Identification and characterization of fungal sesquiterpenoid biosynthetic pathways

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Christopher M. Flynn

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Claudia Schmidt-Dannert, Advisor

May 2016

Acknowledgements

Thank you to everyone in the Schmidt-Dannert lab for their guidance, patience, scientific expertise, and positive outlook. Super special thanks to Claudia and Maureen Quin, and their extra efforts to open doors to work together, provide experimental support, and their tireless assistance in improving my scientific communication skills. Thank you as well to Valeriu Bortnov and Aman Imani, two terrific undergraduate students who did important work, and asked the right questions throughout our time together. I would also like to thank everyone in the BioTechnology Institute. The unique collection of labs and researchers in BTI, all with open doors and smiling faces (primarily Ethan Johnson's), has been invaluable in solving problems and making friends ever since my first undergraduate work in Dr. Friedrich Srienc's lab nearly a decade ago. Finally, I would like to thank my wife, Reyhan, for all of her support during this (very) long and varied journey. Thank you all.

Abstract

Fungi produce terpenoids for a wide variety of functions, primarily as signaling and defense compounds. Several highly bioactive sesquiterpenoids are produced only in fungi, particularly those derived from the protoilludene and hirsutene scaffolds. Many sesquiterpenoids have pharmaceutical applications; however, chemical synthesis of sesquiterpenoids is expensive, and biosynthesis in their native fungal hosts is hindered by low concentration and cultivation that is difficult or impossible. Incredible improvements in engineering terpenoid biosynthetic pathways in heterologous hosts have been made in recent years, rendering terpenoid biosynthesis economically viable. However, developing recombinant engineered systems requires the identification of the enzymes and the respective biosynthetic pathways involved in the production of terpenoids from the native fungal host. The primary goal of this work was to identify and characterize the enzymes producing two classes of sesquiterpenoids, derived from protoilludene and hirsutene scaffolds, from *Stereum hirsutum*. The first step in understanding terpenoid biosynthesis is the identification and characterization of the key scaffold producing enzymes, sesquiterpene synthases (STS). Previously, we developed a predictive framework based on STS protein sequence to accurately guide identification of fungal sesquiterpene synthases based on their initial cyclization mechanism. I successfully implemented this framework to identify three novel protoilludene synthases in the genome of *Stereum hirsutum*. Furthermore, data shows that both gene structure as well as protein sequence is highly conserved between fungal STS, reinforcing the hypothesis that fungal STS have evolved based upon initial cyclization mechanisms.

In addition, application of this same predictive framework led to the identification and cloning of the first hirsutene synthase from *Stereum hirsutum*. Unexpectedly, I discovered that the STS was not a typical, single domain enzyme, but instead an unprecedented two-domain STS, 3-hydroxy-3-methylglutaryl-(HMG)-CoA synthase, HS-HMGS. HMG-CoA synthase is a key enzyme that catalyzes the second step in the isoprenoid-precursor mevalonate (MVA) pathway; duplications of this gene in isoprenoid pathway gene clusters may represent a mechanism for increasing MVA pathway flux in fungi, potentially increasing isoprenoid/sesquiterpenoid yield. Following my discovery of this novel bifunctional HS-HMGS, I conducted a large-scale search of all sequenced fungal genomes for duplications of MVA and isoprenoid pathway genes. I found that duplication of early MVA pathway genes were both common and widespread throughout the fungal kingdom, with many HMG-CoA synthases being found in isoprenoid biosynthetic gene clusters as well. This suggests duplication of early MVA pathway genes may be a common mechanism utilized by fungi to increase production of specific isoprenoids.

Following identification of the the protoilludene and hirsutene synthases, I then cloned the P450 and oxidase enzymes found in their gene clusters. These types of enzymes are predicted to be required for the modification of the sesquiterpene scaffold, yielding the final bioactive sesquiterpenoids. Extensive testing of these potential sesquiterpene modifying enzymes identified two, Omp7a and Omp7b, which modify the protoilludene scaffold. Work to purify sufficient quantities of these compounds to

determine the exact chemical modifications by NMR is ongoing, and would be the first demonstration of protoilludene scaffold modification to date.

Finally, while isolating STS and refactoring their biosynthetic pathways is the major goal of this work, understanding the biological function of sesquiterpenoids in the native host is also key, because it provides insights into the bioactivities of sesquiterpenoids and helps us understand the factors that govern expression of the biosynthetic pathways. Therefore, I initiated a collaboration to identify the sesquiterpenes produced by *Fusarium graminearum*, the cause of the crop disease *Fusarium* Head Blight (FHB). *F. graminearum* is known to produce deoxynivalenol (DON), a toxic sesquiterpenoid derived *via* trichodiene synthase (Tri5). However, genomic analysis identified eight sesquiterpene synthases, while only two STS had been characterized to date, suggesting genomic potential to produce non-trichothecene sesquiterpenoids (NTS) that may affect pathogenesis, a topic that has not been comprehensively studied. I therefore set out to determine what NTS are produced by F. graminearum cultures, and if their production is increased under pathogenic (inducing) conditions. GC/MS analysis identified several NTS produced only in induced cultures. Surprisingly, induced *Δtri5* deletion strains not only ceased production of the anticipated Tri5-derived trichothecenes, but also stopped producing sesquiterpenes not produced by directly by Tri5. Thus, while Tri5 expression is necessary for non-trichothecene sesquiterpene biosynthesis, direct catalysis by Tri5 is not sufficient to explain the observed diversity of sesquiterpenoids. These findings suggest that *tri5* expression, through an as-of-yet unidentified mechanism, is required for production of NTS. To determine the mechanism through which Tri5

influences NTS biosynthesis, either *via* protein:protein interactions, or *via* signaling by trichodiene, a catalytically inactive Tri5 was created that retains its secondary structure, and is currently being prepared for testing. While the role of trichothecenes in phytotoxicity is known, the biological function of non-trichothecene sesquiterpenes and their recently discovered co-regulation has not yet been determined. Understanding the role of NTS in pathogenesis may aid in breeding resistant crops, and prove valuable in controlling FHB.

In summary, this work has identified novel fungal sesquiterpene synthases, and demonstrated success in refactoring sesquiterpenoid biosynthetic pathways. My work has provided insights into evolutionary conservation and adaption of gene and protein structures of STS, which will facilitate future discovery and characterization of novel types of STS. In addition, my work with *F. graminearum* has highlighted that sesquiterpene(oids) may play a potential role in gene regulation of biosynthetic pathways, an important consideration when rebuilding pathways in a heterologous host. Finally, the conservation of early MVA pathway genes in isoprenoid biosynthetic gene clusters suggests a previously neglected mechanism of isoprenoid pathway regulation in fungi, which has wide implications for fungal isoprenoid biosynthesis, genome architecture, and mechanisms pathway regulation.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Sections 1.2 through 1.5 were adapted from the book chapter: Quin M.B., Flynn, C.M., Ellinger, J.J., Schmidt-Dannert, C. (2015) "Complex Natural Products Synthesis: Isoprenoids, Polyketides, Phenylpropanoids, Alkaloids, and Non/-Ribosomal Peptides." In Faber, K., Fessner, W.D., Turner, N.J. Science of Synthesis: Biocatalysis in Organic Synthesis, Vol. 3. 361-402. DOI: 10.1055/sos-SD-216-00220. Only the chapter sections directly relevant to isoprenoid biosynthesis, written by CMF, were included in this volume. The remaining sections regarding non-isoprenoid nautral products, written by M.B. Quin and J.J. Ellinger, are not directly relevant to the thesis topic, and were not included.
Hyperlink to original publication:
https://www.thieme-connect.de/products/ebooks/lookinside/10.1055/sos-SD-216-00220

Section 1.6 was previously published as a review: Quin M.B.*, Flynn, C.M.*, Schmidt-Dannert, C. (2014) "Traversing the fungal terpenome: biosynthetic enzymes and pathways." Natural Products Reports, Sept 10; 31(10):1449-73. *Co-First authors. CMF, MBQ, and CSD, planned, wrote, and edited the manuscript. Sections 1.6.1, 1.6.3, 1.6.4 were written primarily by MBQ, while sections 1.6.5-1.6.7 were written by CMF. All authors contributed significantly to Section 1.6.2. Revisions were performed by CSD.
Hyperlink to original publication:
http://www.ncbi.nlm.nih.gov/pubmed/25171145

1.1 Thesis Overview

The aims of this research were 1) to employ bioinformatics to identify novel sesquiterpene synthases (STS) in Basidiomycota genomes, streamlining cloning and characterization efforts, and 2) to use these newly-discovered STS as anchors to identify sesquiterpenoid biosynthetic gene clusters, which could then be cloned and expressed in a heterologous host, allowing the characterization of (partial) pathways converting linear

farnesyl-pyrophosphate (FPP) into bioactive sesquiterpenoids that are uniquely produced by Basidiomycota. This pathway characterization has the potential to enable biosynthesis of complex sesquiterpenoids for medical use.

Chapter one contains an overview of isoprenoid biosynthesis. Sections 1.1 through 1.5 focus on the state-of-the art in engineering microbial and biocatalytic methods for the production of isoprenoid precursor pathways, as well as subsequent terpenoid and carotenoid biosynthetic pathways. In reviewing these methods, the biochemistry of the Mevalonate (MVA) pathway, as well as the current state of sesquiterpene and carotenoid biosynthetic pathway engineering is introduced. Section 1.6 focuses in more depth on fungal terpenoid pathways; and in the interest of this thesis, on fungal sesquiterpenoid biosynthesis. Details are provided regarding the key enzymes, STS, and the diverse cyclization reactions that they catalyze. Also introduced are the concepts regarding the predictive framework developed by our lab, which enables discovery of STS depending upon cyclization mechanism of choice. Finally, the challenges and opportunities currently faced by researchers studying Basidiomycota biochemical pathways are discussed, and alternative strategies proposed.

Chapter two describes initial work to identify STSs from the Basidiomycota *Stereum hirsutum*, which was chosen based on the high density of putative *trans*-humulyl cation-derived STS found in its genome. This work demonstrated the effectiveness of our predictive framework by guiding the successful identification and cloning of five STS, three of which are protoilludene synthases. Notably, we were unable to clone an STS producing the most well-known *S. hirsutum* compound, the aptly named hirsutene.

Chapter three builds upon chapter two, describing the cloning and characterization of the previously unidentified hirsutene synthase, which is a novel bifunctional STS-3-Hydroxy3-methylglutaryl-CoA (HMG-CoA) synthase fusion. HMG-CoA synthase catalyzes the second step in the MVA pathway, and its activity is required for production of the STS substrate, Farnesyl-pyrophosphate (FPP). Further genomic analysis reveals that HMG-CoA synthase and other early MVA pathway gene duplications are a common occurance in the fungal kingdom, with such duplications commonly being found in isoprenoid biosynthetic gene clusters. We hypothesize that these gene duplications may

be a common mechanism for increasing flux through the MVA pathway, and that clustering with other isoprenoid pathway genes may subsequently result in increased isoprenoid output.

Chapter four describes ongoing work to express and characterize fungal sesquiterpenoid biosynthetic pathways in the heterologus host *S. cerevisiae*. Based on previously characterized trichothecene and other terpenoid biosynthetic pathways, this research focused on co-expression of STS with P450 monooxygenases in yeast to produce modified sesquiterpenoids for structural characterization. Through exhaustive co-expression experiments of hirsutene and protoilludene synthases with their respective clustered genes, as well as *in vitro* microsome testing, I identified two protoilludene-scaffold modifying enzymes (one P450, one FAD-binding oxidoreductase) from the genome of *O. olearius*. This chapter concludes with descriptions of ongoing work to improve total production of these modified sesquiterpenoids to enable structure determination by NMR, followed by suggestions for optimizing sesquiterpenoid refactoring, sesquiterpenoid detection methods, and yield improvements.

Finally, chapter five describes collaborative work on the major plant pathogen *Fusarium graminearum*, a known producer of the toxic trichothecene family of sesquiterpenoids. Genomic and transcriptomic analysis revealed that *F. graminuearm* expresses several STSs during infection, but few sesquiterpenoid products have been described in the literature, but may affect *F. graminearum* virulence. I set out to determine what, if any, non-trichothecene sesquiterpenes (NTS) are produced by the plant pathogen *Fusarium graminearum* with and without trichodiene synthase (Tri5), the first enzyme in the trichothecene biosynthetic pathway. My analysis identified production of seven NTS co-produced with trichothecenes. Surprisingly, my analysis found that the *Δtri5* strain not only eliminated production of downstream trichothecenes, critical compounds to pathogenesis, but also eliminated co-production of other NTS. Our hypothesis was that the *Δtri5* deletion would increase production of NTS. The observed result suggests that either Tri5 regulates other STSs through protein:protein interactions, or *via* a trichothecene signaling molecule. Determining the mechanism through which Tri5 affects total sesquiterpenoid production is the subject of ongoing research.

1.2 Complex Natural Product Synthesis: Isoprenoids, Polyketides, and (Non)ribosomal Peptides[1]

General Introduction

Natural products and their chemical analogues are hugely important as pharmaceutical agents. Nearly half of the new drugs that have been introduced to the market during the last 30 years were discovered or extracted from natural sources. Furthermore, the majority of currently available antibacterial and anti-tumor drugs (78% and 74%, respectively) are derived from natural products. The discovery of several novel antibiotics, mainly produced by the actinomycete *Streptomyces* sp., has also led to a significant increase in life expectancy during the last century.[1] Additionally, several key drugs derived from complex natural products, such as the chemotherapy drug taxol, anti-malarial artemisinin, cholesterol-lowering lovastatin, and immune-supressing cyclosporin A, have improved and extended the lives of millions since their discovery. However, only a small fraction of the existing microbes, plants, and fungi have been identified, and our current understanding of their secondary metabolic pathways is somewhat limited. Therefore, there remains a huge potential to discover an incredible diversity of pharmaceutically relevant natural products.

Yet, despite these past successes and overwhelming future potential, the development of natural products for the pharmaceutical industry has fallen into disfavor. While this can be attributed to several factors, perhaps the most influential reason for the lack industrial interest is cost. The large-scale industrial synthesis of many of these complex compounds is difficult, or in some cases, impossible, and is therefore expensive.[2] Moreover, the full chemical syntheses of complex natural products is limited by multistep and typically low yield methods, by a necessity to resolve numerous chiral centers, and is plagued by wasteful side reactions.[3] The alternative, direct extraction of natural products from host organisms, is equally infeasible. Even though

[1] * Sections 1.2 to 1.3 were previously published as: Quin M.B., Flynn, C.M., Ellinger, J.J., Schmidt-Dannert, C. (2015) "Complex Natural Products Synthesis: Isoprenoids, Polyketides, Phenylpropanoids, Alkaloids, and Non/-Ribosomal Peptides." In Faber, K., Fessner, W.D., Turner, N.J. Science of Synthesis: Biocatalysis in Organic Synthesis, Vol. 3. 361-402. DOI: 10.1055/sos-SD-216-00220. Only the sections written entirely by CMF were included in this volume.

natural products are widely produced by microbial or plant hosts (for example, to inhibit predation, bacterial and viral attack, weaken a parasitized host, or prevent competition from other organisms), direct extraction typically only produces enough of the compound of interest for small scale discovery and screening applications. To put this into context, to prepare a single 300 mg dose of taxol from natural sources would require the weight equivalent of an entire 100 year old *Taxus brevifolia* tree.[4] Natural collection of compounds at such low concentrations could, in short, drive the native producer to extinction. Furthermore, the price of the drug would have to be exorbitant in order to cover the cost of cultivation or collection and subsequent extraction. Finally, many natural producer organisms are not amenable to large scale cultivation for several reasons, including slow growth rates, difficult/impossible monoculture, and expensive growth substrates. Therefore, obtaining significant quantities of purified, complex natural products has historically been inhibited by the difficulty of economically producing sufficient quantities to enable testing and implementation.

Recent advances in enzyme discovery, heterologous expression of natural product biosynthetic pathways in model organisms, and mixed chemo-enzymatic syntheses, has revolutionized the production of several major classes of complex natural products. These advances have been facilitated by several interconnected factors. First, the precipitous decline in the cost of sequencing and DNA synthesis over the last five years has enabled an explosive growth in the number of genome sequencing projects. This, in turn, has led to a massive expansion in the number of identified and characterized enzymes from the biosynthetic pathways of complex natural products. Consequently, this knowledge base has greatly facilitated the mixed chemo-enzymatic synthesis of many new classes of natural products, for example the terpenoids artemisinin and taxol, the alkaloids pyrrolizidine and quinolizidine, and non-ribosomal peptide tetramic acid derivatives.[5-9] In addition, recent advances in synthetic biology have enabled pathway optimization to produce valuable, complex natural products in yields that were unimaginable just a decade ago.[10] Future pathway recombination, coupled with newly discovered or mutated enzymes, holds the key to the biosynthetic production of complex natural and unnatural compounds.[11, 12] The following chapter will describe advances in pathway

assembly and complex natural product (bio)synthesis of isoprenoids, polyketides, phenylpropanoids, alkaloids, and (non)-ribosomal peptides.

1.3 Isoprenoid Biosynthesis: Introduction

Isoprenoids are a vast, structurally diverse group of natural products that includes terpenes, carotenoids, natural rubber, sterols, and prenyl-chain additions. These are critical components of proteins and natural products derived from other biosynthetic pathways. Indeed, more than 23,000 isoprenoids have been isolated and identified from natural sources, many of which have known antibiotic, antimalarial, or cytotoxic bioactivities, making them promising therapeutic compounds.[13-15] Elegant total organic syntheses have been developed for some important terpenoids.[16, 17] However, organic synthesis of terpenoids often suffers from side reactions, low yield, expensive catalysts, and low enantioselectivity which limits wider applications.[3] Similarly, biosynthetic isoprenoid production has historically been limited by low production rates and a lack of identified biosynthetic enzymes.[6] However, recent developments in metabolic engineering of easily cultivated organisms has dramatically increased isoprenoid precursor biosynthesis and resulting isoprenoid concentration, while the advent of the genomic era has resulted in an explosion of identified enzymes from isoprenoid biosynthetic pathways.[18-22] These advances have combined to enable bio- and mixed-biological/synthetic production of several complex isoprenoids. This chapter will describe several key factors affecting the production of isoprenoids using examples of complete and mixed biosyntheses.

1.3.1 Isoprenoid precursor biosynthesis

Efficient biotechnological production of isoprenoids requires obtaining the highest possible *in vivo* flux of the universal isoprenoid precursor, isopentenyl pyrophosphate (IPP) (**1**) from its central metabolic precursors. Furthermore, IPP (**1**) is chemically unstable and membrane-impermeable, meaning that its accumulation should be prevented by suitably high enzymatic conversion to more stable, downstream isoprenoid products such as terpenoids and carotenoids. Similarly, identification and prevention of toxic pathway intermediate accumulation is critical to achieving maximal

isoprenoid yields, and has been the subject of extensive metabolomic analysis.[23, 24] Finally, because IPP (**1**) is a universal metabolite, limiting endogenous IPP (**1**) utilization for biomass or other side-products is also employed whenever possible, typically through pathway deletion or reduction through genetic engineering of the host organism.[25] This section will summarize the native isoprenoid biosythetic pathways, while focusing on the most successful examples of isoprenoid biosynthetic pathway optimization to date, determined by the primary driver of process cost, and the final product concentration or titer.[26, 27]

The biosynthesis of IPP (**1**) and its isomer, dimethylallyl diphosphate (**2**) (DMAPP) is unusual biologically, in that it can be produced by two distinct, generally domain-conserved pathways: the largely gram-positive bacterial, archaeal and eukaryotic mevalonate (MVA) pathway (**Figure 1-1**), and eubacterial and plastid-bound 2-c-methyl-D-erythritol-4-phosphate (MEP) pathway shown in **Figure 1-2**.[28-31] In plastid-containing cells, the MEP and MVA pathways are spatially separated between the plastid and the ER/cytoplasm, respectively.[32-34] In addition, despite the compartmentalization of the MEP and MVA pathways in plants, there is evidence of limited exchange of C_5-C_{15} isoprenyl-diphosphates across the plastid membrane, which can rescue some species inhibited in either pathway.[33, 35, 36] Finally, the MVA and MEP pathways are differently regulated in plants, with the carotenoid- and terpenoid- producing MEP pathway more active under light, while the sterol- and ubiquinone- producing MVA pathway is more active in the dark, consistent with the respective roles of these compounds in photosynthesis and biomass production.[33, 37, 38] The effect of this compartmentalization on plant engineering is the necessity to understand which pathway

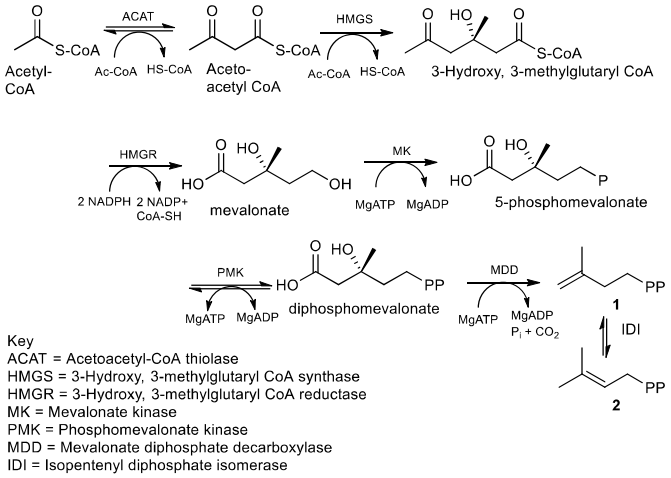


Figure 1-1: The MVA pathway

produces the desired compound, in order to correctly choose the metabolic modifications that will result in the highest production of the desired isoprenoid in plants.

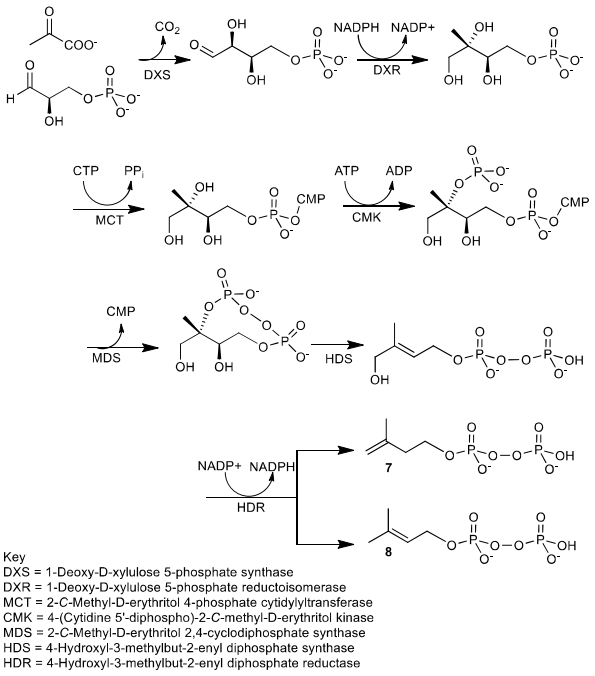


Figure 1-2: The MEP Pathway

1.3.2 Head-to-tail condensation: Isoprenyl diphosphate elongation

All isoprenoids are derived from IPP (**1**). The overwhelming majority of these are derived directly from the C_{10} geranyl diphosphate (**3**), C_{15} farnesyl diphosphate (**4**), and C_{20} geranylgeranyl diphosphate (**5**) **(Figure 1-3)**. Oligomerization of IPP (**1**) into these products is performed by length-specific *E*-isoprenyl diphosphate synthases and can be readily modified through mutagenesis.[39-42] Following reactant binding in the enzyme active site, this head-to-tail prenyl transfer reaction occurs first by abstraction of the terminal pyrophosphate from the C1 carbon of DMAPP (**2**), geranyl diphosphate (**3**), or farnesyl diphosphate (**4**), creating a reactive carbocation. This carbocation is then subject to nucleophilic attack by the terminal (C4) sp^2 carbon, resulting in the formation of a new carbon-carbon bond.[43, 44]

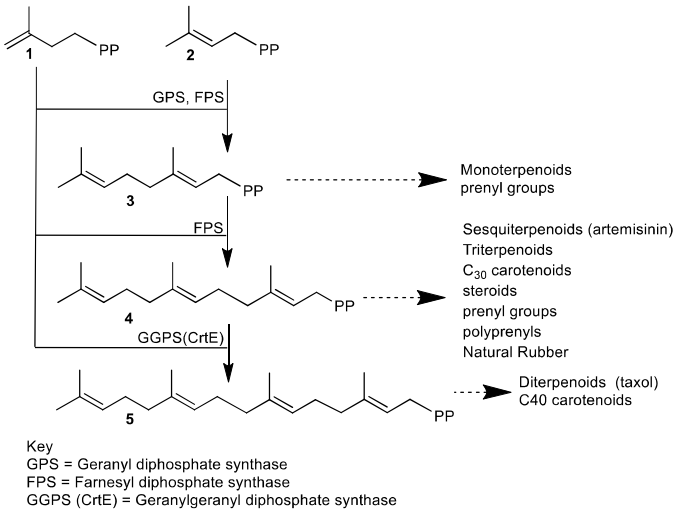


Figure 1-3: Prenyl Diphosphate Elongation

1.3.3 Pathway engineering methods for increasing isoprenoid production in plants[45]

The rate limiting enzyme in the MVA pathway is HMG-CoA reductase (HMGR), while that for the MEP pathway is commonly accepted to be DXP synthase (DXS)[46-49]. As a result, isoprenoid pathway optimization has often focused on increasing the concentration of these enzymes through de-repression, overexpression, or to directly increase their activity through mutation or heterologous expression of enzymes to achieve optimal carbon flux.[50-55] When engineering tomatoes to produce more of the carotenoid β,β-carotene (provitamin A) (**8**) (**Figure 1-4**), constitutive overexpression of the genomically-integrated gene encoding HMGR **(Figure 1.1)** had no effect on carotenoid production, while increasing phytosterol biosynthesis 2.4 fold, as expected based on pathway compartmentalization.[33, 45, 46] Similar genomic integration of the plastid-targeted *E. coli dxs* gene (**Figure 1-2)**, behind the ripening-enhanced fibrillin promoter, resulted in a 60 % increase in total carotenoid accumulation in ripe tomatoes, while having no effect on phytosterol biosynthesis, again consistent with the compartmentalization of carotogenic MEP pathway and generally sterol-producing MVA pathway in plants.[45]

