticancer illudin producer *O. olearius*) was chosen to characterize the STS complement of this fungus with the hope of identifying a drimane STS.

In order to proceed with identification, gDNA from pure cultured *L. uvidus* was subjected to next generation sequencing on an Illumina HiSeq2500 platform at the Mayo Clinic (Rochester, MN). This resulted in raw data of roughly 190 million paired end reads at 101bp in length.

These 190 million reads were refined to roughly 156 million paired end reads via strict quality control measures on the Galaxy platform made available through the Minnesota Supercomputing Institute (MSI). By estimating a genome size of 60 Mbp, the 156million paired end reads at 80bp in length estimated approximately 207-fold total sequence coverage. The processed reads of *L. uvidus* were assembled into 33,058 scaffolds, with an N50 of 3,969 and a total assembly of 60,078,130 bp including gaps. The largest scaffold in the assembly measured 51,876 bp.

By running RepeatMasker [106] on the assembly, 2.8% of the genome was estimated to contain repetitive elements, with the biggest contributor being long-terminal repeats at 2.14%. Open reading frames (ORFs) were predicted on the masked genome with Augustus version 3.0.2[107], using *Laccaria bicolor* as the gene splicing model. The genome contained 26,157 predicted protein-ORFs. This number is almost certainly artificially high due to the dikaryotic nature of gDNA that was isolated from *L. uvidus*.

Indeed, if the number of predicted genes were superficially halved as a monokaryotic genome prediction (30Mbp), they would more closely resemble the number of predicted genes of other Basidiomycota in that genome size range (*Coprinus cinereus*, 38Mbp, ~13,000 predicted genes) [176]. **Table 3.1** below summarizes the assembly of the *L. uvidus* genome in comparison to the genome assembly of *Omphalotus olearius* previously performed by the Schmidt-Dannert laboratory.

Table 3.1 Characteristics of genome assemblies of *L. uvidus* and *O. olearius*.

Parameter	*L. uvidus*	*O. olearius*
Genome Size	60.08 Mbp	28.15 Mbp
Scaffold Number	33,058	868
Scaffold N50	3,969	199,357
Predicted Protein Coding Genes	26,157	8,172

In comparison to the monokaryotic genome assembly of *O. olearius*, the maximum scaffold size of *L. uvidus* was about 4 times smaller than the N50 of *O. olearius*, while the straight comparison of N50s differs by 2 orders of magnitude. Additionally, the number of scaffolds of *L. uvidus* was 2 orders of magnitude greater than *O. olearius*. Much of the discrepancy between the two genomes can be attributed to the dikaryotic nature of gDNA extracted from *L. uvidus* for genome assembly, whereas the *O. olearius* was obtained from a monokaryotic haploid strain. Though the assembly of *L. uvidus* was fragmented, especially compared to *O. olearius*, scaffolds were still large enough to mine for sesquiterpene synthases (STSs), and manually annotate some biosynthetic gene clusters using BLAST.

3.4 Preliminary bioinformatics analysis of putative *Lactarius uvidus* STSs

Using a local BLAST search, with characterized STSs from *O. olearius, S. hirsutum* and *C. cinereus* as query sequences, 23 initial putative STSs were identified in *L. uvidus* However, through a multiple sequence alignment using Muscle in the MEGA suite, the number of putative STSs was manually refined to remove STS sequences lacking the conserved metal binding motifs or not having the characteristic length of other characterized STSs (see **Supplemental Figure S3.3** for a snapshot of the multiple sequence alignment (MSA) and **Supplemental Table S3.1** for details on gene hits). This manual refinement reduced the putative STS count to 12. Phylogenetic analysis using MAFFT, ProtTest and PhyML yielded the following maximum likelihood tree found in **Figure 3.4**.

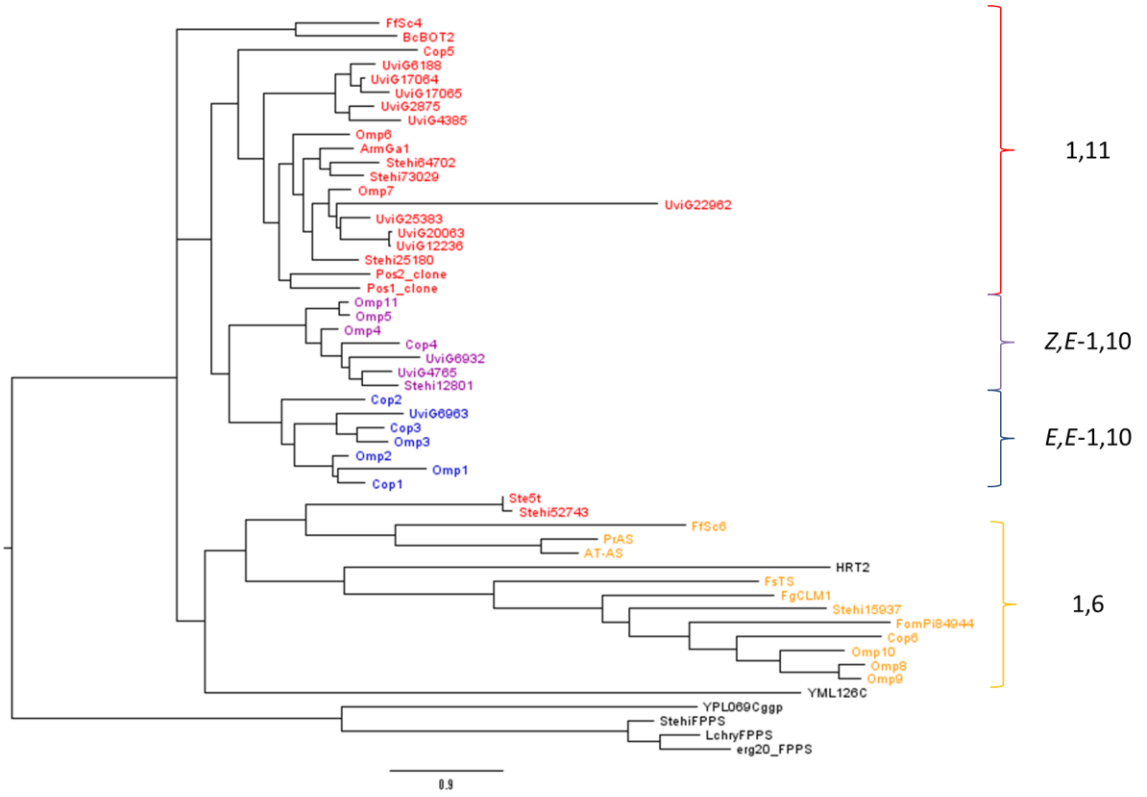
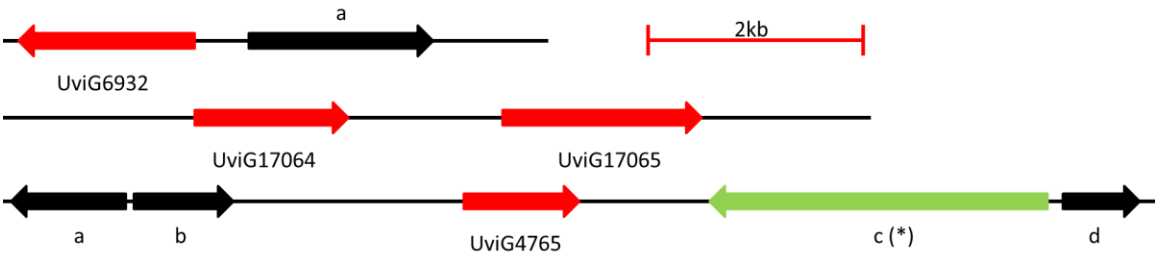


Figure 3.4 Maximum likelihood tree of putative *L uvidus* STSs. STSs are colored according to their inferred cyclization mechanism. Red represents 1,11-cyclization mechanism of 2*E*,6*E*-farnesyl pyrophosphate (*E*,*E*-FPP). Purple represents 1,10-cyclization of (3R)-nerolidyl pyrophosphate (NPP). Blue represents 1,10-cyclization of *E*,*E*-FPP, while orange represents 1,6-cyclization of NPP. Genes of *L. uvidus* are denoted with the notation "UviG" followed by the predicted gene number in the assembly.

Many of the putative *L. uvidus* STSs were grouped into the 1,11-cyclizing clade on the maximum likelihood tree. This finding is not surprising, given the plethora of *trans*-humulyl cation derived scaffolds that have been previously isolated from *Lactarius* species. It is interesting though that many of the putative 1,11-cyclizing STSs of *L. uvidus* form their own sub-clade. Additionally, three STSs grouped in with the germacradienyl cation forming STSs, with genes UviG6932 and UviG4765 resembling Cop4 and UviG6963 resembling Cop3.

Most of the particular scaffolds containing the putative STSs for *L. uvidus* were only large enough to house the predicted STS. However, scaffolds over 5kb in size were

manually annotated by BLAST[108] using the non-redundant protein database at NCBI to determine the presence of any biosynthetic gene clusters. **Figure 3.5** illustrates the scaffolds containing putative STSs UviG4765, UviG6932, UviG17064 and UviG17065.



UviG6932 Scaffold Cluster

Gene	Reference	Top Conserved Domain Hit
UviG6932		
a	NODE_44166_length_5089_cov_70.572609:g2.t1	Recombination factor (PRK13342)

UviG17064 & UviG17065 Scaffold Cluster

Gene	Reference	Top Conserved Domain Hit
UviG17064		
UviG10765		

UviG4765 Scaffold Cluster

Gene	Reference	Top Conserved Domain Hit
a	NODE_30139_length_10936_cov_69.113571:g1.t1	WD40 (cd00200)
b	NODE_30139_length_10936_cov_69.113571:g2.t1	RNA Polymerase Rbp5 (PLN03111)
UviG4765		
c	NODE_30139_length_10936_cov_69.113571:g3.t1	(*) cystathionine beta-lyase (PLN02509)
d	NODE_30139_length_10936_cov_69.113571:g4.t1	HIT zinc-finger (pfam04438)

Figure 3.5 Manually annotated gene clusters of scaffolds containing putative *L. uvidus* sesquiterpene synthases. Predicted ORFs are colored according to their function. Red represents sesquiterpene synthases while green (*) represents hydrocarbon scaffold-modifying enzymes.

It can be seen from **Figure 3.5** that most of the conserved domain hits found near the predicted STS genes are not like any of the biosynthetic genes annotated in the cluster of Stehi1|73029 from Chapter 1. Specifically, no cytochrome P450 enzymes were present and few hydrocarbon scaffold-modifying enzymes were predicted. The results (represented by **Figure 3.4** and **Figure 3.5**) obtained from this first assembly show that the assembly needs to be further improved to obtain larger scaffolds.

Further details of different assembly strategies will be discussed in the Conclusions section of this chapter.

3.5 Preliminary cloning of *L. uvidus* sesquiterpene synthases

Considering the scaffold for UviG4765 was the largest scaffold, and included a predicted gene encoding an enzyme (UviG4765c) involved with isoprenoid precursor biosynthesis pointing to its relevance for the fungus, the predicted STS was selected as a preliminary candidate for cloning from *L. uvidus* cDNA for characterization.

UviG6963 was also selected as a candidate because of the presence of an asparagine instead of a second aspartate within the well conserved D(D/E)xx(D/E) metal binding motif of STSs, which may point to a novel cyclization activity of this enzyme (see **Supplemental Figure S3.3** for a multiple sequence alignment snapshot). In addition, both UviG6963 and UviG4765 clustered most closely with Cop3 and Cop4, which cyclize FPP into a germacradienyl cation, a hypothesized carbocation intermediate to the drimane scaffold.

Libraries of cDNA were prepared from the cap, gills and stem from *L. uvidus* fruiting bodies. Gene specific primers for the ends of UviG4765 and UviG6963, along with gene specific primers for the conserved D(D/E)xx(D/E) and NSE/DTE motifs were used for PCR reactions with the *L. uvidus* cDNA libraries. Weak positive PCR reactions, as seen in **Figure 3.6,** were present for the unspliced versions of the STSs for the cap tissue of UviG6963 and the stem tissue for UviG4765.

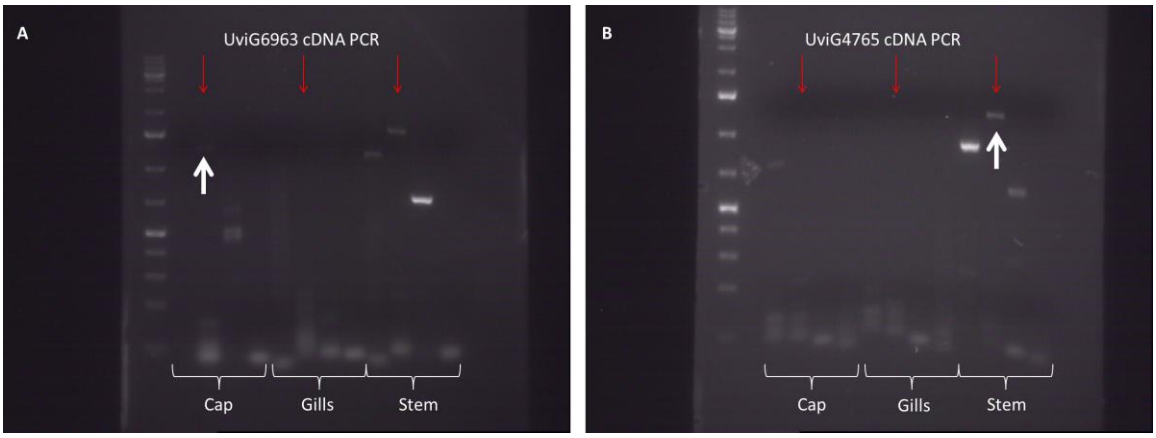


Figure 3.6 Initial PCR screens of putative *L. uvidus* STSs from cDNA libraries. Red arrows indicate expected full length transcripts of putative STSs. The thick white arrows indicate amplicons purified for blunt cloning.

These PCR products were gel purified and blunt cloned into pCR™-BluntII-TOPO® plasmids. Sanger sequencing confirmed correct cloning of unspliced UviG6963 and UviG4765 from each amplicon reaction. Removal of predicted introns from these genes by site-directed mutagenesis is ongoing.

4 Conclusion

Research carried out in this chapter provided the basis for genomics enabled analysis of the biosynthetic potential of *Lactarius* fungi. By adapting cultivation methods previously used in growth study of *Lactarius* species [120], it was determined that MMN media provided the most suitable media for fastest *Lactarius* growth for downstream applications. Additionally, *L. rufus* was determined to be the fastest growing species amongst those sampled, which along with its volatile headspace profile, furthered our interest in the species. Further analysis of *L. rufus* will be outlined in Chapter 4.

Moreover, the volatile headspace from solid and liquid cultures, were sampled from 8 different *Lactarius* species. The headspace analysis was complemented with *in vitro*

assays of latex extracts from available fruiting bodies. Drimane sesquiterpenes, as well as compounds representing all four confirmed cyclization mechanisms for STSs, were identified from these screens. Additionally, the monoterpene linalool was identified throughout various *Lactarius* cultures. No *bona fide* fungal monoterpene synthase has been characterized yet [43]. Sequencing and mining the genome of one of these *Lactarius* species, especially where linalool was majorly abundant, like *L. turpis*, potentially provides an opportunity to identify the first monoterpene synthase in fungi.

While most of the compounds identified in the screens could not be corroborated with the literature, it is important to remember that the samples came from mycelial culture headspaces. Most sesquiterpenoids identified in the literature were extracted from intact fruiting bodies [78, 85, 87, 140-145]. The context for chemical defense is quite different between mycelium and fruiting bodies, where mycelia have to out-compete other microorganisms for valuable soil resources, while fruiting bodies must protect themselves from parasites and predators. Considering the theory of how most *Lactarius* sesquiterpenoids remain as inactive fatty acid esters until fruiting body injury [78, 83, 139], it is not surprising that so few volatile sesquiterpenes were detected in the mycelium headspaces. Furthermore, there is often great variability between results reported from laboratories extracting sesquiterpenoid compounds of the same species due to different extraction and isolation procedures used [78, 139].

However, only the volatile headspaces of *Lactarius* were sampled. Much of the sesquiterpenoids identified from *Lactarius* in the literature are much more heavily modified and much less volatile. In the future sesquiterpenoid screens of *Lactarius*,

hexane or ethyl acetate extracts from the liquid media itself and from intact fruiting bodies should be performed to identify more of the modified sesquiterpenoid scaffolds. Despite this limitation, a number of relevant natural product scaffolds were identified in these species.

The presence of the drimenol and albicanol drimanes in *L. uvidus* latex prompted the further investigation of the *L. uvidus* genome via Illumina sequencing. While the *L. uvidus* genome assembly was not assembled into very large scaffolds, especially when compared to *O. olearius*, and the Illumina short-read assemblies of *L. rufus* and *L. vinaceorufescens* in chapter 4, the scaffolds were large enough to mine for putative sesquiterpene synthases. However, the numerous small scaffolds in the *L. uvidus* genome prompted the adjustment of obtaining long read sequencing data via the PacBio platform (Pacific Biosciences, Menlo Park, CA). At the time of writing this thesis, PacBio data for *L. uvidus* sequencing is still being assembled onto the Illumina reads via the scaffold assembly program PBJelly2[105]. The combined long and short read data of *L. uvidus* should provide larger contigs, especially those involving STSs, to allow for greater annotation of biosynthetic gene clusters.

Finally, unspliced candidate genes, UviG4765 and UviG6963, have been successfully cloned from *L. uvidus* cDNA into pCR™-BluntII-TOPO® plasmids. The correction of these unspliced genes into active spliced variants using Q5 site directed mutagenesis kits (New England Biolabs, Inswich, MA) is ongoing. Further cloning of other putative STSs, particularly those utilizing the *trans*-humulyl cation mechanism will occur in the near future.

5. Supplemental Material

Table S3.1 Initial BLAST hits for putative sesquiterpene synthases for *L. uvidus*. "_length_" in the "Reference" column refers to the length of the assembled scaffold the gene was present on. Subsequent phylogenetic analysis was performed on the bolded genes colored red.

Gene #	Reference	Gene Length	D(D/E)xx(D/E)?	NSE/DTE?
UviG155	>NODE_1082_Gene_155_length_465_cov_33.838711	71	Y	N
UviG17064	**>NODE_117733_Gene_17064_length_8199_cov_38.691181**	**371**	**Y**	**Y**
UviG17065	**>NODE_117733_Gene_17065_length_8199_cov_38.691181**	**476**	**Y**	**Y**
UviG19572	>NODE_138562_Gene_19572_length_3339_cov_64.187782	765	N	N
UviG20063	**>NODE_142772_Gene_20063_length_3976_cov_36.303570**	**323**	**Y**	**Y**
UviG22962	>NODE_170512_Gene_22962_length_4068_cov_34.301624	306	N	N
UviG2875	**>NODE_18718_Gene_2875_length_3101_cov_34.397938**	**342**	**Y**	**Y**
UviG25383	**>NODE_200568_Gene_25383_length_830_cov_19.197590**	**228**	**Y**	**Y**
UviG4385	**>NODE_27867_Gene_4385_length_4432_cov_67.970894**	**346**	**Y**	**Y**
UviG4425	>NODE_28040_Gene_4425_length_1711_cov_72.249565	120	N	N
UviG4765	**>NODE_30139_Gene_4765_length_10936_cov_69.113571**	**341**	**Y**	**Y**
UviG4994	>NODE_31709_Gene_4994_length_6895_cov_52.251343	1216	Y	Y
UviG6188	**>NODE_39123_Gene_6188_length_1668_cov_51.140888**	**278**	**Y**	**Y**
UviG6932	**>NODE_44166_Gene_6932_length_5089_cov_70.572609**	**341**	**Y**	**Y**
UviG6963	**>NODE_44341_Gene_6963_length_1953_cov_67.782898**	**385**	**Y**	**Y**
UviG7472	>NODE_47565_Gene_7472_length_528_cov_72.543564	120	Y	N
UviG9536	>NODE_61243_Gene_9536_length_8493_cov_36.179794	781	Y	Y
UviG10082	>NODE_64946_Gene_10082_length_7602_cov_35.754013	87	N	N
UviG10947	>NODE_70972_Gene_10947_length_9699_cov_35.228374	97	Y	N
UviG12236	**>NODE_80434_Gene_12236_length_1704_cov_22.998827**	**392**	**Y**	**N**
UviG13028	>NODE_85876_Gene_13028_length_24285_cov_32.516411	175	N	Y
UviG13030	>NODE_85876_Gene_13030_length_24285_cov_32.516411	88	N	N
UviG13650	**>NODE_90258_Gene_13650_length_7758_cov_34.644882**	**351**	**Y**	**Y**