1.3.4 MVA pathway engineering to increase isoprenoid production in *S. cerevisiae*; Typical procedure[10]

The most successful application of isoprenoid biosynthetic pathway optimization to date resulted in a titer of 41 g/l (16.98 % Cmol yield) of the antimalarial precursor amorphadiene, a 9,000 fold increase in concentration over unengineered strains.[56] This optimization procedure can be divided into two parts: **1)** genetic engineering of the host strain MVA pathway, and **2)** optimizing bioreactor conditions to maximize recovered amorphadiene concentration. **1)** The MVA pathway of *Saccharomyces cerevisiae* was overexpressed by genomic integration of a single copy of each gene in the yeast MVA and IPP oligomerization pathways (**Figure 1-1, Figure 1-3**) to ERG20 (encoding GGPS) behind the inducible GAL1-10 promoter in CEN.PK2 yeast.[56] The rate-limiting HMGR enzyme was truncated to produce a soluble enzyme (tHMG1) insensitive to sterol-accelerated degradation[50]. The gene encoding this enzyme was then integrated three times into the host genome behind the GAL1 promoter.[50, 57] The promoter for the native squalene synthase-encoding gene, *ERG9,* was replaced with the repressible MET3 promoter to inhibit sterol synthesis under terpene producing conditions. Derepression of the complete pathway from galactose was achieved by deletion of the negative-regulator of galactose-regulon gene expression, *gal80*. The resulting strain was finally transformed with the amorphadiene synthase-containing plasmid pAM426 to produce strain Y293. **2**) The engineered yeast was then cultivated at high density (OD_{600} > 150) in a fed-batch bioreactor at 30ºC, with

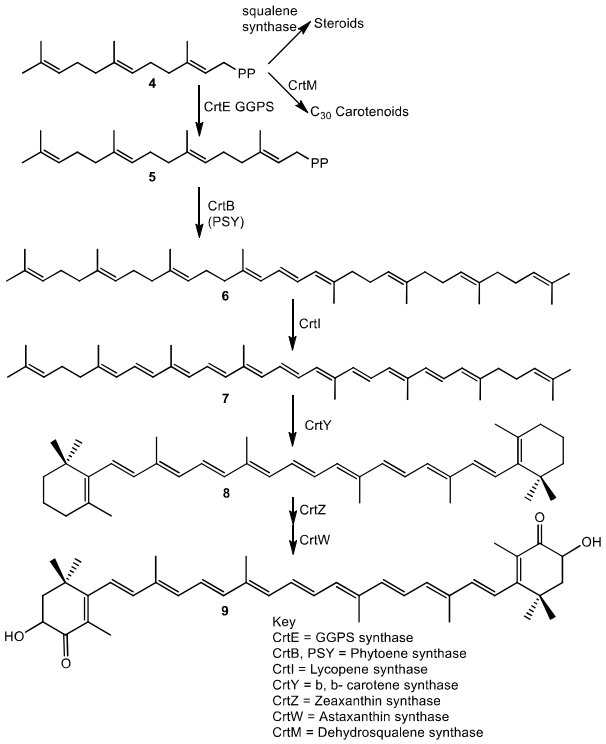


Figure 1-4: Carotenoid Biosynthesis

nitrogen supplementation and pH control using 10 M NH_4OH to maintain a pH of 5.05. Airflow (air or pure O_2 where necessary) supply was constant at 1 L/min, while DO_2 was maintained at 40 % with a variable agitation rate. Batch medium contained glucose (19.5 g/L), methionine (0.25 g/L, added at time of induction 10 H after glucose depletion), 2X phosphate-limiting batch base (500 mL/L), vitamin solution (12 mL/L), trace metals solution (10 mL/L).[58] The 2X batch base contained KH_2PO4 (16 g/L), $(NH_4)_2SO_4$ (30 g/L), and $MgSO_4$•7H2O (12.3 g/L). Batch medium was formulated to be phosphate limited. Seed medium contained tap water (350 mL/L), 2X batch base (500 mL/L), vitamin solution (12 mL/L), trace metals solution (10 mL/L), succinate (0.5 M,pH 5.0; 100 mL/L), Glucose•H_2O (30 mL/L). Strain Y293 was grown overnight in seed medium at 30 ºC and 200 rpm in shake flasks until it reached an OD_{600} = 2.4, was placed on ice for 10 min and frozen at −80ºC in 33 % glycerol in 1 mL aliquots. Sterile stock solutions of glucose monohydrate (715 g/L), succinic acid (60 g/L, 0.5M, pH 5.05), and ethanol (95%) were prepared. Following exhaustion of glucose, ethanol was pulse fed to maintain $[CO_2]$ > 75 % Cmax in 5 minute increments as determined by MS of the offgas. Medium components were adjusted manually once daily to compensate for dilution from the ethanol feed.

1.3.5: MEP pathway engineering to increase isoprenoid production in *E. coli*; Typical procedure[23]

Isoprenoid pathway optimization has also been applied to engineering the MEP pathway, which contrasts with the MVA pathway by being both redox-balanced, and having a 14 % higher theoretical yield from glucose (1.196 glucose for the MEP pathway vs. 1.5 glucose per IPP (**1**) formed for the MVA pathway). The highest titer to date achieved by MEP pathway engineering to date was to produce the C_{20} diterpene taxadiene, the diterpene precursor to taxol (**Figure 1-2**), where a final titer of 1.02 g/L (nearly 15,000-fold improvement *versus* non-MEP modified control) and a glycerol-derived carbon yield of 34 % was obtained. The resulting taxadiene-producing *E. coli* strain (MG1655 *ΔrecA ΔendA* ch1TrcMEPp5t7TG) expressed the *E. coli* MEP pathway module behind a chromosomally integrated Trc promoter, together with taxadiene

synthase and GGPS, expressed behind the strong T5 promoter on the medium-copy number pSC101 plasmid. A net 31X increase in gene expression relative to the native MEP pathway expression level was acheived for GGPS and taxadiene synthase, while integration of the MEP pathway resulted in a 2X increase in MEP pathway gene expression. A 3-L Bioflo bioreactor (New Brunswick) was used for this study. The fermentation was further optimized using a fed batch cultivation with a defined feed medium (13.3 g/L KH_2PO_4, 4 g/L $(NH_4)_2HPO_4$, 1.7 g/L citric acid, 8.4 mg/L EDTA, 2.5 mg/L CoCl2, 15. mg/L MnCl2, 1.5 mg/L CuCl2, 3. mg/L H_3BO_3, 2.5 mg/L Na_2MoO_4, 8. mg/L $Zn(CH_3COO)_2$, 60 mg/L Fe(III) citrate, 4.5 mg/L thiamine, 1.3 g/L MgSO4, 10 g/L glycerol, 5 g/L yeast extract, pH 7.0) and 50 μg/mL spectinomycin for plasmid maintenance. Oxygen was supplied by filtered air at 0.5 (vvm) and agitation was adjusted to maintain DO_2 above 30%. The pH of the culture was controlled at 7.0 using 10 % NaOH. The temperature was controlled at 30ºC until the cells were grown to an OD_{600} of approximately 0.8, whereupon it was reduced to 22ºC and the pathway was induced with 0.1 mM IPTG and cultivated for approximately 120 hours and reaching an OD_{600} of approximately 35. Sterile dodecane was added to 20 % (v/v) of the volume to prevent air stripping of volatile taxadiene. Three grams per liter glycerol was manually added when glycerol concentration dipped to between 0.5 g/L to 1 g/L.

1.4 Head-to-head condensation: Triterpene and carotenoid biosynthesis

Triterpenes and carotenoids are generally derived from the head-to-head condensation of two farnesyl diphosphate (**4**) or geranylgeranyl diphosphate (**5**) molecules, producing precursors such as squalene, dehydrosqualene (C_{30}), or phytoene (C_{40}) (**7**).[40] [59, 60] This section will focus primarily on the production of carotenoids via phytoene production, based on the high level of industrial interest in production of phytoene derived C_{40} carotenoids and due to a relative scarcity of identified biosynthetic enzymes in triterpenoid production.[61] Carotenoids are pigments naturally produced by many microorganisms and plants, many of which may be beneficial to human health.[40] Whether produced from native or non-native hosts, C_{40} carotenoid production requires

the activity of GGPS (CrtE) **(Figure 1-4)**. The resulting geranylgeranyl diphosphate then undergoes a head-to-head condensation reaction, catalyzed by CrtB (PSY), to produce the universal C_{40} carotenoid precursor, phytoene (**6**). Subsequent desaturation and cyclizations produce additional carotenoids, the exact nature of which depends upon the combination of enzymes present.[62] Phytoene-derived carotenoids commonly contain extensive double-bond conjugation, while many also contain cyclohexene functional groups.[61-63] The most notable carotenoids explored for large-scale production are β,β -carotene (**8**), an essential nutrient commonly deficient in many diets potentially alleviated by consumption of „Golden Rice".[64, 65] Astaxanthin (**9**), the most industrially valuable carotenoid, is a red pigment which imbues salmon and other marine animals with their characteristic color, and must be supplied externally as a feed additive in farmed seafoods .[66] Industrially, astaxanthin (**9**) is produced both synthetically and biosynthetically by fungi and algae.[67, 68] Finally, enzyme and pathway engineering techniques have been employed extensively to produce novel carotenoids in non-native hosts with potential applications, often taking advantage of their structure-dependent pigmentation for use in combinatorial biosynthesis and directed evolution strategies to generate new pigments.[12, 40, 69-71]

1.4.1 Carotenoid biosynthesis Introduction

Several methods exist for the total and semi-synthetic production of enantiomerically-impure astaxanthin (**9**) which currently accounts for 97 % of world production, using total organic synthesis methods reviewed elsewhere.[68, 72, 73] Recent biotechnological developments in the fungus *Xanthophyllomyces dendrorhous*, and the algae *Haematococcus pluvialis* have brought the price of biochemically-produced astaxanthin near that of the synthetically-produced compound, with the benefit of being both approved for direct human consumption and entantiomeric purity.[67, 74, 75] In addition, plant strain engineering has successfully increased production of β,β-carotene (**8**) and other carotenoids in plants to increase their micronutrient content.[33, 64, 76]

1.4.2 Carotenoids: Astaxanthin (9) biosynthesis; General procedure

The host organism must either naturally or have been genetically modified to express, the

genes CrtE and CrtB to produce lycopene (**7**). Production of β,β-carotene (**8**) requires additional expresssion of CrtY, while additional co-expression of CrtZ and CrtW are required for the synthesis of astaxanthin (**9**) (**Figure 1-4**). Choice of CrtE and CrtB (PSY) homologue, expression level, and localization significantly affected carotenoid concentration and composition in plants.[32, 64, 69, 74, 77] Exact cultivation procedures vary by organism, while generally requiring two-stage cultivation in single-cell organisms. The first stage requires biomass accumulation under non-expressing conditions (*X. dendrorhous:* darkness, *H. pluvialis:* high-nutrient availability) followed by carotenoid accumulation conditions.[67, 68, 73, 75] *H. pluvialis* forms cysts and accumulates astaxanthin (**10**) under nutrient limitation, while *X. dendrorhous* carotenoid accumuluation can be increased by exposure to light and oxygen.[74] The resulting cells are then pelleted, dried, and pulverized to produce a powdered cell lysate containing 1-3% astaxanthin (**9**) by weight.

1.5 Intramolecular pyrophosphate removal: Terpene cyclization and derivatization

Cyclization of isoprenyl diphosphates is catalyzed by terpene cyclases, found in many bacteria, plants, and fungal species. Terpene cyclases are broadly divided into two classes, based on whether cyclization is initiated by diphosphate removal to generate an unstable, reactive carbocation intermediate (class I), or if the activating step in cyclization is protonation (class II). Terpene cyclase diversity, structure, and reaction mechanisms have been reviewed elsewhere.[78-82] Because a detailed understanding of key residues determining product specificity remains elusive, much of terpene biosynthesis is limited to terpene synthases „mined“ from natural sources, a process greatly accelerated in the last decade by the advent of next-generation genome sequencing.[18, 81, 83-85] The terpene generated by a terpene cyclase is often modified into the final bioactive compound in a multienzymatic pathway, often containing any of several active groups, including epoxides, peroxides, methylene lactones, and other reactive side groups. Despite such complexity, however, the first conversion following cyclization is typically oxidation by a P450 monooxygenase.[21, 24, 61, 83] This section will provide examples

of **1**) terpene biosynthesis utilizing heterologously expressed terpene cyclases, and **2**) the derivatization of these, often biologically inactive, terpenes into bioactive terpenoids by utilizing a combination of enzymatic and organic synthesis methods.

1.5.1 Terpenes: Biosynthesis *in vivo*; General procedure[8, 83, 85, 86]

E. coli strain BL21 (DE3) was transformed with the constitutive-expression plasmid pUCmod, containing the terpene synthase of interest, and cultured 12-18 hours at 30 ºC on LB with appropriate antibiotic for plasmid maintenance. Following incubation, the culture headspace was sampled by Solid Phase MicroExtraction (SPME) for 10 minutes, and analyzed by GC/MS for volatile terpene expression.

1.5.2 Terpenoids: Mixed synthesis of artemisinin (15); Typical procedure [8]

The formation of artemisinin (**15**) from ethanol consists of two distinct steps: first, the formation of artemisinic acid in recombinant *S. cerevisiae*, and second, the four-step chemical conversion of artemisinic acid to artemisinin. Amorphadiene (**10**) was produced in a modified yeast strain similar to that already described, using ethanol-pulse fed bioreactor conditions (section 1.3.4). Yeast expressing *Artemisia annua* amorphadiene synthase (ADS) and cytochrome P450 (CYP71AV1) on the plasmid pAM552 produced artemisinic alcohol (**11**) (**Figure 1-5)**. Subsequent conversion of **11** into artemisinic acid (**12**) was achieved by additional integration of *A. annua*, cytochrome P450-reductase (CPR1), cytochrome b_5 (CYB5), alcohol dehydrogenase (ADH1), and aldehyde dehydrogenase (ALDH1) genes expressed behind the weak GAL3 (CPR1) or strong GAL7 (all others) promoters. Artemisinic acid (**12**) was then chemically reduced with H_2 in the presence of Wilkinson's catalyst to yield (R, R)-dihydroartemisinic acid (**13**) in 94.1:6 EE (**Figure 1-6**). Subsequent protection of the carboxyl group of **13** was mediated by acid-

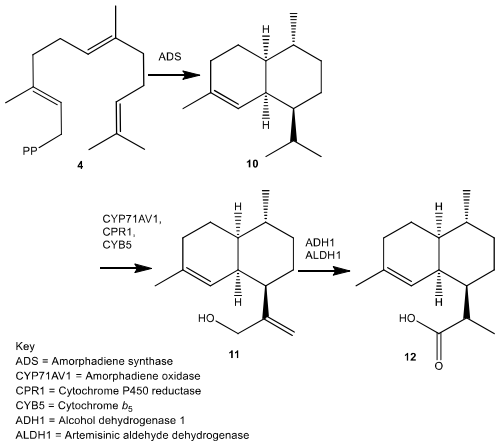


Figure 1-5: Artemisinic Acid Biosynthesis

chloride methyl esterification to produce (R, R)-dihydroartemisinic acid-methyl ester (**14**) in quantitative yields. Finally, a two-step process to first add an allylic 3-hydroperoxide by „ene-type“ reaction of the double bond of **14** was effected by reaction with peroxide-generated singlet oxygen. Subsequent Hock fragmentation and rearrangement produced a ring-opened keto-aldehyde that is reactive with triplet oxygen, forming a 7-membered and a 6-membered lactone ring with correct stereochemistry under acid-catalyzed conditions using Cu(II) DOWEX resin as a solid acid catalyst. This chemical synthetic process resulted in a 40 % yield of artemisinin (**15**) from the biologically produced artemisinic acid precursor.

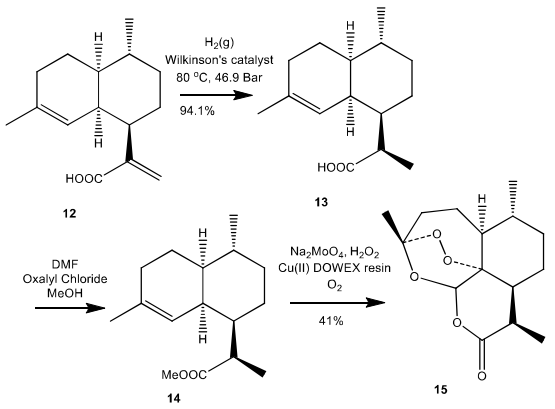


Figure 1-6: Chemical Conversion of Artemisinic Acid to Artemisinin

1.5.3 Isoprenoid biosynthetic methods: Outlook and future directions

The preceding decades of research in isoprenoid biosynthesis have progressed in leaps and bounds. First, the biochemical identification and genetic engineering of the MVA and MEP pathways has resulted in an incredible 4-5 order of magnitude increase in isoprenoid product concentration from whole-cell biocatalysis, raising the economic feasibility of isoprenoid production[10, 23]. The resultant increase in isoprenoid availability made possible by these engineering examples, therefore, will greatly enable isoprenoid study and potential industrial bioproduction of all isoprenoids. Second, the recent explosion in genome sequencing has greatly enabled discovery and characterization of the initial enzymes in isoprenoid biosynthetic pathways. Indeed, the enzyme „toolbox“ now available to the prospective biosynthetic chemist to perform the critical cyclization reaction in terpene biosynthesis has expanded to include examples from nearly every sub-group of terpenes.[19, 22, 82, 83, 87-89] While much exciting work remains for the biosynthetic chemist in MVA, MEP, and isoprenyl diphosphate utilizing enzymes, the complex downstream modification of the initial isoprene is now be

the primary factor limiting large scale production of bioactive isoprenoids. Often this obstacle can be partially overcome by semi-synthetic methods as described above, where intermediate isoprenoids are produced biologically to the greatest extent possible, then chemically modified to yield the final product (e.g. artemisinin, taxol, astaxanthin).[8, 23, 68] However, complete isoprenoid biosynthetic pathway identification and assembly in heterologous hosts faces significant challenges including, but not limited to, incompletely developed methods to predict functional eukaryotic genes, the combinatorial complexity of large pathway assembly, difficulty in expressing active enzymes, and finally the importance of enzyme localization in the efficient production and sequestration of often toxic mature isoprenoids.[22, 33, 90, 91] For example, biosynthesis of deoxynivalenol in the wheat pathogen *Fusarium graminearum* requires the concerted activity of at least 15 enzymes in 3 separate gene clusters, while several steps are likely to be localized to cellular compartments.[92, 93] In addition, correctly balancing metabolite flux and expression of pathway enzymes in heterologous organisms is essential to effective biosynthesis, but difficult to predict *a priori* and may require complex and expensive metabolomic methods.[24, 94] Despite the challenges facing total isoprenoid biosynthesis, recent advances in sequencing, standardization of molecular biology techniques, and progress in enzyme characterization have enabled incredible advances in isoprenoid biosynthesis production and biosynthetic diversity in the last decade.[10, 18, 19, 23, 40, 78]

1.6 Traversing the Fungal Terpenome[2]

1.6.1 Introduction

Fungi are masters of evolution. With a history spanning at least 900 million years,[95] fungi have successfully adapted to almost every habitat on earth.[96] Conservative estimates place global fungal diversity beyond that of land plants by

[2]Section 1.6 was previously published as Quin M.B.*, Flynn, C.M.,* Schmidt-Dannert, C. (2014) "Traversing the fungal terpenome: biosynthetic enzymes and pathways." Natural Products Reports, Sept 10; 31(10):1449-73. See Chapter 1: Introduction for division of labor between MBQ and CMF.

a ratio of 10:1.[97] While there remains some uncertainty regarding the exactnumber of fungal species that exist,[98] it is clear that fungi have been afforded with a unique evolutionary and environmental fitness. Fungi owe this inherent ability to survive to their extensive repertoire of natural product pathways. Many of the natural products from these pathways have antimicrobial, antifungal, immunosuppressive or cytotoxic effects.[99, 100] These bioactive properties enable fungi to successfully take hold of an ecological niche by conferring the ability to compete for nutrients, to deter predators, and to communicate with other organisms in the environment.[101] These same fungal bioactive compounds can also be harnessed by humans to produce valuable medicines such as antimicrobial, anticancer and antiviral agents.[102-109] Natural product pathway discovery and engineering from fungi therefore holds great promise for the pharmaceutical industry.

The importance of natural products for the survival of fungi is evidenced by the fact that they often have many different secondary metabolic pathways that are typically clustered within fungal genomes.[110] One advantage that this feature presents for natural product discovery and engineering is that the physical clustering of genes facilitates genome mining approaches in the identification of biosynthetic pathways. If one gene in a secondary metabolic pathway can be identified, it is highly likely that the other pathway genes will be closely associated and therefore relatively easily identified by bioinformatic methods.[22, 85, 111] The recent decrease in the cost of next-generation DNA sequencing technologies[112] has led to a large increase in the number of sequenced fungal genomes that are publically available through databases such as the Joint Genome Institute (JGI). While some of the fungal species represented in this database are of interest due to their lignin degrading capabilities,[113] recent trends have seen a shift towards genome sequencing and mining strategies to uncover natural product pathways in species that produce medicinally relevant compounds.[89, 114, 115] With such a unique toolbox, those interested in fungal natural product discovery

and biosynthesis can tap into the vast, yet largely unexplored contingent of bioactive natural product pathways.

One of the largest groups of bioactive natural products that have been identified is the terpenoids. With over 55,000 terpene compounds so far described, terpenoids belong to an incredibly structurally diverse class of natural products.[116] Despite this diversity, all terpenoids are derived from the simple five carbon precursor molecules dimethylallyl diphosphate (DMAPP) **1** and isopentenyl diphosphate (IPP) **2** **(Figure 1-7)**. In fungi, these two isomers are produced from acetyl-CoA via the mevalonate pathway.[28] Condensation of IPP **2** and DMAPP **1** monomers results in linear hydrocarbons of varying length: C10 geranyl pyrophosphate (GPP) **3**, C15 (2*E*, 6*E*)-farnesyl pyrophosphate (*(2E, 6E)*-FPP, or FPP) **4**, and C20 geranylgeranyl pyrophosphate (GGPP) **5**. These linear hydrocarbons undergo a dephosphorylation and cyclization cascade to produce terpenes. This highly complex reaction is catalyzed by enzymes known as terpene synthases.[78] Two distinct classes of terpene synthase exist, defined according to substrate activation mechanism. Class I terpene synthases catalyze an ionization-

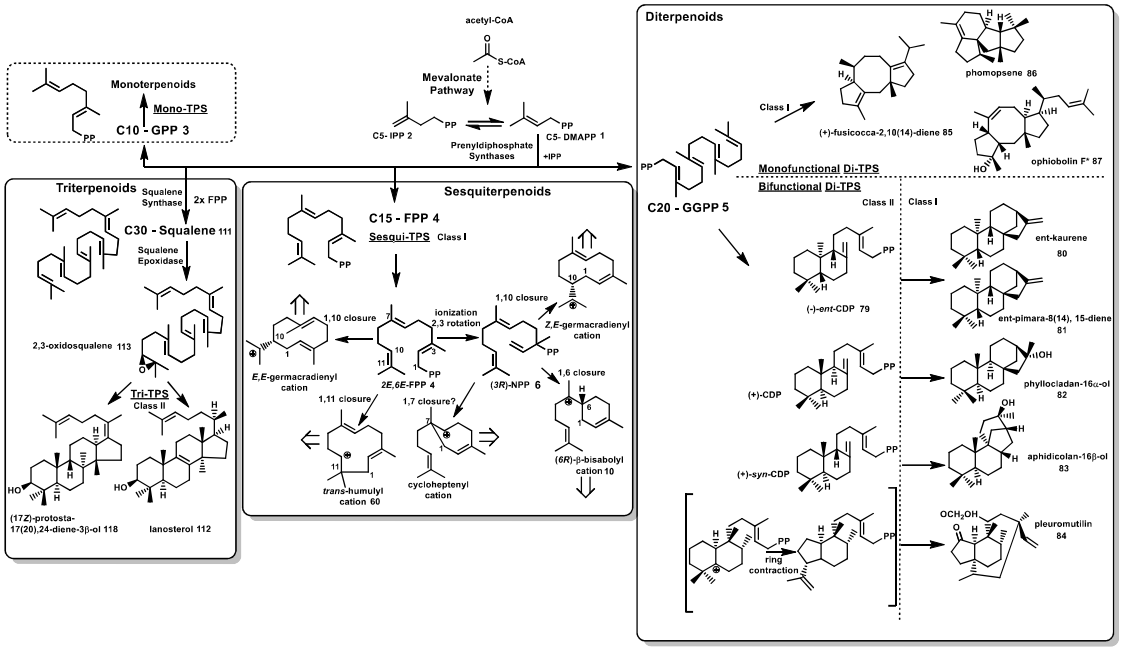


Figure 1-7: Overview of fungal terpene biosynthesis

dependent cyclization of substrate, while class II terpene synthases catalyze a protonation-dependent cascade.[78, 117] [118-120] Depending on the length of the precursor molecule, fungal terpene synthases are known to produce sesquiterpenes (C15), diterpenes (C20) and triterpenes (C30) **(Figure 1-7)**. Further biosynthetic pathway enzymes such as cytochrome P450 monooxygenases, oxidoreductases, and different group transferases modify this initial terpene scaffold, producing the final bioactive terpenoid natural product.

This review will focus on the biosynthesis of the major fungal terpenoid natural product classes that have been described in the last three decades: the sesquiterpenoids, the diterpenoids and the triterpenoids. To date no *bona fide* fungal monoterpene synthases have been described. To keep this contribution focused, natural products of mixed isoprenoid biosynthetic origin (e.g. indole-diterpenoids, meroterpenoids) that involve different enzymatic steps to install the terpenoid moieties in these compounds will not be presented here.

Considering that terpenoids are the focus of a large body of literature, this article will discuss examples of the terpene synthases and cyclases responsible for the generation of the different fungal terpenoid scaffolds as well as examples of known terpenoid biosynthetic pathways. Comparisons will be drawn between the terpenomes of the two major fungal divisions: the Ascomycota which includes many well-known filamentous fungi like *Aspergillus*, *Penicillium* and *Fusarium*, and the Basidiomycota which includes the mushroom-forming fungi. Special emphasis will be placed upon the highly diverse and bioactive *trans*-humulyl cation derived sesquiterpenes produced by the Basidiomycota. Finally, to illustrate the promise of genome mining for the discovery of new bioactive terpenoids from fungi, a description of current bioinformatic approaches used in the identification of sesquiterpenoid biosynthetic pathways in Basidiomycota will be provided as example along with objectives for the development of tools to explore the largely untapped terpenome of this group of fungi.

1.6.2 Fungal sesquiterpenoids

Cyclization of FPP to produce the sesquiterpene scaffold

The sesquiterpenoids are a diverse group of cyclic hydrocarbons, with more than 300 sesquiterpene scaffolds described.[80] All sesquiterpenes share in common a C15 backbone derived from the linear precursor FPP **4**. Typically, FPP **4** is cyclized by class I terpene synthases known as sesquiterpene synthases, which are characterized by the signature active site motifs DDXXD and NSE.[121-123] These amino acid residues play an important role in coordinating the catalytically essential divalent metal ions that stabilize the pyrophosphate group of FPP **4** within the active site cavity.[122, 123] Cyclization is initiated by a metal ion induced ionization of the substrate and departure of inorganic pyrophosphate (PPi), which promotes structural displacements including closure of the active site lid.[122] The resulting highly reactive carbocation undergoes an initial ring closure at the 1,10 or the 1,11 position. Some enzymes catalyze first the *trans-cis* isomerization of the 2,3-double-bond of (*2E,6E*)-FPP **4** to (3*R*)-nerolidylpyrophosphate **6** prior to generating a cisoid, allylic nerolidyl-carbocation upon PPi cleavage.[124, 125] This nerolidyl-carbocation results in a known initial ring closure at either the 1,6 or the 1,10 position; 1,7 ring closure may be possible but has not yet been mechanistically shown (Scheme 1). Subsequent proton shifts, methyl shifts, and complex ring rearrangements are stabilized by the aromatic residues that line the active site cavity of all sesquiterpene synthases.[78] The reaction cascade is quenched either by attack by a water molecule,[126-128] or by deprotonation which, based on the fact that no active site base has yet been confirmed,[129-131] is believed to be mediated by the leaving PPi group.[126, 132] The final terpene product and divalent metal ions are released by the enzyme upon opening of the active site lid. Comprehensive mutational analyses and structural studies have provided detailed insights into the catalytic mechanisms used by fungal sesquiterpene synthases, therefore this class of terpene synthase is relatively well understood (reviewed in: [78, 80, 118]).

Structures are known for several microbial and plant sesquiterpene synthases.[121, 122, 126, 133-138] However, crystal structures have only been solved for two fungal sesquiterpene synthases; aristolochene **7** synthase from *Penicillium roqueforti* and *Aspergillus terreus*;[134, 138-140] and trichodiene **8**

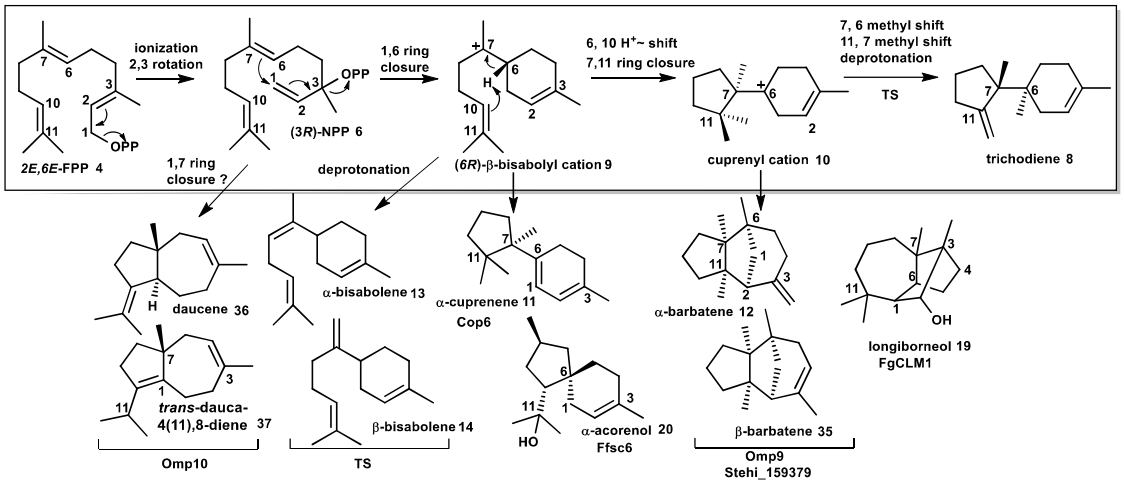


Figure 1-8: Examples of sesquiterpenoids generated by 1,6-cyclization of FPP upon isomerization to NPP

synthase from *Fusarium sporotrichioides*.[129, 137, 141, 142]

Trichodiene **8** synthase (TS) from the plant pathogens *Trichothecium roseum*[125] and *Fusarium sporotrichioides* was the first characterized fungal enzyme.[143] Production of trichodiene **8 (Figure 1-8)** is the first committed step in the pathway to the trichothecene mycotoxins (see "modifying the sesquiterpene scaffold" below) that cause the major cereal crop disease *Fusarium* head blight (FHB).[144] The mechanistic details of TS have been deduced by site directed mutagenesis[123] and structure solution of recombinant wild type and D100E enzyme in complex with PPi resulting from incubation with the substrate analogue 2-fluorofarnesyl diphosphate, and in complex with benzyl triethylammonium.[129, 137, 141] Removal of PPi from FPP **4** is initiated by three Mg^{2+} ions, two of which are coordinated by D100, and a third which is coordinated by the triad N225, S229 and E233.[137, 141] The resulting cation is isomerized to (3*R*)-nerolidyl pyrophosphate (NPP) **6** upon recapture of PPi.[145] Rotation about the C2-C3 bond allows the intermediate to adopt a *cisoid* conformation, and departure of PPi

results in a 1,6 ring closure to yield a *(6R)*-β-bisabolyl cation **9**. Residues K232, R304 and Y305 act to stabilize PPi, which remains bound at the active site throughout product generation.[146] Subsequent hydride transfer and a 7,11 ring closure produces a secondary cuprenyl cation **10**.[147, 148] Finally, two methyl shifts and deprotonation, possibly mediated by the remaining PPi, results in trichodiene **8** as the major product **(Figure 1-8)**, and several minor products including α-cuprenene **11**, α-barbatene **12,** α- and β-bisabolene **13** and **14**.[125, 129, 141]

The structure of aristolochene **7** synthase (AS) from *Penicillium roqueforti* (PR-AS)[134] and *Aspergillus terreus* (AT-AS) is also known.[138-140] AS belongs to the subclass of 1,10 cyclizing sesquiterpene synthases, catalyzing the ionization of FPP **4** and an initial ring closure to yield germacrene A **15** (Scheme 3). Protonation at C6 leads to the eudesmane carbocation **16**,[149] which following a methyl migration, hydride transfer, and deprotonation, yields **(+)**-aristolochene **7**.[150, 151] Aristolochene **7** is the sesquiterpenoid scaffold of several mycotoxins, including the PR-toxin **17**, sporogen-AO1 and phomenone.[152]

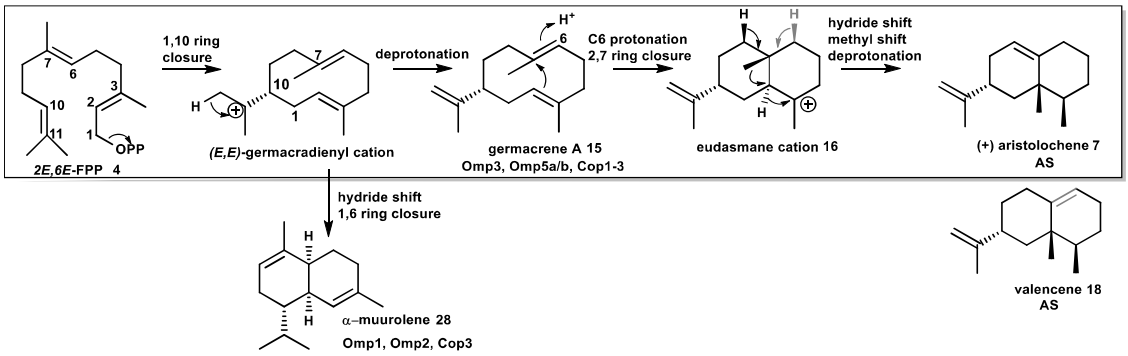


Figure 1-9: Examples of sesquiterpenoids generated by 1, 10-cyclization of FPP

Detailed characterization of AS from the two distinct fungal sources has provided a unique perspective upon the evolution of sesquiterpene synthases. Both PR-AS and AT-AS follow the same cyclization mechanisms and display very similar three dimensional folds, indicating that the two enzymes share a common evolutionary ancestor. However, the underlying amino acid sequence of the two enzymes differs significantly, and this divergence in sequence has resulted in a

divergence in fidelity. AT-AS produces (+)-aristolochene **7 (Figure 1-9)** as its sole product, while PR-AS is product promiscuous, producing (+)-aristolochene **7** as its major product and (-)-valencene **18** and germacrene A **15** as minor products (at a ratio of 94:2:4). Comparative site directed mutagenesis of the active site metal coordinating sites of both enzymes highlighted the importance of these conserved motifs in product fidelity, in particular the NSE motif. In PR-AS, the product ratio of mutants N244D, S248A and E252D shifted to 20:80 (+)-aristolochene **7**: germacrene A **15**, while E252Q produced only germacrene A **15**. However, none of the mutants reported for PR-AS produced solely (+)-aristolochene **7**. The product ratio of the corresponding AT-AS mutants N219D and E227D shifted from (+)-aristolochene **7** as the only product to 44:56 and 26:74 (+)-aristolochene **7**: germacrene A **15**, respectively.[130] It would therefore appear that while the NSE metal coordinating motif serves as the modulator of product fidelity in AT-AS, other amino acids beyond these conserved residues play a role in determining the specific production of (+)-aristolochene **7** by PR-AS.

Three sesquiterpene synthases have recently been described from *Fusarium* strains. One terpene synthase (longiborneol **19** synthase FgCLM1, **Figure 1-8**) from *F. graminearum* is presumed to catalyze a 1,6 cyclization to synthesize longiborneol **19** as the precursor of the antifungal compound culmorin.[153] Gene deletion studies in the rice pathogen *F. fujikuroi* identified two terpene synthases

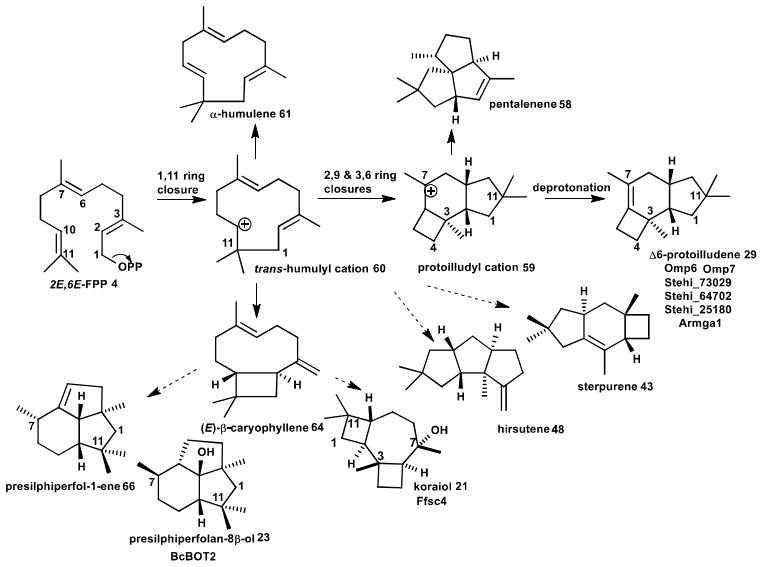


Figure 1-10: Examples of sesquiterpenoids generated by 1, 11-cyclization of FPP

involved in the production of α-acorenol **20** (acorenol synthase Ff_sc6, **Figure 1-8**) and koraiol **21** (koraiol synthase Ff_sc4, **Figure 1-10**) via a 1,6 and presumably a 1,11-cyclization reaction, respectively, followed by quenching of the final carbocation with water to yield the sesquiterpenoid alcohol products.[154] The botrydial **22** biosynthetic gene cluster (see section 2.3) identified in the grey mold *Botrytis cinerea* encodes yet another 1,11-cyclizing enzyme (BcBOT2) which generates the tricyclic alcohol presilphiperfola-8β-ol **23** (**Figure 1-10**) as precursor for this phytotoxin.[155, 156]

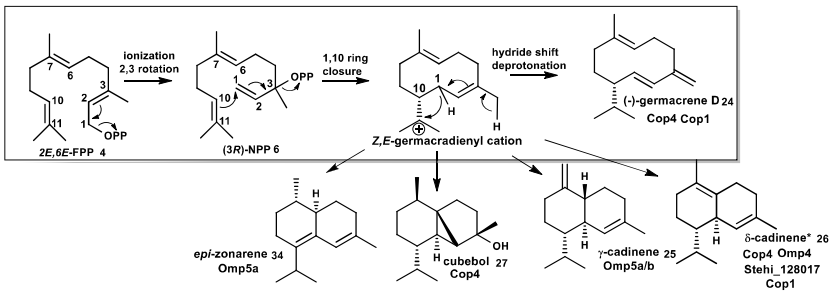


Figure 1-11: Examples of sesquiterpenoids generated by 1, 10 cyclization of FPP upon isomerization to NPP

Surprisingly, until a few years ago no sesquiterpene synthases were known from Basidiomycota, despite the fact that these fungi are prolific producers of bioactive sesquiterpenoids, many of which are derived from the *trans*-humulyl cation,[102] which will be discussed in more detail in the next section. Mining of the first genome of a mushroom-forming Basidiomycota, *C. cinerea*, led to the cloning and biochemical characterization of six enzymes (Cop1-6) catalyzing 1,10 and 1,6 cyclization reactions of *3R*-NPP **6** to α-cuprenene **11**, germacrene D **24**, γ-cadinene **25**, δ-cadienene **26**, cubebol **27** shown in **Figure 1-8** and **1-11**, and 1,10 cyclization of FPP **4** to α-muurolene **28** and germacrene A **15** shown in Scheme 3. One of the enzymes, Cop6, catalyzes highly selective synthesis of α-cuprenene **11** as the precursor of the antimicrobial compound lagopodin.[83, 157, 158]

Of particular interest are enzymes that generate the 1,11-cyclization product Δ6-protoilludene **29 (Figure 1-10)**, the precursor to many pharmaceutically relevant compounds,[102] including the cytotoxic compounds illudin M **30** and S **31** (Scheme 6) that are being developed as anticancer therapeutics.[108, 159] With

the objective of characterizing the biosynthesis of the anticancer illudins and their derivatives (compounds **30-33**, **Figure 1-12**) made by the Jack-O-Lantern mushroom *Omphalotus olearius*,[160, 161] we sequenced its genome to apply a similar genome mining approach to find the Δ6-protoilludene **29** and presumably co-localized illudin biosynthetic genes. We discovered a surprisingly large complement of ten sesquiterpene synthases (Omp1-10) in this fungus. The recombinant enzymes catalyze all cyclization reactions of FPP shown in Scheme 1 (Schemes 2-5 show the major products for each Omp sesquiterpene synthase)**.** The products made by of Omp1-5a/b are similar to those obtained with the *C. cinere*a enzymes Cop1-4, although Omp5a/b also makes as major products *epi*-zonarene **34** and γ-cadinene **25** (**Figure 1-11**). Omp6 and Omp7 cyclize FPP highly selective into Δ6-protoilludene **29** (**Figure 1-4**) while the other enzymes also display new cyclization activities, producing barbatenes **12, 35** (Omp9) and daucenes **36, 37** (Omp 10).[22, 89] Another Δ6-protoilludene synthase was cloned from the honey mushroom *Armillaria gallica* (ArmGa1) which makes antimicrobial melleolide I **38** (Scheme 6) sesquiterpenoids.[162]

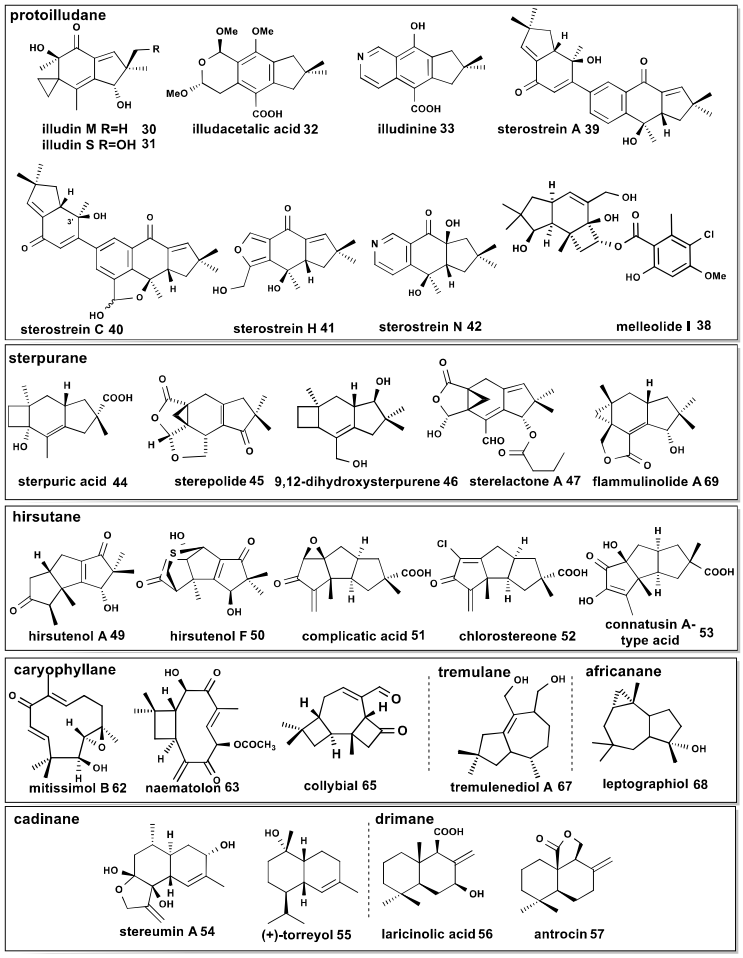


Figure 1-12: Representative examples of modified sesquiterpenoids isolated from Basidiomycota.

The biochemical characterization of the identified sesquiterpene synthases from *Coprinus cinereus* (Cop1-6)[83, 157, 158] and of *Omphalotus olearius* (Omp1-10)[22, 89] guided the subsequent development of *in silico* approaches for the directed discovery of new sesquiterpene synthases and their associated biosynthetic genes based upon cyclization mechanism of choice.[22] Initiatives such as the 1000 Fungal Genomes Project[163] have led to a substantial increase in the number of sequenced genomes that are publically available.[164] The increase in the number of sequenced Basidiomycota genomes has been particularly steep – from less than a handful in 2008 to more than 100 genomes listed in the genome database of the Joint Genome Institute (JGI) at the time of writing. Our own BLAST searches for terpene synthases across Basidiomycota with sequenced genomes led to the identification of hundreds of putative sesquiterpene synthases, as well as associated biosynthetic gene clusters.[22, 83, 89]

The genome sequence of the wood-rotting mushroom *Stereum hirsutum* offers a unique opportunity to explore the terpenoid structural diversity of a fungal genus that is well known to produce a number of bioactive sesquiterpenoid natural products. [90, 165-177] These include the sterostreins (e.g. **39-42**), sterpurene **43** and its derivatives (e.g. **44-47**), hirsutene **48** and its derivatives (e.g. **49-53**), cadinane (**25, 26**) (e.g. **54-55**) and drimane (e.g. **56, 57**) compounds of which representative examples are shown in **Figure 1-12**. We identified and predicted the cyclization mechanisms of 16 sesquiterpene synthases in this fungus. Presently, five of the predicted enzymes were cloned and biochemically characterized. Stehi1|25180, 64702, 73029 are highly product specific Δ6-protoilludene **29** synthases (Scheme 4) that are located in large biosynthetic gene clusters presumably involved in sterostrein (**Figure 1-12**, e.g. compounds **39-42)** biosynthesis.[174, 175] Stehi1|159379 and Stehi1|128017 are 1,6- and 1,10-cyclizing enzymes, respectively (**Figures 1-8 and 1-11**). Within the clade of not yet characterized 1,11-cyclizing sesquiterpene synthases sequences identified in *S. hirsutum*, there may be enzymes that synthesize hirsutene **48** and sterpurene **43**

(**Figure 1-10**) for the above mentioned modified sesquiterpenoid compounds isolated from this fungal genus.

Furthermore, pentalenene **58** (**Figure 1-10**) has been observed as a minor volatile product in the culture headspace of *S. hirsutum* and *O. olearius*, and as a major volatile product of *C. cinerea*.[22, 83, 89] However, no sesquiterpene synthase responsible for the production of pentalenene **58** has yet been identified from these fungi. Site directed mutagenesis and quantum-chemical modelling studies with bacterial pentalenene synthase indicate that the pathway to pentalenene **58** occurs through rearrangement of a protoilludyl cation **59**.[131, 178] It is therefore likely that one or more of the multiple Δ6-protoilludene synthases characterized from *S. hirsutum* and *O. olearius* are responsible for the production of pentalenene **58**, under unidentified reaction conditions.[22, 89] On the other hand, Cop5 is likely the sesquiterpene synthase responsible for the production of pentalenene **58** by *C. cinerea*, which does not produce Δ6-protoilludene **29**. However, an active form of the predicted sesquiterpene synthase could not be cloned from *C. cinerea* cDNA,[83] presumably due to incorrect prediction of gene borders, and/or the presence of frame shifting introns in all amplification products (see section 5 for more information).

The t*rans*-humulyl cation derived sesquiterpenoids

While it is difficult to pin down the exact number of sesquiterpenoid-derived compounds characterized from the two major fungal divisions, searches in SciFinder™ and Web of Science™ return ten times more reports (~500) on sesquiterpenoids isolated from Basidiomycota than from Ascomycota. A diverse range of highly bioactive sesquiterpenoids with cytotoxic, antibacterial, nematocidal, and antiviral activities have been isolated from Basidiomycota (for reviews see:[102, 179]). Many of these compounds are thought to be made to protect fungal fruiting bodies from consumption.[14, 180]

Basidiomycota are known to synthesize sesquiterpenoids derived from 1,6 and 1,10 cyclized cations (**Figure 1-7**), including germacranes,[179] bisabolanes,[180] cadinanes[90, 171, 176] and drimanes,[169, 181] including the mTOR signaling pathway inhibitor antrocin **57** produced by *Antrodia camphorata*[182] (**Figure 1-12**) However, the vast majority of sesquiterpenoids with medicinal or cytotoxic properties isolated from Basidiomycota are derived from the 1,11 cyclized *trans*-humulyl cation **60** as shown in Schemes 4 and 7, which will be the emphasis of this section.[102] Of particular interest are derivatives of the tricyclic protoilludyl cation[14, 102] which have potential applications as anticancer, antifungal and antibiotic agents.[102]

The *trans*-humulyl cation can be directly deprotonated to produce α-humulene **61** (**Figure 1-10**), which has been observed in *F. fujikorii* and *S. hirsutum.*[22, 154] Modified humulanes, such as the potential antitumor compound mitissimol B **62** (**Figure 1-12**), have also been isolated from the fruiting bodies of several members of the genus *Lactarius*.[183]

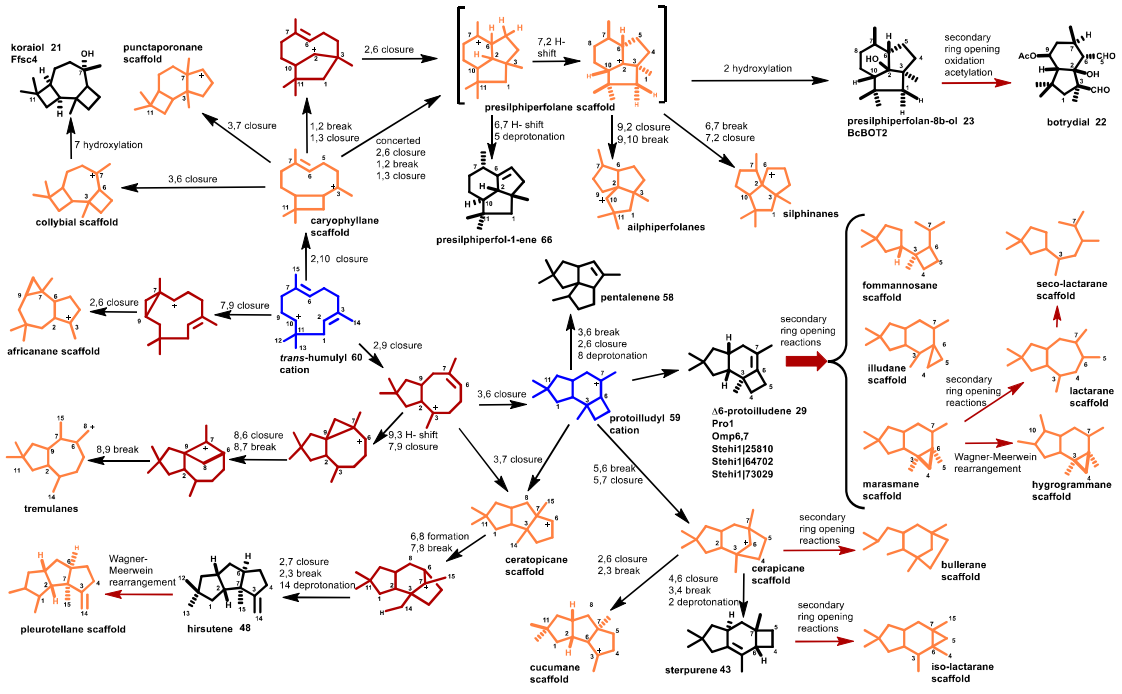


Figure 1-13: Major fungal transformations of the trans-humulyl cation.

Transformations of the *trans*-humulyl cation include a 2,10 ring closure, creating the distinctive dimethylated cyclobutyl ring of the caryophyllane scaffold. The mildly antibacterial sesquiterpenoid naematolon **63** (**Figure 1-12**) is an oxidized, acetylated derivative of (*E*)-β-caryophyllene **64** produced by several members of the *Naematoloma* (*Hypholoma*) genus.[180]

In addition, the caryophyllane scaffold can be further cyclized to produce the tricyclic 4:7:4 collybial scaffold by a secondary 3,5 closure. The collybial scaffold is the precursor to the antibiotic collybial **65** (**Figure 1-12**), produced by *Collybia confluens.*[184] Hydroxylation at the C7 of the collybial scaffold results in koraiol **21**. Alternatively, the caryophyllane scaffold can be cyclized to produce the triquinane presilphiperfolane scaffold, either via a stepwise cyclization,[156] or via a concerted carbocation rearrangement.[120] The resulting presilphiperfolane scaffold may then proceed through a 1,2 hydride shift and deprotonation to yield presilphiperfolan-1-ene **66**, a minor volatile product of *S. hirsutum*.[22] The presilphiperfolane scaffold may also undergo a 1,3 hydride shift and hydroxylated to produce **23**,[156] or be rearranged further to produce the silphiperfolane and silphinane families of triquinane sesquiterpenoids.[185]

Several tremulanes have been characterized from *Phellinus tremulae*, including tremulenediol A **67** (Scheme 6).[186, 187] The potential route to the tremulane scaffold was deduced by ^{13}C NMR analysis of tremulanediol A **67** formation by *P. tremulae*, and follows an initial 2,9 closure of the *trans*-humulyl cation **60**. Following this, an additional cyclobutyl ring formation occurs at C6-C9. A secondary ring-opening rearrangement yields the fifth methyl group on C6 (**Figure 1-13**).[187]

The africananes are defined by a cyclopropyl group fused to a central cycloheptane ring, formed by a 7,9 cyclization of the *trans*-humulyl cation **60**. Examples of modified africanane alcohols are leptographiol **68** (**Figure 1-12**), isoleptographiol, and isoafricanol, isolated from *Leptopgraphium lundbergii*.[14, 188] Interestingly, unmodified africanenes have been identified in the headspace of

S. hirsutum and *O. olearius* liquid cultures.[22, 89] However, the sesquiterpene synthase(s) responsible for their production have not yet been identified.