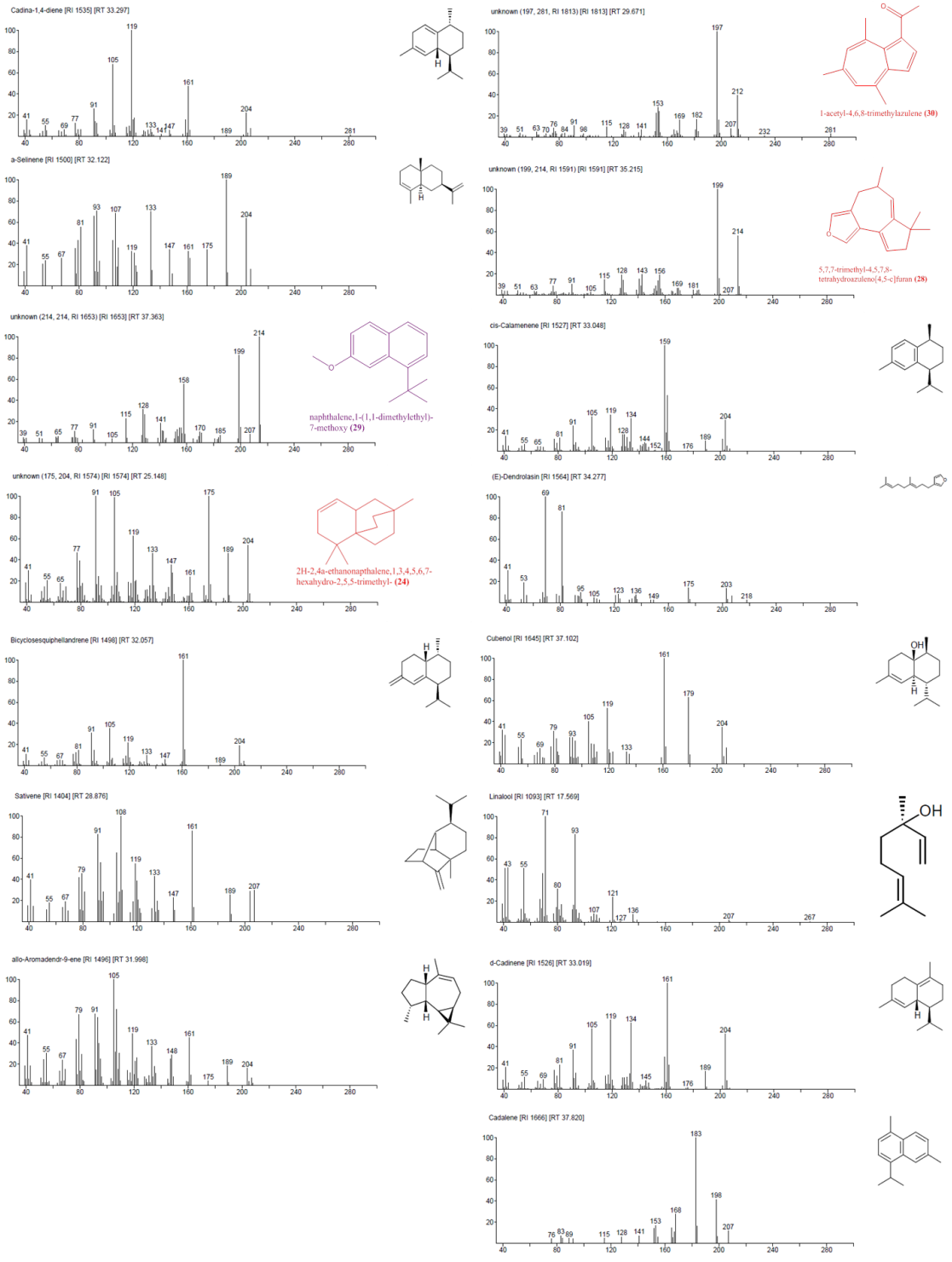


Figure S3.1 Mass spectra of identified mono- and sesquiterpenoid peaks from *Lactarius* mycelial culture headspace and latex assays.

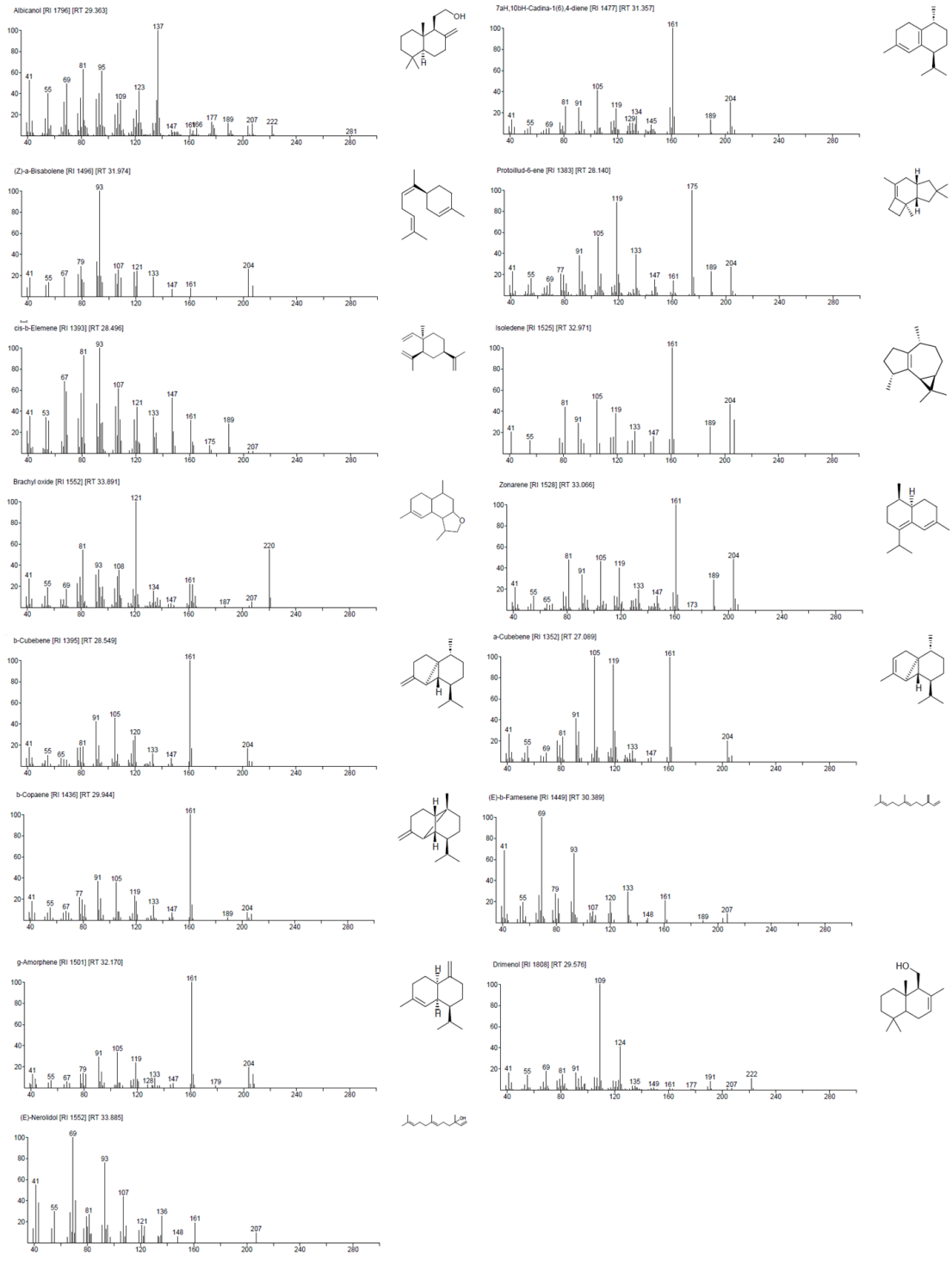


Figure S3.2 Mass spectra of identified mono- and sesquiterpenoid peaks from *Lactarius* mycelial culture headspace and latex assays.

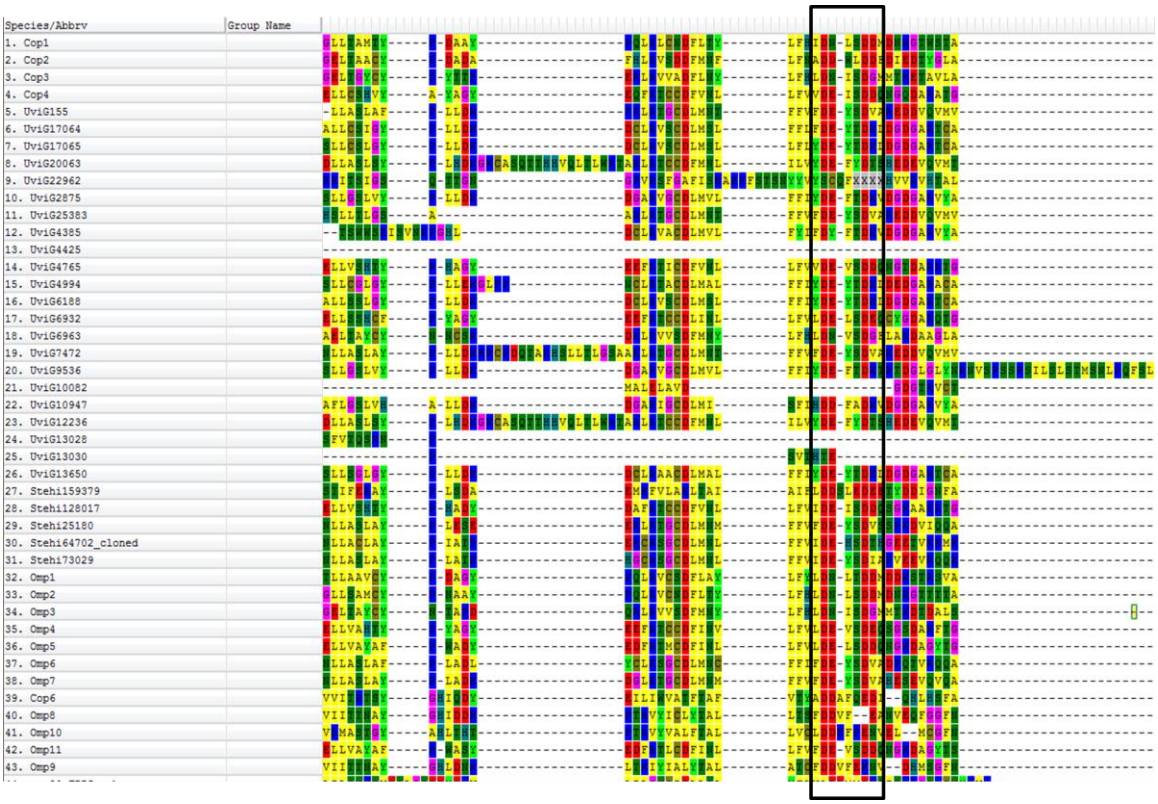


Figure S3.3 Snapshot of multiple sequence alignment of putative *Lactarius* STSs. D(D/E)xx(D/E) motif boxed in black.

Chapter 4: Towards the genome assemblies of *Lactarius rufus* and *L. vinaceorufescens* for higher molecular weight isoprenoid biosynthesis elucidation

1. Introduction

1.1 The state of the natural rubber industry

Natural rubber, or *cis*-1,4-polyisoprene, is a biopolymer of vast importance in the modern industrial world,[47, 146, 147] particularly because of its superior properties to synthetic rubber. These properties include resilience, elasticity, abrasion resistance, impact resistance, heat dispersion, and cold temperature malleability [146, 147]. Natural rubber is used in well over 40,000 commercial products including aircraft tires, surgical gloves and medical devices [47]. The superior qualities of natural rubber have allowed it to remain between 30 and 40% of the U.S. market share over the past 30 years [47].

The cost in US dollars per kg of natural rubber has been less consistent over the same time period. **Figure 4.1** below represents the cost of natural rubber in US dollars per kg over the past 30 years. From 1985 to 2005, the cost of natural rubber had mostly maintained a price of approximately $1 per kg, with slight spikes in price to $1.46 per kg in June of 1988 and $1.84 per kg in April 1995. However, by March 2005, the price of natural rubber followed a large general trend in increasing price to $6.18 per kg in February 2011. This spike in price has since recovered to $1.32 per kg at the time of writing this thesis[148]. However, the large price hike between 2006 and 2011 was one

reason that facilitated the interest in exploring alternative natural rubber producing sources.

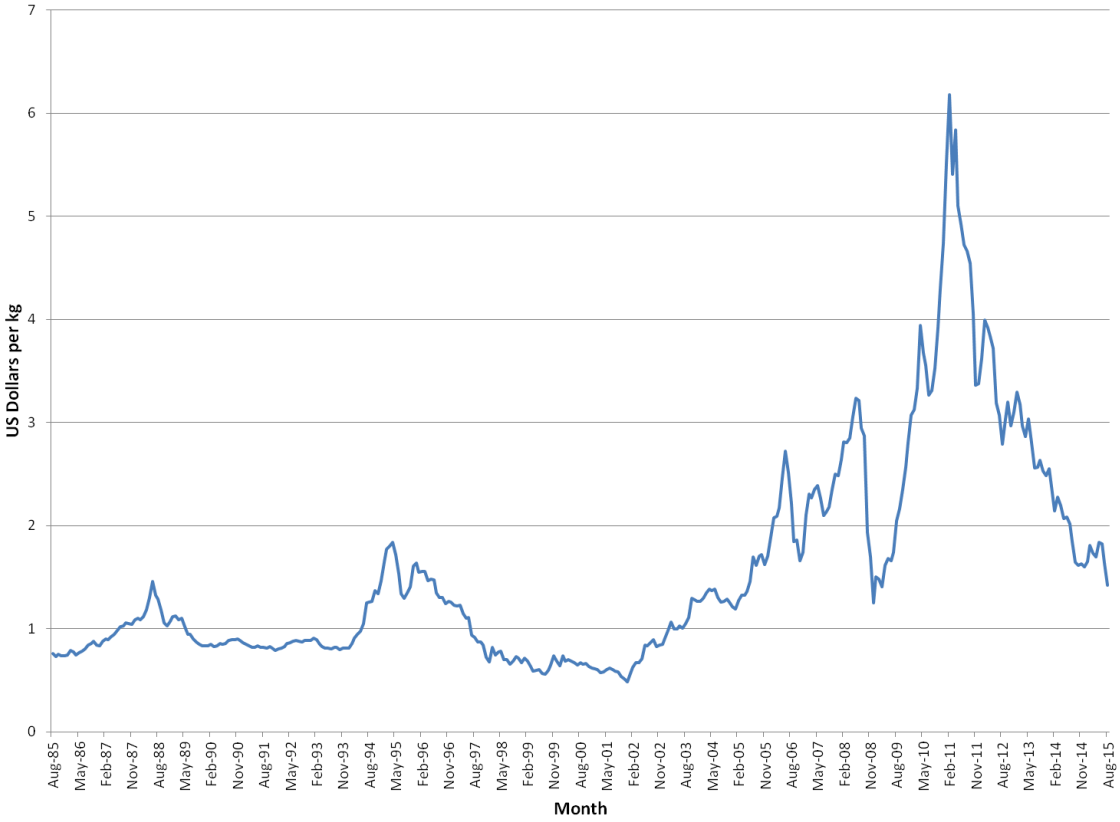


Figure 4.1. Price of natural rubber in US dollars per kg over the past 30 years.

The call for alternative producing platforms of natural rubber intensified between 2006 and 2011. Greater than 90% of all natural rubber is produced from the Brazilian Rubber Tree, *Hevea brasiliensis*; mostly from plantations in Southeast Asia. The prevalence of South American Leaf Blight (SALB) precludes the commercial production of natural rubber from *Hevea* trees in South America [149, 150]. This disease, caused by the fungus *Microcylus ulei*, has a high risk of spreading to Southeast Asia as well [47]. The high risk of complete loss due to crop failure, along with increasing demand in

rapidly expanding economies, has intensified efforts towards production of natural rubber from alternative sources.

Another drawback towards natural rubber production with *H. brasiliensis* includes the emergence of allergic reactions to co-purified proteins associated with *H. brasiliensis* latex, with between 1-6% of the general population affected [47, 151]. Additionally, *H. brasiliensis* plantations are experiencing competition for valuable land against palm-oil plantations, which profit better from the biofuel business [152]. Finally, the daily physical tapping of *Hevea* trees by hand is a labor intensive process [47].

While natural rubber from *H. brasiliensis* dominates the natural rubber commercial market, it is far from the only organism to produce it, with 2500 other plant species [153] and *Lactarius* mushrooms [48, 154] also known to produce natural rubber. The two most promising alternative natural rubber producing plants include the guayule plant, *Parthenium argentatum,* and the Russian dandelion, *Taraxacum koksaghyz* [47]. However, what distinguishes the natural rubber of *H. brasiliensis*, *P. argentatum* and *T. koksaghyz* from other plants sources is their propensity to produce large quantities of high molecular weight rubber [155], which provides a higher quality product [146, 156].

1.2 The synthesis of higher molecular weight isoprenoids

There has also been a push towards engineering microbial fermentation platforms for the production of natural rubber utilizing cloned genes involved in natural rubber biosynthesis from *Hevea* and some *Taraxacum* species (with further polyisoprene synthesis characterization from *Arabidopsis thaliana* and *Saccharomyces cerevisiae*

species) [157-163]. However, understanding of natural rubber biosynthesis at the molecular scale is still poor, hindering progress for a microbial fermentation approach. While the molecular mechanisms of rubber biosynthesis are not well understood, the basic enzymes involved have been identified. Natural rubber itself is the biopolymer *cis*-1,4-polyisoprene, belonging to the isoprenoid class of natural products. A number of *in vitro* studies have shown that FPP appears to be the initiating molecule for *cis*-1,4-polyisoprene biosynthesis [146, 154, 164]. Continual additions of IPP onto the initiator FPP molecule lead to the *cis*-1,4-polyisoprene biopolymer by a *cis*-prenyl transferase (CPT) enzyme [157]. CPTs are further characterized with 5 conserved amino acid regions CSI, CSII, CSIII, CSIV and CSV [157, 163].

Attempts to synthesize natural rubber from cloned *Hevea, Taraxacum, Arabidopsis* and *Saccharomyces* CPTs have not succeeded [157-162], producing polyisoprene with no greater than 20 isoprene units. In contrast, polyisoprene from *Hevea* latex contains up to 18000 isoprene units in a single polymer [47]. However, a cloned CPT from *Hevea* exhibited greater activity for synthesizing higher molecular weight polyisoprene by the addition of a latex fraction to it in *in vitro* assays [165, 166]

CPT enzymes are thought to be bound to rubber particles while they polymerize IPP units onto FPP. However, recent evidence shows that the CPT enzyme generally exists more as a cytosolic enzyme, and is instead recruited to the endoplasmic reticulum of cells by a CPT-like protein (human Nogo-B receptor homolog) for polyisoprene synthesis [167]. This data was further corroborated with a similar mechanism proposed from co-expressing *Taraxacum brevicorniculatum* CPTs with TbRTA, also a human Nogo-B

receptor homolog [168]. This data supports the hypothesis that other rubber associated proteins are necessary for long chain polyisoprene production.

Identification of additional rubber particle associated proteins has proven to be difficult in plants though, as biosynthetic genes in plants are difficult to locate, despite exhaustive genomic, transcriptomic and proteomic approaches [169-171].

1.3 *Lactarius* mushrooms also produce *cis*-1,4-polyisoprene

As reviewed in Chapter 1 of this thesis, a number of *Lactarius* mushroom species have also been proven to produce *cis*-1,4-polyisoprene [48, 154, 164]. That study (Tanaka *et al.* 1994) was able to demonstrate that the tested *Lactarius* mushroom species produced *cis*-1,4-polyisoprene with roughly 300 isoprene units. This is still two orders of magnitude below the 18,000 isoprene units found in *Hevea* latex, but is an order of magnitude better than *in vitro* rubber biosynthesis to date. However, an additional study characterizing the latex of six different *Lactarius* mushrooms provided evidence that latex extracts showed *in vitro* rubber transferase activity in the presence of *E,E*-FPP [164], thereby linking it to plant rubber biosynthesis by merit of the shared priming molecule.

A unique advantage that *Lactarius* mushrooms provide in elucidating natural rubber biosynthesis is the propensity for mushrooms to physically cluster related biosynthetic genes together in their respective genomes [24, 38, 43, 46, 172], unlike plants. By identifying *cis*-polyisoprene biosynthetic genes in *Lactarius*, namely CPTs, other genes physically related to polyisoprene biosynthesis should be clustered nearby. Furthermore, it is hypothesized that an FPP synthase (FPPS) would likely be clustered closely to a CPT

in the *Lactarius* genome. FPPSs can also readily be identified via genome mining by virtue of having two DDxxD motifs. By characterizing the relevant biosynthetic gene clusters of *Lactarius* species, the foundation could potentially be set for finding genes with similar functions in plants like *Hevea*.

1.4 Continuing genome mining fundamentals for *cis*-prenyl transferases

As part of a collaboration with Bridgestone Tires Inc., it was the goal of this study to adapt the *in silico* genome mining framework for identifying fungal STSs established by the Schmidt-Dannert laboratory to mine for CPTs and FPPSs in order to better understand natural rubber biosynthesis. Identification of related genes physically clustered together with CPTs and FPPSs in *Lactarius* species would possibly reveal homologous rubber associated proteins in plants. *Lactarius rufus* was selected for this study because of its relatively fast growth rate compared to the other cultivated *Lactarius* species in this thesis (see Chapter 3 for more details). *L. vinaceorufescens* and *L. xanthogalactus* were also selected to use in this study due to their close taxonomic relationship with *L. chrysorrheus*, a *Lactarius* species proven to contain *cis*-1,4-polyisoprene in its latex [48, 154, 173]. Genomic analysis was also complemented by proteomics data obtained from latex extracts of various *Lactarius* species. The predictive framework established by the Schmidt-Dannert laboratory was also utilized to further discover novel STSs in *Lactarius* species.

2. Materials and Methods

2.1 *Lactarius rufus, L. vinaceorufescens* and *L. xanthogalactus* cultivation and tissue prep

Establishment of *L. vinaceorufescens* and *L. xanthogalactus* MMN cultures from fruiting bodies was performed as described in Chapter 3. All genomic work on *L. rufus* in this chapter was derived from one mycelial culture given to us from Dr. Peter Kennedy's lab. Serial culture inoculations were performed as previously described in Chapter 3 by agar plug inoculations. Tissue from the cap, gills and stem from fruiting bodies were prepared for RNA extraction (previously described in Chapter 3).

2.2 *Lactarius rufus, L. vinaceorufescens* and *L. xanthogalactus* gDNA, RNA, and latex extractions

Genomic DNA extractions for *L. rufus*, and *L. vinaceorufescens* were performed as previously described in Chapter 3. Briefly, gDNA extraction was performed on lyophilized mycelium tissue cultivated from an MMN plate covered with a nitrocellulose membrane (see Chapter 3 for MMN recipe) using OmniPrep for Fungus commercial kit (G Biosciences, St. Louis, MO). Genomic DNA from *L. xanthogalactus* was directly extracted from lyophilized cap tissue. Purified gDNA samples were sent overnight on dry ice to the Molecular Biology Core at the Mayo Clinic's Medical Genome Facility (Rochester, MN). RNA extraction was carried out with the TRIzol® reagent (Invitrogen, Carlsbad, CA) on prepared fruiting body tissue, and latex extractions from fruiting body gills were performed as previously described in Chapter 3.

2.3 Illumina sequencing of *Lactarius* species

101bp Paired-end libraries for *L. rufus, L. vinaceorufescens,* and *L. xanthogalactus* were constructed with NEBNext® Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) and subsequent sequencing on an Illumina HiSeq2500 platform was performed as previously described in Chapter 3. Raw forward and reverse .fastq files were transferred to MSI using FTP.

2.4 Illumina quality control, genome assembly and functional annotation for *L. rufus, L. vinaceorufescens,* and *L. xanthogalactus*

Quality control of Illumina reads was performed on the Galaxy platform using the established workflow outlined on the *L. uvidus* data described in Chapter 3. Processed Illumina reads for *L. rufus, L. vinaceorufescens,* and *L. xanthogalactus* were exported for genome assembly using Velvet [37]. RepeatMasker [106] and Augustus [107] were respectively used for repeat masking and gene prediction for the *L. rufus* and *L. vinaceorufescens* assembly. Poor assembly quality precluded *L. xanthogalactus* from further bioinformatics analysis.

2.5 PacBio® SMRTbell™ library construction and PacBio® Sequencing of *L. rufus*

An additional purified *L. rufus* gDNA sample was shipped overnight to the Molecular Biology Core at the Mayo Clinic Medical Genome Facility for PacBio SMRTbell™ library construction and subsequent PacBio® sequencing (Pacific Biosciences, Menlo Park, CA). At Mayo, *L. rufus* gDNA was processed into SMRTbell™ (Pacific

Biosciences, Menlo Park, CA) template strands, and size selected for 10kb inserts with the BluePippen™ Size-Selection system (Sage Sciences, Beverly, MA) in accordance with the suggested protocols. 5 x 10kb size-selected SMRTbell™ libraries from *L. rufus* were loaded onto 5 single molecule real-time (SMRT) sequencing cells at respective concentrations of 10, 20, 25, 30, and 40nM. Ensuing sequencing runs were monitored as 240min movies, with all subsequent data being output in .h5, .fasta, .fastq, .log and .xml formats. Outputted files were transferred to MSI for user analysis using FTP as previously described.

2.6 *Lactarius rufus* PacBio read quality control, hybrid scaffolding genome assembly, masking, and gene prediction

Raw .fastq files were processed into filtered subreads by the Subread Filtering tool from the PacBio® SMRT Analysis Portal (Pacific Biosciences, Menlo Park, CA) provided by MSI. Filtered subread .fastq files were exported for genome assembly. A hybrid scaffolding genome assembly strategy was used for *L. rufus* genome assembly by mapping the filtered PacBio subreads onto the previous Velvet assembled Illumina reads using PBJelly2 [105]. Assembly statistics were gathered on the Illumina only assembly, and of the Illumina-PacBio hybrid assembly of *L. rufus* using summarizeAssembly.py available from the PBJelly2 suite. RepeatMasker version 4.0.5 [106] and Augustus version 3.0.2 [107] were run on the *L. rufus* scaffolded genome as previously described.

2.7 TCA/Acetone Precipitation of *Lactarius* latex

Latex extracts from *L. rufus, L. turpis, L. uvidus, L. vinaceorufescens,* and *L. xanthogalactus* were subjected to TCA/Acetone precipitation [174]. In short, 750µL of

chilled 13.3% w/v trichloroacetic acid in acetone were added to 250µL latex extract aliquots and incubated overnight at -20^{o}C before centrifugation at 16,000xg for 10min at 4^{o}C. After removal of supernatant, 1mL of chilled acetone containing 0.2% w/v dithiothreitol was added to the pellet, and the pellet was broken up using a glass rod. Over the course of 1 hour, the tube containing the broken pellet in acetone was vortexed 4 times for 30sec each. The centrifugation, acetone wash, breakage and vortexing steps were repeated once before a final centrifugation at the same speed, time and temperature to obtain a protein pellet. The pellet was allowed to air dry for 5 minutes before being resuspended in 40µL sodium dodecyl sulfate (SDS) sample-loading buffer, and subjected to SDS polyacrylamide gel electrophoresis on a 15% SDS gel. Gel bands of interest that were heavily stained by Bio-Safe™ Coomassie Blue (Bio-Rad, Hercules, CA) were extracted from the gel via scalpel and diced into ~2mm x 2mm cubes and submitted to the University of Minnesota Center for Mass Spectrometry and Proteomics (CMSP).

2.8 Proteomics analysis of *Lactarius* latex

At CMSP, gel samples were subjected to in-gel trypsin digestion according to the following protocol [175] to achieve peptide samples. Peptide samples were then subjected to desalting according to the STAGE-Tip protocol [176] and submitted for tandem mass spectrometry. Tandem MS was performed on all peptide samples on an LTQ Orbitrap Velos* Mass Spectrometer with an ESI (nano-spray) ion source and high energy CID (y and b ion) fragmentation mode. The MS Scan and subsequent MS/MS scan modes were performed in FT-ICR/Orbitrap modes.

Raw tandem mass spectra data for each peptide sample were packaged as zipped PEAKS7.0 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) project files and submitted to the University of Minnesota's TINT server for user access. From PEAKS viewer, mapped peptide database hits on the Russulales family and *de novo* peptides were exported into .csv files for filtering. A perl script called peptideFilter.pl was developed, and used to filter out any peptides that were the result of human keratin contamination and exported the remaining peptide sequences into a .fasta file. The .fasta files containing the filtered peptides were then aligned to predicted amino acid sequences of *L. rufus* and *L. vinaceorufescens* to determine any overlap between the genome and latex proteome.

2.9 Functional annotation analysis for *L. rufus* and *L. vinaceorufescens*

Output .gff3 files for both *L. rufus* and *L. vinaceorufescens* gene predictions were parsed into amino acid .fasta files using AugustusParserv3.0.pl as previously described. The predicted amino acid .fasta files were subjected to a local alignment with BLAST [108] using *Lactarius chrysorrheus* Farnesyl diphosphate synthase, *Hevea brasiliensis* rubber transferase 2, and a library of previously characterized sesquiterpene synthases (STSs) as queries to identify candidate genes involved in natural rubber biosynthesis and putative STSs.

2.10 Bioinformatics analysis of candidate genes

Reference libraries containing previously characterized CPTs and FPPSs were created for multiple sequence alignments with candidate genes in both the *L. rufus* and *L. vinaceorufescens* genomes. Multiple sequence alignment was performed with Muscle in the MEGA version 6 bioinformatics suite as previously described to refine putative gene

candidates. After refinement, maximum likelihood trees of putative CPTs, FPPSs and STSs of *L. rufus* and *L. vinaceorufescens* were constructed against established CPTs, FPPSs, and STSs according to the previously mentioned MAFFT[111], ProtTest3.2[112] and PhyML[119] workflow for the *L. uvidus* genome.

2.11 Cloning of candidate genes

cDNA construction from *L. rufus* and *L. vinaceorufescens* extracted RNA was performed with the Superscript® III First-Strand Synthesis System for RT-PCR Kit using oligo(dT) primers to enrich cDNA from mRNA. PCR scans using Phusion high fidelity polymerase were performed on *L. rufus* cDNA using primers adhering to the 2 conserved DDxxD motifs for FPPSs, 4 separate conserved regions for CPTs (CSI, CSIII-CSV), and the gene-specific ends for each predicted gene. PCR reactions were examined via gel electrophoresis on a 2% agarose gel. Amplicon bands, at their respective predicted full length gene size, were extracted using the Wizard Gel Extraction and PCR Cleanup Kit (Promega, Madison, WI) according to the manufacturers' suggested protocol. Gel extracted amplicons were blunt cloned into pCR™-BluntII-TOPO® plasmids (abbreviated as TOPO) using the Zero Blunt® TOPO® PCR Cloning Kit according to the manufacturer's instructions (Invitrogen, Carlsbad, CA), and submitted to the University of Minnesota Genomics Center for sequencing verification. Incorrectly spliced gene inserts were corrected for using the Q5 Site Directed Mutagenesis Kit (New England Biolabs, Inswich, MA) according to the manufacturer's instructions. Correctly spliced genes were subcloned from TOPO constructs and into pET32b expression vectors using

BamHI and HindIII restriction sites, and also into our in-house pCuminBB expression vector (unpublished data) using NdeI and XhoI restriction sites.

2.12 Complementation Assays

Three samples of chemically competent *E. coli* JM109 cells were differentially transformed with pAC-crtM-crtN-crtOx [177], pCuminBB-LrufCPT2 (candidate *cis*-prenyl transferase from *L. rufus*), or both pAC-crtMNOX and pCuminBB-LrufCPT2 respectively. Transformant plates were incubated overnight at 37^{o}C before being incubated at room temperature in the dark for two more days. After 2 days, nitrocellulose membranes were lightly placed over the growing colonies to provide a white background, and transformant plates were photographed with a Casio Commando G'zOne Smartphone (Shibuya, Tokyo, Japan).

3 Results and Discussion

3.1 Genome sequencing and analysis of *L. rufus, L. vinaceorufescens* and *L. xanthogalactus*

To determine the identity of natural rubber biosynthetic genes from the genus *Lactarius,* we initially extracted genomic DNA from cultivated samples of *L. rufus,* and also directly from the interior of *L. xanthogalactus* fruiting bodies. Genomic DNA was then subjected to next-generation sequencing on an Illumina HiSeq2500 platform. This resulted in approximately 161million and 186million 101 bp paired end reads of *L. rufus* and *L. xanthogalactus*, respectively.

Using the established Illumina quality control workflow outlined for *L. uvidus* in Chapter 3 of this thesis, the 161million, and 186million paired end reads were processed to approximately 125million and 152million high quality 80bp paired end reads on the Galaxy platform. These refined reads represented 167-fold and 203-fold coverage for *L. rufus* and *L. xanthogalactus* genomes respectively (using an estimated dikaryon genome size of 60Mbp).

The *L. xanthogalactus* assembly with Velvet [37] was problematic, as it resulted in 80,417 scaffolds, with a scaffold N50 of 670bp. The high number of Illumina reads, and small scaffolds were likely due to microbial gDNA contamination that was also isolated from the fruiting body tissue. This highlighted the importance of isolating genomic DNA from pure culture. Given that the N50 of *L. xanthogalactus* was smaller than 1kb, further work in mining its genome was suspended.

However, by that time we had successfully cultivated a close relative to *L. xanthogalactus—L. vinaceorufescens*. Illumina sequencing of *L. vinaceorufescens* gDNA yielded approximately 137 million raw 101bp reads that were refined to 112 million reads at 80bp, representing 149-fold genome coverage for assembly with Velvet [37]. **Table 4.1** summarizes respective genome assemblies of *L. rufus* and *L. vinaceorufescens* in comparison to the *L. uvidus* and *O. olearius* genome assemblies in addition to initial bioinformatics analyses using RepeatMasker[106] and Augustus [107].

Table 4.1 Illumina assembly stats of *L. rufus, L. vinaceorufescens* and *L. xanthogalactus*. Illumina assembly stats of *L. uvidus* and *O. olearius* from chapter 3 also included for comparison.

Parameter	*L. rufus*	*L.vinaceorufescens*	*L. xanthogalactus*	*L. uvidus*	*O. olearius*
Genome Size	35.90 Mbp	61.81 Mbp	50.77 Mbp	60.08 Mbp	28.15 Mbp
Scaffold Number	13,799	26,016	80,417	33,058	868
Scaffold N50	15,537	6,169	670	3,969	199,357
Max Scaffold Size	225,733	126,981	21,711	51,876	N/A
Percent of Repetitive Elements	3.09%	5.47%	N/A	2.80%	N/A
Predicted Protein Coding Genes	15,071	22,590	N/A	26,157	8,172

The short-read genome assemblies of *L. rufus* and *L. vinaceorufescens* yielded more useful scaffolds for subsequent analysis, with respective scaffold N50s of 15,537 and 6,169, and maximum scaffold sizes of 225kb and 126kb. While these assembly stats were still inferior to the genome assembly of *Omphalotus olearius*, their scaffold size was more amenable to genome mining and manual biosynthetic gene cluster annotation than either *L. xanthogalactus* or *L. uvidus*.

The genomes of *L. rufus* and *L. vinaceorufescens* were also more repetitive than *L. uvidus*, with 3.09% and 5.47% of each respective genome being masked by RepeatMasker[106] before gene prediction with Augustus [107]. The genomes of *L. rufus* and *L. vinaceorufescens* contained 15,071 and 22,590 predicted genes respectively. The number of predicted genes of *L. rufus* falls within the general realm of predicted genes for fungal genomes of that size [178], while *L. vinaceorufescens* seems to be somewhat artificially high for the same reasons outlined for the predicted number of genes for *L. uvidus* in Chapter 3.

Another interesting attribute for these genome assemblies was the total assembly size of *L. rufus* and *L. vinaceorufescens.* At respective sizes of 35.90 Mbp and 61.81 Mbp, it appeared that the Velvet assembly program was able to better resolve the dikaryon

characteristics of *L. rufus* than *L. vinaceorufescens* or even *L. uvidus*. Another reason for this could be that the two nuclei in the *L. rufus* strain were more similar to each other than for the *L. vinaceorufescens* and *L. uvidus* nuclei. The total number of predicted genes for *L. rufus* further supports this assertion.

While the overall assembly of *L. rufus* was superior to the other *Lactarius* genome assemblies, it was still inferior to the *O. olearius* assembly. It was hypothesized that utilizing long read sequence data would greatly improve the genome assembly of *L. rufus*. Thus, we obtained long-read sequence data from *L. rufus* gDNA using the PacBio sequencing platform (Pacific Biosciences, Menlo Park, CA) available at the Molecular Biology Core of the Mayo Clinic Medical Genome Facility (Rochester, MN). PacBio sequencing of *L. rufus* on 5 single-molecule real-time cells yielded 57,599 reads, calling 617,984,429 bases in total with a read length N50 of 16,208 in raw data.

Using the PacBio® SMRT Analysis Portal available from MSI, these 57,599 reads were processed into 115,368 filtered subreads, to correct for the presence of ligated SMRT-bell adapters from the SMRT-bell library prep. These filtered subreads possessed a read N50 of 11,068 and a total number of bases of 613,876,755. Using an estimated genome size of 60Mbp, approximately 10-fold coverage in PacBio sequencing data was achieved. This coverage precluded the *de novo* assembly of *L. rufus* from PacBio sequencing data alone. The HGAP3 program [179] available in the PacBio® SMRT Analysis Portal, needs to be at the very least roughly 50-fold coverage to be able to perform a *de novo* assembly [179].

Alternative assembly strategies involving long-read data were available though. The program PBJelly2, provides gap filling and interscaffolding algorithms to map PacBio reads onto preassembled contigs [105]. By utilizing PBJelly2, we were able to markedly improve the assembly of *L. rufus* by mapping the filtered subreads of the PacBio data onto the preassembled contigs file from the Velvet assembly. The comparison of the *L. rufus* assemblies is illustrated in **Table 4.2**.

Table 4.2 Comparison of assembly statistics between short-read and short/long-read hybrid assembly strategies for *L. rufus* genome.

Parameter	*L. rufus Short-Read Assembly*	*L. rufus Hybrid Short and Long-Read Assembly*	*O. olearius*
Genome Size	35.90 Mbp	41.92 Mbp	28.15 Mbp
Scaffold Number	13,799	11,283	868
Scaffold N50	15,537	46,813	199,357
Max Scaffold Size	225,733	454,504	N/A
Percent of Repetitive Elements	3.09%	3.51%	N/A
Predicted Protein Coding Genes	15,071	14,951	8,172

The number of scaffolds between the two *L. rufus* assemblies decreased 18% from 13,799 to 11,283 between the two assemblies. Additionally, the scaffold N50 improved 3-fold from 15.5kb to 46.8kb. Similarly, the maximum scaffold size between the two assemblies doubled. Finally, the number of predicted protein coding genes remained relatively constant between the two assemblies, at 15,071 and 14,951 respectively.

The improved assembly of *L. rufus* using a hybrid approach was found to greatly improve the manual annotation of biosynthetic gene clusters, as discussed in Sections 3.2 and 3.4.

3.2 The *cis*-prenyl transferase and farnesyl pyrophosphate synthases of *L. rufus* and *L. vinaceorufescens* are in different biosynthetic gene clusters

Local BLAST searches, using *Hevea brasiliensis* rubber transferase 2 (HRT2) as a query sequence, were performed against the predicted amino acid sequences of *L. rufus*

and *L. vinaceorufescens*. Using this approach, 2 predicted genes (gene numbers 6147 and 6745) from *L. rufus* were identified as putative CPTs. Both putative *L. rufus* CPTs contained all 5 conserved regions found in CPTs. Additionally, 1 putative CPT (gene number 2431) was identified in *L. vinaceorufescens*. This putative CPT contained the first four conserved motifs present in CPTs CSI-CSIV. However, upon manual inspection, the last conserved region, CSV was found in a predicted intron for this particular gene. Thus, the splicing model prediction for this putative CPT was inferred to be incorrect. *L. rufus* putative CPTs 6147 and 6745 were renamed LrufCPT1 and LrufCPT2, respectively, for further analysis. Similarly, the *L. vinaceorufescens* putative CPT gene 2431 was renamed to LvinCPT1 for further analysis (see **Supplemental Figure S4.1** for a snapshot of the multiple sequence alignment (MSA) of putative CPTs and **Supplemental Table S4.1** for gene details of putative CPTs).

Similarly, using the cloned and characterized farnesyl pyrophosphate synthase from *L. chysorrheus* [180], 3 putative FPPSs were identified in *L. rufus* (gene numbers 8769, 10620, and 11419). Additionally, 4 FPPSs were initially identified in *L. vinaceorufescens*, however, the candidates were manually reduced to 2 (gene numbers 12973 and 13101) as the other two candidates were not of characteristic length of FPPSs [180]. Gene numbers 8769, 10620, and 11419 from *L. rufus* were renamed LrufFPPS1, LrufFPPS2 and LrufFPPS3, respectively, for further analysis. Similarly, *L. vinaceorufescens* gene numbers 12973 and 13101 were renamed LvinFPPS1 and LvinFPPS2 for subsequent analysis (see **Supplemental Figure S4.2** and **Supplemental Table S4.2** for respective MSA and putative FPPS details). A dendrogram containing the

following putative CPTs and FPPSs versus established CPTs and FPPSs can be viewed in **Figure 4.2.**

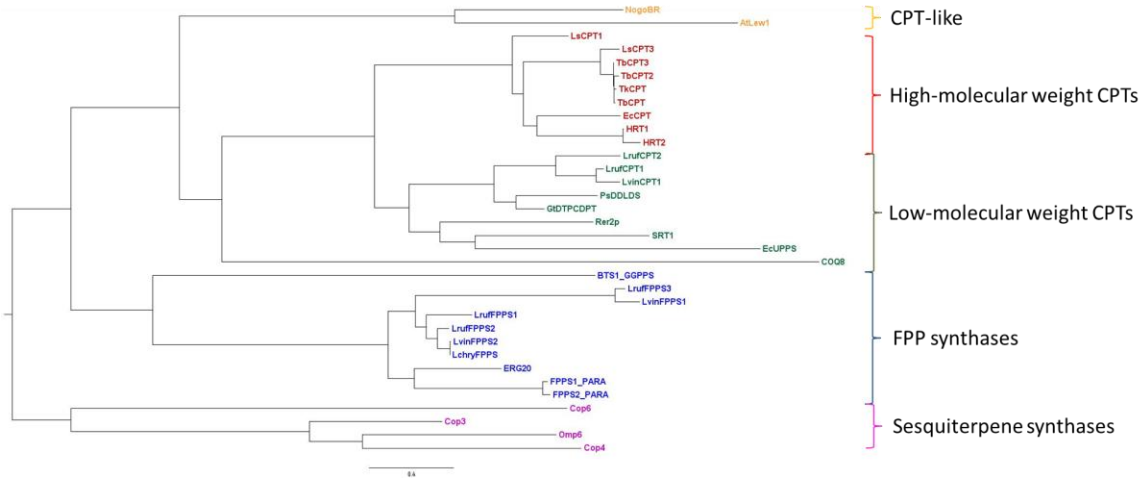
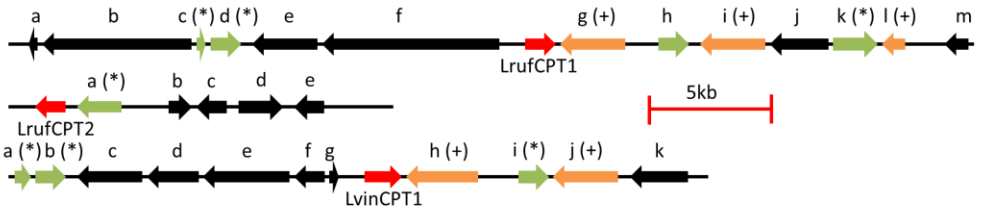


Figure 4.2. Dendrogram of putative CPTs and FPPSs from *L. rufus* and *L. vinaceorufescens*. CPT-like rubber transferase activators are colored orange. Rubber transferase CPTs are colored red. CPTs colored green are dihydrodolichol diphosphate synthases and undecaprenyl transferases. Trans prenyl transferases are colored blue. Sesquiterpene synthase outgroups are colored pink. At, *Arabidopsis thaliana*; Ls, *Lattica sativa*; Tb, *Taraxacum brevicorniculatum*; Tk, *Taraxacum koksaghyz*; Ec, *Euphorbia characias*; PsDDLDS, *Punctularia strigosozonata* dihydrodolichyl diphosphate synthetase; GtDTPCDPT, *Gloeophyllum trabeum* di-trans-poly-cis-decaprenylcistransferase; Rer2p and SRT1, *Saccharomyces cerevisiae* CPTs; EcUPPS, *Escherichia coli* undecaprenyl diphosphate synthase; COQ8, *S. cerevisiae* ubiquinone prenyl transferase; BTS1, *S. cerevisiae* GGPP synthase; ERG20, *S. cerevisiae* FPPS; FPPS1_PARA, *Parthenium argentatum* FPPS.