The largest and most diverse group of *trans*-humulyl cation **60** derived sesquiterpenoids produced by Basidiomycota are those derived from the 4:6:5 tricyclic protoilludyl cation **59**.[102] This scaffold is a key intermediate in the production of a large number of sesquiterpenes including e.g. sterpurene **43**, hirsutene **48**,[189] Δ6-protoilludene **29** and thereof derived scaffolds,[89] and of pentalenene **58**[190] as described below and shown in **Figure 1-13**.

Δ6-protoilludene **29** is the precursor to diverse bioactive sesquiterpenoids, including the marasmanes, fommanosanes, (seco)lactaranes, hygrogammane and the illudanes (reviewed in: [102]). These compounds are hypothesized to be produced by secondary ring opening and concurrent oxidations of Δ6-protoilludene **29**, with the initial steps likely catalyzed by one or more cytochrome P450 monooxygenases and/or other oxygenases that are typically located in biosynthetic gene clusters of fungal sesquiterpenoid pathways.[22, 89, 191, 192] In the case of the illudins **30** and **31**, a ring opening and subsequent contraction of the Δ6-protoilludene **29** cyclobutyl ring yields the reactive cyclopropyl ring that defines the illudanes. Presently, no biosynthetic enzymes have been characterized that initiate these ring opening reactions.

Reopening of the protoilludyl cation **59** at C5-C6, and a subsequent 5,7 cyclization may afford the cerapicane scaffold which is the precursor to the rare cucumane and bullerane-type sesquiterpenes.[102] The cerapicane scaffold is transformed to the sterpurane scaffold by an unusual ring rearrangement reaction that was deduced by ^{13}C NMR analysis of 9,12-dihydroxysterpurene **46** synthesized by *Stereum purpureum*.[166] Further ring opening modifications can then contract the cyclobutyl ring of sterpurene **43** into the defining dimethylated cyclopropyl ring of the *iso*-lactaranes, several of which have been identified from *Lactarius rufus,*[193] *S. purpureum*,[194] *Phlebia uda,*[195] *Flammulina velutipes,*[196] and *Merulius tremillosus.*[197] The *iso*-lactaranes are often co-produced with sterpurenes, supporting the hypothesis of a shared biosynthetic

pathway.[165] Examples of *iso*-lactaranes include the cytotoxic flammulinolide A **69**,[196] sterepolides (e.g. **45**),[194] and the sterelactones (e.g. **47**) (**Figure 1-12**).[172]

Sterostrein A **39** produced by *S. ostrea* is a Δ6-protoilludene **29** dimer hypothesized to be formed by a Diels-Alder reaction that installs the central benzene ring.[174] Biosynthesis of the melleolides (e.g. **38** in **Figure 1-12**) by *Armillaria gallica* requires the esterification Δ6-protoilludene **29** with orsillinic acid.[162, 198]

The bioactive hirsutanes (e.g. compounds **49-53**, **Figure 1-12**) produced by *S. hirsutum* and other Basidiomycota[22, 177] [173, 199] may be synthesized either through 3,7 rearrangement of the protoilludyl cation **59** or directly via 3,7 cyclization of a bicyclic precursor shown in **Figure 1-13**. While the intermediacy of a protoilludyl cation **59** has not been confirmed, a ceratopicane intermediate is consistent with the ^{13}C labelling pattern observed in hirsutene **48** biosynthesis by *Stereum complicatum*.[189] The ceratopicane intermediate undergoes multiple rearrangements, shifting the C3 (and its attached C14 methyl group) to its final position, and is then deprotonated to produce hirsutene **48**.[189] Hirsutene **48** may be converted by a Wagner-Meerwein rearrangement to produce the pleurotellane scaffold.[102]

Modifying the sesquiterpene scaffold

Compared to pathways responsible for the biosynthesis of diterpenoid natural products discussed in the next section, very few fungal sesquiterpenoid biosynthetic pathways have been characterized on a gene level. One of the most extensively characterized fungal sesquiterpenoid biosynthetic pathways is the route to the trichothecene mycotoxins (**Figure 1-14**). Over 200 trichothecenes have been identified from diverse sources including the plant pathogens *Fusarium* and *Myrothecium*, the soil associated filamentous fungi *Trichoderma* and the mold

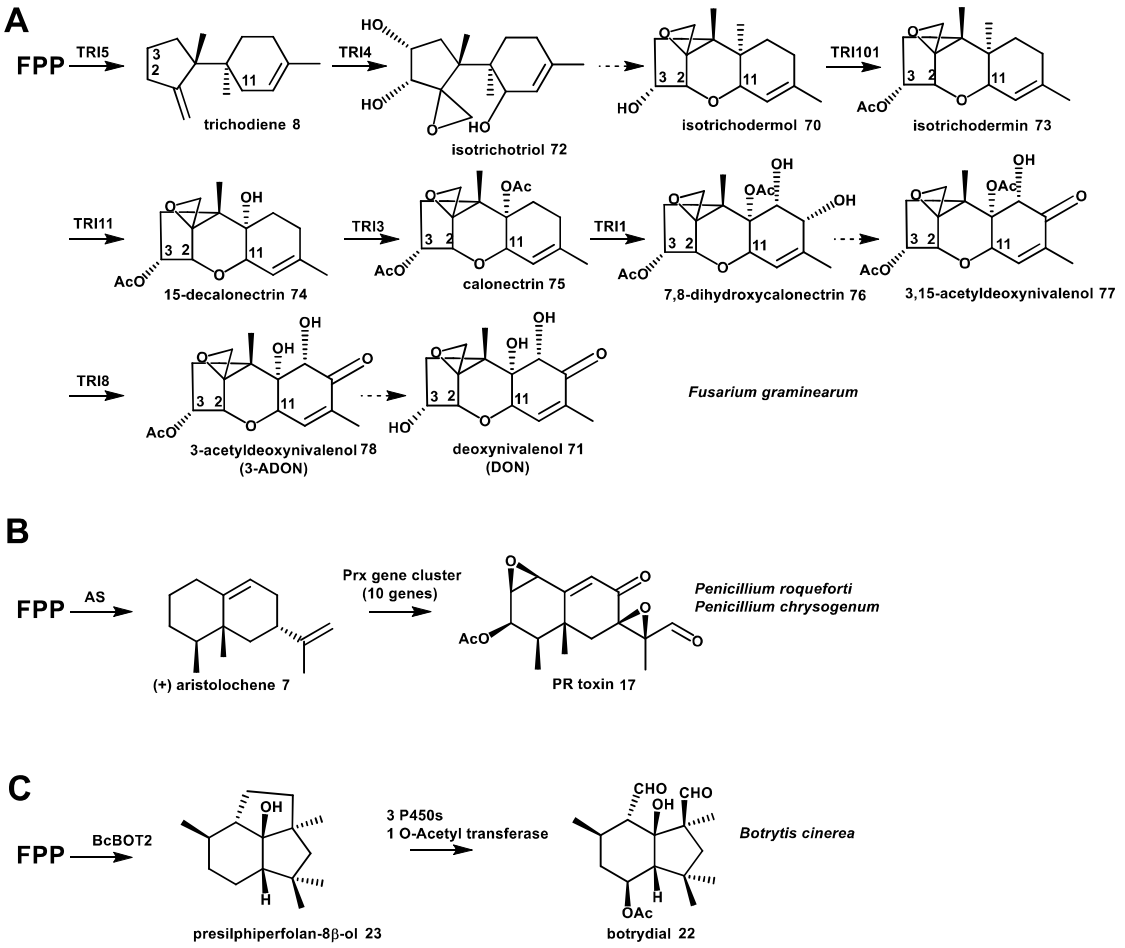


Figure 1-14: Identified sesquiterpenoid biosynthetic pathways in Ascomycota

Stachybotrys.[200, 201] This particular pathway has been the focus of a great deal of research by the USDA during the past 30 years due to the significant economic impact that trichothecene related blight and ear rot has on global cereal crops each

year, as well as the toxic effects of trichothecene contaminated feed on livestock.[201-203] The trichothecene family of toxins is subclassified as Type A, B, C, and D depending on the chemical modifications to the common tricyclic core structure 12,13-epoxytrichothec-9-ene (EPT), exemplified by isotrichodermol **70** in Scheme 8. Of particular interest is deoxynivalenol (DON) **71**, the Type B trichothecene that is most commonly associated with *Fusarium* head blight in cereal crops. DON **71** plays a vital role in activating fungal pathogenicity and virulence as part of a complex feedback exchange between plant, pathogen and environment.[204-206]

The initial, common trichothecene biosynthetic steps in the pathway to DON **71** were established in the model fungus *Fusarium sporotrichioides*.[143, 207] Later steps specific to the production of DON **71** were studied in a *F. graminearum* strain that was isolated from infected crops.[208] Activation of biosynthesis of DON **71** is controlled by the global transcriptional regulator *TRI6*, which autoregulates in response to differences in nutrient levels.[209] The first step on the biosynthetic pathway to all trichothecenes is the cyclization of FPP **4** to trichodiene **8** by the *TRI5* encoded trichodiene **8** synthase,[143] as described in section 2.1 and in Scheme 2. Trichodiene **8** is subjected to multiple oxidations by the cytochrome P450 monooxygenase TRI4, which installs oxygens at C2, C3, C11 and the C12-C13 epoxy group to give the intermediate isotrichotriol **72**.[210] Isotrichotriol **72** undergoes a spontaneous dual cyclization to yield isotrichodermol **70**, a modified form of EPT.[211] An acetyltransferase TRI101 acetylates the hydroxyl group at C3 to produce isotrichodermin **73**, reducing toxicity of the intermediate to the host.[212, 213] C15 of **73** is hydroxylated by the cytochrome P450 monoxygenase TRI11[214] to form intermediate **74** and acetylated by TRI3[215] yielding calonectrin **75**, a key branch point on the pathway to biosynthesis of Type A and Type B trichothecenes by *F. sporotrichioides* and *F. graminearum*, respectively.[201] For DON **71** production, calonectrin **75** is hydroxylated at positions C7 and C8 by TRI1 to give 7,8-dihydroxycalonectrin **76**.[216] The C8 hydroxyl group of **76** is then converted to a keto group by an unidentified enzyme,

leading to 3,15-acetyldeoxynivalenol **77**. Different variants of the deacetylase TRI8 remove the acetyl group from either the C3 or the C15 position, resulting in 15-acetyldeoxynivalenol (15-ADON) or 3-ADON **78**, (the latter compound is shown in **Figure 1-14**), respectively.[217] The final biosynthetic step, deacetylation of ADON to DON **71**, has yet to be elucidated (**Figure 1-14**).

The ability of the fungal host to tolerate the toxic effects of intermediates and products of its own secondary metabolic pathways is key to survival. TRI12 is an integral membrane protein that acts as an efflux pump to export trichothecenes from hyphal cells, removing toxic products that could otherwise damage the cell.[218] Subcellular localization of TRI12 to large motile vesicles[219] indicated that *F. graminearum* could rely on an encapsulation mechanism to sequester trichothecene biosynthesis. The cytochrome P450 monooxygenases TRI1 and TRI4 co-localize with TRI12, suggesting that at least part of the trichothecene biosynthetic pathway is integrated within cellular targeting hubs, "toxisomes", preventing toxic intermediates from causing unwanted effects.[92] Whether this compartmentalization strategy is unique to *F. graminearum*, or whether it is a global mechanism employed by fungi producing bioactive secondary metabolites, remains to be established. Similarly, vesicles have been shown to also play an important role aflatoxin biosynthesis by *Aspergillus parasiticus.*[220, 221]

A much less complex gene cluster has been identified in the mold *Botrytis cinerea* that is responsible for the biosynthesis of the phytotoxin botrydial **22** (**Figure 1-14**).[155, 156] Presilphiperfolan-8β-ol **23**, the product of terpene synthase BcBOT2 (**Figure 1-10**), is converted by three cytochrome P450 monoxygenases and an acetyltransferase to the final phytotoxin. Most recently, screening of a genomic phage library and comparison with orthologous genes identified in the sequenced genome of a related strain, *P. chysogenum,* led to the identification of a 10-gene cluster encoding the biosynthetic pathway for the aristolochene **7**-derived PR-toxin **17** in the blue cheese mold *P. roqueforti* (**Figure 1-14**).[222]

As discussed in the previous section, multiple copies of Δ6-protoilludene **29** synthase are encoded in the genomes of *O. olearius* (Omp6 and Omp7) and *S. hirsutum* (Stehi1|25180, 64702, 73029) which are each located in predicted biosynthetic gene clusters responsible for the biosynthesis of illudin derived compounds (e.g. compounds **30-33**) and potentially the sterostreins (e.g. compounds **39-42**) isolated from *O. olearius* and *Stereum ostrea* BCC22955, respectively (**Figure 1-12**).[22, 89] The clusters share in common several cytochrome P450 monooxygenases, as well as further scaffold decorating enzymes including different types of oxygenases, oxidoreductases, group transferases and membrane transporters.

In *O. olearius,* Omp7 is located in a mini-gene cluster, including one P450 and an FAD-type oxidoreductase, while Omp6 is part of a large gene cluster.[89] It appears that the two copies of protoilludene **29** synthase may have arisen from a gene duplication event, thereby boosting the production of illudin compounds.[89] In the case of *S. hirsutum*, the three Δ6-protoilludene **29** synthases may have resulted from gene duplication, or they may have diverged from a common ancestor. Differences in catalytic efficiencies of the enzymes, as well as a diversity in product fidelity, could represent a spectrum of evolutionary time points on the course to a specific Δ-6 protoilludene **29** synthase.[22] Furthermore, each of the three Δ-6 protoilludene **29** synthases is located in a distinct biosynthetic gene cluster which appear to be expressed at the same time under standard growth conditions. Heterologous refactoring of the entire gene clusters (see section 1.6.6) will be required to understand the role that each of these pathways plays in defining the natural product landscape of this fungus. It remains to be seen if all three biosynthetic gene clusters are responsible for the biosynthesis of the different Δ6-protoilludene **29** derived sterostreins (**Figure 1-14**) isolated from *Stereum*, make yet uncharacterized compounds, or may even carry out a sophisticated combinatorial scheme among the three different clusters that each act on the same protoilludene scaffold.

1.6.3 Fungal diterpenoids

Cyclization of GGPP to produce the diterpene scaffold

Diterpenoids are a diverse class of natural products derived from the C20 precursor GGPP **5**, with at least 12,000 compounds already described.[223] The first committed step in diterpenoid biosynthesis is the cyclization of GGPP **5** to produce the diterpene scaffold, which occurs via a carbocation cascade. Classically, activation of the carbocation cascade by terpene synthases corresponds to the removal of the pyrophosphate group from the linear substrate (as described for the sesquiterpene synthases in section 1.6.2). This ionization-dependent reaction is catalyzed by class I terpene synthases.[78]

Notably, the initial generation of the reactive carbocation follows a different path in the biosynthesis of labdane-related diterpenes.[224] Here, the carbocation cascade is preceded by a protonation-dependent bicyclization reaction to install the labdane bicycle. This reaction is catalyzed by class II diterpene synthases, more akin to the triterpene synthases (described in section 1.6.4).[225] In this case, the pyrophosphate group of GGPP **5** remains intact and is later removed in a class I cyclization reaction (**Figure 1-15**).[224] Thus, this type of diterpene synthase appears to have evolved as an amalgamation of ancestral bacterial class I and class II terpene synthases.[224]

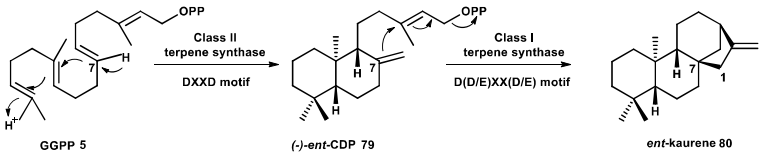


Figure 1-15: Cyclization mechanism of the bifunctional CPS/KS type diterpene synthases

The first bifunctional fungal diterpene synthase was isolated and characterized from the gibberellin producer *Phaeosphaeria* sp. L487.[226] The putative gene encoding the diterpene synthase was isolated using degenerate primers based upon conserved sequence motifs of plant copalyl diphosphate synthases. Characterization of the gene product revealed a conglomerate enzyme: that is, a bifunctional copalyl diphosphate synthase/*ent*-kaurene synthase (CPS/KS), which catalyzed a sequential class II followed by class I cyclization of

GGPP **5** to produce (-)-*ent*-CDP **79** as an intermediate on the path to *ent*-kaurene **80** (**Figure 1-15**).[227] The class I activity is located in the N-terminal region, while the C-terminal region contains the class II activity. This differed from the previously characterized plant diterpene synthases, which appeared to involve separate mono-functional CPS and KS enzymes.[228, 229]

Several structures of plant diterpene synthases have recently been solved and shown to contain both an N-terminal α-helical domain (α-domain) typical of class I terpenoid synthases and a C-terminal α-barrel (or γβ-) domain typical of class II terpenoid synthases.[230-232] Many of these modular γβα-domain plant diterpene synthases, however, are monofunctional in that only one of the cyclization activities (CPS or KS) is functional (reviewed in: [224]). However, some plant enzymes (e.g. abietadiene synthase[232]) are bifunctional, like the fungal enzymes. The γβα-domain taxadiene synthase, however, converts GGPP **5** at its class I domain directly into the tricyclic taxadiene scaffold.[231] Although protein sequences of diterpene synthases from plants and fungi share little similarity except for some conserved motifs, homology modelling suggests that they share the same domain organization.

A number of CPS/KS-type fungal diterpene synthases have been identified in addition to the one from *Phaeosphaeria* sp. These include the *ent*-kaurene **80** synthase from the gibberellic acid producers *Fusarium fujikuroi* (teleomorph: *Gibberella fujikuroi*),[226, 227, 233] the *ent*-primara-8(14),15-diene **81** synthase AN1495 from *Aspergillus nidulans*,[234] phyllocladen-16α-ol **82** synthase (PaDC1) and a CPS/KS synthase (PaDC2) with only a functional CPS domain[235] from *Phomopsis amygdali,* and aphdicolan-16β-ol **83** synthase (PbACS) from *Phoma betae*.[236] Only one labdane-type diterpene synthase has thus far been cloned from a Basidiomycota: pleuromutilin **84** synthase from *Clitopilus passeckerianu*s.[237] (see **Figure 1-7** for cyclization reactions).

Typically, these bifunctional diterpene synthases and their associated biosynthetic gene clusters (see section 1.6.2) have been identified by a sequence

analysis strategy that first involved the identification of a GGPP **5** synthase (GGS) via genomic amplification followed by genome walking to identify flanking regions. GGPP **5** utilizing biosynthetic pathways require a dedicated GGS that extends the common isoprene precursor FPP **4** by one IPP **2** unit.[238, 239] The gene encoding GGPP **5** synthase, which is essential for production of the precursor molecule is commonly clustered with other pathway genes.[240] This gene finding strategy has led to the discovery two unusual monofunctional, non-labdane-type diterpene synthases from *Phoma betae*.[241, 242] PaFS and PaPS, which synthesize fusicocca-2,10(14)-diene **85** and phomopsene **86** (**Figure 1-7**), respectively, deviate from other CPS/KS-type fungal diterpene synthases in that they display multiple DDXXD motifs, as opposed to the prototypical N-terminal DXDD and C-terminal DEXXE motifs common to the dual-function diterpene synthases that carry out sequential type II and type I cyclization of GGPP **5** (**Figure 1-15**).[243] The DXDD motif is required for the class II type protonation dependent cyclization reaction, like that carried out by squalene hopene cyclase,[244] while the DDXXD motif is associated with the class I type ionization dependent reaction, which has been described for terpene synthases and prenylchain synthases like GGPP **5** and FPP **4** synthases.[121, 245]

Both PaFS and PaPS appear to derive from a gene fusion, resulting in chimeric proteins with an N-terminal class I terpene synthase domain and another C-terminal class I domain that contains the GGPP **5** synthase activity. This physical co-localization of diterpene synthase and GGPP **5** synthase may have evolved to improve flux of precursor molecules towards downstream modifying enzymes. A chimeric sesterterpene (C25) synthase AcOS was recently identified in *Aspergillus clavatus.* Here a GFPP (geranylfarnesyl diphosphate) synthase domain supplies the substrate for subsequent cyclization by the terpene synthase domain into ophiobolin F **87** (**Figure 1-7**).[246] Structural analyses of plant and bacterial diterpene synthases indicate that modularity in domain architecture and diversity in mechanisms are a common theme for diterpene synthases.[230-232]

Table 1-1: representative examples of fungal diterpenoids with characterized bioactivities

Biological activity	Fungal species	Compound name(s)	Reference
Ascomycota			
Cytotoxic	*Alternaria brassicola*	Brassicicene **C (109)** †	[243, 247, 248]
Cytotoxic	*Cladosporium sp.*	Cotylenin **A**-E **(88)**	[249-251]
Phytotoxin	*Phomopsis amygdali*	Fusicoccins **A (108)**†, F	[242]
DNA-Pol. inhibitor	*Phoma betae*	Aphidicolin (**106**) †	[236, 252]
Anti-inflammatory	*Bipolaris coicis**	Coicenals **A(89)**-D	[253]
Cytotoxic	*Acremonium striatisporum*	Virescenosides **A(90)**, B, C, M, N	[254]
Cytotoxic	*Aspergillus wentii* EN-48*	Asperolides **A (91)**-C	[81]
Cytotoxic	*Cercospora* sp.*	Cercosporenes **A (92)**-F	[255]
Cytotoxic	*Geopyxis* sp.*	Geopyxins **A (93)**-D	[256]
Cytotoxic	*Smardaea* sp.*	Smardaesidins **A (94)** -G	[257]
Growth hormone	*Fusarium fujikuroi*	Gibberellins **GA14 (102),4 (103)**,3, **1 (104)**	[233]
Growth hormone	*Phaeosphaeria* sp. L487	Gibberellins **GA12 (101),4 (103)**,9,20,**1 (104)**	[226, 258]
Growth hormone	*Sphaceloma manihoticola*	Gibberellins **GA14 (102),4 (103)**	[259]
Basidiomycota			
Antibacterial	*Clitopilus* sp. & *C. passeckeranus*	Pleuromutilin **(84)**†	[237, 260]
Antibacterial	*Sarcodon scabrosus*	Sarcodonin **L (95)**, M	[261]
Antifungal	*Coprinus heptemerus*	Heptemerones A-**G (96)**	[262]
Cytotoxic	*Coprinus plicatilis*	Plicatilisin **A (97)**-D	[263]
Cytotoxic	*Coprinus radians*	Radianspene **A (98)**-M	[264]
Cytotoxic	*Crinipellis* sp.	Crinipellins A-**C (99)**, D	[265]
Cytotoxic	*Lepista sordida*	Lepistol **(100)**	[266]

Modifying the diterpene scaffold

The diversity of diterpenoid compounds stems from the large number of modifications to the diterpene scaffold, typically catalyzed by decorating enzymes such as P450 monoxygenases, different types of oxidases and oxidoreductases as well as trasnferases. Diterpenoids are widely produced by plants, functioning as signaling compounds, as photosynthetic pigments, and in defense against infection.[223] While a number of enzymes involved in the biosynthesis of diterpenoids in plants and bacteria have been characterized in detail, fewer diterpenoid biosynthetic enzymes have been characterized from fungi.[223, 267] Except for pleuromutilin **84** biosynthesis by *Clitopilus passeckeranus*,[237] all other characterized fungal diterpene pathways are from Ascomycota. Yet, diverse bioactive diterpenoids have been isolated from Basidiomycota, although their biosynthetic pathways remain to be identified (**Table 1-1**, see **Figure 1-16** for modified diterpenoid structures discussed below).[268]

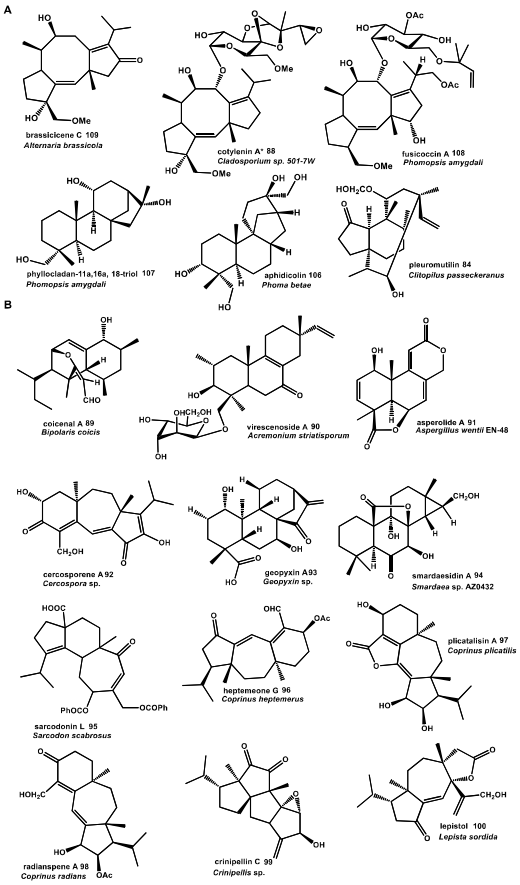


Figure 1-16: Examples of fungal diterpenoid natural products and their producer organisms

One of the first fungal species identified as a diterpenoid producer was the rice plant pathogen *F. fujikuroi*. This fungus causes stems of plants to become hyperelongated and etiolated, leading to what is commonly known as "foolish seedling" disease. The causative agent of this phenotype was discovered to be the gibberellin (GA) diterpenoids (GA compounds **101-104 Figure 1-17**) produced by *F. fujikuroi*. The biosynthetic pathway to the production of GAs by *F. fujikuroi*,

which is clustered on the genome, has since been fully elucidated by Tudzynski and co-workers[269] (**Figure 1-17**). Detailed characterization of the enzymes involved in GA biosynthesis by *F. fujikuroi* has highlighted key differences in sequence and function from that of plant GA pathways,[270] such that it has been suggested that GA biosynthesis in bacteria, plants and fungi evolved separately.[240]