Figure 4.2 shows that the *Lactarius* putative CPTs were found in the clade containing the dihydrodolichol diphosphate synthases and undecaprenyl transferases. These types of *cis*-prenyl transferases are responsible for making C55-C110 polyisoprene which are ubiquitously present in all living organisms [181, 182]. While it is not surprising to see these CPTs grouped with lower molecular weight CPTs, as *Lactarius* are only capable of producing *cis*-polyisoprene with 300 isoprene units[154], it would be more encouraging to see these putative CPTs grouped within the higher molecular weight producing CPT clade, or at least within a subclade of their own to denote mid-molecular weight polyisoprene CPTs.

Conversely, the putative *Lactarius* FPPSs grouped with characterized FPP synthases as expected. From this phylogenetic analysis, it appears that LrufFPPS2 and LvinFPPS2 are homologs of each other, as are LrufCPT1 and LvinCPT1.

Given that a number of studies have shown FPP is an initiator molecule in *cis*-polyisoprene biosynthesis, it was hypothesized that FPPSs and CPTs would reside in the same biosynthetic gene clusters within the respective *Lactarius* genomes. Because the genome assemblies of *L. rufus* and *L. vinaceorufescens* were assembled into larger scaffolds than *L. uvidus*, it was determined that manual annotation of gene clusters using BLAST would be more informative. Indeed, all putative genes identified in the genomes resided in scaffolds greater than 5kb in length, allowing for a more complete manual annotation set. However, the hypothesis that the FPPSs and CPTs cluster together in *Lactarius* genomes was not supported with what was observed, as all identified CPTs and FPPSs were found within their own respective clusters. **Figure 4.3** below represents manually annotated biosynthetic gene clusters of the putative CPTs, while **Figure 4.4** represents the manually annotated biosynthetic gene clusters of the putative FPPSs.



LrufCPT1 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig419:g114.t1	Ribosomal protein S16 (TIGR00002)
b	Contig419:g115.t1	Armadillo/beta-catenin repeats (cd00020)
c	Contig419:g116.t1	(*) Cytochrome c oxidase subunit VIa (pfam02046)
d	Contig419:g117.t1	(*) Ribulose-5-phosphate 3-epimerase (cd00429)
e	Contig419:g118.t1	No hit
f	Contig419:g119.t1	Pyruvate kinase (PRK05826)
LrufCPT1		
g	Contig419:g121.t1	(+) MFS (cd06174)
h	Contig419:g122.t1	(*) Oxidoreductase (pfam01266)
i	Contig419:g123.t1	(+) Sugar transporter (pfam00083)
j	Contig419:g124.t1	PPR repeat (TIGR00756)
k	Contig419:g125.t1	(*) Cyclophilin cis-trans isomerase (cd01923)
l	Contig419:g126.t1	(+) Snf7 Transporter (pfam03357)
m	Contig419:g127.t1	G-protein coupled glucose receptor (pfam11970)

LrufCPT2 Gene Cluster

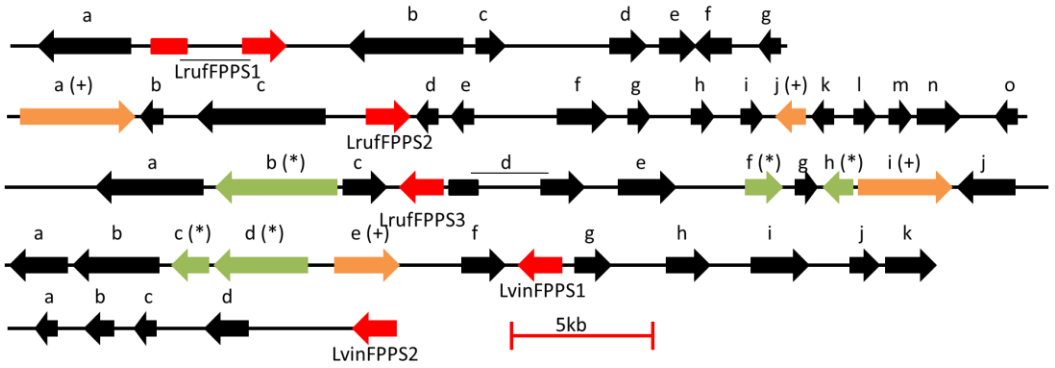
Gene	Reference	Top Domain Hit
LrufCPT2		
a	Contig521:g2.t1	(*) Dehydrogenase (cd08297)
b	Contig521:g3.t1	No hit
c	Contig521:g4.t1	Phosphatase regulatory subunit (pfam03370)
d	Contig521:g5.t1	DUF4097 (COG3595)
e	Contig521:g6.t1	Phosphatase regulatory subunit (pfam03370)

LvinCPT1 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_20756_length_58314_cov_44.604008:g1.t1	(*) Cytochrome c oxidase subunit VIa (pfam02046)
b	NODE_20756_length_58314_cov_44.604008:g2.t1	(*) Ribulose-5-phosphate 3-epimerase (cd00429)
c	NODE_20756_length_58314_cov_44.604008:g3.t1	No hit
d	NODE_20756_length_58314_cov_44.604008:g4.t1	Pyruvate kinase (PRK05826)
e	NODE_20756_length_58314_cov_44.604008:g5.t1	DNA binding protein (pfam09271)
f	NODE_20756_length_58314_cov_44.604008:g6.t1	Ring finger domain (pfam13639)
g	NODE_20756_length_58314_cov_44.604008:g7.t1	No hit
LvinCPT1		
h	NODE_20756_length_58314_cov_44.604008:g8.t1	(+) MFS (cd06174)
i	NODE_20756_length_58314_cov_44.604008:g9.t1	(*) Oxidase (COG0665)
j	NODE_20756_length_58314_cov_44.604008:g10.t1	(+) MFS (cd06174)
k	NODE_20756_length_58314_cov_44.604008:g11.t1	PPR repeat (TIGR00756)

Figure 4.3. Biosynthetic gene clusters of putative Lactarius CPTs. Transporters are colored orange (+) while hydrocarbon scaffold modifying enzymes are colored green (*).

There appears to be a number of hydrocarbon scaffold modifying enzymes and transporters closely associated with the putative CPTs. The presence of transporters nearby is especially encouraging as well, furthering the evidence of a rubber particle or membrane associated form of polyisoprene synthesis. Additionally, **Figure 4.3** further supports the assertion that LrufCPT1 and LvinCPT1 are structural homologs, given the similarity of associated genes clustered nearby.



LrufFPPS1 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig928:g1.t1	Ankyrin repeats (cd00204)
LrufFPPS1		
b	Contig928:g3.t1	Hydroxylase (pfam02373)
c	Contig928:g4.t1	Large tegument protein UL36 (PHA03247)
d	Contig928:g5.t1	Late embryogenesis abundant protein (pfam 03168)
e	Contig928:g6.t1	DUF4341 (pfam14241)
f	Contig928:g7.t1	RNA recognition motif (cd12224)
g	Contig928:g8.t1	No hit

LrufFPPS2 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig1902:g9.t1	(+) Mg-transporting ATPase (PRK15122)
b	Contig1902:g10.t1	No hit
c	Contig1902:g11.t1	Ankyrin repeats (PHA03095)
LrufFPPS2		
d	Contig1902:g13.t1	No hit
e	Contig1902:g14.t1	Rnse P (pfam01900)
f	Contig1902:g15.t1	F-box-like protein (pfam12937)
g	Contig1902:g16.t1	H-lectin (pfam09458)
h	Contig1902:g17.t1	H-lectin (pfam09458)
i	Contig1902:g18.t1	H-lectin (pfam09458)
j	Contig1902:g19.t1	(+) Mg-transporter (cd12829)
k	Contig1902:g20.t1	No hit
l	Contig1902:g21.t1	H-lectin (pfam09458)
m	Contig1902:g22.t1	No hit
n	Contig1902:g23.t1	No hit
o	Contig1902:g24.t1	No hit

LrufFPPS3 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig2635:g1.t1	Leucine-rich repeat (sd00031)
b	Contig2635:g2.t1	(*) 2-methylaconitate isomerase (COG2828)
c	Contig2635:g3.t1	Nuclease (pfam08652)
LrufFPPS3		
d	Contig2635:g5.t1	Replication factor C small subunit (PRK0440)
e	Contig2635:g6.t1	Large tegument protein (PHA03247)
f	Contig2635:g7.t1	(*) PEP Lyase (cd00377)
g	Contig2635:g8.t1	Cytochrome binding protein (pfam00173)
h	Contig2635:g9.t1	(*) Dihydrodipicolinate synthase (cd00408)
i	Contig2635:g10.t1	(+) Serine incorporator (pfam 03348)
j	Contig2635:g11.t1	Protein kinase (cd00180)

LvinFPPS1 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_138734_length_35075_cov_43.478603:g1.t1	DNA polymerase subunit Cdc27 (pfam09507)
b	NODE_138734_length_35075_cov_43.478603:g2.t1	No hit
c	NODE_138734_length_35075_cov_43.478603:g3.t1	(*) 2-methylaconitate cis-trans isomerase (COG2828)
d	NODE_138734_length_35075_cov_43.478603:g4.t1	(*) Enoyl-CoA hydratase (PRK08260)
e	NODE_138734_length_35075_cov_43.478603:g5.t1	(+) MFS (cd06174)
f	NODE_138734_length_35075_cov_43.478603:g6.t1	No hit
LvinFPPS1		
g	NODE_138734_length_35075_cov_43.478603:g8.t1	No hit
h	NODE_138734_length_35075_cov_43.478603:g9.t1	Replication factor C subunit (PRK00440)
i	NODE_138734_length_35075_cov_43.478603:g10.t1	Transcriptional regulator ICP4 (PHA03307)
j	NODE_138734_length_35075_cov_43.478603:g11.t1	Cytochrome binding domain (pfam00173)
k	NODE_138734_length_35075_cov_43.478603:g12.t1	Protein kinase (smart00239)

LvinFPPS2 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_140709_length_14195_cov_33.274532:g1.t1	H-type lectin (pfam09458)
b	NODE_140709_length_14195_cov_33.274532:g2.t1	H-type lectin (pfam09458)
c	NODE_140709_length_14195_cov_33.274532:g3.t1	H-type lectin (pfam09458)
d	NODE_140709_length_14195_cov_33.274532:g4.t1	No hit
LvinFPPS2		

Figure 4.4 Biosynthetic gene clusters of putative *Lactarius* FPPSs. Transporters are colored orange (+) while hydrocarbon scaffold modifying enzymes are colored green (*). LrufFPPS1 and LrufFPPS3d were predicted to have very large introns, illustrated by the bridging line between exons.

In general, there does not appear to be many hydrocarbon scaffold modifying enzymes or transporters associated with the putative *Lactarius* FPPSs, save for LvinFPPS1. Interestingly, the most common conserved domain hit for genes associated with the putative FPPSs were H-type lectin proteins. These proteins have generally been shown to be involved with cell recognition through the binding of carbohydrates [183].

Perhaps these H-type lectins could be involved with recruiting glycosylated proteins for polyisoprene biosynthesis, assuming these FPPSs are involved in polyisoprene synthesis.

It is known that many more proteins are necessary for the synthesis of high-molecular weight *cis*-polyisoprene than just CPTs and FPPSs, as accessory proteins from *Hevea* [165, 166], *Taraxacum* [168] and *Latuca sativa* [184](lettuce) rubber particles have been characterized as activators, or recruiters for higher-molecular weight biosynthesis. Examples of these proteins include rubber elongation factor (REF), small rubber particle protein (SRPP), rubber transferase activator (TbRTA), and *cis*-prenyltransferase-like 2 (CPTL2). However, searching the predicted amino acid space of *L. rufus* and *L. vinaceorufescens* using these accessory proteins as queries yielded no candidate hits.

3.3 ***L. rufus*** **and** ***L. vinaceorufescens*** **contain many putative 1,11-cyclizing STSs**

The STS profile of *L. rufus* and *L. vinaceorufescens* were also explored in addition to the genome mining for natural rubber biosynthetic genes. The superior genome assemblies of *L. rufus* and *L. vinaceorufescens* compared to *L. uvidus* allowed for more optimal gene cluster annotation with the conserved domain database available from BLAST [185]. Using the same established genome mining workflow presented in this thesis, 12 and 20 initial STS hits were identified from *L. rufus* and *L. vinaceorufescens* predicted amino acid sequences, respectively (see **Supplemental Figure S4.3** for MSA alignments and **Supplemental Table S4.3** for putative STS details). Superficial analysis of the predicted STSs with a multiple sequence alignment prompted the reduction of candidates down to 7 and 5 genes for *L. rufus* and *L. vinaceorufescens*, respectively. Gene names for these putative STSs were denoted LrSTS<gene number> or LvSTS<gene

number> for *L. rufus* and *L. vinaceorufescens* STSs respectively, where the gene number is the predicted gene number from the Augustus annotation.

Judging from the screened sesquiterpenoid profiles of mycelium cultures of these species, in addition to the previously reported amount of *trans*-humulyl cation derived scaffolds isolated from *Lactarius* fruiting bodies from the literature[78], it was expected that a number of 1,11-cyclizing STSs would be present amongst the candidates. Furthermore, given the abundance of (*Z,E*)-germacradienyl derived sesquiterpenes identified from the headspace of *L. rufus* cultures, it was expected there would also be a substantial amount of STSs utilizing the 1,10-cyclization of (*3R*)-nerolidyl pyrophosphate mechanism. **Figure 4.5** illustrates the maximum likelihood tree generated from the putative STSs from *L. rufus* and *L. vinaceorufescens*.

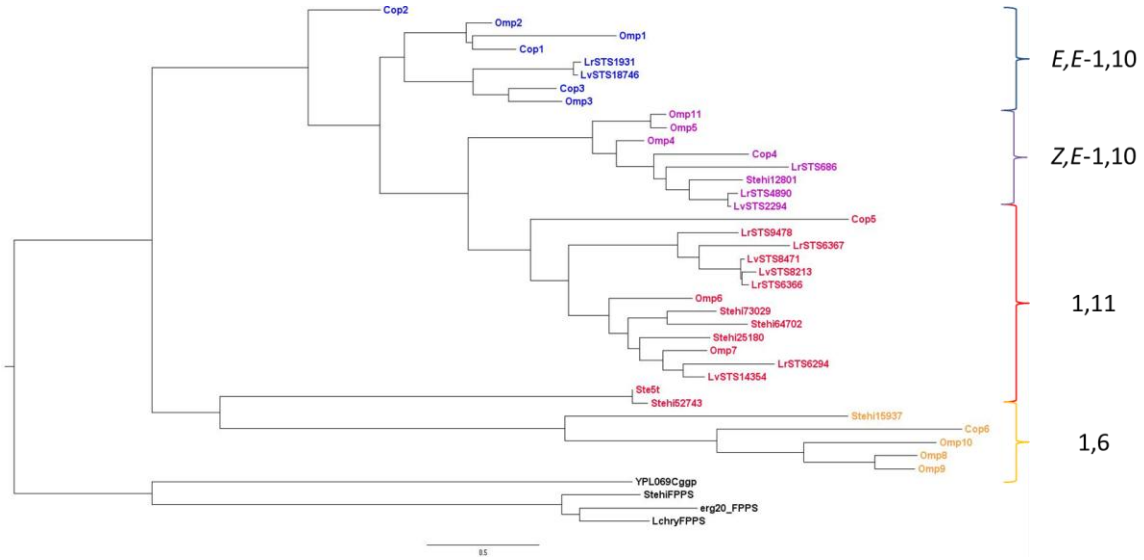
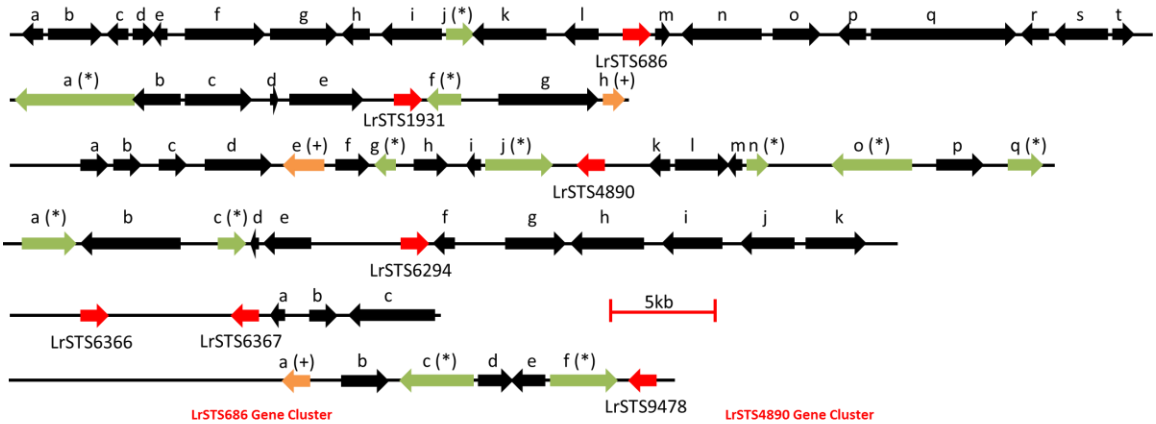


Figure 4.5 Maximum likelihood tree of putative STSs from *L. rufus* and *L. vinaceorufescens*. Blue STSs perform a 1,10-cyclization of *2E,6E*-FPP. Pink STSs perform a 1,10-cyclization of (*3R)*-NPP. Red STSs perform a 1,11-cyclization of *2E,6E*-FPP. Orange STSs perform a 1,6-cyclization of (*3R)*-NPP.

As expected, a large amount of the putative STSs from *L. rufus* and *L. vinaceorufescens* clustered with the established 1,11-cyclizing STS clade. Furthermore,

the same sub-clustering of putative *Lactarius* STSs within the 1,11-cyclization clade was observed like that found in the *L. uvidus* STS bioinformatics analysis from Chapter 3. Finally, 2 *Lactarius rufus* STSs where found to reside in the 1,10-cyclizing of (3*R*)-NPP clade. Based on the volatile headspace of *L. rufus* mycelial cultures, it was unexpected to see only 2 putative STSs in this clade. However, STSs possess a broad range of fidelity [43], so it is not unreasonable to postulate that all of the *Z,E*-germacradienyl derived terpenes found in the *L. rufus* headspace were derived solely from these two putative terpene synthases. It also plausible that the abundance of *Z,E*-germacradienyl derived terpenes found in the mycelial culture headspaces of *L. rufus* from Chapter 3 could be the products of STSs from other clades. Recent evidence shows that STSs can illicit different cyclization mechanisms in the same STS depending upon the divalent metal cofactor bound to the enzyme [186]. Given that the screened culture headspaces were in complex media containing multiple divalent metal ions, STS promiscuity appears to be the more plausible explanation.

The superior assemblies of *L. rufus* and *L. vinaceorufescens* compared to the *L. uvidus* assembly also yielded better manual annotation of biosynthetic gene clusters containing putative STSs. The biosynthetic gene clusters of putative STSs from *L .rufus* and *L. vinaceorufescens* can be viewed in **Figure 4.6** and **Figure 4.7** respectively.



LrSTS686 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig41:g70.t1	DUF1681 (pfam07933)
b	Contig41:g71.t1	Poly(A) polymerase (COG5186)
c	Contig41:g72.t1	Phospholipase (pfam02230)
d	Contig41:g73.t1	tRNA methyltransferase (pfam01746)
e	Contig41:g74.t1	No hit
f	Contig41:g75.t1	Large tegument protein (PHA03247)
g	Contig41:g76.t1	Transcriptional regulator (PHA03307)
h	Contig41:g77.t1	Transcription factor (smart00225)
i	Contig41:g78.t1	Inositol polyphosphate kinase (pfam03770)
j	Contig41:g79.t1	(*) Enoyl-CoA hydratase (PRK13342)
k	Contig41:g80.t1	Sec10 (pfam07393)
l	Contig41:g81.t1	Recombination factor RarA (PRK13342)
LrSTS686		
m	Contig41:g83.t1	RNA polymerase I initiation factor (pfam04090)
n	Contig41:g84.t1	50S ribosomal protein (PRK09374)
o	Contig41:g85.t1	No hit
p	Contig41:g86.t1	Ran GTPase-activating protein (COG5238)
q	Contig41:g87.t1	RNA polymerase subunit A (TIGR02390)
r	Contig41:g88.t1	Alloslipin (cd13437)
s	Contig41:g89.t1	Kinase (cd14709)
t	Contig41:g90.t1	RNA-binding protein (COG2947)

LrSTS1931 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig95:g14.t1	(*) δ-aminolevulinic acid dehydratase (cd04824)
b	Contig95:g15.t1	Chromatin (pfam02731)
c	Contig95:g16.t1	No hit
d	Contig95:g17.t1	GPI-anchored membrane protein (pfam10342)
e	Contig95:g18.t1	WD40 (cd00200)
LrSTS1931		
f	Contig95:g20.t1	(*) Acyl-CoA thioesterase (COG1946)
g	Contig95:g21.t1	CDC/Septin GTPase (cd01850)
h	Contig95:g22.t1	(+) ATPase (TIGR01652)

LrSTS6366 & LrSTS6367 Gene Cluster

Gene	Reference	Top Domain Hit
LrSTS6366		
LrSTS6367		
a	Contig452:g3.t1	No hit
b	Contig452:g4.t1	No hit
c	Contig452:g5.t1	Ankyrin repeat (cd00204)

LrSTS4890 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig299:g1.t1	No hit
b	Contig299:g2.t1	No hit
c	Contig299:g3.t1	No hit
d	Contig299:g4.t1	Urease (pfam00449)
e	Contig299:g5.t1	(+) MFS (cd06174)
f	Contig299:g6.t1	WD40 (cd00200)
g	Contig299:g7.t1	(*) Sphingolipid hydroxylase (COG3000)
h	Contig299:g8.t1	Pepsin protease (cd05471)
i	Contig299:g9.t1	No hit
j	Contig299:g10.t1	(*) Cystathionine β-lyase (PLN02509)
LrSTS4890		
k	Contig299:g12.t1	RNA polymerase Rpb5 (pfam01191)
l	Contig299:g13.t1	WD40 (cd00200)
m	Contig299:g14.t1	No hit
n	Contig299:g15.t1	(*) Enoyl-CoA hydratase (cd06558)
o	Contig299:g16.t1	(*) Acetyltransferase (PLN00104)
p	Contig299:g17.t1	Protein kinase (cd14134)
q	Contig299:g18.t1	(*) Homocitrate synthase (TIGR02146)

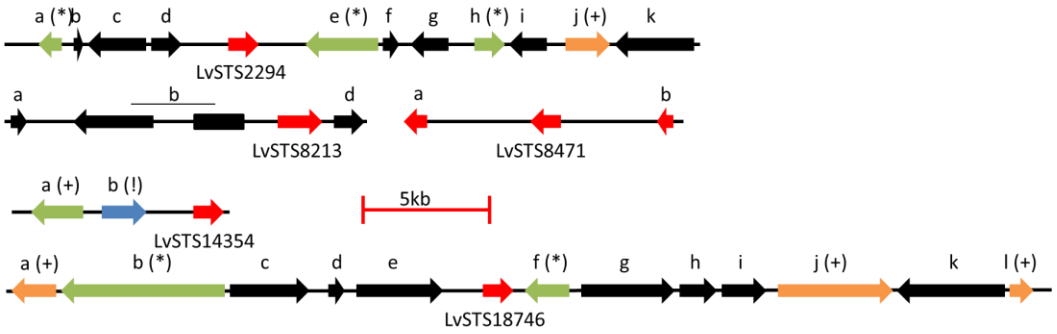
LrSTS6294 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig438:g28.t1	(*) Carbonyl reductase (cd05324)
b	Contig438:g29.t1	Ankyrin repeats (cd00204)
c	Contig438:g30.t1	(*) Oxidoreductase (cd05233)
d	Contig438:g31.t1	No hit
e	Contig438:g32.t1	Membrane Protein (pfam03006)
LrSTS6294		
f	Contig438:g34.t1	No hit
g	Contig438:g35.t1	Large tegument protein (PHA03247)
h	Contig438:g36.t1	Transcriptional regulator (PHA03307)
i	Contig438:g37.t1	No hit
j	Contig438:g38.t1	Coiled-coil protein (pfam15254)
k	Contig438:g39.t1	Atrophin-1 (pfam03154)

LrSTS9478 Gene Cluster

Gene	Reference	Top Domain Hit
a	Contig1254:g1.t1	(+) Iron transport protein (COG0390)
b	Contig1254:g2.t1	50S ribosomal protein (PRK04042)
c	Contig1254:g3.t1	(*) Coq4 (pfam05019)
d	Contig1254:g4.t1	RNA recognition motif (cd12360)
e	Contig1254:g5.t1	Zinc Finger (smart00356)
f	Contig1254:g6.t1	(*) Dehydrogenase (PRK02106)
LrSTS9478		

Figure 4.6. Biosynthetic gene clusters of putative *L. rufus* STSs. Transporters are colored orange (+) while hydrocarbon scaffold modifying enzymes are colored green (*).



LvSTS2294 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_20233_length_28123_cov_46.616505:g1.t1	(*) Enoyl-CoA hydratase (cd06558)
b	NODE_20233_length_28123_cov_46.616505:g2.t1	No hit
c	NODE_20233_length_28123_cov_46.616505:g3.t1	WD40 (cd00200)
d	NODE_20233_length_28123_cov_46.616505:g4.t1	RNA polymerase Rpb5 (pfam01191)
LvSTS2294		
e	NODE_20233_length_28123_cov_46.616505:g6.t1	(*) Cystanthionine β-lyase (PLN02509) & (*) mevalonate kinase (PLN02677)
f	NODE_20233_length_28123_cov_46.616505:g7.t1	No hit
G	NODE_20233_length_28123_cov_46.616505:g8.t1	Pepsin protease (cd05471)
h	NODE_20233_length_28123_cov_46.616505:g9.t1	(*) Sphingolipid Hydroxylase (COG3000)
i	NODE_20233_length_28123_cov_46.616505:g10.t1	WD40 (cd00200)
j	NODE_20233_length_28123_cov_46.616505:g11.t1	(+) MFS (cd06174)
k	NODE_20233_length_28123_cov_46.616505:g12.t1	Urease (cd00375)

LvSTS8213 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_78168_length_14618_cov_35.214531:g1.t1	No hit
b	NODE_78168_length_14618_cov_35.214531:g2.t1	Fungal transcription factor (cd12148)
LvSTS8213		
c	NODE_78168_length_14618_cov_35.214531:g4.t1	Transcriptional regulator Hpr (PRK13777)

LvSTS8471 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_81206_length_10955_cov_43.050755:g1.t1	STS domain (cd00687)
LvSTS8471		
b	NODE_81206_length_10955_cov_43.050755:g3.t1	STS domain (cd00687)

LvSTS14354 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_159368_length_9369_cov_41.093925:g1.t1	(*) FAD binding protein (pfam01565)
b	NODE_159368_length_9369_cov_41.093925:g2.t1	(!) P450 (pfam00067)
LvSTS14354		

LvSTS18746 Gene Cluster

Gene	Reference	Top Domain Hit
a	NODE_220021_length_58512_cov_43.853363:g4.t1	(+) Protein import complex subunit Tim54 (pfam11711)
b	NODE_220021_length_58512_cov_43.853363:g5.t1	(*) Δ-aminolevulinic acid dehydratase (smart01004)
c	NODE_220021_length_58512_cov_43.853363:g6.t1	VRR-nuclease (smart00990)
d	NODE_220021_length_58512_cov_43.853363:g7.t1	GPI-anchored membrane (pfam10342)
e	NODE_220021_length_58512_cov_43.853363:g8.t1	WD40 domain (cd 00200)
LvSTS18746		
f	NODE_220021_length_58512_cov_43.853363:g10.t1	(*) Acyl-CoA thioesterase (pfam02551)
g	NODE_220021_length_58512_cov_43.853363:g11.t1	DUF3429 (pfam11911)
h	NODE_220021_length_58512_cov_43.853363:g12.t1	SRPBCC_Ahal (cd08892)
i	NODE_220021_length_58512_cov_43.853363:g13.t1	CDC/Septin GTPase (cd01850)
j	NODE_220021_length_58512_cov_43.853363:g14.t1	(+) Mg-transporting ATPase (COG0474)
k	NODE_220021_length_58512_cov_43.853363:g15.t1	DNA Helicase (COG0553)
l	NODE_220021_length_58512_cov_43.853363:g16.t1	(+) Mitochondrial import protein Pam17 (pfam08566)

Figure 4.7. Biosynthetic gene clusters of putative L. vinaceorufescens STSs. Transporters are colored orange (+) while hydrocarbon scaffold modifying enzymes are colored green (*) and cytochrome P450s are colored blue (!) LvSTS8213b was predicted to have a large intron, illustrated by the line bridging the exons.

In comparison to the biosynthetic gene clusters found previously for *Stereum hirsutum* (see **Figure 1.4** in Chapter 1 for comparison) [38], there are unexpectedly few cytochrome P450 enzymes present. The LvSTS14354 cluster is the only example between both *L. rufus* and *L. vinaceorufescens* assemblies containing a cytochrome P450. However, cytochrome P450 enzymes are involved in lignin degradation [187], which could explain the high amount of P450s found in the genome of the lignin degrading mushroom of *S. hirsutum* compared to *Lactarius*.

Another cluster of interest is the LvSTS8471 cluster, where fragments of other STSs were identified. Indeed, these genes were initially identified in the initial BLAST search

of *L. vinaceorufescens* predicted amino acids. However, they were removed from consideration after discovering they only were fragments of STS genes. Predicted gene LvSTS8471a could be another STS, but it resides on the end of the assembled scaffold. Thus, the C-terminus won't be resolved until an improved assembly of *L. vinaceorufescens* with larger scaffolds occurs. This will be discussed more in the conclusion section.

The abundance of DNA regulatory elements and transcription factors amongst the *L. rufus* and *L. vinaceorufescens* STS gene clusters further contrast with the previously annotated *Stereum* cluster. The presence of these factors could further support the chemical defense model for *Lactarius*, where these factors could be used to activate transcription of STSs in response to an environmental factor.

3.4 Proteomics analysis of *Lactarius sp.*

Upon initial fruiting body collection, latex was extracted from the gills of *L. rufus, L. turpis, L. uvidus, L. vinaceorufescens* and *L. xanthogalactus* for the purpose of proteomics analysis. This complementary proteomics analysis would potentially identify proteins in the latex related to natural rubber biosynthesis, and perhaps be mapped to accessory proteins in the putative CPT and FPPS biosynthetic gene cluster. Proteins in the extracted latex were concentrated by trichloroacetic acid precipitation and acetone as previously described [174]. Protein-enriched latex samples were further separated through SDS-polyacrylamide gel electrophoresis. Heavily stained protein bands from the resulting gel were cut out and submitted to the University of Minnesota's Center for Mass Spectrometry and Proteomics, where samples underwent further processing into peptides

via trypsin digestion. **Figure 4.8** illustrates the gel bands submitted for proteomics analysis.

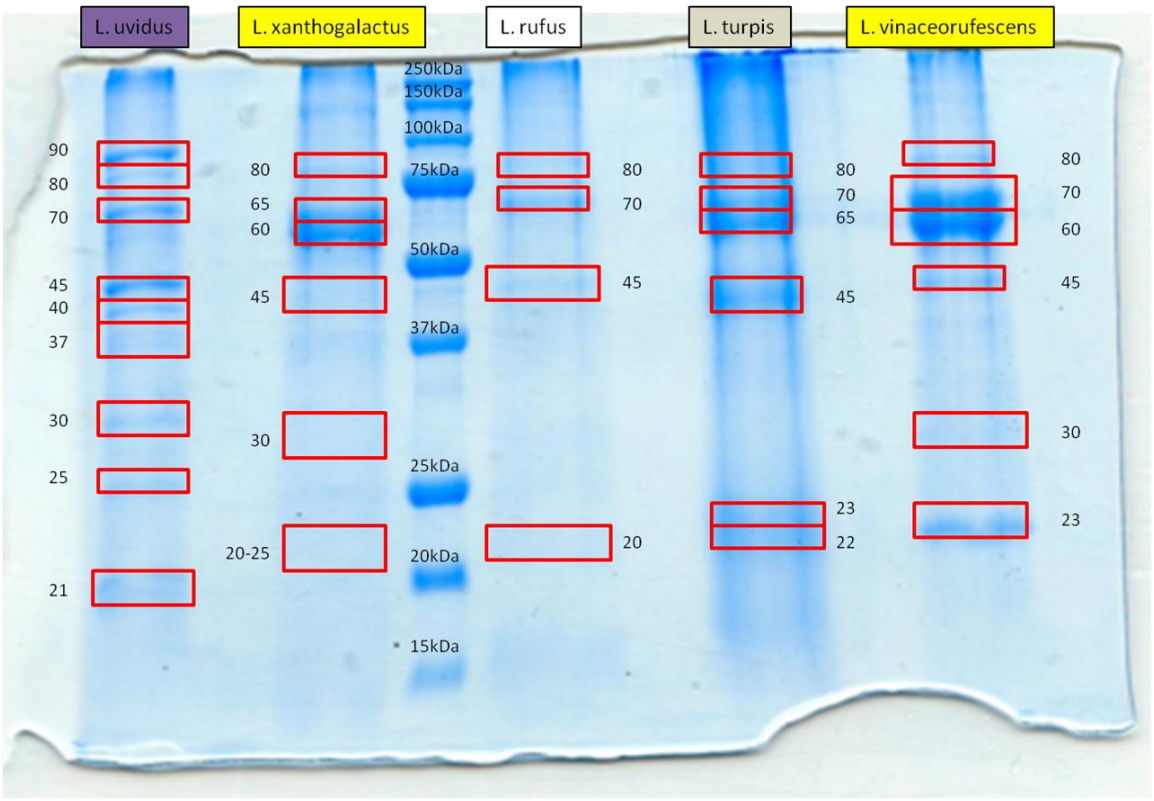


Figure 4.8. 15% SDS gel of enriched proteins from *Lactarius* fruiting body latex. Protein bands submitted for tandem mass spectrometry are boxed in red and given a name corresponding to their estimated protein size.

After processing into peptides, samples underwent peptide sequencing via tandem mass spectrometry. Each sample run yielded an average of approximately 1100 *de novo* peptide sequences an average of 12 amino acids long (see **Supplementary Table S4.4** for *de novo* sequences generated for each sample). However, a disadvantage of cutting gel bands out for mass spectrometry experiments is the inevitable introduction of contaminant keratin proteins. The presence of contaminating keratin peptides was mitigated by authoring a Perl script called peptidefilter.pl that would remove the contaminant peptides. The remaining peptide sequences were written to a .fasta file for

subsequent BLAST analysis against predicted amino acid sequences. This strategy of complementing genomics data with proteomics data using BLAST was inherently flawed though, as BLAST is not optimized to run with short peptides as queries. The average peptide length from tandem mass spectrometry was 12 amino acids in length. Thus, parsing out meaning from this analysis was too difficult and tedious to justify using BLAST.

However, the proteomics data was collected before the first *Lactarius* draft genome assemblies were obtained. Because PEAKS7.0 is optimized to deal with mapping short peptides using sequence tag homology searches, plans are in place to use PEAKS7.0 to map *de novo* peptide sequences onto predicted amino acids from assembled *Lactarius* genomes, much like it would use databases from Genbank. Thus, the integration of *Lactarius* proteomics data with respective genomic data is still ongoing.

3.5 Preliminary cloning and characterization of *L. rufus* CPTs, and FPPSs

As previously stated, the *L. rufus* genome was the first genome assembled in this project. Thus, attempts were made to PCR clone out each predicted CPT and FPPS from *L. rufus* cDNA libraries prepared from the cap, gills, and stem using gene specific primers. Positive amplicons were obtained for the predicted genes LrufFPPS2 and LrufCPT2 between all three tissue types. Amplicons were subsequently gel purified and blunt cloned into pCR™-BluntII-TOPO® plasmids (Invitrogen, Carlsbad, CA). Presence of LrufFPPS2 and LrufCPT2 in pCR™-BluntII-TOPO® (abbreviated as TOPO) were confirmed via Sanger sequencing.

Sequenced confirmed constructs on the pCR™-BluntII-TOPO® cloning plasmids were selected to be subcloned onto the pET32b expression vector (Novagen, Madison, WI) and on our in house *E. coli* expression plasmid vector pCuminBB. pET32b constructs would be used for high level protein expression for purification and subsequent *in vitro* assays with the PiPer™ Pyrophosphate Assay Kit (ThermoFisher Scientific, Waltham, MA). Use of this kit would confirm diphosphate cleavage activity. pCuminBB is a pUC BioBrick™ plasmid that has been modified in the promoter region to become cumate inducible. pCuminBB constructs would be used in tandem with our in-house pAC-crtM-crtN-crtOx (abbreviated as pAC-crtMNOx) plasmid in carotenoid-based screens to infer activity. pAC-crtMNOx is a low copy number plasmid housing three carotenoid genes: crtM, crtN, and crtOx, capable of synthesizing at least 11 different carotenoid molecules [177]. Recombinant *E. coli* cultures turn pink as a result of the carotenoids expressed from pAC-crtMNOx.

Plans are in place to co-express pCuminBB-FPPS constructs with pAC-crtMNOx in a ΔispA *E. coli* strain [188] that we possess in-house. ispA is the farnesyl pyrophosphate synthase in *E. coli*. Recombinant ΔispA *E. coli* co-expressing both pAC-crtMNOx and pCuminBB-FPPS should only turn pink if the FPPS in the construct is functional, as it would be the only major source of FPP in the recombinant strain. This screen has not been performed yet for reasons described later in this section.

A carotenoid-based screen was also used to infer CPT activity from pCuminBB-CPT constructs in *E. coli* JM109 cells [189]. However, both carotenoids and *cis*-1,4-polyisoprene utilize FPP as a substrate in their biosynthesis. Thus, an active CPT would

reduce the amount pink coloration in recombinant *E. coli* colonies co-expressing pAC-crtMNOx and pCuminBB-CPT constructs. See **Figure 4.9** below for a scheme of these two screens.

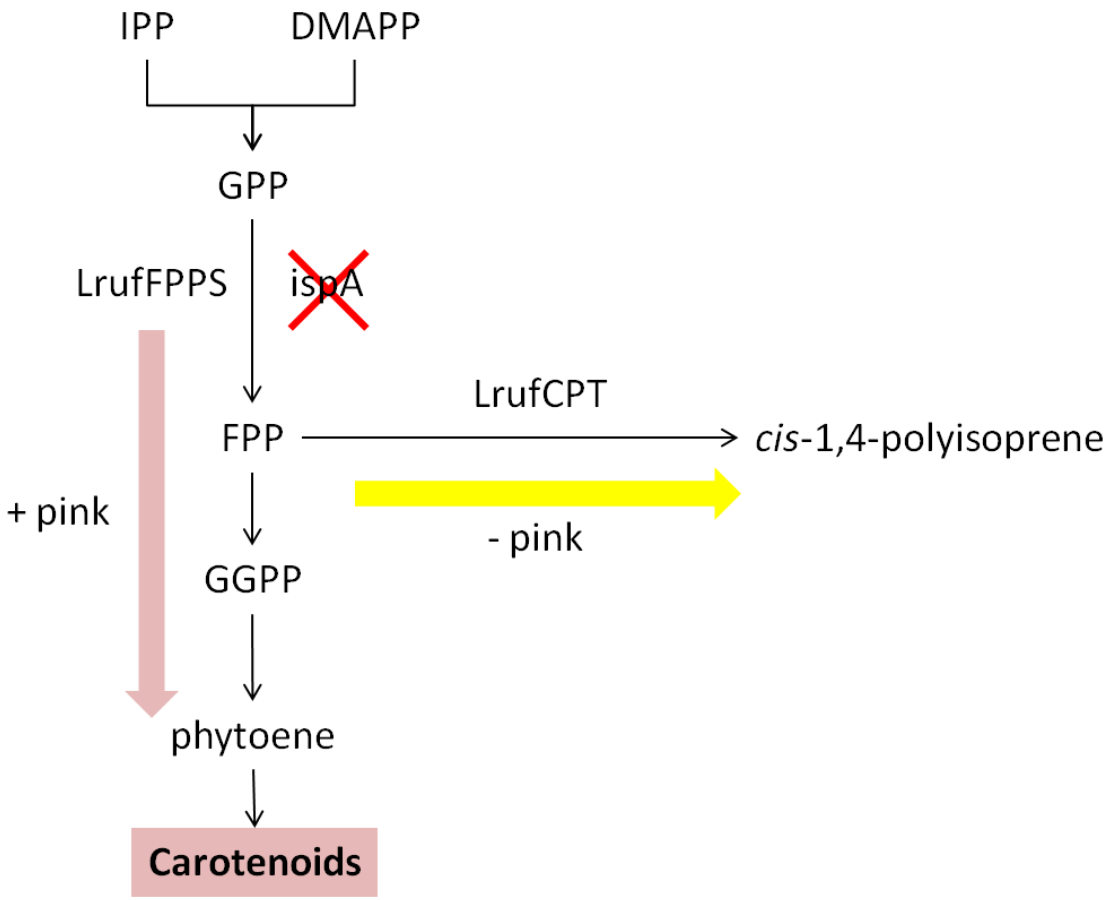


Figure 4.9 Carotenoid screens for FPPS and CPT activity. Co expression of pAC-crtMNOx with active FPPS in Δisp *E. coli* will result in pink colonies. Co-expression of pAC-crtMNOx with active CPT in *E. coli* JM109 cells will result in yellow or less pink colonies. The color change is detectable by naked eye.