In GA biosynthesis, GGPP **5** is the substrate for a bifunctional diterpene synthase CPS/KS.[233] The resulting tetracyclic *ent*-kaurene **80** diterpene is converted to *ent*-kaurenoic acid **105** through oxidation by the cytochrome P450 monooxygenase P450-4.[239] Hydroxylation at the C7 position of *ent*-kaurenoic acid **105** is carried out by P450-1, resulting in aldehyde GA12 **101** (note GA numbering is based on the order the compounds were identified), with a characteristic 6-5-6-5 ring structure. A further hydroxylation at the C3 position and an oxidation by the multifunctional P450-1 results in GA14 **102**, a key intermediate in the pathway. A closely related cytochrome P450, P450-2, catalyzes the essential oxidation of GA14 **102** at C20 to yield the biologically active form,

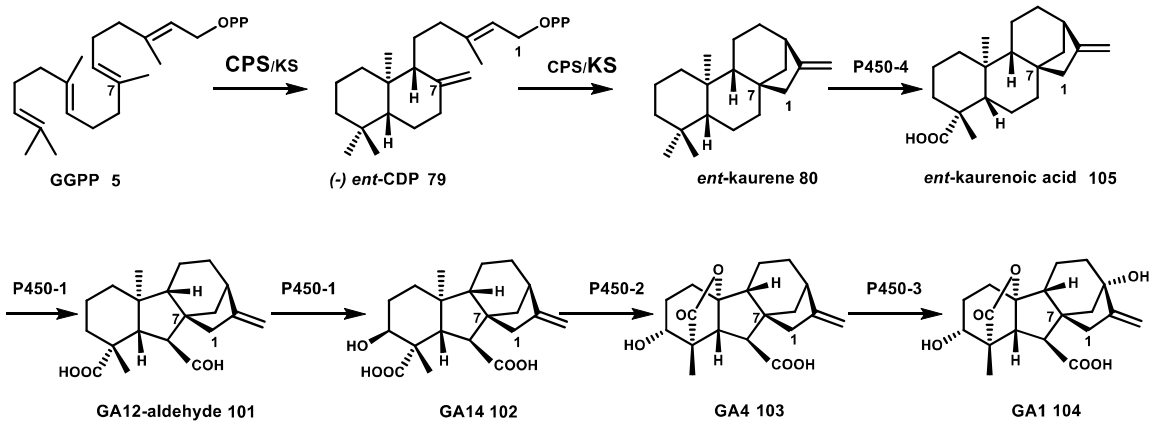


Figure 1-17: Proposed biosynthetic pathway to gibberellic acids (GA) in *F. fujikuroi*

GA4 **103**.[233, 271] Finally, P450-3 hydroxylates GA4 to GA1 **104**, another bioactive form of the growth hormone.[272] Each of the steps catalyzed by the cytochrome P450s depends on a cytochrome P450 oxidoreductase (CPR) to transfer electrons from the cofactor NADPH via FAD and FMN.[273]

Subsequent identification and characterization of the GA biosynthetic pathway from *S. manihoticola* revealed that the organization of the gene cluster, the underlying gene sequences, and the function of the encoded GA enzymes, were similar to that of *F. fujikuroi*.[259] Despite the fact that *S. manihoticola* lacked enzymes responsible for GA1 **104**, 3, and 7 production, complementation with its gene cluster was sufficient to restore GA biosynthesis in *F. fujikuroi* mutants. This indicates that the GA pathways of these two distantly related fungi may share a common evolutionary ancestor. By contrast, characterization of the GA biosynthetic pathway in *Phaeosphaeria* sp. L487 shows that it is more similar to that of plants.[258, 274] Here, the GA pathway branch point is GA12 **101** rather than GA14 **102**, which is the key intermediate linking the GA pathways of *F. fujikuroi* and *S. manihoticola*. Additionally, the C3 hydroxylation takes place later in the GA pathway of *Phaeosphaeria* sp. L487 than that of the other fungi.

Diversification of ancestral GA-pathways likely led to new diterpene pathways that produce different bioactive compounds. Several dozens of such putative diterpene biosynthetic gene clusters can be readily identified in published Ascomycota genome sequences. However, only very few diterpene biosynthetic gene clusters have been characterized so far. The only labdane-type diterpene pathway from a Basidiomycota has been disclosed in a patent application for pleuromutilin antibiotic biosynthesis in *Clitopilus passeckeranu*s.[237]

Identification of the aphdicolan-16β-ol **83** synthase[236] from *Phoma betae* (**Figure 1-7**) facilitated subsequent identification and heterologous expression in *Aspergillus oryzae* of the aphidicolin **106** biosynthetic pathway (Scheme 6).[275] In addition to the bifunctional diterpene synthase, this cluster encompasses two P450 enzymes that carry out three hydroxylations of the diterpenoid alcohol scaffold, as well as a transporter and the aforementioned GGPP **5** synthase. The two CPS/KS homologs PaDC1 phyllocladen-16α-ol **82** synthase) and PaDC2 ((+)-CDP synthase) characterized from *P. amygdali* are each part of separate gene clusters that are involved in the biosynthesis of phyllocladan-11α,16α,18-triol

107.[235] Interestingly, in addition to different types of oxygenases, each of two clusters includes its own GGPP synthase.

Genome mining recently led to the identification of an e*nt*-primara-8(14),15-diene **81** synthase AN1495 in *Aspergillus nidulans* which is part of a gene cluster that has yet to be characterized.[234] Encoded within this cluster are not only a GGPP synthase, but also a HMG-CoA synthase, a mevalonate pathway specific enzyme, which may boost precursor supply for diterpenoid biosynthesis by this pathway.

The monofunctional diterpene synthase, fusicocca-2,10(14)diene **85** synthase (PaFS), from *P. amygdali* is part of a five gene cluster,[242] including three enzymes that catalyze oxidative modifications to the tricyclic diterpene scaffold.[248] This cluster together with a later discovered second, nine-gene cluster are required to synthesize the bioactive fusiccocin A **108** diterpenoid.[276] The presence of a prenyltransferase previously shown to catalyze the prenylation of glucose[277] which is a moiety also present in fusiccocin A **108**, led to the identification of this second fusiccocin cluster in the draft genome sequence of this fungus. Biochemical studies of recombinant enzymes together with gene deletions were carried to complete characterization of fusiccocin biosynthesis.[276] The same group also identified and dissected another fusicoccadiene-type pathway for brassicicene C **109** biosynthesis in *A. brassicola;* as seen with the other diterpene pathways, a contingent of oxidizing enzymes (five cytochrome P450 enzymes, an oxidoreductase and a dioxygenase) are involved in modifying the terpene hydrocarbon scaffold.[243, 247, 248]

1.6.4 Fungal triterpenoids

Oxidosqualene/lanosterol cyclases

The triterpenoids are C30 prenyl chain derived compounds that are found widely in nature as steroids and sterols.[278] The biosynthetic pathway to ergosterol **110**, the major component of the plasma membrane in fungi, begins

with the cyclization of squalene **111** to lanosterol **112**.[279] Epoxidation of squalene **111** yields 2,3-oxidosqualene **113** which adopts a *chair-boat-chair* conformation.[280] Cyclization of the linear polyisoprene chain is catalyzed by a Class II type terpene synthase oxidosqualene/lanosterol cyclase (OSC) as part of a protonation-dependent reaction.[244, 281] Activation of the epoxide in **113** is initiated by a conserved Asp, and a sequential ring-forming cascade results in a 6-6-6-5 tetracylic protosteryl cation **114**. Two methyl migrations and hydride shifts, followed by a final deprotonation, results in lanosterol **112** (**Figure 1-18**), the last common intermediate to the lanostane triterpenoids.[225, 282, 283]

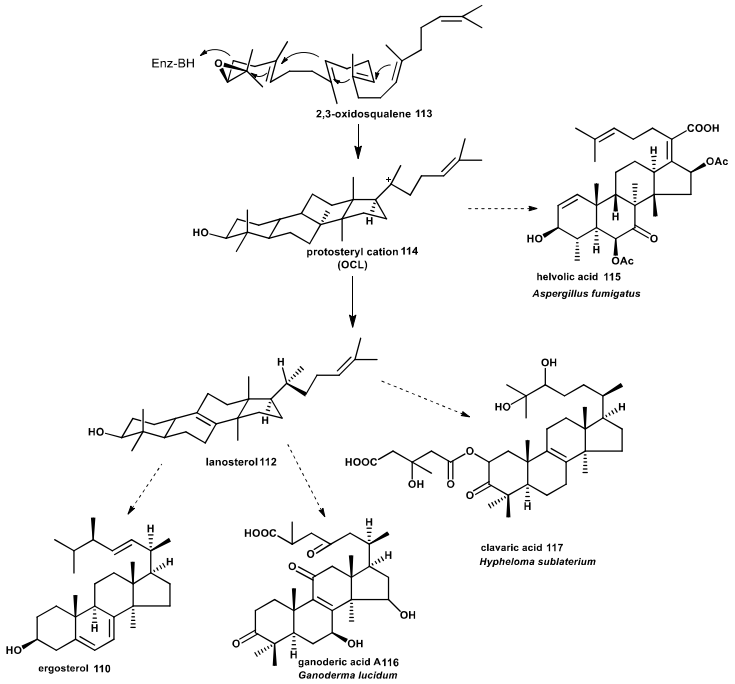


Figure 1-18: Triterpenoid biosynthesis in Fungi

Prokaryotic and eukaryotic triterpene cyclases share a similar overall protein fold, with two α-helical domains linked by a membrane spanning channel. The large aromatic-lined active site cavity spans the two domains, and the protonating group that initiates cyclization is located at the polar region at the top of the central channel.[244, 281] Triterpene cyclases are also defined by conserved 16 amino acid long tandem repeats known as QW sequence motifs. These Gln and Trp residues provide a bonding network between the α-helices, which stabilizes the enzyme during the exergonic reaction, allowing stringent control over product formation.[284] Beyond these structural similarities, bacterial squalene hopene cyclases (SHC) and eukaryotic OSC cyclases have diverged both in sequence and in mechanism. Notably, SHC is characterized by the presence of the protonating

Asp in a conserved DXDD motif, while the protonating Asp of OSC is located in the motif XXDCX.[283] This branching in the evolutionary tree of SHC and OSC has resulted in the mechanistically defining feature of the enzymes: SHC accepts squalene **111** as a substrate, while OSC accepts 2,3-oxidosqualene **113** as a substrate.[283, 285]

The first eukaryotic OSC to be cloned and characterized was the ergosterol **110** biosynthetic pathway enzyme ERG7 from *Saccharomyces cerevisiae*.[179, 286] The catalytic mechanism of ERG7 has since been elucidated by mutational studies. Substrate activation is highly dependent upon the maintenance of a protonating environment, and aromatic residues play a key role in the stabilization of carbocation intermediates, which funnels the product trajectory of ERG7 specifically towards lanosterol **112**.[186, 187, 287, 288] Further fungal OSCs have been characterized from diverse species, including the helvolic acid **115** antibiotic producers *Cephalosporium caerulens*[181] and *Aspergillus fumigatus*,[183] as well as the cytotoxic ganoderic acid (e.g. **116**) producer *Ganoderma lucidum*.[289] Notably, characterization of an oxidosqualene-clavainone cyclase from *Hypholoma sublateritium* revealed the presence of two OSCs: one specifically for primary and one dedicated to secondary metabolism for clavaric acid **117** production.[290]

More recently, bioinformatic approaches led to a wider scale identification of putative OSCs from Ascomycota and Basidiomycota.[291] This database represents a diversity of OSC sequences that could potentially be used to engineer a product promiscuous fungal OSC, offering biosynthetic routes to a greater diversity of fungal triterpenoids, like those produced by plant OSCs.[193, 196, 292]

1.6.5 Discovery of triterpenoid biosynthetic gene clusters

Triterpenoid biosynthesis is relatively well understood and characterized in plants.[292-295] Comparatively little, however, is known about fungal triterpenoid

biosynthesis; despite the isolation of a large number of bioactive fungal triterpenoids mostly from Basidiomycota.[296, 297]

Recent genome sequencing and transcriptome projects have provided the first insights into the biosynthetic pathways to the triterpenoids.[114, 298-300] *G. lucidum* is a prolific producer of bioactive ganoderic acid **116** triterpenoids, correspondingly it has an OSC and a large complement of cytochrome P450s encoded within its genome. Almost half of the genes encoding P450s are upregulated during the transition from primordia to fruiting bodies, which correlates with lanosterol **112** expression and triterpenoid production. Therefore, developmental stage of fungal growth plays an important role in regulation of secondary metabolite genes. Some of these same genes are physically clustered on the genome, albeit not with the gene encoding OSC.[114] Likewise, transcriptome analysis of the medicinally relevant *Wolfiporia cocos* identified several P450s that were upregulated simultaneously with its OSC.[298] Contrastingly, the genome of *A. fumigatus* shows that it contains multiple genes encoding OSC as opposed to a single copy. Furthermore, the putative enzymes related to helvolic acid **115** production are located in a gene cluster encoding protostadienol **118** synthase, multiple P450s, a reductase, acyltransferases and a dehydrogenase. Together, these enzymes are predicted to represent the full helvolic acid **115** biosynthetic pathway.[301]

1.6.6 Mining the fungal (sesqui)terpenome

Terpene synthase identification

Numerous databases exist to search the rapidly increasing number of fungal genomes for biosynthetic gene candidates, and for predicting putative coding sequences. For example, the Joint Genome Institute (JGI) listed 362 fungal genomes (117 Basidiomycota) at the time of writing, with a goal of 1000 sequenced genomes from all families in the fungal kingdom in the next few years.[163] Users of the JGI web interface (http://genome.jgi-psf.org)[302] benefit

from standardized EST and/ or RNAseq data available for many of the fungal genomes that are that are displayed along with gene annotations.

Putative terpene synthases can be identified by performing a Basic Local Alignment Search Tool (BLAST)[303] search of target genome(s) using characterized terpene synthases sequences as query sequences. We have used BLAST searches of fungal genomes, combined with biochemical knowledge of the initial FPP **4** cyclization reaction (see Scheme 1) catalyzed by the *O. olearius* enzymes (Omp1-10, section 2.1) to develop a predictive framework for further, targeted discovery of additional fungal sesquiterpene synthases and their associated biosynthetic gene clusters.[22, 89]

We found that sesquiterpene synthase sequences identified by BLAST analysis in the sequenced Basidiomycota genomes (presently more than 1000

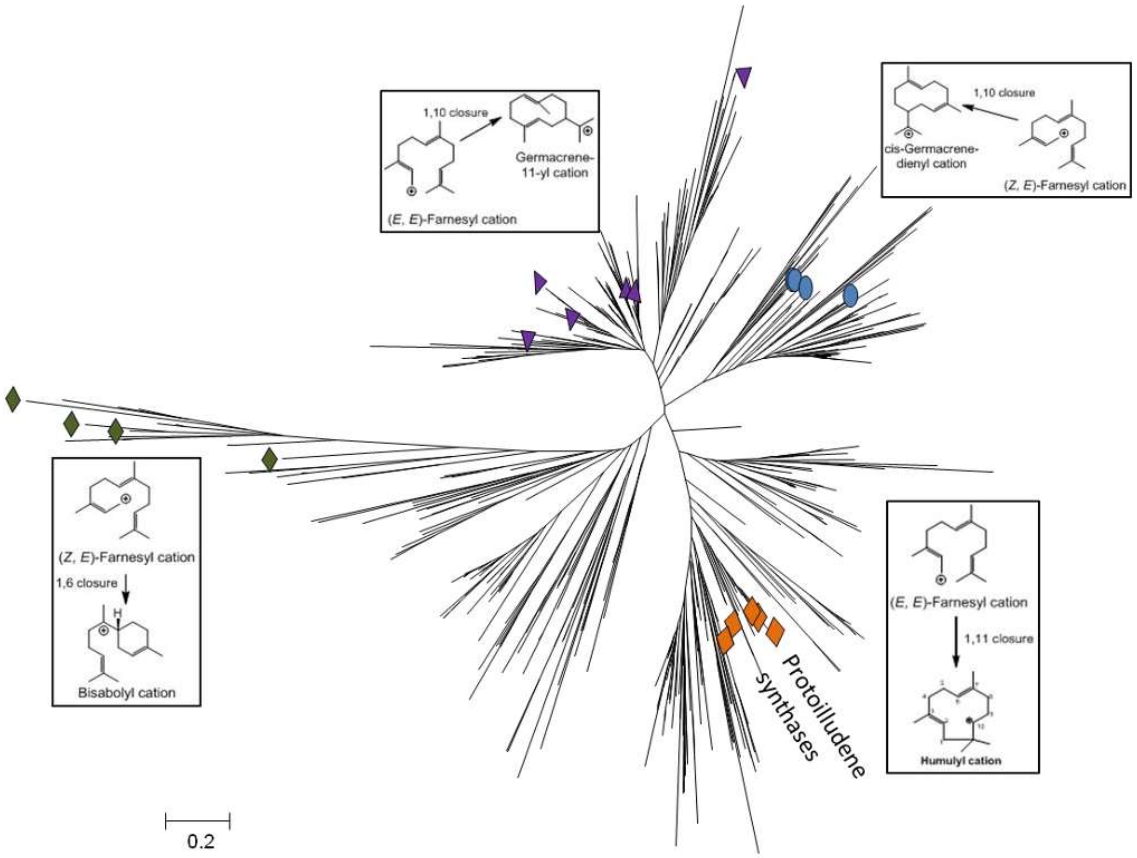


Figure 1-19: Putative Basidiomycota sesquiterpene synthases form clades consistent with their initial cyclization reaction.

sequences in ~100 genomes) cluster in distinct clades according to initial cyclization mechanism (**Figure 1-19**).[89] This method led to the focused discovery and characterization of several 1,11-cyclizing terpene synthases, which produce *trans*-humulyl cation **60** derived sesquiterpenes (section 1.6.2).[22, 89] Notably, these characterized Δ6-protoilludene synthases form a single branch within the 1,11 cyclizing clade (**Figure 1-19**). Many other predicted *trans*-humulyl cation **60** producing sesquiterpene synthases remain yet to be discovered, potentially producing the diverse scaffolds shown in **Figure 1-13**.

Identification of terpenoid biosynthetic gene clusters

Discovery of fungal terpenoid biosynthetic pathways by genome mining approaches is facilitated by the fact that fungi tend to cluster genes encoding biosynthetic pathways within their genome. Clusters can be identified by manually scanning the genome sequence ±20 kb upstream and downstream of predicted terpene synthase sequences e.g. identified using the strategy described in the previous section. BLAST searches of the NCBI protein database and the NCBI Conserved Domain Database (CDD) allows annotation of predicted genes surrounding the terpene synthase. Typical enzyme classes and proteins expected in such clusters are cytochrome P450 monooxygenases, oxidoreductases and oxygenases, group transferases, transporters, transcription factors and enzymes required for isoprenoid precursor biosynthesis. The likely boundaries of such biosynthetic clusters are predicted by the presence of large gaps between genes (>10 kb), or the presence of several consecutive genes that are not likely to be associated with natural product biosynthesis.[85]

However, manual identification of biosynthetic gene clusters is subject to user interpretation, is extremely time consuming, and as a result is a significant bottleneck in the identification of putative biosynthetic pathways. Therefore, automated search algorithms are required for the large scale identification of fungal natural product pathways and several web-based analysis tools have been

developed for this purpose (reviewed in: [304, 305]). When using these platforms, however, one has to keep in mind that these tools have often been trained for the identification of specific natural products gene clusters (e.g. non-ribosomal peptide (NRP) and polyketide synthase (PKS) pathways) common in certain groups of organisms (e.g. Ascomycota, bacteria). From our own experience we have found that these tools are not particularly adept at identifying terpenoid clusters in Basidiomycota. For example, antiSMASH 2.0[306] identified four of the eleven manually identified putative terpene synthases in the genome of *O. olearius*.[89] Similarly, five out of eighteen manually predicted terpene synthase sequences were identified in *S. hirsutum*.[22] This and other tools also significantly underestimated biosynthetic gene cluster size compared to our manual annotations, probably due to the high degree of complexity and variability of fungal gene clusters, as well that fact that few Basidiomycota clusters have been fully characterized.[307] As more natural product pathways from Basidiomycota are characterized, these tools can be trained for more accurate, automated cluster prediction in these types of fungi.

Characterization of terpenoid biosynthetic gene clusters

Characterization of fungal terpenoid gene clusters identified in Ascomycota described throughout this contribution has relied on a combined approach of biochemical characterization of recombinant enzymes and specific gene deletions in the native producer host, as well as transfer of biosynthetic cluster genes into closely related fungal hosts for which genetic tools are available.[308-310] In order to access the incredible diversity of Basidiomycota terpenoid natural product pathways encoded in the large (and increasing) number of fungal genomes, different tools and methods need to be developed to go from genome mining to terpenoid biosynthesis. Some of the problems that need to be overcome for this to happen will be discussed briefly using the Basidiomycota as an example, but similar difficulties may also be encountered with other, less well-studied classes of fungi.

Several major challenges significantly impede efficient characterization of terpenoid biosynthetic gene clusters in Basidiomycota: 1. Correct cDNA prediction, 2. Therefore, necessity to clone genes from cDNA and thus the need to establish laboratory growth of strains under conditions where target genes/pathways are expressed, 3. Fungal strains are not genetically tractable, and therefore require heterologous gene expression and pathway assembly.

The first step following the identification of putative terpene synthases and the associated gene cluster in a genome requires a refined prediction of the correct coding sequences for each gene. However, this is in our experience not trivial for Basidiomycota, which tend to have have very intron-rich genomes.[311] We have found several very small and unpredicted exons (sometimes only 11-18 bp in length [312] and introns/exons 27 and 28 bp in length[83, 89]) in genes cloned from cDNA. Genes synthesized based on cDNA predictions are therefore often non-functional. **Figure 1-20** illustrates different structural gene predictions, alternative splice variants and the functional cDNAs of sesquiterpene synthases cloned from *S. hirsutum*. The gene prediction program Augustus (http://bioinf.uni-greifswald.de/augustus/)[313, 314] can be used to identify the most likely cDNA sequences for a biosynthetic enzyme. Gene predictions created from the putative terpene synthase ±10 kb can

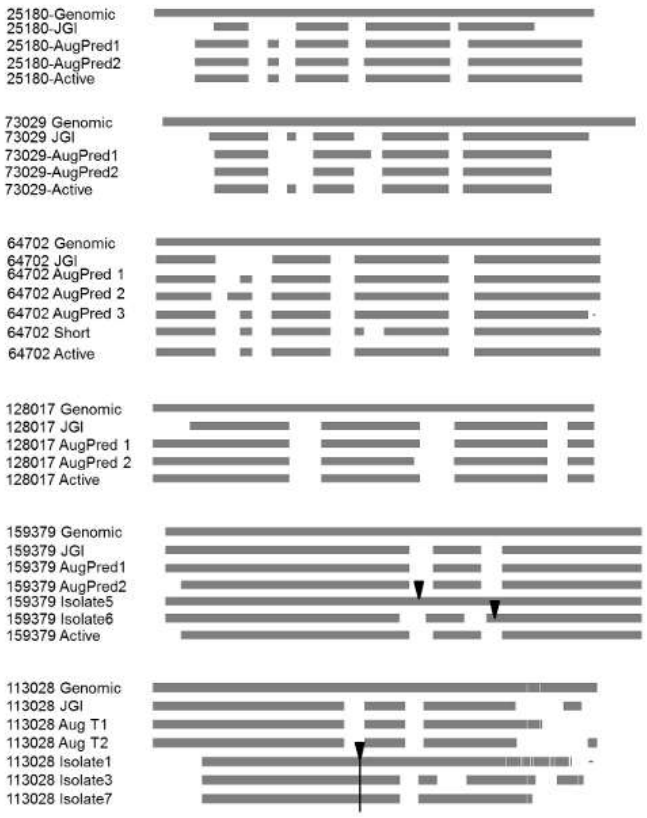


Figure 1-20: Comparison of the intron architecture predicted by JGI and Augustus, and that found in cloned and characterized sesquiterpene synthases from *S. hirsutum*.

then be used to manually identify potential gene start/stop codons by alignment in MEGA6.[315] This allows the design of a small set of cloning primers to attempt amplification of the correct ORF from cDNA. In addition, the program FGENESH (http://linux1.softberry.com/) by SoftBerry (Mount Kisco, NY) offers multiple methods to predict gene models.[316] However, these programs have been trained to a small set of fungal strains and often do not correctly predict the less common splicing pattern present in Basidiomycota.

Collection of fungal tissue sufficient for cDNA isolation is often a significant barrier, as many fungi, and especially many Basidiomycota, are difficult or impossible to culture in the laboratory and may be not available from field collections. Also, the target genes for cloning must be expressed at the time of cDNA preparation. Fungal gene expression is highly complex, subject to RNAi silencing and influenced by trans-acting elements of the genome architecture.[317, 318] Therefore, a large number of biosynthetic gene clusters remain silent.[111] Strategies have been developed to activate fungal natural product pathways including modification of culture conditions, co-culture,[319] and epigenetic manipulation, typically in high-throughput assays to produce novel fungal compounds for isolation.[111, 320-323] However, the conditions required to express specific genes are generally unknown, and cannot be predicted with confidence.

Most Basidiomycota are not genetically tractable and no suitable tractable Basidiomycota model systems are available that could correctly splice and functionally express heterologous biosynthetic gene clusters cloned from other members of this fungal division. Hence, characterization of terpenoid natural product pathways identified in these fungi has to rely on the tedious assembly method of one gene at a time, in alternative hosts, such as yeast. Successful expression of pathways in these hosts can be challenging, as some enzymes e.g. cytochrome P450 monoxygenases, require a specific reductase for optimal activity[8] and modified terpenoids may be toxic to the expression host, therefore complex cellular sequestration and secretion machinery could be required to

prevent reduced cell growth.[92] In the future, additional development of genetic platforms specifically designed for natural product pathway expression in Basidiomycota, such as *Ustilago maydis,* is needed,[324, 325] similar to the advanced tools currently available for Ascomycota.[326]

1.6.7 Conclusions and outlook: Fungal terpenoid natural products

Fungi have an enormous capacity for terpenoid natural product biosynthesis, and therefore represent an incredible resource for the discovery of new biosynthetic pathways. Identification and characterization of some of the pathways responsible for terpenoid production has revealed a diversity in structure, chemistry and function, including new classes of enzymes. The increasing availability of fungal genome sequences is streamlining the identification of novel pathways and enzymes to a wide range of bioactive terpenoids. The majority of such biosynthetic studies have focused on Ascomycota and only recently have we begun to look into the terpenome of Basidiomycota. Genome surveys clearly show that we have barely begun to unlock some of this biosynthetic potential encoded in the relatively small number of fungi with sequenced genomes that represent just a tiny fraction of the fungal diversity (**Figure 1-19**).