Sequencing of the TOPO construct confirmed correct cloning of LrufFPPS2, which was successfully subcloned into pET32b. Further subcloning of LrufFPPS2 into our in house *E. coli* expression vector pCuminBB is ongoing, which has prevented the carotenoid screen of this gene to date.

Sequencing of the TOPO-LrufCPT2 construct identified a small intron located in the C-terminal region of cloned LrufCPT2. This intron introduces a premature stop codon in the LrufCPT2 amino acid sequence, removing conserved regions CSIV and CSV from the protein. Thus, the intron was removed using a Q5 Site Directed Mutagenesis Kit (New England Biolabs, Inswich, MA) from the TOPO plasmid yielding a full length LrufCPT2 construct. The original LrufCPT2 gene cloned from cDNA was denoted "truncated LrufCPT2" in subsequent analyses, while the gene construct with the removed intron was denoted "full-length LrufCPT2". Full-length LrufCPT2 was successfully subcloned into pET32b, while subcloning of full length LrufCPT2 into pCuminBB is ongoing.

It is hypothesized that the initial PCR amplicon from cDNA came from an RNA template that had not been fully spliced into its active form, or initial amplification occurred from residual gDNA. However, truncated LrufCPT2 was successfully subcloned onto pET32b and pCuminBB. Continued cloning of the truncated construct was justified in the unlikely event the small predicted intron in LrufCPT2 was not actually an intron. Co-expression of pCuminBB-LrufCPT2 (truncated) with pAC-crtMNOx in *E. coli* JM109 cells demonstrated competition for FPP as seen in **Figure 4.10**. This reduction in color infers potential *cis*-prenyl transferase activity in truncated LrufCPT2.

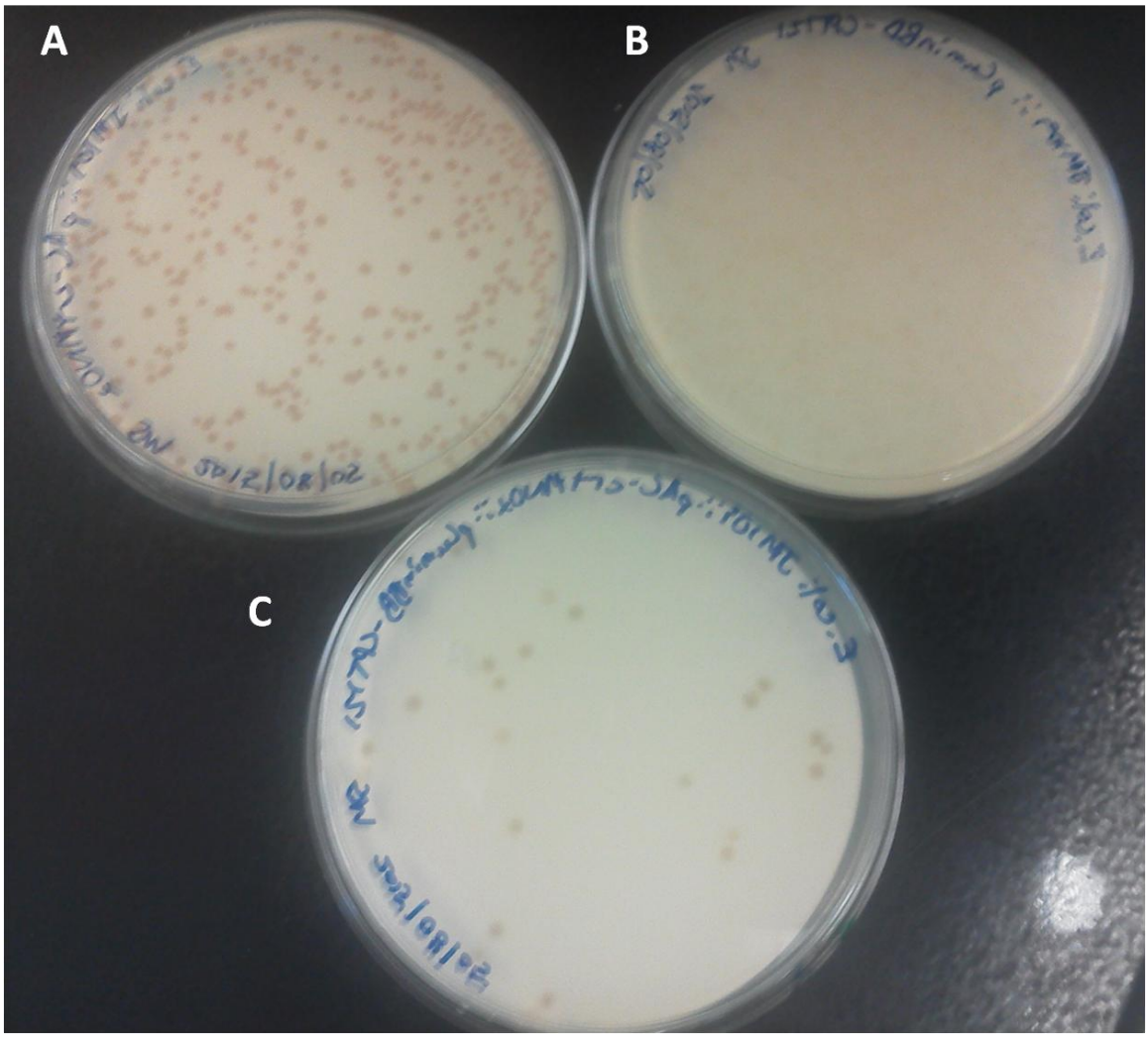


Figure 4.10. Co-transformation carotenoid screen of LrufCPT2 (truncation) in *E. coli* JM109 cells. (A): pAC-crtMNOx transformants; **(B):** pCuminBB-LrufCPT2 (truncation) transformants; **(C):** Co-transformants of pAC-crtMNOx and pCuminBB-LrufCPT2 (truncation).

This co-transformation was performed without using cumate to induce expression of LrufCPT2 (truncation) from the pCuminBB plasmid. The results from the above screen are encouraging considering no induction was performed on a putative truncated CPT. It will be quite interesting to see if results will be replicated with the full length LrufCPT once it is successfully cloned into pCuminBB.

4 Conclusion

The research conducted in this chapter provided encouraging results towards better understanding natural rubber biosynthesis in *Lactarius* mushrooms. The genome assemblies of *L. rufus* and *L. vinaceorufescens* were superior to the genome assembly of *L. uvidus*. The superior genome assembly of *L. rufus* was improved even more with the addition of long-read PacBio sequencing data. This spurred the interest in obtaining long-read PacBio sequencing data for *L. vinaceorufescens* as well. At the time of writing this thesis, long-read PacBio sequencing data of *L. vinaceorufescens* is currently being assembled onto its respective short read assembly via PBJelly2 [105].

The improved draft genomes of *L. rufus* and *L. vinaceorufescens* allowed for improved biosynthetic gene cluster annotation for putative CPTs, FPPSs and STSs in these two species. The manual annotation of these biosynthetic gene clusters yielded encouraging, but also unexpected results.

The presence of predicted transporters clustered nearby the putative CPTs encourage the rubber particle based synthesis of *cis*-1,4-polyisoprene. Unexpectedly, the hypothesis that an FPPS would physically cluster close with a CPT in *Lactarius* was not supported by the scaffolds annotated in this chapter. Another surprise was the lack of P450 enzymes clustered with predicted STSs in *Lactarius* compared to *Stereum*, though the lignin-degrading abilities of *Stereum* may justify the presence of more P450s throughout the *Stereum* genome.

One cause for concern in the characterization of *L. rufus* and *L. vinaceorufescens* CPTs was their grouping with CPTs that make lower molecular weight polyisoprene,

rather than with the *bona fide* rubber transferases. Additionally, Bridgestone Tires Inc. was unable to detect polyisoprene through ^{1}H and ^{13}C NMR analysis in any latex samples we sent from *L. rufus, L. turpis, L. vinaceorufescens* or *L. xanthogalactus*. Conversely, Bridgestone was able to detect polyisoprene in *L. volemus* latex from fruiting bodies we sent them. However, cultivation of *L. volemus* in the laboratory has so far been unsuccessful, precluding genomic analysis. This leads to the hypothesis that not all *Lactarius* species produce *cis*-1,4-polyisoprene in their latex, which the results from the research in this chapter cannot refute.

However, preliminary cloning of truncated LrufCPT2 has demonstrated competition for FPP, indicating possible *cis*-prenyl transferase activity, which is encouraging. It will be especially promising if the full length LrufCPT2 will replicate these results, once it is cloned into pCuminBB. Furthermore, it will also be reassuring to see confirmed FPPS activity in LrufFPPS2 upon its successful cloning into pCuminBB. Cloning of these constructs is currently an ongoing process.

Finally, further analysis of purified LrufCPT2 and LrufFPPS2 proteins from their pET32b constructs with the PiPer™ Pyrophosphate Assay Kit would provide more quantitative insights into the activity of these cloned enzymes, which will be performed soon.

5. Supplemental Material

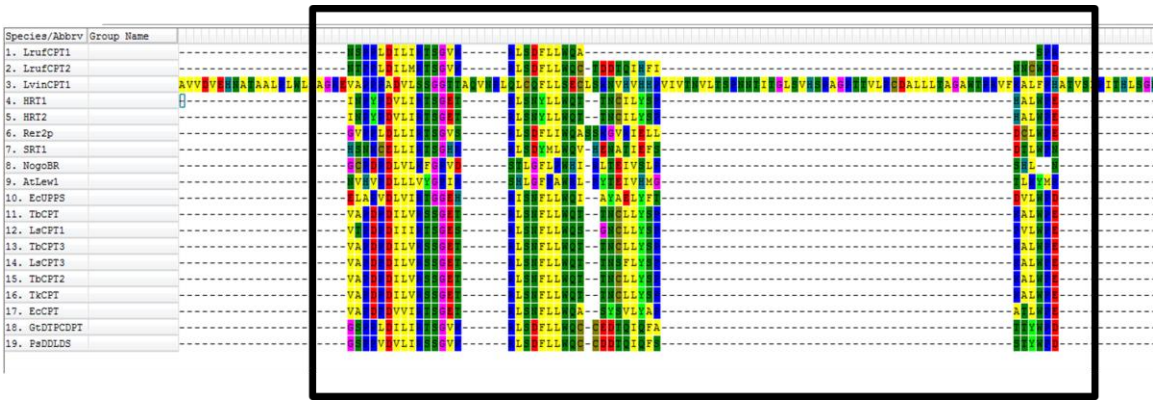


Figure S4.1 Snapshot of multiple sequence alignment of putative CPTs in *L. rufus* and *L. vinaceorufescens*. Conserved region CSV illustrated in the black box**.**

Table S4.1 Initial BLAST hits for putative *cis*-prenyl transferases for *L. rufus and L. vinaceorufescens*. "_length_" in the "Reference" column refers to the length of the assembled scaffold the gene was present on. Gene length refers to the amino acid length of the gene. Subsequent phylogenetic analysis was performed on all genes listed.

Species	Gene #	Reference	Gene Length	CSI?	CSII?	CSIII?	CSIV?	CSV?
L. rufus	6174	Contig419_Gene_6174	247	Y	Y	Y	Y	Y
	6745	Contig521_Gene_6745	290	Y	Y	Y	Y	Y
L. vinaceorufescens	2431	NODE_20756_Gene_2431_length_58314_cov_44.604008	517	Y	Y	Y	Y	N

Figure S4.2 Snapshot of multiple sequence alignment of putative FPPSs in *L. rufus* and *L. vinaceorufescens*. Conserved DDxxD illustrated in the black box.

Table S4.2 Initial BLAST hits for putative *farnesyl pyrophosphate synthases* for *L. rufus and L. vinaceorufescens*. "_length_" in the "Reference" column refers to the length of the assembled scaffold the gene was present on. Gene length refers to the amino acid length of the gene. Subsequent phylogenetic analysis was performed on genes bolded in red.

Species	Gene #	Reference	Gene Length	DDxxD 1?	DDxxD2?
L. rufus	**8769**	**Contig928_Gene_8769**	**869**	**N**	**Y**
	10620	**Contig1902_Gene_10620**	**406**	**Y**	**Y**
	11419	**Contig2635_Gene_11419**	**420**	**Y**	**Y**
L. vinaceorufescens	6922	NODE_63805_Gene_6922_length_11967_cov_38.464027	109	N	N
	12973	**NODE_138734_Gene_12973_length_35075_cov_43.478603**	**378**	**Y**	**Y**
	13101	**NODE_140709_Gene_13101_length_14195_cov_33.274532**	**370**	**Y**	**Y**
	18923	NODE_222067_Gene_18923_length_5553_cov_28.477579	294	Y	Y
	18924	NODE_222067_Gene_18924_length_5553_cov_28.477579	296	N	N

Figure S4.3 Snapshot of multiple sequence alignment of putative STSs in *L. rufus* and *L. vinaceorufescens*. Conserved D(D/E)xx(D/E) illustrated in the black box.

Table S4.3 Initial BLAST hits for putative sesquiterpene synthases for *L. rufus and L. vinaceorufescens*. "_length_" in the "Reference" column refers to the length of the assembled scaffold the gene was present on in bp. Gene length refers to the amino acid length of the gene. Subsequent phylogenetic analysis was performed on all genes bolded in red.

Species	Gene Number	Reference	Gene Length	D(D/E)xx(D/E)?	NSE/DTE?
L. rufus	**686**	**Contig41_Gene_686**	**341**	**Y**	**Y**
	1931	**Contig95_Gene_1931**	**385**	**Y**	**Y**
	4846	Contig296_Gene_4846	356	Y	Y
	4890	**Contig299_Gene_4890**	**343**	**Y**	**Y**
	6294	**Contig438_Gene_6294**	**355**	**Y**	**Y**
	6366	**Contig452_Gene_6366**	**346**	**Y**	**Y**
	6367	**Contig452_Gene_6367**	**351**	**Y**	**Y**
	9478	**Contig1254_Gene_9478**	**629**	**Y**	**Y**
	9742	Contig1520_Gene_9742	593	N	N
	10378	Contig1799_Gene_10378	133	Y	N
	10404	Contig1801_Gene_10404	126	Y	N
	12742	Contig4167_Gene_12742	954	Y	Y
L. vinaceorufescens	**2294**	**NODE_20233_Gene_2294_length_28123_cov_46.616505**	**341**	**Y**	**Y**
	4559	NODE_39720_Gene_4559_length_2587_cov_46.838036	104	N	Y
	6299	NODE_56511_Gene_6299_length_2245_cov_52.483742	127	Y	N
	7033	NODE_65216_Gene_7033_length_1456_cov_41.956730	213	N	Y
	7693	NODE_71904_Gene_7693_length_10224_cov_28.075899	405	Y	N
	8213	**NODE_78168_Gene_8213_length_14618_cov_35.214531**	**461**	**Y**	**Y**
	8470	NODE_81206_Gene_8470_length_10955_cov_43.050755	144	Y	N
	8471	**NODE_81206_Gene_8471_length_10955_cov_43.050755**	**346**	**Y**	**Y**
	8472	NODE_81206_Gene_8472_length_10955_cov_43.050755	214	N	Y
	9162	NODE_89100_Gene_9162_length_596_cov_16.699665	147	Y	N
	9219	NODE_89938_Gene_9219_length_11270_cov_26.645252	193	N	Y
	11742	NODE_122308_Gene_11742_length_1895_cov_55.605804	267	Y	Y
	12120	NODE_127247_Gene_12120_length_2787_cov_44.876209	227	N	Y
	14354	**NODE_159368_Gene_14354_length_9369_cov_41.093925**	**365**	**Y**	**Y**
	14666	NODE_163991_Gene_14666_length_24826_cov_25.188068	127	N	Y
	14668	NODE_163991_Gene_14668_length_24826_cov_25.188068	151	N	Y
	14671	NODE_163991_Gene_14671_length_24826_cov_25.188068	443	Y	Y
	14673	NODE_163991_Gene_14673_length_24826_cov_25.188068	121	N	Y
	18414	NODE_214816_Gene_18414_length_6348_cov_53.663673	208	N	Y
	18746	**NODE_220021_Gene_18746_length_58512_cov_43.853363**	**385**	**Y**	**Y**

Table S4.4 Number of *de novo* sequences identified for each gel slice submitted to the UMN CMSP facility.

Species	Sample Name	*de novo* Peptide Sequences Identified
L. rufus	R20	996
	R45	1349
	R70	1230
	R80	1459
L. turpis	T20	1409
	T23	1192
	T45	1199
	T65	1114
	T70	921
	T80	1178
L. uvidus	U21	876
	U25	984
	U30	1219
	U37	330
	U40	1280
	U45	331
	U70	1191
	U80	1301
	U90	1031
L. vinaceorufescens	V23	1427
	V30	1538
	V45	1404
	V60	1114
	V70	1018
	V80	1984
L. xanthogalactus	X20_25	970
	X30	1413
	X45	1329
	X60	1027
	X65	1072
	X80	1465

Chapter 5: Conclusion and Future Directions

1. Concluding Remarks

The work presented in this thesis continues to represent the value of the predictive framework established by the Schmidt-Dannert laboratory towards fungal sesquiterpenoid biosynthetic gene identification and characterization. The initial aim of this thesis was to fully understand the cyclization mechanism and variable product specificity of Stehi7, a Δ6-protoilludene synthase from the wood-rotting fungus *Stereum hirsutum*. While Stehi7 crystals were ultimately not obtained, precluding a solved fungal STS crystal structure, this aim still represented the tail end of the workflow from gene prediction to structural characterization.

The research presented in this thesis also exemplifies the multidisciplinary background required of contemporary scientists, as the applicable skills required for this project changed with the aims of this thesis. The second aim of this thesis was to further demonstrate the predictive power of the isoprenoid gene prediction pipeline established by the Schmidt-Dannert laboratory towards STS identification in *Lactarius* mushrooms, and adapt it further to identify genes involved in natural rubber biosynthesis. In order to pursue this aim, ecology skills were necessary for taxonomic identification of *Lactarius* species in the field. Furthermore, skills in microbiology were necessary for the successful cultivation of 8 *Lactarius* species. Additional knowledge was necessary in biochemistry discipline in the screening of mycelial culture headspace and latex analysis by GC-MS. Most importantly, skills in bioinformatics were necessary to systematically refine raw genomic data into predicted genes of interest. Finally, skills in molecular biology were

necessary to generate constructs for the biochemical characterization of these predicted genes.

The application of these multiple disciplines has demonstrated the tremendous progress made in updating the established predictive foundation for fungal STS discovery and characterization into a broader scope to include natural rubber biosynthetic genes. Indeed, two equally insightful Ph.D. projects could be performed on the respective sesquiterpenoid, or higher molecular weight isoprenoid biosynthetic genes identified in the three *Lactarius* genomes presented in this thesis alone. Further factoring in the general decline in genome sequencing costs, the sequencing of the other 5 *Lactarius* genomes we have in culture would also constitute an insightful project.

2. Future Directions

2.1 Continuing crystallization screening of Stehi7

There remains a fundamental lack in understanding in what dictates the variable fidelity found in STSs, aside from the residues lining the hydrophobic cavity of the active site. There also remains few crystal structures on sesquiterpene synthases, with only one solved STS structure utilizing the 1,11-cyclization mechanism[64]. The major setback of attempting to crystallize Stehi7 was not attempting to replicate the sitting-drop vapor diffusion method performed in the initial crystallization screens during scale-up. Furthermore, only 4 out of the 79 initial crystallization hits were subjected to further screening, representing only 5% of the initial screen hits. Finally, the reproducible crystalline precipitate of Stehi7 that was observed in Chapter 2 demonstrated that Stehi7 was very close to growing large protein crystals. Given the right time commitment, it

seems plausible that large Stehi7 crystals can be obtained. Otherwise, the abundance of putative 1,11-cyclizing STSs identified in the 3 *Lactarius* genome assemblies could provide other more suitable crystallization targets for a 1,11-cylizing fungal STS.

2.2 Expanding sesquiterpenoid screens of *Lactarius* to liquid and fruiting body extracts

The work presented in Chapter 3 illustrated the differences between sesquiterpenoids identified from *Lactarius* fruiting bodies in the literature, and the volatile headspace profiles of *Lactarius* mycelial cultures. The differences between the two profiles can be explained by the different environments each respective cell type faces. However, the main limitation in screening *Lactarius* mycelial culture headspace was the potential to miss more modified sesquiterpenoids. Sesquiterpenoids lose their volatility as more functional groups are added to the terpene scaffold. Performing solvent extractions on the liquid media of *Lactarius* cultures would potentially allow for a greater pathway elucidation, by virtue of identifying more structural intermediates and final products.

2.3 Achieving *de novo* assemblies solely from long-read data

As was hinted in both Chapter 3 and Chapter 4, PacBio long-read sequencing data is currently being refined with the short-read assemblies of *L. uvidus* and *L. vinaceorufescens.* However, not mentioned was how the raw data output of *L. uvidus* and *L. vinaceorufescens* PacBio sequencing runs were far superior to the *L. rufus* sequencing runs. **Figure 5.1** illustrates the subread filtering curves of the accumulated PacBio SMRTcell sequencing data for each *Lactarius* species.

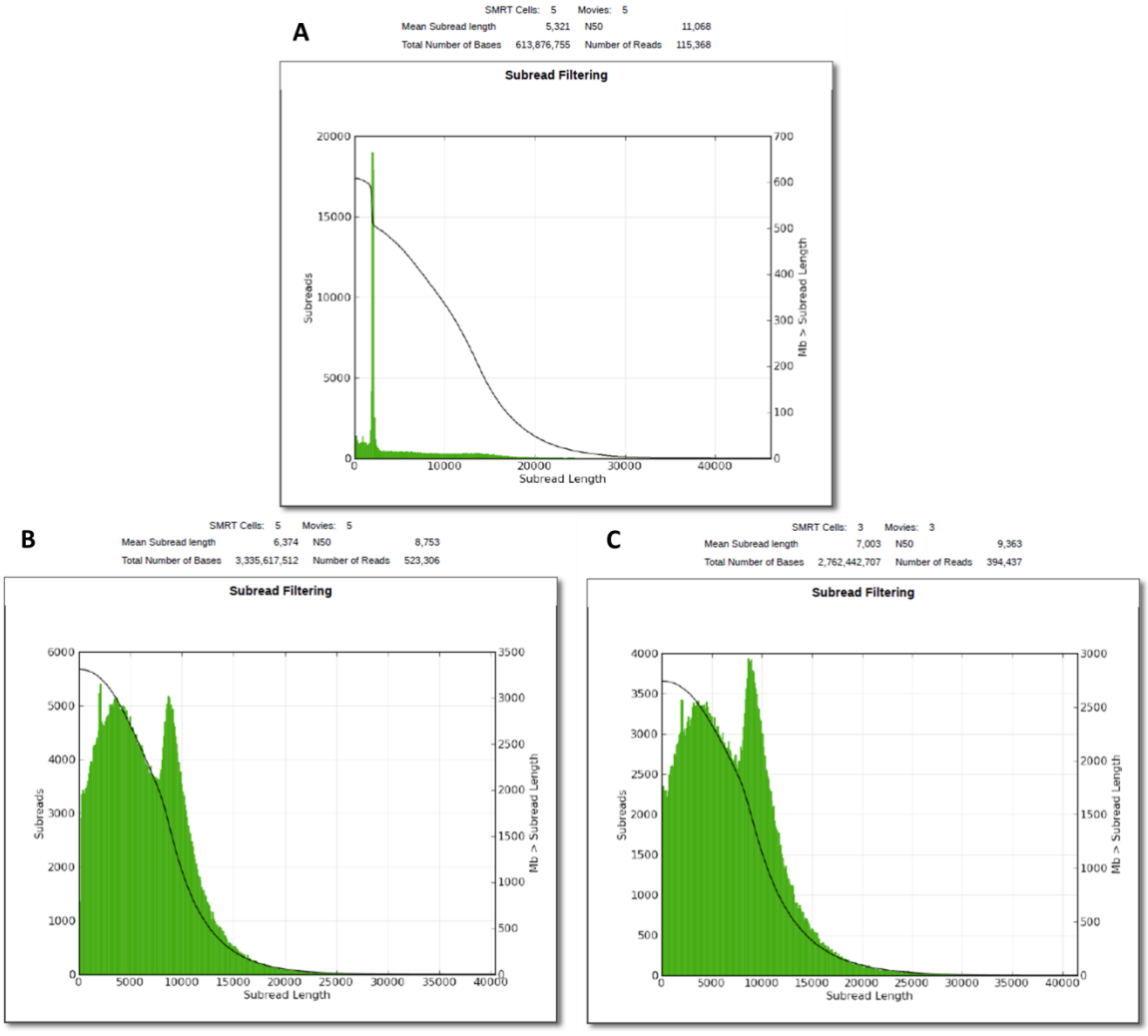


Figure 5.1. Filtered subreads comparison of *Lactarius* long-read PacBio data. (A) *L. rufus*; **(B)** *L. uvidus*; **(C)** *L. vinaceorufescens.*

The subread filtering curve of *L. rufus* indicated a very low yield in the SMRTbell library prep of its sample. This low library yield subsequently resulted in low sequencing coverage. Conversely, the coverage of *L. uvidus* and *L. vinaceorufescens* are on the 50x coverage threshold for a *de novo* assembly from long read sequencing data alone with HGAP3 [179]. Also, the emergence of the diploid-aware assembly program, Falcon [190]

would greatly help resolve the complexities of assembling a dikaryotic *Lactarius* genome. Falcon is said to be in the process of becoming available on MSI soon. Performing assemblies with the above listed software has the potential of reducing the current number of scaffolds in for both respective assemblies by 3 orders of magnitude.

2.4 Refined mining of *L. uvidus* genome with recently published *Antrodia* genome for drimane synthase elucidation

It was very briefly mentioned in Chapter 3 that certain drimanes have demonstrated mTOR inhibiting activity. Antrocin, from the fungus *Antrodia camphorata* in particular is a potent anticancer compound that functions as an mTOR inhibitor [92]. The genome of *Antrodia cinnamomea* has recently been published in PNAS [191]. This genome could potentially be useful in comparative analysis between two drimane synthesizing fungal species, or perhaps help define an antrocin or antrocin-like pathway in *L. uvidus*.

2.5 Establishment of a natural rubber platform in fungal production host

All characterization of the *Lactarius* putative natural rubber biosynthetic genes to this point has been performed in *E. coli*. However, LrufCPT2 and LrufFPPS2 have been successfully cloned into the pESC-Ura yeast expression vector by other members of the Schmidt-Dannert laboratory. Soon, we will have further characterization of these genes in *Saccharomyces cerevisiae*, a more suitable host for fungal biosynthetic gene expression. Should the characterization of *Lactarius* natural rubber synthesis continue to gather promising results, (e.g. express well in yeast), the next step logical step could be establishing a heterologous fungal production host for large-scale polyisoprene production. As filamentous fungi are excellent at secretion of secondary metabolites, they

could provide an ideal production host. For instance, *Fusarium graminearum* has been thought to compartmentalize isoprenoid derived mycotoxins into small vesicles from toxisomes in its Tri5 pathway [192, 193]. The encapsulated mycotoxins are then thought to be removed from the cell by fusion of the small vesicles with the plasma membrane. If natural rubber could instead be produced and encapsulated in these smaller vesicles in place of the isoprenoid mycotoxins, the potential exists for a largely industrially relevant natural rubber production platform.

References

1. Withers, S.T. and J.D. Keasling, *Biosynthesis and engineering of isoprenoid small molecules.* Applied Microbiology and Biotechnology, 2007. **73**(5): p. 980-990.
2. Gershenzon, J. and N. Dudareva, *The function of terpene natural products in the natural world.* Nature Chemical Biology, 2007. **3**(7): p. 408-414.
3. Langenheim, J.H., *Higher-Plant Terpenoids - a Phytocentric Overview of Their Ecological Roles.* Journal of Chemical Ecology, 1994. **20**(6): p. 1223-1280.
4. Justicia, J., et al., *Total synthesis of 3-hydroxydrimanes mediated by titanocene(III) - Evaluation of their antifeedant activity.* European Journal of Organic Chemistry, 2005(4): p. 712-718.
5. Chen, W.Y., I. Vermaak, and A. Viljoen, *Camphor-A Fumigant during the Black Death and a Coveted Fragrant Wood in Ancient Egypt and Babylon-A Review.* Molecules, 2013. **18**(5): p. 5434-5454.
6. Morisaki, N., et al., *Conversion of 6-Protoilludene into Illudin-M and Illudin-S by Omphalotus-Olearius.* Tetrahedron Letters, 1985. **26**(39): p. 4755-4758.
7. Rohmer, M., et al., *Isoprenoid Biosynthesis in Bacteria - a Novel Pathway for the Early Steps Leading to Isopentenyl Diphosphate.* Biochemical Journal, 1993. **295**: p. 517-524.
8. Takahashi, S., et al., *A 1-deoxy-D-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-C-methyl-D-elythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(17): p. 9879-9884.
9. Rohdich, F., et al., *Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(3): p. 1158-1163.
10. Rohdich, F., et al., *Cytidine 5 '-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of Escherichia coli catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(21): p. 11758-11763.
11. Luttgen, H., et al., *Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(3): p. 1062-1067.
12. Hecht, S., et al., *Studies on the nonmevalonate pathway to terpenes: The role of the GcpE (IspG) protein.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(26): p. 14837-14842.
13. Barkovich, R. and J.C. Liao, *Metabolic engineering of isoprenoids.* Metabolic Engineering, 2001. **3**(1): p. 27-39.
14. Schmidt-Dannert, C., *Biosynthesis of terpenoid natural products in fungi.* Adv Biochem Eng Biotechnol, 2015. **148**: p. 19-61.
15. Cao, R., et al., *Diterpene cyclases and the nature of the isoprene fold.* Proteins-Structure Function and Bioinformatics, 2010. **78**(11): p. 2417-2432.

16. Gao, Y., R.B. Honzatko, and R.J. Peters, *Terpenoid synthase structures: a so far incomplete view of complex catalysis.* Natural Product Reports, 2012. **29**(10): p. 1153-1175.
17. Oldfield, E. and F.Y. Lin, *Terpene Biosynthesis: Modularity Rules.* Angewandte Chemie-International Edition, 2012. **51**(5): p. 1124-1137.
18. Christianson, D.W., *Structural biology and chemistry of the terpenoid cyclases.* Chemical Reviews, 2006. **106**(8): p. 3412-3442.
19. Cane, D.E. and I. Kang, *Aristolochene synthase: Purification, molecular cloning, high-level expression in Escherichia coli, and characterization of the Aspergillus terreus cyclase.* Archives of Biochemistry and Biophysics, 2000. **376**(2): p. 354-364.
20. Vedula, L.S., et al., *Structural and mechanistic analysis of trichodiene synthase using site-directed mutagenesis: Probing the catalytic function of tyrosine-295 and the asparagine-225/serine-229/glutamate-233-Mg-B(2+) motif.* Archives of Biochemistry and Biophysics, 2008. **469**(2): p. 184-194.
21. Tantillo, D.J., *Biosynthesis via carbocations: Theoretical studies on terpene formation (vol 28, pg 1035, 2011).* Natural Product Reports, 2011. **28**(12): p. 1956-1956.
22. Abraham, W.R., *Bioactive sesquiterpenes produced by fungi: are they useful for humans as well?* Curr Med Chem, 2001. **8**(6): p. 583-606.
23. Kramer, R. and W.-R. Abraham, *Volatile sesquiterpenes from fungi: what are they good for?* Phytochemistry Reviews, 2011. **11**(1): p. 15-37.
24. Wawrzyn, G.T., et al., *Draft Genome of Omphalotus olearius Provides a Predictive Framework for Sesquiterpenoid Natural Product Biosynthesis in Basidiomycota.* Chemistry & Biology, 2012. **19**(6): p. 772-783.
25. Elisashvili, V., *Submerged Cultivation of Medicinal Mushrooms: Bioprocesses and Products (Review).* International Journal of Medicinal Mushrooms, 2012. **14**(3): p. 211-239.
26. Wasser, S.P., *Current findings, future trends, and unsolved problems in studies of medicinal mushrooms.* Applied Microbiology and Biotechnology, 2011. **89**(5): p. 1323-1332.
27. Zhong, J.J. and J.H. Xiao, *Secondary Metabolites from Higher Fungi: Discovery, Bioactivity, and Bioproduction.* Biotechnology in China I, 2009. **113**: p. 79-150.
28. Lindequist, U., T.H.J. Niedermeyer, and W.D. Julich, *The pharmacological potential of mushrooms.* Evidence-Based Complementary and Alternative Medicine, 2005. **2**(3): p. 285-299.
29. Blackwell, M., *The Fungi: 1, 2, 3 ... 5.1 Million Species?* American Journal of Botany, 2011. **98**(3): p. 426-438.
30. Hawksworth, D.L., *Fungal diversity and its implications for genetic resource collections.* Studies in Mycology, 2004. **50**: p. 9-18.
31. Altshuler, D.M., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
32. Chooi, Y.H. and Y. Tang, *Navigating the fungal polyketide chemical space: from genes to molecules.* J Org Chem, 2012. **77**(22): p. 9933-53.

33. Evans, B.S., S.J. Robinson, and N.L. Kelleher, *Surveys of non-ribosomal peptide and polyketide assembly lines in fungi and prospects for their analysis in vitro and in vivo.* Fungal Genet Biol, 2011. **48**(1): p. 49-61.
34. Boettger, D. and C. Hertweck, *Molecular diversity sculpted by fungal PKS-NRPS hybrids.* Chembiochem, 2013. **14**(1): p. 28-42.
35. Donadio, S., P. Monciardini, and M. Sosio, *Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics.* Nat Prod Rep, 2007. **24**(5): p. 1073-109.
36. Education, P.; Available from: http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-18/18_11.jpg.
37. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
38. Quin, M.B., et al., *Mushroom Hunting by Using Bioinformatics: Application of a Predictive Framework Facilitates the Selective Identification of Sesquiterpene Synthases in Basidiomycota.* Chembiochem, 2013. **14**(18): p. 2480-2491.
39. Agger, S., F. Lopez-Gallego, and C. Schmidt-Dannert, *Diversity of sesquiterpene synthases in the basidiomycete Coprinus cinereus.* Molecular Microbiology, 2009. **72**(5): p. 1181-95.
40. Lopez-Gallego, F., et al., *Sesquiterpene Synthases Cop4 and Cop6 from Coprinus cinereus: Catalytic Promiscuity and Cyclization of Farnesyl Pyrophosphate Geometric Isomers.* Chembiochem, 2010. **11**(8): p. 1093-1106.
41. Lopez-Gallego, F., G.T. Wawrzyn, and C. Schmidt-Dannert, *Selectivity of Fungal Sesquiterpene Synthases: Role of the Active Site's H-1 alpha Loop in Catalysis.* Applied and Environmental Microbiology, 2010. **76**(23): p. 7723-7733.
42. Anchel, M., A. Hervey, and W.J. Robbins, *Antibiotic Substances from Basidiomycetes: VII. Clitocybe Illudens*.* Proceedings of the National Academy of Sciences of the United States of America, 1950. **36**(5): p. 300.
43. Quin, M.B., C.M. Flynn, and C. Schmidt-Dannert, *Traversing the fungal terpenome.* Natural Product Reports, 2014. **31**(10): p. 1449-1473.
44. Agger, S., F. Lopez-Gallego, and C. Schmidt-Dannert, *Diversity of sesquiterpene synthases in the basidiomycete Coprinus cinereus (vol 72, pg 1181, 2009).* Molecular Microbiology, 2009. **72**(5): p. 1307-1308.
45. Lopez-Gallego, F. and C. Schmidt-Dannert, *Multi-enzymatic synthesis.* Current Opinion in Chemical Biology, 2010. **14**(2): p. 174-183.
46. Wawrzyn, G., *Discovery and Characterization of Sesquiterpenoid Biosynthetic Pathways from Basidiomycota*, in *Department of Biochemistry, Molecular Biology and Biophysics*. 2014, Univeristy of Minnesota.
47. van Beilen, J.B. and Y. Poirier, *Establishment of new crops for the production of natural rubber.* Trends Biotechnol, 2007. **25**(11): p. 522-9.
48. Tanaka, Y., et al., *Structure of cis-polyisoprene from Lactarius mushrooms.* ACTA BIOCHIMICA POLONICA-ENGLISH EDITION-, 1994. **41**: p. 303-303.
49. Starks, C.M., et al., *Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase.* Science, 1997. **277**(5333): p. 1815-1820.

50. Lesburg, C.A., et al., *Crystallization and preliminary-X-Ray diffraction analysis of recombinant pentalenene synthase.* Protein Sci., 1995. **4**(11): p. 2436-2438.
51. Lesburg, C.A., et al., *Crystal structure of pentalenene synthase: Mechanistic insights on terpenoid cyclization reactions in biology.* Science, 1997. **277**(5333): p. 1820-1824.
52. Caruthers, J.M., et al., *Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti.* J Biol Chem, 2000. **275**(33): p. 25533-9.
53. Rynkiewicz, M.J., D.E. Cane, and D.W. Christianson, *Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade.* Proc Natl Acad Sci USA, 2001. **98**(24): p. 13543-8.
54. Shishova, E.Y., et al., *X-ray crystal structure of aristolochene synthase from Aspergillus terreus and evolution of templates for the cyclization of farnesyl diphosphate.* Biochemistry, 2007. **46**(7): p. 1941-51.
55. Shishova, E.Y., et al., *X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis.* The Journal of Biological Chemistry, 2008. **283**(22): p. 15431-9.
56. Aaron, J.A., et al., *Structure of epi-isozizaene synthase from Streptomyces coelicolor A3(2), a platform for new terpenoid cyclization templates.* Biochemistry, 2010. **49**(8): p. 1787-97.
57. McAndrew, R.P., et al., *Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production.* Structure, 2011. **19**(12): p. 1876-84.
58. Vetting, M.W., et al., *Experimental Strategies for Functional Annotation and Metabolism Discovery: Targeted Screening of Solute Binding Proteins and Unbiased Panning of Metabolomes.* Biochemistry, 2015. **54**(3): p. 909-931.
59. Li, J.-X., et al., *Rational engineering of plasticity residues of sesquiterpene synthases from Artemisia annua: product specificity and catalytic efficiency.* Biochemical Journal, 2013. **451**(3): p. 417-426.
60. Baer, P., et al., *Induced-fit mechanism in class i terpene cyclases.* Angew Chem Int Ed Engl, 2014. **53**(29): p. 7652-6.
61. PDB.
62. Quin, M.B., G. Wawrzyn, and C. Schmidt-Dannert, *Purification, crystallization and preliminary X-ray diffraction analysis of Omp6, a protoilludene synthase from Omphalotus olearius.* Acta Crystallographica Section F-Structural Biology and Crystallization Communications, 2013. **69**: p. 574-577.
63. Seemann, M., et al., *Pentalenene synthase. Analysis of active site residues by site-directed mutagenesis.* J Am Chem Soc, 2002. **124**(26): p. 7681-9.
64. Lesburg, C.A., et al., *Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology.* Science, 1997. **277**(5333): p. 1820-4.
65. XTalPred.; Available from: http://ffas.burnham.org/XtalPred-cgi/xtal.pl.
66. Mcpherson, A., *Current Approaches to Macromolecular Crystallization.* European Journal of Biochemistry, 1990. **189**(1): p. 1-23.

67. Bergfors, T., *Seeds to crystals.* Journal of Structural Biology, 2003. **142**(1): p. 66-76.
68. Stewart, P.D.S., et al., *Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization.* Crystal Growth & Design, 2011. **11**(8): p. 3432-3441.
69. Obmolova, G., et al., *Promoting crystallization of antibody-antigen complexes via microseed matrix screening.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 927-933.
70. D'Arcy, A., V.A. Frederic, and M. Marsh, *An automated microseed matrix-screening method for protein crystallization.* Acta Crystallographica Section D-Biological Crystallography, 2007. **63**: p. 550-554.
71. Ireton, G.C. and B.L. Stoddard, *Microseed matrix screening to improve crystals of yeast cytosine deaminase. (vol D60, pg 601, 2004).* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 801-801.
72. D'Arcy, A., A. Mac Sweeney, and A. Haber, *Using natural seeding material to generate nucleation in protein crystallization experiments.* Acta Crystallographica Section D-Biological Crystallography, 2003. **59**: p. 1343-1346.
73. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2005. **1751**(2): p. 119-139.
74. Whittington, D.A., et al., *Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15375-80.
75. Koksal, M., et al., *Structure of Isoprene Synthase Illuminates the Chemical Mechanism of Teragram Atmospheric Carbon Emission.* Journal of Molecular Biology, 2010. **402**(2): p. 363-373.
76. Koksal, M., et al., *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis.* Nature, 2011. **469**(7328): p. 116-20.
77. Koksal, M., et al., *Structure of 2-Methylisoborneol Synthase from Streptomyces coelicolor and Implications for the Cyclization of a Noncanonical C-Methylated Monoterpenoid Substrate.* Biochemistry, 2012. **51**(14): p. 3011-3020.
78. Clericuzio, M., et al., *Sesquiterpenes of Lactarius and Russula (mushrooms): An update.* Natural Product Communications, 2008. **3**(6): p. 951-974.
79. Bergendorff, O. and O. Sterner, *The Sesquiterpenes of Lactarius-Deliciosus and Lactarius-Deterrimus.* Phytochemistry, 1988. **27**(1): p. 97-100.
80. Suortti, T., A. Vonwright, and A. Koskinen, *Necatorin, a Highly Mutagenic Compound from Lactarius-Necator.* Phytochemistry, 1983. **22**(12): p. 2873-2874.
81. Becerra, A., et al., *Anatomical and molecular characterization of Lactarius aff. omphaliformis, Russula alnijorullensis and Cortinarius tucumanensis ectomycorrhizae on Alnus acuminata.* Mycologia, 2005. **97**(5): p. 1047-1057.
82. Spiteller, P., *Chemical Defence Strategies of Higher Fungi.* Chemistry-a European Journal, 2008. **14**(30): p. 9100-9110.