Our ability to characterize the fungal terpenome is dependent on, and currently severely limited by, our capability to rapidly identify and subsequently biochemically characterize terpenoid biosynthetic genes and their diverse functions. Current approaches of first finding potential pathways and then using molecular biology strategies to tease out the functions of individual enzymes in a biosynthetic pathway are slow, cumbersome and inadequate in the face of rapidly increasing genomic information. The development of better bioinformatic and genetic tools for the mining and heterologous assembly of large natural product pathways will be absolutely essential to enable high-throughput pathway discovery. For this to occur, it will be important for the natural products community to adopt synthetic biology and modular pathway assembly principles

along with suitable platform organisms for process automation. The ultimate goal for the development of such discovery pipelines will be the ability to quickly move from *in silico* pathway prediction to high-throughput biosynthetic pathway assembly, heterologous expression, and analytical profiling of products followed by bioactivity screening. Revitalization of the natural products drug discovery pipeline will critically depend on the implementation of such strategies.

Chapter 2 Mushroom hunting using bioinformatics: Application of a predictive framework facilitates the selective identification of sesquiterpene synthases in Basidiomycota.

The following is a reprint of the article Quin M.B.*, Flynn, C.M.,* Wawrzyn, G.T., Choudhary, S. Schmidt-Dannert, C. (2013) "Mushroom hunting using bioinformatics: application of a predictive framework facilitates the selective identification of sesquiterpene synthases in Basidiomycota" *ChemBioChem*, Dec 16; 14(18):2480-91. *co-first authors.

Paper hyperlink:

http://onlinelibrary.wiley.com/doi/10.1002/cbic.201300349/abstract

CMF performed strain growth and GC/MS compound identification, cDNA preparation, genomic & phylogenetic data collection and analysis, and cloned and characterized Stehi1|159379, Stehi1|128017. MBQ purified and characterized the three protoilludene synthases, and led paper structure and writing. SC cloned the protoilludene synthases. GTW helped design figures 2 and 6 using data collected by CMF and SC.

2.1 Summary

Basidiomycota represent a diverse source of natural products, particularly the sesquiterpenoids. Recently, the genome sequencing, mining, and subsequent discovery of a suite of sesquiterpene synthases was described in *Omphalotus olearius*. A predictive framework was developed to facilitate the discovery of sesquiterpene synthases in Basidiomycota. Phylogenetic analyses indicated a conservation of both sequence and initial cyclization mechanisms used. Here, the first robust application of this predictive framework is reported. It is used to pursue and selectively identify sesquiterpene synthases that follow a 1,6-, 1,10-, and 1,11-cyclization mechanism in the crust fungus Stereum hirsutum. The successful identification and characterization of a 1,6- and a 1,10-cyclizing sesquiterpene synthase, as well as three 1,11-cyclizing Δ-6 protoilludene synthases, is described. This study verifies the accuracy and utility of the predictive

framework as a roadmap for the discovery of specific sesquiterpene synthases from Basidiomycota, representing an important step forward in natural product discovery.

2.2 Introduction

Basidiomycota are a rich source of natural products. Mushroom-forming fungi are widespread in nature; current estimates of fungal diversity exceed that of land plants by a ratio of 10:1.[97, 327] The ubiquitous nature of mushrooms can be attributed to their highly active secondary metabolic pathways. Many of the bioactive secondary metabolites produced serve to protect the mushroom from predators, enhance survival under harsh growth conditions, and regulate cell proliferation and differentiation.[328] The isolation, chemical characterization, and chemical synthesis of bioactive compounds from mushrooms has led to the production of pharmaceuticals with antimicrobial, antifungal, and anticancer applications.[106, 328]

Sesquiterpenoids represent a major group of secondary metabolites isolated from Basidiomycota. These natural products are bioactive compounds due to their unique chemical structures, which result from complex modifications and rearrangements of the precursor sesquiterpene scaffold by accessory modifying enzymes.[102, 329] Sesquiterpenes are produced by a class of enzymes known as sesquiterpene synthases, which cyclize the 15-carbon linear molecule farnesyl pyrophosphate ((2*E*,6*E*)-FPP) via a specific ring closure, producing hundreds of different types of volatile cyclic scaffolds.[78, 80] Despite this diversity of products, all sesquiterpene synthases share a common three dimensional fold and catalytic mechanism. The cyclization cascade is initiated by a divalent metal-ion mediated dephosphorylation of (2*E*,6*E*)-FPP to produce a highly reactive carbocation, which is subjected to a series of rearrangement reactions until a final deprotonation or attack by water. The resulting volatile cyclic product is then released from the catalytic active site.[330, 331] The robust nature and cyclization products of these enzymes makes them ideal candidates for study, not only to further our understanding of carbocation chemistry, but also to develop a toolbox

for the biosynthetic production of pharmaceutically relevant compounds that may not be accessible by traditional chemical syntheses.[332, 333]

Only a handful of fungal sesquiterpene synthases have been cloned, purified, and enzymatically characterized.[83, 89, 134, 137, 140, 150, 155, 158, 162, 334-336] Comprehensive studies are challenging, not only because it can be relatively difficult to cultivate mushrooms under laboratory conditions,[103] but also because few systematic surveys of sequence information have been conducted. As such, the identification of novel fungal sesquiterpene synthases has been a labor-intensive and unpredictable process. However, the relative affordability of next-generation sequencing technologies [112] has led to a large increase in the number of sequenced fungal genomes. The availability of sequence information facilitates mining of genomes for the discovery of new sesquiterpene synthases and pathways, as has been applied to the well-studied bacterial [19, 337, 338] and plant terpene synthases.[87] Our group recently carried out a comprehensive phylogenetic analysis of all sequenced Basidiomycota genomes. It was discovered that sesquiterpene synthases appeared to cluster, not only according to sequence conservation, but also by cyclization mechanism.[89] If correct, these phylogenetic analyses offer a predictive framework, allowing the focused discovery of novel sesquiterpene synthases based upon cyclization mechanism of choice, streamlining the identification and cloning of novel sesquiterpene synthases which produce desirable natural products.

The goal of this work was to test the reliability of the predictive framework to selectively identify novel sesquiterpene synthases from the crust fungus *Stereum hirsutum*. This fungus was chosen because it has a sequenced genome and according to our bioinformatics analyses, it possesses several putative sesquiterpene synthases. Further, a search in SciFinder® (www.cas.org) shows that compared to other genome sequenced Basidiomycota a number of bioactive natural products have been isolated and characterized from *Stereum* strains. The majority of sesquiterpenoids previously isolated from *Stereum* strains derive from a *trans*-humulyl cation generated from (2*E*,6*E*)-FPP by an initial 1,11-cyclization reaction.[167-170, 173-175, 339] The *trans*-humulyl cation is the first committed step in the production of a wide swathe of pharmaceutically relevant

sesquiterpene scaffolds made by Basidiomycota.[102] It was therefore hypothesized that

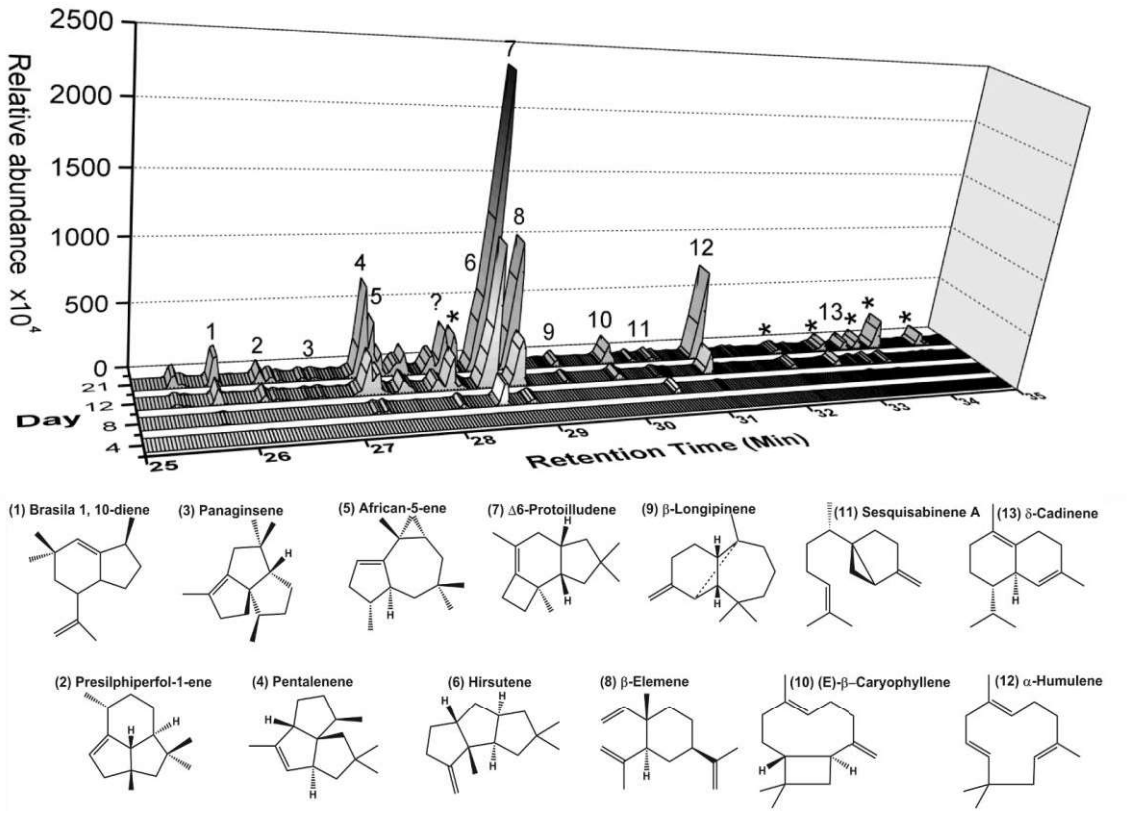


Figure 2-1 The headspace of a liquid culture of S. hirsutum was sampled and analyzed for the production of volatile sesquiterpenes by GC/MS over a period of 21 days.
Identified sesquiterpene products are numbered, and their relative stereochemical structures are shown. One peak at m/z 204 could not be identified (labeled with a question mark). Peaks labeled with an asterisk are likely modified terpenoids, each with *m/z* 218-22.

S. hirsutum would also be a prolific producer of sesquiterpenes, including those derived from the *trans*-humulyl cation. This work shows that not only was this hypothesis correct, but also that by using the predictive framework, it was possible to selectively clone, heterologously express, and enzymatically characterize 1,6-, 1,10- and 1,11-cyclizing sesquiterpene synthases, including three novel Δ-6 protoilludene synthases. These data serve to validate the predictive framework as a practical and reliable guide, facilitating the directed discovery of sesquiterpene synthase activities in Basidiomycota. The predictive framework will be a valuable resource for future natural product discovery in Basidiomycota, which will become increasingly important as additional fungal

genomes are sequenced.

2.3 Results and Discussion

2.3.1 *S. hirsutum* produces a wide range of sesquiterpenes

Several sesquiterpenoids derived from the *trans*-humulyl cation have been isolated and characterized from the fruiting bodies of *Stereum* sp. [173-175, 339] and from culture media,[168, 170] indicating that members of this genus produce 1,11-cyclizing sesquiterpene synthases. Additionally, sesquiterpenoids derived from a 1,6-cyclization [169] as well as a 1,10-cyclization mechanism [167, 339] have also been isolated from *Stereum* sp. (**Figure 2-1**). No sesquiterpene synthase has yet been described or characterized from *Stereum*.

To confirm that the genome sequenced strain of *S. hirsutum* is a prolific producer of sesquiterpene scaffolds, the headspace of a liquid culture was analyzed for the presence of volatile hydrocarbons by gas chromatography/mass spectrometry (GC/MS)

(**Figure 2-1**, Figure S2-1). Volatile sesquiterpene production was apparent after 8 days, and the relative abundance of these products increased over the sampling period of 21 days. The major sesquiterpene produced was Δ-6 protoilludene **7**, a 1,11-cyclization product, and the second most abundant sesquiterpene was β-elemene **8**, a heat-induced Cope rearrangement product of the 1,10-cyclization product germacrene A.[340] Other abundant volatiles included α-humulene **12**, hirsutene **6**, and pentalenene **4**, all of which are 1,11-cyclization products.[102] Other less abundant products included the 1,10-cyclization product δ-cadinene **13**, and the 1,6-cyclization product sesquisabinene A **11**. These findings confirmed that *S. hirsutum* possesses a number of as-of-yet uncharacterized sesquiterpene synthases that follow a 1,6-, a 1,10- and a 1,11-cyclization mechanism.

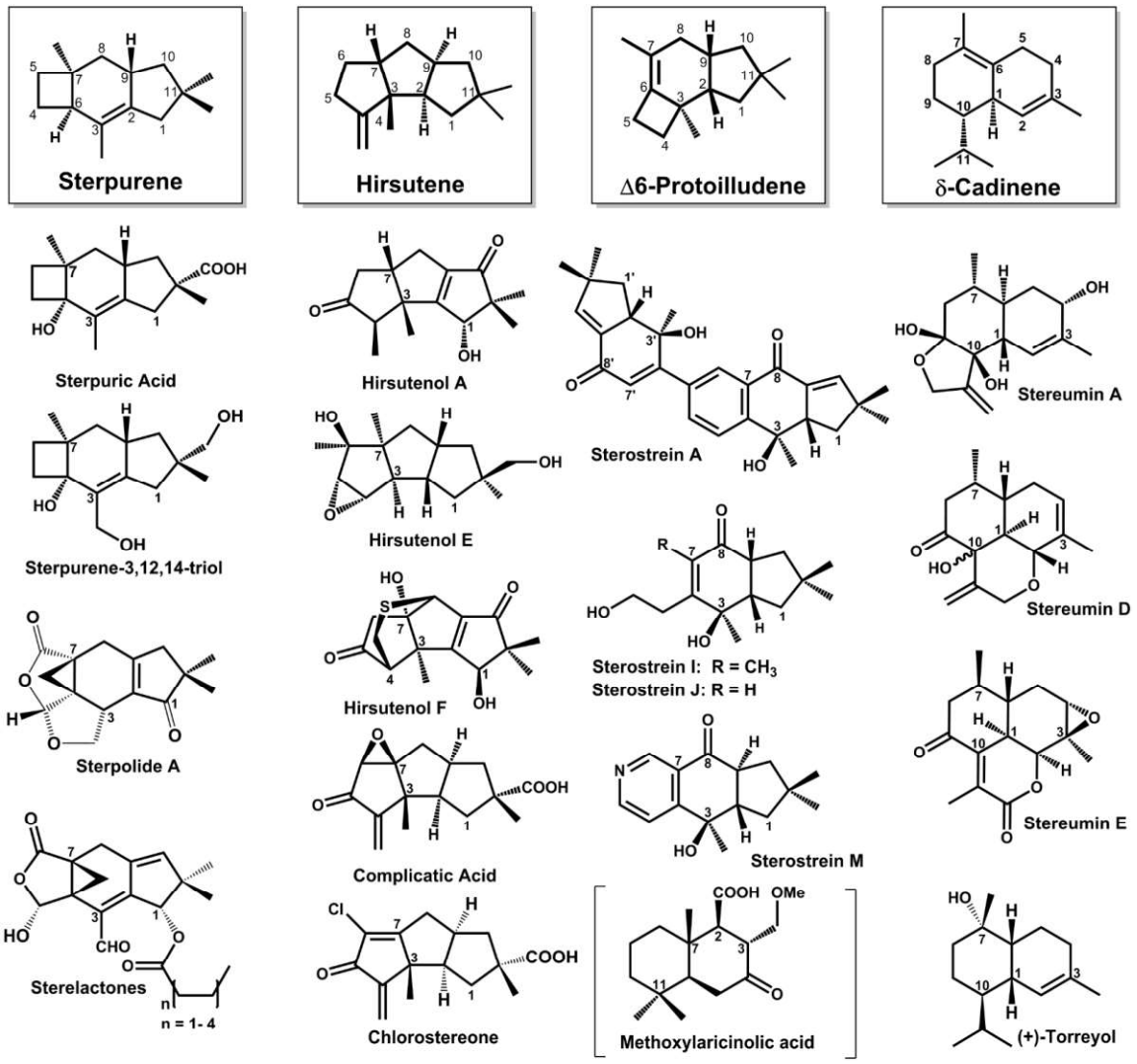


Figure 2-2: Examples of modified sesquiterpenoids previously isolated from *Stereum* sp.
Most sesquiterpenoids isolated from *Stereum* sp. are derived from the *trans*-humulyl cation, a product of the (2*E*,6*E*)-FPP 1,11-cyclization mechanism. Atom numbering is based upon designations in the FPP precursor. Sterpuric acid, sterpolide A, sterpurene-3,12,14-triol, and the sterelactones result from the modification of the sterpurene scaffold.[165, 166, 172, 194, 341] Modification of the hirsutene scaffold yields hirsutenols A-F,[168, 170] as well as complicatic acid and chlorostereone.[173] The sterostreins [174, 175] are modified products of the Δ-6 protoilludene scaffold. Additionally, the stereumins are derived from the (3*R*)-NPP 1,10-cyclization product δ-cadinene,[167, 339] (+)-torreyol is a product of attack by water on either the (2*E*,6*E*)-FPP 1,10-cyclization intermediate muurolyl cation or the (3*R*)-NPP 1,10-cyclization intermediate cadinyl-cation.[128] Finally, methoxylaricinolic acid results from a putative 2,7-cyclization product drimane [169, 181] (shown in parentheses). Relative stereochemistries of sesquiterpenoids are shown.

2.3.2 Using phylogenetic analyses to predict 1,6-, 1,10-, and 1,11-cyclizing sesquiterpene synthases in *S. hirsutum*

Previously, our group carried out a comprehensive bioinformatic survey and phylogenetic analysis of 40 Basidiomycota genomes using sequence and function data for characterized sesquiterpene synthases from *C. cinereus* and *O. olearius* as a guide.[89] A total of 542 putative sesquiterpene synthases were found. The enzymes formed five distinct clades in a phylogenetic tree, apparently clustering by sequence conservation and cyclization mechanism. Clade I consisted of enzymes (Omp1-3, Cop 1-3) that utilize a 1,10-cyclization of (2*E*,6*E*)-FPP to produce sesquiterpenes derived from a *E*,*E*-germacradienyl cation.[83, 89] Clade II consisted of enzymes (Cop4, Omp4, Omp5a and 5b) that shared a 1,10-cyclization of (3*R*)- nerolidyl diphosphate (NPP) mechanism,[132, 135] producing sesquiterpenes derived from a *Z*,*E*-germacradienyl cation. Clade III consisted of enzymes believed to share a common 1,11-cyclization of (2*E*,6*E*)-FPP mechanism, producing the *trans*-humulyl cation. Omp6 and Omp7, both of which are Δ-6 protoilludene synthases that were characterized from *O. olearius*, clustered in this group. Cop5, which is a putative pentalenene synthase from *C. cinereus,* also clustered in Clade III. Clade IV consisted of enzymes that shared a 1,6-cyclization of (3*R*)-NPP or (3*S*)-NPP mechanism,[150, 342] producing a bisabolyl cation. The α-cupranene synthase Cop6 from *C. cinereus*, and the α/β-barbatene synthase Omp9 as well as the daucene/*trans*-dauca-4(11),8-diene synthase Omp10 from *O. olearius* clustered in this group. Finally, Clade V consisted of enzymes believed to share a 1,6-cyclization mechanism.

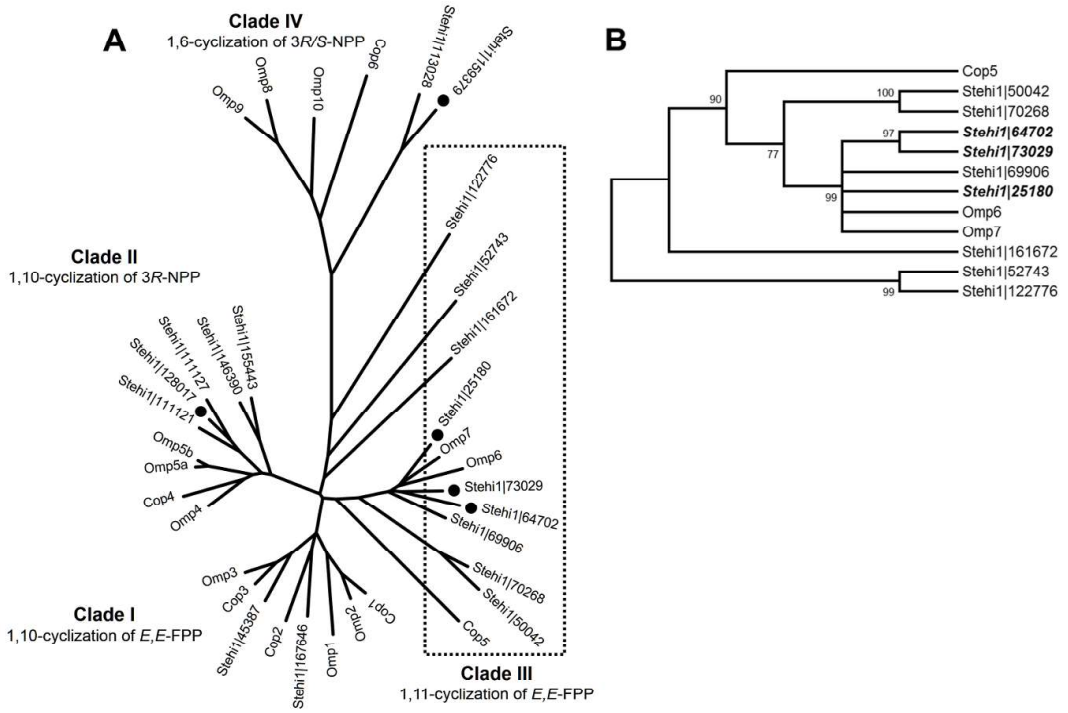


Figure 2-3: Phylogenetic analysis of sesquiterpene synthase homologues.
Sesquiterpene synthase homologs from *Coprinus cinereus* (Cop), *Omphalotus olearius* (Omp) and *Stereum hirsutum* (Stehi1) are labeled. A) Sesquiterpene synthases form distinct clades in an unrooted neighbor-joining phylogram. Clade III, described previously,[89] is highlighted with a box. Gene accession numbers highlighted with a circle indicate cloned and characterized sesquiterpene synthases. B) An unrooted neighbor-joining tree of members of Clade III. Branches are labeled with their bootstrap values. Gene accession numbers highlighted in bold indicate cloned and characterized sesquiterpene synthases.

We set out to establish whether this apparent phylogenetic clustering according to cyclization mechanism could be used to predict enzyme function from a genomic perspective for *S. hirsutum*. Multiple sequence alignments and phylogenetic analyses revealed a total of 18 putative sesquiterpene synthases from *S. hirsutum* located in Clades I-IV. Two of the sesquiterpene synthases were clustered in Clade I, five of the sesquiterpene synthases clustered together in Clade II, nine of the sesquiterpene synthases were clustered with sequences in Clade III, and two of the sesquiterpene synthases were located in Clade IV (**Figure 2-3A**). Note that none of the sesquiterpene synthases were located in the previously described Clade V.[89]

The two Clade I putative sesquiterpene synthases, Stehi1|45387 and Stehi1|167646, clustered together with Omp3, Cop2 and Cop3, likely utilizing a 1,10-

cyclization of (2*E*,6*E*)-FPP to yield a *E,E*-germacradienyl cation-derived product, such as β-elemene **8**, which is a heat-induced Cope rearrangement product of the 1,10-cyclization product germacrene A.[340]

Of the Clade II putative sesquiterpene synthases, Stehi1|111121, Stehi1|111127 and Stehi1|128017 are clustered most closely with Cop4, Omp4, Omp5a and Omp5b. It was predicted that these would likely be responsible for the production of the δ-cadinene **13** detected in the headspace of *S. hirsutum* cultures, via a *Z*,*E*-germacradienyl cation from a postulated 1,10-cyclization of (3*R*)-NPP.[83, 89] The other predicted 1,10-cyclizing sesquiterpene synthases Stehi1|155443 and Stehi1|146390 are located together on a separate branch.

Half of the 18 predicted sesquiterpene synthases are located in Clade III (1,11-cyclization mechanism), substantiating the finding that *S. hirsutum* is a prolific producer of *trans*-humulyl cation derived sesquiterpenes (**Figure 2-1**). Stehi1|25180, Stehi1|73029, Stehi1|64702 and Stehi1|69906 cluster together, and are closely related to the Δ-6 protoilludene synthases Omp6 and Omp7 [89] (**Figure 2-3B**). It was therefore predicted that these putative sesquiterpene synthases would produce the Δ-6 protoilludene **7** detected in the culture headspace, utilizing a 1,11-ring closure mechanism to yield a *trans*-humulyl cation. Stehi1|70268 and Stehi1|50042 are more closely related to the putative pentalenene synthase Cop5, and therefore may produce pentalenene **4**. The other three putative 1,11-cyclizing sesquiterpene synthases in Clade III, Stehi1|122776, Stehi1|52743, and Stehi1|161672, did not cluster tightly with other members, and are instead located on separate branches. One or more of these enzymes may be responsible for the production of α-humulene **12** or hirsutene **6** by *S. hirsutum*.

Finally, the two members of Clade IV, Stehi1|113028 and Stehi1|159379, are located on a separate branch and are not as tightly clustered as members of other clades. However, they clustered closest to Cop6, and were therefore predicted to utilize a 1,6-cyclization of (3*R*)-NPP or (3*S*)-NPP to yield a bisabolyl cation-derived product such as sesquisabinene A **11**, in a similar fashion to the α-cuprenene synthase Cop6.[89, 157]

2.3.3 Cloning the suite of predicted sesquiterpene synthases from *S. hirsutum* cDNA

To test the accuracy of our predictive framework, the genes encoding the putative enzymes were cloned and expressed heterologously. Gene predictions of the 18 putative sesquiterpene synthases from the *S. hirsutum* genome sequence were refined by manual re-annotation. All of the potential transcripts obtained for each sesquiterpene synthase-encoding gene were first aligned against known sequences of previously isolated sesquiterpene synthases to identify the most likely transcript(s) encoding a functional sesquiterpene synthase. Using gene specific primers designed for the manually predicted gene models, PCR amplification products of several splice variants were successfully obtained from cDNA for Stehi1|159379, Stehi1|113028, Stehi1|128017, Stehi1|25180, Stehi1|64702 and Stehi1|73029. Incorrectly spliced isoforms were excluded from further analysis due to the presence of internal stop

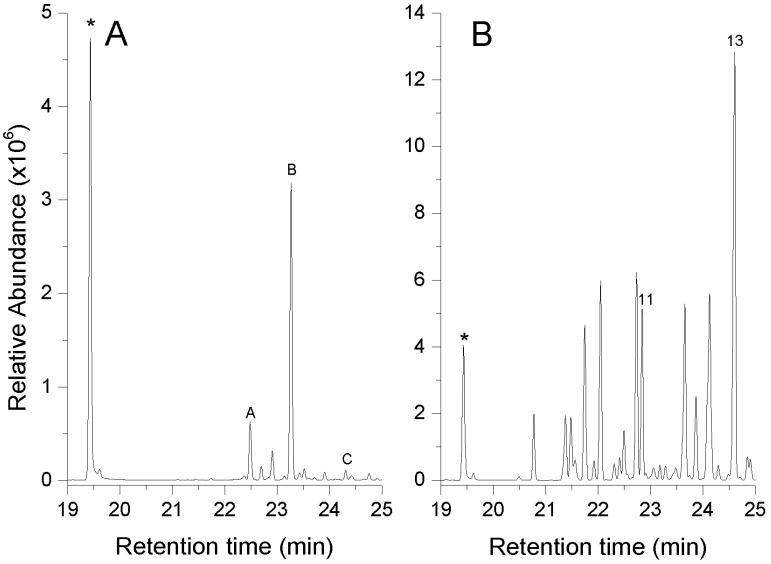


Figure 2-4: Volatile sesquiterpene production by E. coli cultures expressing Stehi1|159379 and Stehi1|128017.
The headspace of cultures was sampled and analyzed for production of sesquiterpenes by GC/MS. *E. coli* cells constitutively expressing A) Stehi1|159379 produced primarily the following identified products: β-barbatene **B**, α-barbatene **A**, and small quantities of α-cuprenene **C**. All of these sesquiterpenes are products of a 1,6-cyclization of (3*R*)-NPP. (Note that peaks identified in the headspace of *E. coli* cells expressing Stehi1|159379 were not detected in the culture headspace of *S. hirsutum* and peak labeling follows instead that of Figure S2 showing Cop6 and Omp9 products). *E. coli* cells constitutively expressing B) Stehi1|128017 produced as major products that were identified in the *S. hirsutum* culture headspace (Figures 2-2 and S2-1): δ–cadinene **13**, a putative 1,10-cyclization of (3*R*)-NPP product, as well as sesquisabinene A **11**, which is believed to result from a 1,6-cyclization of (3*R/S*)-NPP. A number of other sesquiterpenoid products are also produced and are identified and labeled in Figure 2-5. Indole (labeled with an asterisk *), a breakdown product of tryptophan, is naturally produced by *E. coli* and serves as an internal standard.

codons or frameshift mutations. Several unsuccessful attempts were made to obtain a correctly spliced version of Stehi1|113028, which is located adjacent to Stehi1|159379, leading the authors to believe that either the gene prediction is incorrect or that it is a pseudogene. Notably, despite exhaustive efforts, PCR amplification products could not be obtained from cDNA for any of the other 12 predicted sesquiterpene synthases. These genes may not be expressed under the growth conditions used,[343] may be pseudogenes,[344] or may not be accurately predicted using the fungal gene prediction models available (data not shown). Furthermore, Basidiomycota genes typically have many small introns and exons that are difficult or impossible to predict using current fungal gene models, making the task of correctly predicting splice variants of genes particularly tedious and challenging,[344] and in most cases preclude the feasibility of gene synthesis approaches.

In order to confirm that members of certain clades would follow the predicted cyclization mechanisms, the PCR products of the putative sesquiterpene synthases were cloned into our in-house expression vector pUCBB,[345] which allows for constitutive expression of genes in the heterologous host *E. coli*. The putative sesquiterpene synthases were functionally characterized by GC/MS analysis of the volatile components of the headspace from recombinant BL21 (DE3) *E. coli* cultures.[85, 127]

2.3.4 Predicted 1,6- and 1,10-cyclizing sesquiterpene synthases produce bisabolyl cation and putative *Z,E*-germacradienyl cation-derived terpenes

The culture headspace of cells expressing the Clade IV member Stehi1|159379 contained one major peak which was identified as β-barbatene, as well as minor peaks which were identified as α-barbatene and α-cuprenene (**Figure 2-4A**). Although these peaks were not identified in the *S. hirsutum* headspace analysis perhaps due to further modification into a non-volatile compound(s) (**Figure 2-1**), identical retention indices and mass fragmentation patterns were obtained for the major peaks from the culture headspace of *E. coli* cells expressing the previously characterized α/β-barbatene synthase

Omp9 from *O. olearius,*[89] as well as the α-cuprenene synthase Cop6 from *C. cinereus* [158] (**Figure S2-2**). These data confirm that Stehi1|159379 utilizes a 1,6-ring closure of (3*R*)-NPP or (3*S*)-NPP to yield bisabolyl cation-derived products,[150, 342] as had been predicted using our phylogenetic analyses.

The headspace of cells expressing the Clade II member Stehi1|128017 contained one major peak, δ-cadinene **13,** as well as several minor peaks including β-copaene, sativene, γ-muurolene, α-muurolene, β-cubebene, germacrene D, and sesquisabinene A **11** (**Figure 2-3B, Figure S2-3**). Of these, δ-cadinene **13** and sesquisabinene A **11** were identified in the headspace of *S. hirsutum* cultures (**Figure 2-1,** Figure S1). Retention indices and mass fragmentation patterns of the δ-cadinene **13** peak matched those obtained for the δ-cadinene synthases Omp4 [89] and Cop4,[83, 157] and the γ-cadinene synthase Omp5a [89] (**Figure S2-3**). These findings suggest that Stehi1|128017 utilizes a putative 1,10-cyclization of (3*R*)-NPP to yield sesquiterpenes derived from a *Z*,*E*-germacradienyl cation. However, sesquisabinene A **11** is proposed to derive from a 1,6-cyclization mechanism.[346] Also, a recent study raised into question whether δ-cadinene

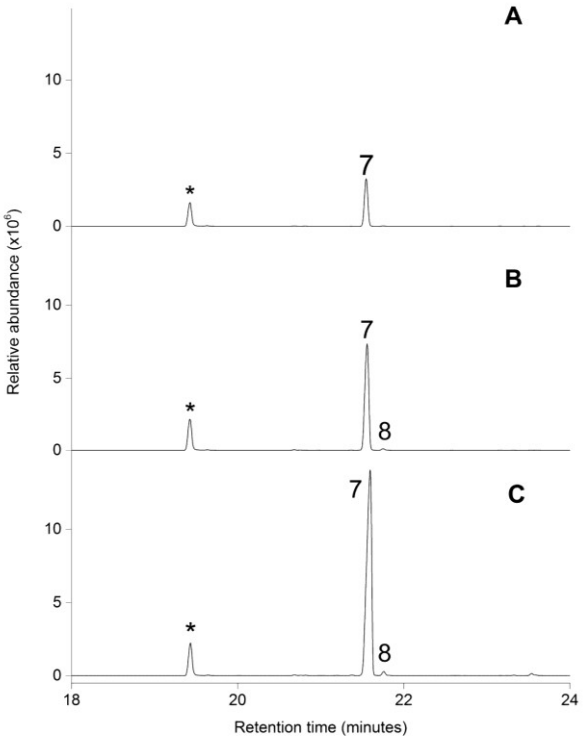


Figure 2-5: Volatile sesquiterpene production by E. coli cultures expressing Stehi1|25180, Stehi1|64702 and Stehi1|73029.
The headspace of *E. coli* cultures expressing A) Stehi1|25180, B) Stehi1|64702 and C) Stehi1|73029 was sampled and analyzed for production of sesquiterpenes by GC/MS. The major product of cells expressing Stehi1|25180, Stehi1|64702 and Stehi1|73029 is Δ-6 protoilludene **7**, a 1,11-cyclization of (2*E*,6*E*)-FPP. The minor peak observed for Stehi1|64702 and Stehi1|73029 was identified as β-elemene **8**, which results from a 1,10-cyclization of (2*E*,6*E*)-FPP. Indole (labeled with an asterisk *), a breakdown product of tryptophan, is naturally produced by *E. coli* and serves as an internal standard.

synthase utilized a 1,10- or a 1,6-ring closure, although conclusive evidence could not be obtained to confirm which mechanism was correct.[150] It is therefore not clear whether Stehi1|128017 follows dual cyclization mechanisms, or whether the putative 1,6-cyclization mechanism is actually a 1,10-ring closure, or *vice versa*. These unanswered questions highlight the versatile nature of the terpene synthases as an enzyme class. Taken together, these data confirm the accuracy of our predictive framework in the selective identification of sesquiterpene synthases that follow the currently accepted 1,6- and 1,10-cyclization mechanisms of (3*R*)-NPP and (3*S*)-NPP.

2.3.5 Predicted 1,11-cyclizing sesquiterpene synthases produce *trans*-humulyl cation derived terpenes

The *trans*-humulyl cation is the precursor to a wide range of pharmaceutically relevant sesquiterpenoids,[102] making the directed discovery of 1,11-cyclizing sesquiterpene synthases particularly relevant. The culture headspace of *E. coli* cells expressing the Clade III members Stehi1|25180, Stehi1|64702 and Stehi1|73029 all the accuracy of our predictive framework (**Figure 2-5**). Identical retention indices and mass fragmentation patterns were obtained for the major peak from the culture headspace of *E. coli* cells expressing the previously characterized Δ-6 protoilludene synthases Omp6 and Omp7 from *O. olearius* [89] (Figure S4). Other minor peaks were also observed in the culture headspace of Stehi1|64702 and Stehi1|73029. One of these peaks was identified as β-elemene **8** (**Figure 2-5**), also produced by Omp3 [89] (Figure S5).

These results support our hypothesis that putative sesquiterpene synthases in Clade III would likely follow a 1,11-cyclization mechanism, producing *trans*-humulyl derived products. Quantum-chemical calculations and isotopic labeling experiments indicate that a 1,11-ring closure of (2*E*,6*E*)-FPP yields a highly reactive *trans*-humulyl carbocation. A 1,2-hydride shift followed by further ring closures results in the protoilludyl cation, which undergoes a final deprotonation, resulting in Δ-6 protoilludene **7** (**Figure 2-6**).[178, 347] Interestingly, the presence of a small amount of β-elemene **8** in the headspace of Stehi1|64702 and Stehi1|73029 expressing cultures suggests that these sesquiterpene synthases are less specific than Stehi1|25180, and the Δ-6 protoilludene synthases Omp6 and Omp7 from *O. olearius*, which display only one single peak when analyzed by GC/MS, both *in vivo* and *in vitro*. Likewise, heterologous expression of the Δ-6 protoilludene synthase Pro1 from *A. gallica* resulted in one peak, suggesting that it is also a highly specific enzyme, although the activity of this enzyme has not yet been fully studied *in vitro*.[162] Stehi1|25180, Stehi1|64702 and Stehi1|73029 sesquiterpene synthases appear to preferentially utilize a 1,11-cyclization mechanism, but Stehi1|64702 and Stehi1|73029 also simultaneously follow a 1,10-cyclization

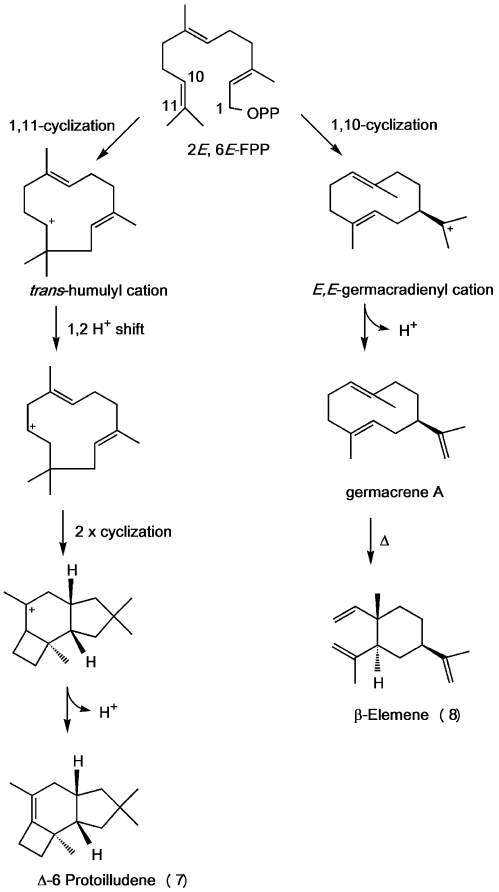


Figure 2-6: Cyclization pathways for the production of Δ-6 protoilludene, germacrene A and β-elemene.
A metal ion mediated dephosphorylation of (2*E*,6*E*)-FPP yields a reactive carbocation, followed by a specific ring closure. A 1,11-cyclization mechanism produces a *trans*-humulyl cation, a precursor to a wide range of sesquiterpenes. The *trans*-humulyl cation can undergo a 1,2 hydride shift followed by two cyclizations and loss of a proton, resulting in Δ-6 protoilludene **7**. A 1,10-cyclization mechanism yields a *E,E*-germacradienyl cation, and loss of a proton leads to germacrene A. β-elemene **8** is the heat induced Cope rearrangement product of germacrene A. Relative stereochemistries of sesquiterpenes are shown.

pathway to yield a small amount of β-elemene **8**, a germacrene A-derived product (**Figure 2-6**), again highlighting the mechanistic versatility of these enzymes.

2.3.6 Kinetic characterization of Stehi1|25180, Stehi1|64702 and Stehi1|73029 highlights differences in catalytic efficiencies

To further understand the cyclization mechanism employed by Stehi1|25180, Stehi1|64702 and Stehi1|73029, the sesquiterpene synthases were purified and their activities were characterized *in vitro*. The enzymes were incubated with (2*E*,6*E*)-FPP under standard assay conditions and the product profiles where determined by GC/MS. Stehi1|25180, Stehi1|64702 and Stehi1|73029 all maintained their preference for a 1,11-cyclization mechanism, producing high levels of Δ-6 protoilludene **7** as their sole sesquiterpene product (**Figure 2-5**). Therefore, the low level of the 1,10-cyclization derived product β-elemene **8**, which was produced during *in vivo* heterologous expression of Stehi1|64702 and Stehi1|73029 must be the result of specific, unknown intracellular conditions. Future studies to determine the conditions that regulate the cyclization mechanism of these Δ-6 protoilludene synthases would provide novel insights into the biochemical mechanisms underlying the product specificity of sesquiterpene synthases.

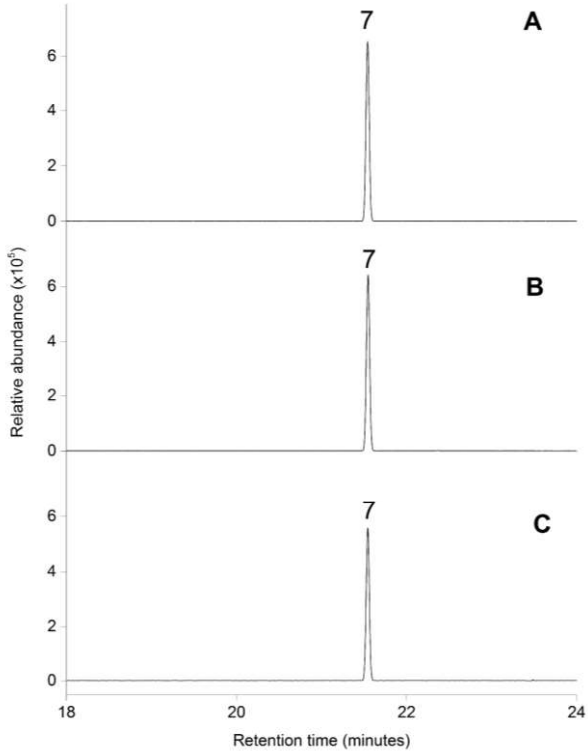


Figure 2-7: *In vitro* activities of purified Stehi1|25180, Stehi1|64702 and Stehi1|73029.
Purified enzymes were incubated with (2*E*,6*E*)-FPP and the headspace of vials was sampled and analyzed for production of sesquiterpenes by GC/MS. A single peak, Δ-6 protoilludene **7** was observed for A) Stehi1|25180, B) Stehi1|64702, and C) Stehi1|73029.

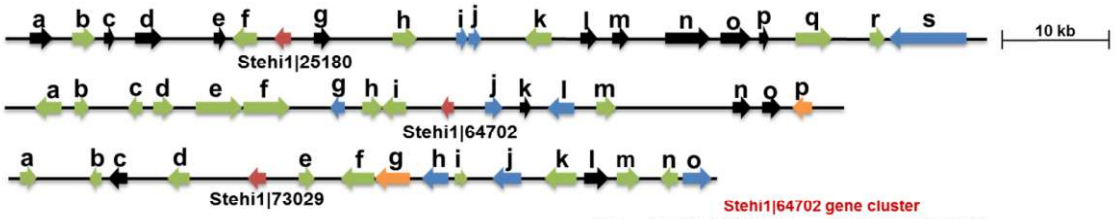
Additionally, the kinetic parameters for purified Stehi1|25180, Stehi1|64702 and Stehi1|73029 were calculated with (2*E*,6*E*)-FPP as a substrate (**Table 2-1**). Catalytic turnover rates are similar between the three enzymes, however their affinities for (2*E*,6*E*)-FPP varied such that the catalytic efficiency of Stehi1|73029 was two-fold greater than that of Stehi1|64702, and four-fold greater than that of Stehi1|25180. These findings emphasize the relative importance of these apparently identical enzymes to the fungal host. Δ-6 protoilludene **7** is the major volatile sesquiterpene product of *S. hirsutum* (**Figure 2-1**). It could be envisioned that *in vivo* the three Δ-6 protoilludene synthases act in a cooperative fashion to maintain sufficient levels of this key intermediate sesquiterpene scaffold, which is a precursor to a wide swathe of bioactive sesquiterpenoids produced by *Stereum* sp. [173-175, 339] (**Figure 2-2**).

Table 2-1: Kinetic parameters determined for Stehi1|25180, Stehi1|64702 and Stehi1|73029. Data were collected using (2*E*,6*E*)-FPP as a substrate in a coupled spectrophotometric assay.			
	K_m [M]	k_{cat} [s^{-1}]	k_{cat}/K_m [$M^{-1} s^{-1}$]
Stehi1|25180	(5.02 ± 0.9) x 10^{-6}	(4.4 ± 0.8) x 10^{-3}	(8.89 ± 0.7) x 10^{2}
Stehi1|64702	(1.91 ± 0.3) x 10^{-6}	(3.7 ± 0.6) x 10^{-3}	(19.51 ± 1.5) x 10^{2}
Stehi1|73029	(1.52 ± 0.2) x 10^{-6}	(6.3 ± 0.7) x 10^{-3}	(41.84 ± 5.1) x 10^{2}

2.3.7 Insights into the production of *trans*-humulyl-derived sesquiterpenoids by Basidiomycota

The majority of sesquiterpenoids isolated from *Stereum* strains are derived from a modified *trans*-humulyl cation scaffold (**Figure 2-2**),[102] indicating that *Stereum* is not only a prolific producer of 1,11-cyclizing sesquiterpene synthases, but also of accessory scaffold modifying enzymes. *Stereum* strains also produce the stereumins, which are derived from a modified δ-cadinene scaffold,[167, 339] and methoxylaricinolic acid, which is derived from a modified drimane scaffold,[169] suggesting that *Stereum* strains have several highly active secondary metabolic pathways. Gene clustering of biosynthetic secondary metabolic pathways is a hallmark characteristic of fungi.[348] It was therefore predicted that the *S. hirsutum* sesquiterpene synthases would be located in gene clusters with the accessory scaffold modifying enzymes that are responsible for the production of the final bioactive sesquiterpenoids.

To identify potential sesquiterpene biosynthetic gene clusters, sequence regions 10-15 kb upstream and downstream of the genes encoding the functionally cloned and expressed sesquiterpene synthases were manually reannotated. The β-barbatene synthase Stehi1|159379 and the δ-cadinene synthase Stehi1|128017 are located in putative biosynthetic gene clusters, containing oxidoreductases, but lacking P450



Stehi1|25180 gene cluster

Gene	Sup. Materials Reference	Top Conserved Domain Hit
a	scaffold4_130k-330k:g19.t1	WD40 domain (cd00200)
b	scaffold4_130k-330k:g20.t1	Dehydrogenase (cd08284)
c	scaffold4_130k-330k:g21.t1	Hydroxylase (PRK06475)
d	scaffold4_130k-330k:g22.t1	Hydroxylase (PRK06475)
e	scaffold4_130k-330k:g23.t1	Peroxidase (pfam01328)
f	scaffold4_130k-330k:g24.t1	HMG-CoA synthase (pfam08540)
Stehi1|25180		
g	scaffold4_130k-330k:g26.t1	ER lumen protein (pfam00810)
h	scaffold4_130k-330k:g27.t1	HMG-CoA synthase (pfam08540)
i	scaffold4_130k-330k:g28.t1	P450 (pfam00067)
j	scaffold4_130k-330k:g28.t1	P450 (pfam00067)
k	scaffold4_130k-330k:g29.t1	Oxidoreductase (pfam13738)
l	scaffold4_130k-330k:g30.t1	No hit
m	scaffold4_130k-330k:g31.t1	Deacetylase (cd10980)
n	scaffold4_130k-330k:g32.t1	Forkhead domain (cd00060)
o	scaffold4_130k-330k:g33.t1	No hit
p	scaffold4_130k-330k:g34.t1	No hit
q	scaffold4_130k-330k:g35.t1	Cu-Oxidase (pfam00394)
r	scaffold4_130k-330k:g36.t1	Cu-Oxidase (pfam07731)
s	scaffold4_130k-330k:g37.t1	P450 (pfam00067)

Stehi1|64702 gene cluster

Gene	Sup. Materials Reference	Top Conserved Domain Hit
a	scaffold15_10k-210k:g9.t1	Aldehyde dehydrogenase (cd07106)
b	scaffold15_10k-210k:g10.t1	Flavoprotein oxygenase (COG1853)
c	scaffold15_10k-210k:g11.t1	Aldehyde reductase (cd05227)
d	scaffold15_10k-210k:g12.t1	O-methyltransferase (pfam00891)
e	scaffold15_10k-210k:g13.t1	Oxidoreductase (pfam05199)
f	scaffold15_10k-210k:g13.t1	Reductase (cd05259)
g	scaffold15_10k-210k:g14.t1	P450 (pfam00067)
h	scaffold15_10k-210k:g15.t1	FAD binding protein (pfam01565)
i	scaffold15_10k-210k:g16.t1	Oxidoreductase (pfam05199)
Stehi1|64702		
j	scaffold15_10k-210k:g18.t1	P450 (pfam00067)
k	scaffold15_10k-210k:g19.t1	No hit
l	scaffold15_10k-210k:g20.t1	P450 (pfam00067)
m	scaffold15_10k-210k:g21.t1	Reductase (cd08249)
n	scaffold15_10k-210k:g22.t1	BTB protein (smart00225)
o	scaffold15_10k-210k:g23.t1	NAD+ binding protein (pfam07993)
p	scaffold15_10k-210k:g24.t1	Multidrug efflux transporter (pfam13536)

Stehi1|73029 gene cluster

Gene	Sup. Materials Reference	Top Conserved Domain Hit
a	scaffold2_2070k-2270k:g16.t1	Glycosyl hydrolase (cd08998)
b	scaffold2_2070k-2270k:g17.t1	Acyl-CoA transferase (COG1804)
c	scaffold2_2070k-2270k:g18.t1	Outer membrane protein (pfam08219)
d	scaffold2_2070k-2270k:g19.t1	Aspartate aminotransferase (cd01494)
Stehi1|73029		
e	scaffold2_2070k-2270k:g21.t1	Aldo-keto reductase (cd06660)
f	scaffold2_2070k-2270k:g22.t1	Oxidoreductase (pfam05199)
g	scaffold2_2070k-2270k:g22.t1	MFS (cd06174)
h	scaffold2_2070k-2270k:g23.t1	P450 (pfam00067)
i	scaffold2_2070k-2270k:g24.t1	Reductase (cd05259)
j	scaffold2_2070k-2270k:g25.t1	P450 (pfam00067)
k	scaffold2_2070k-2270k:g26.t1	Oxidoreductase (pfam05199)
l	scaffold2_2070k-2270k:g27.t1	Hydroxylase (PRK08163)
m	scaffold2_2070k-2270k:g28.t1	FAD binding protein (pfam01565)
n	scaffold2_2070k-2270k:g29.t1	Aldo-keto reductase (cd06660)
o	scaffold2_2070k-2270k:g30.t1	P450 (pfam00067)

Figure 2-8: Biosynthetic gene clusters of 1,11-cyclizing sesquiterpene synthases.
Stehi1|25180, Stehi1|64702 and Stehi1|73029 are located in putative biosynthetic gene clusters, which are manually annotated. Predicted open reading frames are colored according to function. Sesquiterpene synthases are represented in red, P450 monooxygenases are blue, hydrocarbon scaffold modifying enzymes are green, and transporters are orange. The peptide reference numbers are included for each open reading frame (Data S1). Also included in parentheses is the top conserved domain hit at NCBI. Stehi1|25180 is located in a large gene cluster with three putative P450 monooxygenases, one reductase, two HMG-CoA synthases, and two Cu-oxidases. Note that Stehi1|25180 –i and –j appear to have resulted from a gene duplication event (Data S1), and are identical. Stehi1|64702 is also located in a large gene cluster with three putative P450 monoxygenases, five reductases, an aldehyde dehydrogenase and a multidrug efflux transporter. Stehi1|73029 is located in a gene cluster with three putative P450 monoxygenases, four reductases, a FAD binding protein, two transferases, a hydrolase and a multiple drug transporter.

monooxygenases typically associated with the biosynthesis of bioactive natural products (Figure S6). The Stehi1|128017 cluster may be involved in the biosynthesis of the stereumins which are known to be produced by *Stereum* strains and are proposed to be derived from δ-cadinene (**Figure 2-2**).[167, 339, 349] Although no P450 monoxygenases are associated with this cluster, and are most likely required for stereumin biosynthesis, additional scaffold modifying enzymes could be located on satellite clusters known from other fungal natural product biosynthetic pathways such as for trichothecene mycotoxin biosynthesis by *Fusarium graminearum*.[202]

The three Δ-6 protoilludene synthases, Stehi1|25180, Stehi1|64702 and Stehi1|73029 are located in large gene clusters which encode for a range of scaffold modifying enzymes, including several P450 monooxygenases (**Figure 2-8**, Data S1). Stehi1|73029 is located in the smallest biosynthetic gene cluster, with three putative P450 monoxygenases, two oxidoreductases, two aldo-keto reductases, a FAD binding protein, a hydrolase, two transferases and a transporter. Stehi1|64702 biosynthetic gene cluster also contains three P450 monooxygenases, two oxidoreductases, two reductases, an O-methyltransferase, a dehydrogenase and a transporter. Stehi1|25180 is located in the largest biosynthetic gene cluster, with three P450 monoxygenases, although two of these may result from a gene duplication event. Additionally, this gene cluster has fewer oxidoreductases, no transferases, and no transporter (**Figure 2-8**). To date, the Δ-6 protoilludene-derived bioactive sesquiterpenoids that have been isolated from *Stereum* strains are the stereostreins (**Figure 2-2**).[174, 175] However, the identification of three large gene clusters associated with the three functional Δ-6 protoilludene synthases, Stehi1|25180, Stehi1|64702 and Stehi1|73029 would suggest that *Stereum hirsutum* has the potential of producing a number of yet-to-be identified bioactive sesquiterpenoids from the Δ-6 protoilludene scaffold. Biosynthesis of other *trans*-humulyl cation-derived bioactive sesquiterpenoids such as the sterpuranes isolated from *Stereum* sp. [165, 172, 194, 341] and the hirsutenes from *Stereum hirsutum* [168, 170] (**Figure 2-2**) likely involve putative Clade III sesquiterpene synthase members that could not been cloned from cDNA (**Figure 2-3**).