83. Sterner, O., et al., *Velutinal Esters of Lactarius-Vellereus and Lactarius-Necator - the Preparation of Free Velutinal.* Tetrahedron Letters, 1983. **24**(13): p. 1415-1418.
84. Daniewski, W.M. and G. Vidari, *Constituents of Lactarius (Mushrooms).* Progress in the Chemistry of Organic Natural Products, 1999. **77**: p. 69-171.
85. Kim, K.H., et al., *Lactarane sesquiterpenoids from Lactarius subvellereus and their cytotoxicity.* Bioorganic & Medicinal Chemistry Letters, 2010. **20**(18): p. 5385-5388.
86. Malagon, O., et al., *Structures and biological significance of lactarane sesquiterpenes from the European mushroom Russula nobilis.* Phytochemistry, 2014. **107**: p. 126-34.
87. Jansen, B.J.M. and A. Degroot, *The Occurrence and Biological-Activity of Drimane Sesquiterpenoids.* Natural Product Reports, 1991. **8**(3): p. 309-318.
88. Liermann, J.C., et al., *Drimane Sesquiterpenoids from Marasmius sp Inhibiting the Conidial Germination of Plant-Pathogenic Fungi.* Journal of Natural Products, 2012. **75**(11): p. 1983-1986.
89. Montenegro, I., et al., *Study on the cytotoxic activity of drimane sesquiterpenes and nordrimane compounds against cancer cell lines.* Molecules, 2014. **19**(11): p. 18993-9006.
90. Derita, M., et al., *Structural requirements for the antifungal activities of natural drimane sesquiterpenes and analogues, supported by conformational and electronic studies.* Molecules, 2013. **18**(2): p. 2029-51.
91. Erkel, G., et al., *Kuehneromycin-a and Kuehneromycin-B, 2 New Biological Active Compounds from a Tasmanian Kuehneromyces Sp (Strophariaceae, Basidiomycetes).* Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1995. **50**(1-2): p. 1-10.
92. Rao, Y.K., et al., *Identification of antrocin from Antrodia camphorata as a selective and novel class of small molecule inhibitor of Akt/mTOR signaling in metastatic breast cancer MDA-MB-231 cells.* Chem Res Toxicol, 2011. **24**(2): p. 238-45.
93. Paza, C., et al., *Drimendiol, a drimane sesquiterpene with quorum sensing inhibition activity.* Nat Prod Commun, 2013. **8**(2): p. 147-8.
94. Kwon, M., et al., *Molecular cloning and characterization of drimenol synthase from valerian plant (Valeriana officinalis).* FEBS Lett, 2014. **588**(24): p. 4597-603.
95. Tariq, V. *Formation of Dikaryons*. 2014; Available from: http://www.fungionline.org.uk/7sexual/5dikaryon.html.
96. Ekblom, R. and J.B.W. Wolf, *A field guide to whole-genome sequencing, assembly and annotation.* Evolutionary Applications, 2014. **7**(9): p. 1026-1042.
97. Ullrich, R., et al., *Novel haloperoxidase from the agaric basidiomycete Agrocybe aegerita oxidizes aryl alcohols and aldehydes.* Applied and environmental microbiology, 2004. **70**(8): p. 4575-4581.

98. Anh, D.H., et al., *The coprophilous mushroom Coprinus radians secretes a haloperoxidase that catalyzes aromatic peroxygenation.* Applied and Environmental Microbiology, 2007. **73**(17): p. 5477-5485.
99. Grobe, G., et al., *High-yield production of aromatic peroxygenase by the agaric fungus Marasmius rotula.* Amb Express, 2011. **1**.
100. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biology, 2010. **11**(8).
101. Blankenberg, D., et al., *Manipulation of FASTQ data with Galaxy.* Bioinformatics, 2010. **26**(14): p. 1783-1785.
102. Giardine, B., et al., *Galaxy: A platform for interactive large-scale genome analysis.* Genome Research, 2005. **15**(10): p. 1451-1455.
103. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014: p. btu170.
104. Andrews, S. *FastQC*. Available from: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/.
105. English, A.C., et al., *Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology.* Plos One, 2012. **7**(11).
106. Smit, A., Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015; Available from: http://wwww.repeatmasker.org.
107. Stanke, M., et al., *AUGUSTUS: a web server for gene finding in eukaryotes.* Nucleic Acids Research, 2004. **32**: p. W309-W312.
108. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
109. Larkin, M.A., et al., *Clustal W and clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.
110. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Mol Biol Evol, 2013. **30**(12): p. 2725-9.
111. Katoh, K. and D.M. Standley, *MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.* Molecular Biology and Evolution, 2013. **30**(4): p. 772-780.
112. Darriba, D., et al., *ProtTest 3: fast selection of best-fit models of protein evolution.* Bioinformatics, 2011.
113. Akaike, H., *Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models.* Biometrika, 1973. **60**(2): p. 255-265.
114. Schwarz, G., *Estimating Dimension of a Model.* Annals of Statistics, 1978. **6**(2): p. 461-464.
115. Sugiura, N., *Further Analysis of Data by Akaikes Information Criterion and Finite Corrections.* Communications in Statistics Part a-Theory and Methods, 1978. **7**(1): p. 13-26.
116. Rangel, L.A.D., L.F.A.M. Gomes, and R.A. Moreira, *Decision theory with multiple criteria: an aplication of ELECTRE IV and TODIM to SEBRAE/RJ.* Pesquisa Operacional, 2009. **29**: p. 577-590.

117. Le, S.Q. and O. Gascuel, *An improved general amino acid replacement matrix.* Molecular Biology and Evolution, 2008. **25**(7): p. 1307-1320.
118. Yang, Z., *Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites.* Molecular Biology and Evolution, 1993. **10**(6): p. 1396-1401.
119. Guindon, S. and O. Gascuel, *A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.* Systematic Biology, 2003. **52**(5): p. 696-704.
120. Beyhan, K., *Mycelial growth requirements of Lactarius pyrogalus and Lactarius controversus.* African Journal of Microbiology Research, 2011. **5**(28).
121. Breheret, S., et al., *Monoterpenes in the aromas of fresh wild mushrooms (Basidiomycetes).* Journal of Agricultural and Food Chemistry, 1997. **45**(3): p. 831-836.
122. Munzenberger, B., et al., *Detoxification of ferulic acid by ectomycorrhizal fungi.* Mycorrhiza, 2003. **13**(2): p. 117-21.
123. Mihailovic, M., et al., *Protective Effects of the Mushroom Lactarius deterrimus Extract on Systemic Oxidative Stress and Pancreatic Islets in Streptozotocin-Induced Diabetic Rats.* J Diabetes Res, 2015. **2015**: p. 576726.
124. Muhlmann, O. and F. Gobl, *Mycorrhiza of the host-specific Lactarius deterrimus on the roots of Picea abies and Arctostaphylos uva-ursi.* Mycorrhiza, 2006. **16**(4): p. 245-50.
125. Grdovic, N., et al., *The protective effect of a mix of Lactarius deterrimus and Castanea sativa extracts on streptozotocin-induced oxidative stress and pancreatic beta-cell death.* Br J Nutr, 2012. **108**(7): p. 1163-76.
126. Sarikurkcu, C., B. Tepe, and M. Yamac, *Evaluation of the antioxidant activity of four edible mushrooms from the Central Anatolia, Eskisehir - Turkey: Lactarius deterrimus, Suillus collitinus, Boletus edulis, Xerocomus chrysenteron.* Bioresour Technol, 2008. **99**(14): p. 6651-5.
127. LeonGuzman, M.F., I. Silva, and M.G. Lopez, *Proximate chemical composition, free amino acid contents, and free fatty acid contents of some wild edible mushrooms from Queretaro, Mexico.* Journal of Agricultural and Food Chemistry, 1997. **45**(11): p. 4329-4332.
128. Velisek, J. and K. Cejpek, *Pigments of Higher Fungi: A Review.* Czech Journal of Food Sciences, 2011. **29**(2): p. 87-102.
129. Kakou, Y., P. Crews, and G.J. Bakus, *Dendrolasin and Latrunculin A from the Fijian Sponge Spongia mycofijiensis and an Associated Nudibranch Chromodoris lochi.* Journal of Natural Products, 1987. **50**(3): p. 482-484.
130. Kuo, M. *Lactarius psammicola.* . 2011; Available from: http://www.mushroomexpert.com/lactarius_psammicola.html.
131. Daniewski, W.M., M. Kocor, and S. Thoren, *Constituents of Higher Fungi .8. Isolactarorufin, a Novel Tetracyclic Sesquiterpene Lactone from Lactarius Rufus.* Heterocycles, 1976. **5**: p. 77-84.

132. Luo, D.Q., et al., *Rufuslactone, a new antifungal sesquiterpene from the fruiting bodies of the basidiomycete Lactarius rufus.* Journal of Antibiotics, 2005. **58**(7): p. 456-459.
133. Fugmann, B., B. Steffan, and W. Steglich, *Necatorone, an alkaloidal pigment from the gilled toadstool Lactarius necator (agaricales).* Tetrahedron letters, 1984. **25**(33): p. 3575-3578.
134. Sterner, O., *The Co-Formation of Sesquiterpene Aldehydes and Lactones in Injured Fruit Bodies of Lactarius-Necator and Lactarius-Circellatus - the Isolation of Epi-Piperalol.* Acta Chemica Scandinavica, 1989. **43**(7): p. 694-697.
135. Daniewski, W.M., et al., *8-Epi-Pipertriol, a Lactarane Sesquiterpene from Lactarius-Necator.* Phytochemistry, 1988. **27**(10): p. 3315-3317.
136. Garlaschelli, L., et al., *Fungal Metabolites .32. New Fatty-Acid Esters of Drimane Sesquiterpenes from Lactarius-Uvidus.* Journal of Natural Products, 1994. **57**(7): p. 905-910.
137. Debernardi, M., et al., *Fungal Metabolites .15. Structure and Chemical Correlations of Uvidin-C, Uvidin-D, and Uvidin-E, New Drimane Sesquiterpenes from Lactarius-Uvidus Fries.* Journal of the Chemical Society-Perkin Transactions 1, 1983(11): p. 2739-2743.
138. Debernardi, M., et al., *Fungal Metabolites .5. Uvidins, New Drimane Sesquiterpenes from Lactarius-Uvidus Fries.* Journal of the Chemical Society-Perkin Transactions 1, 1980(1): p. 221-226.
139. Debernardi, M., et al., *Fungal Metabolites .29. The Chemical Basis of Hot-Tasting and Yellowing of the Mushrooms Lactarius-Chrysorrheus and L-Scrobiculatus.* Tetrahedron, 1993. **49**(7): p. 1489-1504.
140. Favrebonvin, J. and K. Gluchofffiasson, *Structures of 2 Glutinopallal Esters, New Natural Sesquiterpenoids from Lactarius-Glutinopallens.* Phytochemistry, 1988. **27**(1): p. 286-287.
141. Luo, D.Q., et al., *Humulane-type sesquiterpenoids from the mushroom Lactarius mitissimus.* J Nat Prod, 2006. **69**(9): p. 1354-7.
142. Pyysalo, H. and K.G. Widen, *Sesquiterpenoids in Finnish Lactarius Mushrooms .3. Application of Gas-Chromatography to the Analysis of Sesquiterpene Lactones from Lactarius (Russulaceae) Mushrooms.* Journal of Chromatography, 1980. **190**(2): p. 466-470.
143. Vlad, P.F., *Synthetic Investigations in the Field of Drimane Sesquiterpenoids.* Bioactive Natural Products (Pt M), Vol 33, 2006. **33**: p. 393-432.
144. Widen, K.G. and E.L. Seppa, *Sesquiterpenoids in Finnish Lactarius Mushrooms .1. 15-Hydroxyblennin a, a New Lactarane-Type Sesquiterpene Lactone Isolated from Lactarius-Torminosus.* Phytochemistry, 1979. **18**(7): p. 1226-1227.
145. Yaoita, Y., K. Machida, and M. Kikuchi, *Studies on the constituents of mushrooms, part V - Structures of new marasmane sesquiterpenoids from Lactarius piperatus (SCOP.: Fr.) S. F. Gray.* Chemical & Pharmaceutical Bulletin, 1999. **47**(6): p. 894-896.
146. Cornish, K., *Similarities and differences in rubber biochemistry among plant species.* Phytochemistry, 2001. **57**(7): p. 1123-1134.

147. Cataldo, F., *Guayule rubber: A new possible world scenario for the production of natural rubber?* Progress in rubber and plastics technology, 2000. **16**(1): p. 31-59.
148. IndexMundi. *Rubber Daily Price*. Available from: http://www.indexmundi.com/commodities/?commodity=rubber&months=360.
149. Davis, W., *The rubber industry's biological nightmare.* Fortune, 1997. **136**(3): p. 86.
150. Le Guen, V., et al., *Bypassing of a polygenic Microcyclus ulei resistance in rubber tree, analyzed by QTL detection.* New Phytologist, 2007. **173**(2): p. 335-345.
151. Bousquet, J., et al., *Natural rubber latex allergy among health care workers: a systematic review of the evidence.* Journal of Allergy and Clinical Immunology, 2006. **118**(2): p. 447-454.
152. Belcher, B., N. Imang, and R. Achdiawan, *Rattan, rubber, or oil palm: cultural and financial considerations for farmers in Kalimantan.* Economic Botany, 2004. **58**(1): p. S77-S87.
153. Bowers, J.E., *Natural rubber-producing plants for the United States.* Natural rubber-producing plants for the United States., 1990.
154. Ieda, T., et al., *Rubber Transferase Activity and Rubber Particle Sizes Derived from Lactarius Mushrooms.* Asian Journal of Chemistry, 2013. **25**(9): p. 5185-5188.
155. Bushman, B.S., et al., *Identification and comparison of natural rubber from two Lactuca species.* Phytochemistry, 2006. **67**(23): p. 2590-2596.
156. Puskas, J.E., et al., *Natural rubber biosynthesis—A living carbocationic polymerization?* Progress in polymer science, 2006. **31**(6): p. 533-548.
157. Cornish, K. and W.S. Xie, *Natural Rubber Biosynthesis in Plants: Rubber Transferase.* Natural Product Biosynthesis by Microorganisms and Plant, Pt A, 2012. **515**: p. 63-82.
158. Collins-Silva, J., et al., *Altered levels of the Taraxacum kok-saghyz (Russian dandelion) small rubber particle protein, TkSRPP3, result in qualitative and quantitative changes in rubber metabolism.* Phytochemistry, 2012. **79**: p. 46-56.
159. Post, J., et al., *Laticifer-specific cis-prenyltransferase silencing affects the rubber, triterpene, and inulin content of Taraxacum brevicorniculatum.* Plant physiology, 2012. **158**(3): p. 1406-1417.
160. Rojruthai, P., et al., *In vitro synthesis of high molecular weight rubber by Hevea small rubber particles.* Journal of bioscience and bioengineering, 2010. **109**(2): p. 107-114.
161. Schmidt, T., et al., *Molecular cloning and characterization of rubber biosynthetic genes from Taraxacum koksaghyz.* Plant molecular biology reporter, 2010. **28**(2): p. 277-284.
162. Schmidt, T., et al., *Characterization of rubber particles and rubber chain elongation in Taraxacum koksaghyz.* BMC biochemistry, 2010. **11**(1): p. 11.
163. Asawatreratanakul, K., et al., *Molecular cloning, expression and characterization of cDNA encoding cis-prenyltransferases from Hevea brasiliensis.* European Journal of Biochemistry, 2003. **270**(23): p. 4671-4680.

164. Tanaka, Y., et al., *Initiation of Biosynthesis in Cis Polyisoprenes.* Phytochemistry, 1995. **39**(4): p. 779-784.
165. Berthelot, K., et al., *Hevea brasiliensis REF (Hev b 1) and SRPP (Hev b 3): An overview on rubber particle proteins.* Biochimie, 2014. **106**: p. 1-9.
166. Berthelot, K., et al., *Rubber particle proteins, HbREF and HbSRPP, show different interactions with model membranes.* Biochim Biophys Acta, 2014. **1838**(1 Pt B): p. 287-99.
167. Qu, Y., et al., *A lettuce (Lactuca sativa) homolog of human Nogo-B receptor interacts with cis-prenyltransferase and is necessary for natural rubber biosynthesis.* Journal of Biological Chemistry, 2015. **290**(4): p. 1898-1914.
168. Epping, J., et al., *A rubber transferase activator is necessary for natural rubber biosynthesis in dandelion.* Nature Plants, 2015. **1**(5).
169. Mantello, C.C., et al., *De novo assembly and transcriptome analysis of the rubber tree (Hevea brasiliensis) and SNP markers development for rubber biosynthesis pathways.* PLoS One, 2014. **9**(7): p. e102665.
170. Rahman, A.Y., et al., *Draft genome sequence of the rubber tree Hevea brasiliensis.* BMC Genomics, 2013. **14**: p. 75.
171. Dai, L., et al., *In-depth proteome analysis of the rubber particle of Hevea brasiliensis (para rubber tree).* Plant molecular biology, 2013. **82**(1-2): p. 155-168.
172. Wawrzyn, G.T., S.E. Bloch, and C. Schmidt-Dannert, *Discovery and Characterization of Terpenoid Biosynthetic Pathways of Fungi.* Natural Product Biosynthesis by Microorganisms and Plant, Pt A, 2012. **515**: p. 83-105.
173. De Gussem, K., et al., *Raman spectroscopic monitoring of Lactarius latex.* Phytochemistry, 2006. **67**(23): p. 2580-2589.
174. Link, A.J. and J. LaBaer, *Trichloroacetic acid (TCA) precipitation of proteins.* Cold Spring Harbor Protocols, 2011(2011): p. 993-4.
175. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels.* Analytical chemistry, 1996. **68**(5): p. 850-858.
176. Rappsilber, J., Y. Ishihama, and M. Mann, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.* Analytical chemistry, 2003. **75**(3): p. 663-670.
177. Mijts, B.N., P.C. Lee, and C. Schmidt-Dannert, *Identification of a carotenoid oxygenase synthesizing acyclic xanthophylls: Combinatorial biosynthesis and directed evolution.* Chemistry & Biology, 2005. **12**(4): p. 453-460.
178. Ter-Hovhannisyan, V., et al., *Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.* Genome research, 2008. **18**(12): p. 1979-1990.
179. Chin, C.-S., et al., *Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.* Nature methods, 2013. **10**(6): p. 563-569.
180. Mekkriengkrai, D., et al., *Cloning and characterization of farnesyl diphosphate synthase from the rubber-producing mushroom Lactarius chrysorrheus.* Biosci Biotechnol Biochem, 2004. **68**(11): p. 2360-8.

181. Spiro, R.G., *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds.* Glycobiology, 2002. **12**(4): p. 43R-56R.
182. Teng, K.-H. and P.-H. Liang, *Structures, mechanisms and inhibitors of undecaprenyl diphosphate synthase: A cis-prenyltransferase for bacterial peptidoglycan biosynthesis.* Bioorganic chemistry, 2012. **43**: p. 51-57.
183. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic Acids Res, 2015. **43**(Database issue): p. D222-6.
184. Qu, Y., et al., *A lettuce (Lactuca sativa) homolog of human Nogo-B receptor interacts with cis-prenyltransferase and is necessary for natural rubber biosynthesis.* J Biol Chem, 2015. **290**(4): p. 1898-914.
185. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic Acids Research, 2015. **43**(D1): p. D222-D226.
186. Quin, M.B., S.N. Michel, and C. Schmidt-Dannert, *Moonlighting Metals: Insights into Regulation of Cyclization Pathways in Fungal Δ6-Protoilludene Sesquiterpene Synthases.* ChemBioChem, 2015.
187. Martinez, D., et al., *Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78.* Nature biotechnology, 2004. **22**(6): p. 695-700.
188. Fujisaki, S., et al., *Cloning and nucleotide sequence of the ispA gene responsible for farnesyl diphosphate synthase activity in Escherichia coli1.* Journal of biochemistry, 1990. **108**(6): p. 995-1000.
189. Furubayashi, M., et al., *A high-throughput colorimetric screening assay for terpene synthase activity based on substrate consumption.* PloS one, 2014. **9**(3).
190. Chin, J. *Falcon Genome Assembly Toolkit Manual for v0.2 Branch*. 2014; Available from: https://github.com/PacificBiosciences/FALCON/wiki/Manual.
191. Lu, M.-Y.J., et al., *Genomic and transcriptomic analyses of the medicinal fungus Antrodia cinnamomea for its metabolite biosynthesis and sexual development.* Proceedings of the National Academy of Sciences, 2014. **111**(44): p. E4743-E4752.
192. Menke, J., Y. Dong, and H.C. Kistler, *Fusarium graminearum Tri12p influences virulence to wheat and trichothecene accumulation.* Mol Plant Microbe Interact, 2012. **25**(11): p. 1408-18.
193. Menke, J., et al., *Cellular Development Associated with Induced Mycotoxin Synthesis in the Filamentous Fungus Fusarium graminearum.* PLoS One, 2013. **8**(5): p. e63077.