The strained cyclobutyl ring of Δ-6 protoilludene **7** is believed to undergo different secondary rearrangements upon activation by scaffold modifying enzymes such as P450 monooxygenases, yielding the final bioactive sesquiterpenoid.[83, 102, 192, 329] Significantly, all of the *S. hirsutum* Δ-6 protoilludene synthase gene clusters share in common the presence of several P450 monooxygenases. Additionally, the Δ-6 protoilludene-derived sterostreins isolated from *Stereum* sp. have had at least one oxygen atom inserted, typically at the C-8 position of the cyclohexene ring, yielding a ketone (**Figure 2-2**). Likewise, the previously described biosynthetic gene clusters associated with the Δ-6 protoilludene synthases Omp6 and Omp7 both contain at least one P450 monooxygenase, and many of the illudin sesquiterpenoids isolated from *O. olearius* are also oxygenated at this same C-8 position.[89] The similarity of the modifications to the Δ-6 protoilludene scaffold produced by *O. olearius* and *Stereum* strains may represent a common functionality of their sesquiterpene synthase associated P450 monooxygenases, perhaps indicating a common evolutionary ancestor.

In order to establish whether P450 monooxygenases responsible for modifying the Δ-6 protoilludene scaffold could share a common ancestor, a comparative sequence analysis was undertaken. Phylogenetic analyses were carried out by aligning the amino acid sequences of the predicted P450 monooxygenases associated with the

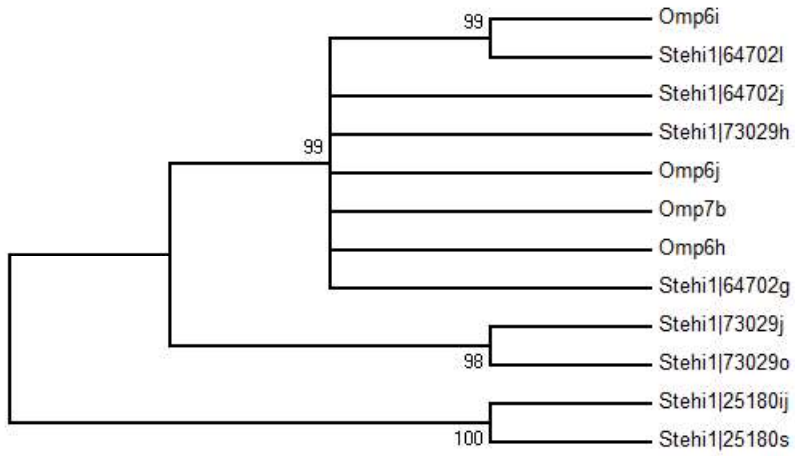


Figure 2-9: Phylogenetic analysis of putative Δ-6 protoilludene-associated P450 monooxygenase homologs.
An unrooted neighbor-joining tree of P450 monoxygenase homologs from the *Omphalotus olearius* (Omp) and *Stereum hirsutum* (Stehi1) Δ-6 protoilludene synthase associated biosynthetic gene clusters. Branches are labeled with their bootstrap values. P450 labeling is consistent with that of Figure 2-6 and previously described biosynthetic gene clusters.[89] (Note that Stehi|25180 -i and -j appear to result from a gene duplication event (Figure 2-8, Data S2-1)).

Δ-6 protoilludene synthases Omp6 and Omp7 [89] with those identified in the Stehi1|25180, Stehi1|64702 and Stehi1|73029 clusters (**Figure 2-8**). Interestingly, the P450 monooxygenases associated with Stehi1|64702 are all closely related to those associated with Omp6 and Omp7. Additionally, one of the P450 monooxygenases from the Stehi1|73029 gene cluster (Stehi1|73029h) is closely related to the Omp6

and Omp7 P450 monooxygenases, while the other two P450 monooxygenases (Stehi1|73029j and Stehi|73029o) are more distantly related. Notably, the Stehi1|25180 P450 monooxygenases are all located on a separate branch, and are not closely associated with other P450 monoxygenases, resulting from a divergence in sequence, and potentially in function (**Figure 2-9**). Given the degree of sequence conservation between the predicted Δ-6 protoilludene synthase associated P450 monoxygenases, as well as the three characterized Δ-6 protoilludene synthases in *S. hirsutum*, it could be hypothesized that their presence results from a process of horizontal gene transfer. However, the question as to whether the apparent manifold nature of the Δ-6 protoilludene synthases and their associated biosynthetic gene clusters indicates a difference in functionality in the production of the different bioactive compounds derived from a modified Δ-6 protoilludene scaffold, known to be produced by *Stereum* sp. (**Figure 2-2**), remains open to speculation.

Knowledge of the sequence conservation of the P450 monoxygenases associated with Δ-6 protoilludene synthases could be used to guide biosynthetic pathway elucidation and in particular, identification of the steps leading to the rearrangement of the cyclobutyl ring of Δ-6 protoilludene that give rise to the diverse humulyl-derived bioactive sesquiterpene scaffolds made by Basidiomycota. In addition, knowledge of key P450 function and sequence conservation could open paths for the engineering of biosynthetic routes to novel bioactive sesquiterpenoids through combinatorial approaches and/or P450 protein engineering.

2.4 Conclusion

This work serves as a proof-of-concept of our predictive framework as an accurate tool for the directed discovery of sesquiterpene synthases based upon cyclization mechanism of choice. Phylogenetic analyses have uncovered several novel sesquiterpene synthases in the Basidiomycete *S. hirsutum*. These bioinformatic studies enabled the selective identification of sesquiterpene synthases which were predicted to follow a 1,6- and a 1,10-cyclization mechanism, producing bisabolyl cation and *Z,E*-germacradienyl cation derived products, respectively, as well as enzymes that utilize a 1,11-cyclization mechanism, producing *trans*-humulyl cation derived scaffolds. Biosynthetic gene clusters encoding for the 1,11-cyclizing sesquiterpene synthases and their accessory modifying enzymes have been described. These sesquiterpene synthases have been cloned, expressed, and their activities characterized. GC/MS analysis confirmed the discovery of a β-barbatene synthase, which follows the predicted 1,6-cyclization mechanism, and a δ-cadinene synthase, which follows the predicted 1,10-cyclization mechanism. Additionally, three novel Δ-6 protoilludene synthases, which follow the predicted 1,11-cyclization mechanism, have been characterizied. The kinetic parameters show these enzymes to be highly active, producing significant amounts of Δ-6 protoilludene **7**. However, Stehi1|64702 and Stehi1|73029 appear to be less specific than other previously described Δ-6 protoilludene synthases, and are capable of simultaneously producing a 1,10-cyclization derived scaffold, β-elemene **8** when expressed in a heterologous host. The discovery and characterization of these Δ-6 protoilludene synthases, which are located in large biosynthetic gene clusters, furthers the development of a biosynthetic toolkit. It is envisioned that this toolkit could be applied in a combinatorial approach to evolve and build robust biocatalysts for the production of specific protoilludene-derived bioactive compounds, such as new anticancer illudin derivatives.[190] At present, many of these compounds are not accessible due to the inhibitory cost and difficulty in chemical syntheses. Producing the compounds biologically is a much more cost effective and efficient option. These results represent an important step for natural product

discovery in fungi, which with the predictive framework described herein, has the potential to become a much more efficient, streamlined, and strategic enterprise.

2.5 Experimental Section

2.5.1 Chemicals and reagents

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO), unless indicated otherwise.

2.5.2 Homolog identification, phylogenetic tree construction and biosynthetic cluster prediction

Sequences for the biochemically characterized sesquiterpene synthases from *C. cinereus* and *O. olearius* [83, 89] were aligned with the putative sesquiterpene synthases from *S. hirsutum* FP-91666 SS1 v1.0 using the gene models associated with this genome, provided by the Joint Genome Institute (JGI) [113]. Alignments were computed using ClustalW [350] and phylogenetic analyses were conducted using MEGA version 5.2 [351] using the default parameters for the Neighbor-Joining method [352] with a bootstrap test of phylogeny (500 replicates). Alignments were manually examined for the presence and proper alignment of the conserved metal-binding motifs characteristic for sesquiterpene synthases.[78] For the identification of putative sesquiterpene biosynthetic gene clusters, each of the scaffolds containing the 18 sesquiterpene synthases from *S. hirsutum* were analyzed for the presence of nearby enzymes implicated in secondary metabolite biosynthesis. A previously described workflow was used for gene identification and splicing predictions.[85] For prediction of the putative function of cluster genes, the top conserved domain hit (CDD) at NCBI was used. For the comparative analysis of P450s, the sequences of Omp6-i, -j, -h, and Omp7-b [89] were used in BLASTp searches of the *S. hirsutum* FP-91666 SS1 v1.0 genome via the JGI server, using an E value threshold of 10. The results were manually scanned 20 kb up/downstream to confirm association with putative sesquiterpene synthases using Augustus.[314] The putative amino acids sequences were aligned and phylogenetic

analyses were carried out using MEGA version 5.2 [351] using the default parameters for the Neighbor-Joining method [352] with a bootstrap test of phylogeny (500 replicates).

2.5.3 Growth of *S. hirsutum* and headspace analysis of volatile compounds

Stereum hirsutum FP-91666 SS1 was obtained as a gift from Dr. Robert Blanchette (University of Minnesota, USA) and was grown on potato dextrose agar (PDA) for 2-3 weeks in the dark prior to inoculating liquid cultures with a agar plug (0.5 cm^2). Liquid cultures (100 mL) were cultivated in foil-sealed flasks (500 mL) for up to 21 days in Rich Medium (20 g/L malt extract, 20 g/L glucose, 5 g/L peptone) at room temperature, in the dark. Volatile terpenoids which had accumulated in the headspace of flasks containing *S. hirsutum* growing on Rich Medium were sampled every 4-5 days for 15 minutes by inserting a 100 µm polydimethylsiloxane fiber (Supelco, Bellfonte, PA) solid phase microextraction fiber (SPME) through the aluminum foil seal of the culture flask, as performed previously.[83, 89] Similarly, following 21 days of growth, the culture supernatant was clarified by centrifugation at 4000 rpm for 10 minutes, and sampled for non-volatile terpenoid accumulation using a 100 µm polydimethylsiloxane-divinylbenzene SPME fiber. Terpenoids were separated and analyzed by GC/MS, described below.

2.5.4 *S. hirsutum* mRNA extraction and cDNA preparation

Mycelial *S. hirsutum* cultivated for 21 days in liquid Rich Medium was dried manually by compression 4-5 times on filter paper, dissected, aliquoted into eppendorf tubes, flash-frozen in liquid N_2, and stored at -80 º C. Frozen tissue was ground into a fine powder using a N_2 cooled, sterile mortar and pestle. mRNA was then extracted using TRIzol® Reagent (Life Technologies, Grand Island, NY) following the manufacturer's procedures. Briefly, powdered tissue was incubated in TRIzol Reagent for 5 minutes, separated by centrifugation from insoluble components, extracted with chloroform, and precipitated with isopropanol. The resulting precipitated nucleotides were washed with 75 % ethanol, and resuspended in DNA/RNA-free water. Single-stranded *S. hirsutum*

cDNA was synthesized using SuperScript™ III First Strand Synthesis System for RT-PCR (Life Technologies, Grand Island, NY) utilizing oligo$(dT)_{20}$ primers for RT-PCR, followed by RNase H treatment.

2.5.5 Cloning of sesquiterpene synthases from *S. hirsutum* cDNA

Following initial phylogenetic analyses to predict sesquiterpene synthases in *S. hirsutum*, gene predictions were further refined by manually re-scanning the genomic sequence of each putative sesquiterpene synthase and 10-15 kb up/downstream of open reading frames using Augustus.[314] Potential sesquiterpene synthase-encoding transcripts were then manually aligned against previously isolated sesquiterpene synthases using MEGA version 5.2 to identify the most likely transcript(s) encoding the functional sesquiterpene synthase. A comparison between initial JGI predictions and cloned ORF and protein sequences are presented in Tables S1 and S2, respectively. Finally, gene specific primers (Table S1) were used to amplify sesquiterpene synthase-encoding genes from previously prepared *S. hirsutum* cDNA. Genes were PCR amplified from *S. hirsutum* cDNA using PfuUltra High Fidelity Polymerase (Agilent Technologies, Inc., Santa Clara, CA). PCR products were cloned using the Zero Blunt® TOPO system (Life Technologies, Grand Island, NY) and were transformed into the supplied MachI™ competent cells. Positive clones were isolated by blue-white colony screening, and plasmids were isolated using the Wizard® *Plus* SV Miniprep kit (Promega, Madison, WI). Expected DNA sequences were confirmed by single pass DNA sequencing (ACGT Inc., Wheeling, IL). The genes were then subcloned into the pUCBB vector [345], to allow constitutive expression of genes in the heterologous host *E. coli*, and into the pET32b(+) vector system (EMD Millipore, Billerica, MA), to allow inducible expression in frame with a thioredoxin tag to improve solubility and a His_6-tag for purification purposes.

2.5.6 Sampling of volatile compounds produced by *E. coli* cultures expressing putative sesquiterpene synthases

Activity of putative sesquiterpene synthases was confirmed by heterologous expression in *E. coli* BL21 (DE3) cells, which were transformed with pUCBB constructs. Single colonies were isolated on Lysogeny Broth (LB) agar plates supplemented with ampicillin (100 µg/mL), and were used to inoculate LB broth (50 mL). Cultures were incubated for 16 hours at 30 ºC with shaking at 220 rpm. The headspace of the cultures was sampled for 10 minutes by SPME followed by GC/MS analysis as described below.

2.5.7 Gas chromatography/mass spectrometry analysis

GC/MS analysis was conducted on an HP GC 7890A coupled to an anion-trap mass spectrometer HP MSD triple axis detector (Agilent Technologies, Santa Clara, CA). Separation of compounds was performed using a HP-5MS capillary column (30 m × 0.25 mm × 1.0 µm) with an injection port temperature of 250 ºC and helium as a carrier gas. The oven temperature started at 60 ºC and was increased 10 ºC min^{-1} to a final oven temperature of 250 ºC. Mass spectra were scanned in the range of 5 – 300 atomic mass units at 1 s intervals. Compounds produced were identified by first calibrating the GC/MS with a C8-C20 alkane mix. Retention indices and mass spectra of compound peaks were compared to reference spectra in MassFinder's (software version 4) terpene library [353] and in the National Institute of Technology (NIST) standard reference database as described previously.[85, 89]

2.5.8 Expression and purification of Stehi1|25180, Stehi1|64702 and Stehi1|73029

Rosetta (DE3) pLysS cells were transformed with pET32b (+) constructs of Stehi1|25180, Stehi1|64702 and Stehi1|73029. Single colonies were isolated on LB agar supplemented with ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL), and were used to inoculate LB starter cultures (50 mL). Large-scale expression was carried out in LB (1000 mL). Cultures were incubated at 30 ºC with shaking at 220 rpm until an OD_{600} 0.6 was reached. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM), and cells were incubated for a further 3 hours at

30 ºC. Cells were harvested by centrifugation, lysed by sonication and soluble proteins were separated from the cell slurry by centrifugation. The His_6-tagged proteins were bound to a HisTrap™ FF column (GE Healthcare, Pittsburgh, PA) in Buffer A (50 mM Tris.HCl, 100 mM NaCl, 5 mM imidazole (pH 7.5)), and were eluted in Buffer B (50 mM Tris.HCl, 100 mM NaCl, 250 mM imidazole (pH 7.5)). Following overnight cleavage of the thioredoxin tag by enterokinase (EMD Millipore, Billerica, MA), the tags were separated from the sesquiterpene synthase by passage over a HisTrap™ FF column. The proteins were purified to homogeneity by passage over a S200 10/300 GL size exclusion column preequilibrated with Buffer C (50 mM Tris.HCl pH 7.5, 100 mM NaCl, 1 mM β-mercaptoethanol, 10 % (v/v) glycerol).

2.5.7 Enzyme characterization of purified Stehi1|25180, Stehi1|64702 and Stehi1|73029

In vitro activity assays with purified sesquiterpene synthases were performed in Buffer D (20 mM Tris-HCl, 200 mM NaCl, 10 mM $MgCl_2$ (pH 8.0)) in a final reaction volume of 100 µl. Sesquiterpene synthases (10 µg) were incubated with (2*E*,6*E*)-FPP (2 µM) in a sealed glass vial, and reactions were carried out at 21 ºC for 16 hours. The headspace of the reaction vessel was sampled for 10 min by SPME followed by GC/MS analysis as described above. The kinetic parameters of Stehi1|25180, Stehi1|64702 and Stehi1|73029 were determined using previously published protocols.[89, 158] Briefly, sesquiterpene synthases were incubated with (2*E*,6*E*)-FPP at a range of concentrations (1-100 µM). Enzyme activity was monitored using the PiPer Pyrophosphate™ enzyme coupled spectrophotometric assay (Invitrogen, Carlsbad, CA), which measures the release of pyrophosphate (PPi). Control reactions without enzyme and without substrate were also carried out, and all reactions were performed in triplicate. The kinetic parameters were determined by nonlinear regression analysis of the data in Origin 9.0.

2.6 Supplementary Information

Figure S2-1: Volatile sesquiteprene production by *Stereum hirsutum* after 21 days of cultivation in rich liquid medium

The headspace of a liquid culture was sampled and analyzed for the production of volatile sesquiterpenes by GC/MS. Identified sesquiterpene products are numbered and their relative stereochemical structures are shown. One peak (labeled with a question mark ?) which has a *m/z* 204 could not be identified. Peaks labeled with an asterisk * are likely modified terpenoids, each with a *m/z* of 218-222 consistent with hydroxylation. Hirsutene (retention time: 28.51 minutes) appears as a shoulder of the larger Δ-6 protoilludene peak (28.61 minutes); both terpenoids were clearly distinguished based on their MS fragmentation pattern (inset). The retention indices of these products are consistent with retention indices in MassFinder.[353]

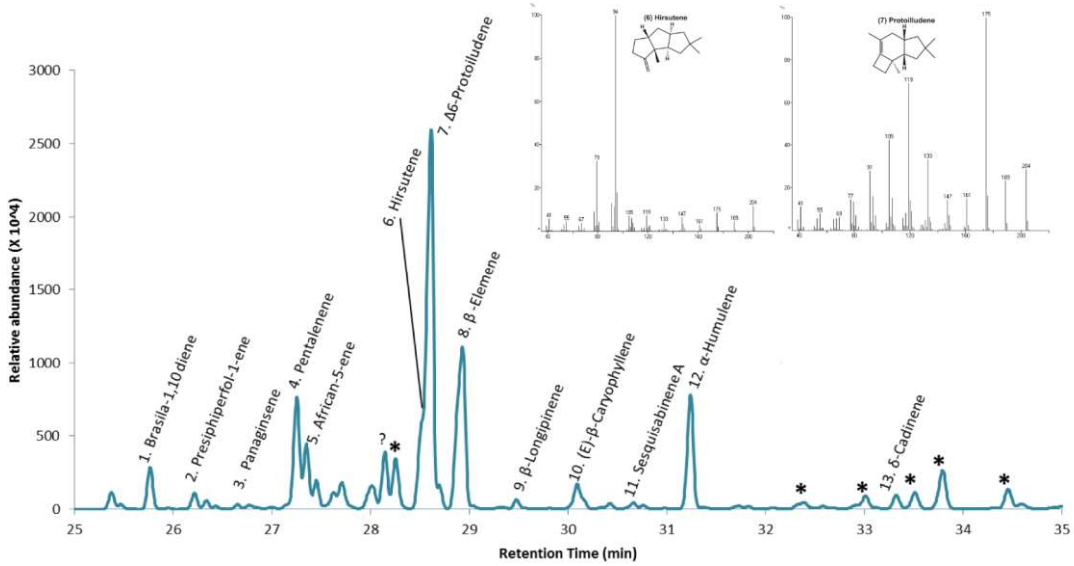


Figure S2-2: Volatile sesquiterpene production by E. coli cultures expressing Cop6, Omp9 and Stehi1|159379.

The headspace of cultures was sampled and analyzed for production of sesquiterpenes by GC/MS. Indole *, a breakdown product of tryptophan, is naturally produced by *E. coli* and acts as an internal standard. Only peaks that shared comparable retention indices with the previously characterized Cop6 and Omp9 have been labeled. Cells expressing Omp9 accumulated α-barbatene (A) and β-barbatene (B), while those expressing Cop6 produced α-cuprenene (C).

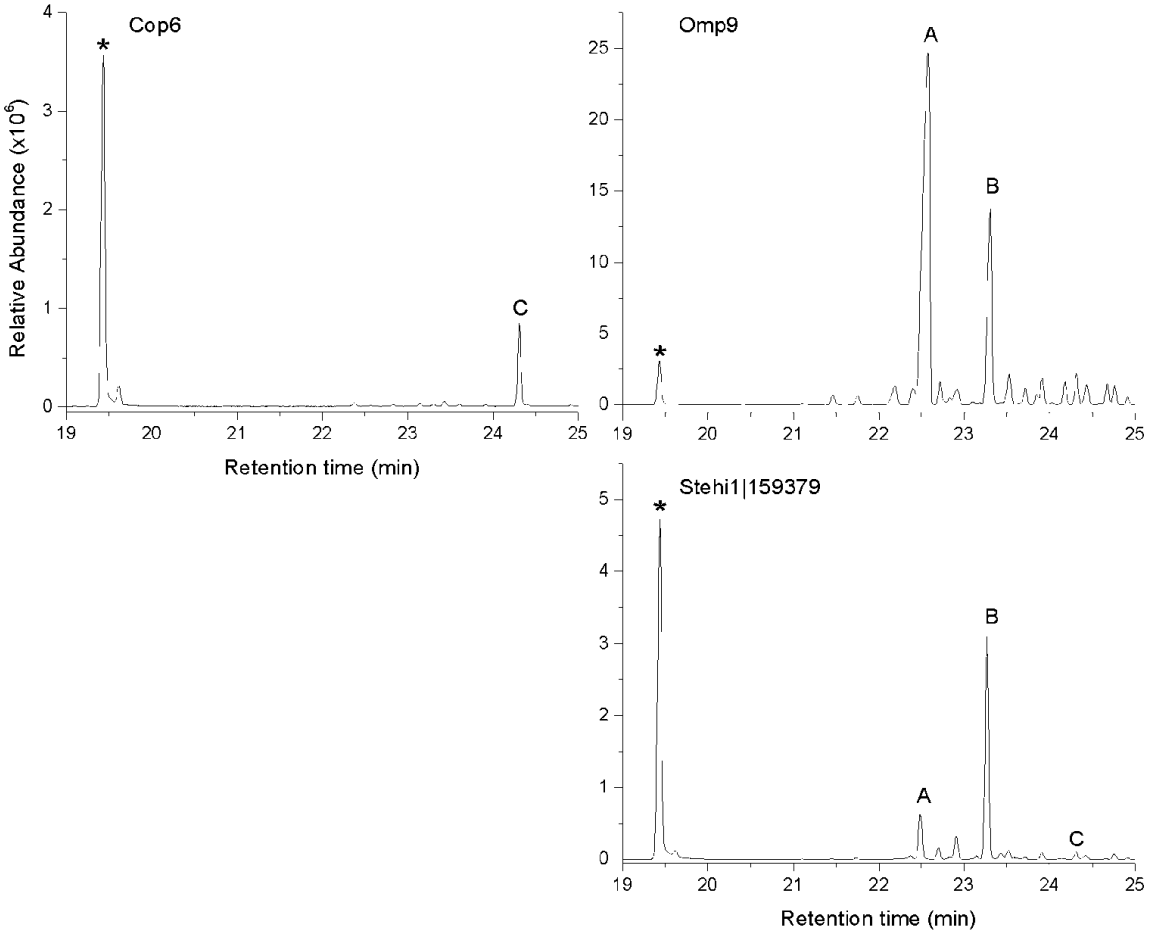


Figure S2-3: Volatile sesquiterpene production by *E. coli* cultures expressing Omp3, Omp4, Omp5a, Cop3, Cop4, and Stehi1|128017.
The headspace of cultures was sampled and analyzed for production of sesquiterpenes by GC/MS. Indole *, a breakdown product of tryptophan, is naturally produced by *E. coli* and acts as an internal standard. Only those peaks which were identified in the headspace of cells expressing Stehi1|128017 have been labeled in the Omp and Cop headspaces. Cells expressing each sesquiterpene synthase produced δ-cadinene (**K**) (corresponding to **13** in Figures 1 and 3). Cop4 also produced the β-cubebene (**D**), sativene (**E**), β-copaene (**F**), and germacrene D (**I**). Additionally, Omp3 and Cop3 produced α-muurolene (**J**). Cop3 produced γ-muurolene (**H**) and germacrene D (**I**) as well. Finally, sesquisabinene A (**G**) (corresponding to **11** in Figures 1 and 3) was identified in the headspace of cells expressing Stehi1|128017 based upon comparative retention indices with those reported in MassFinder. [353]

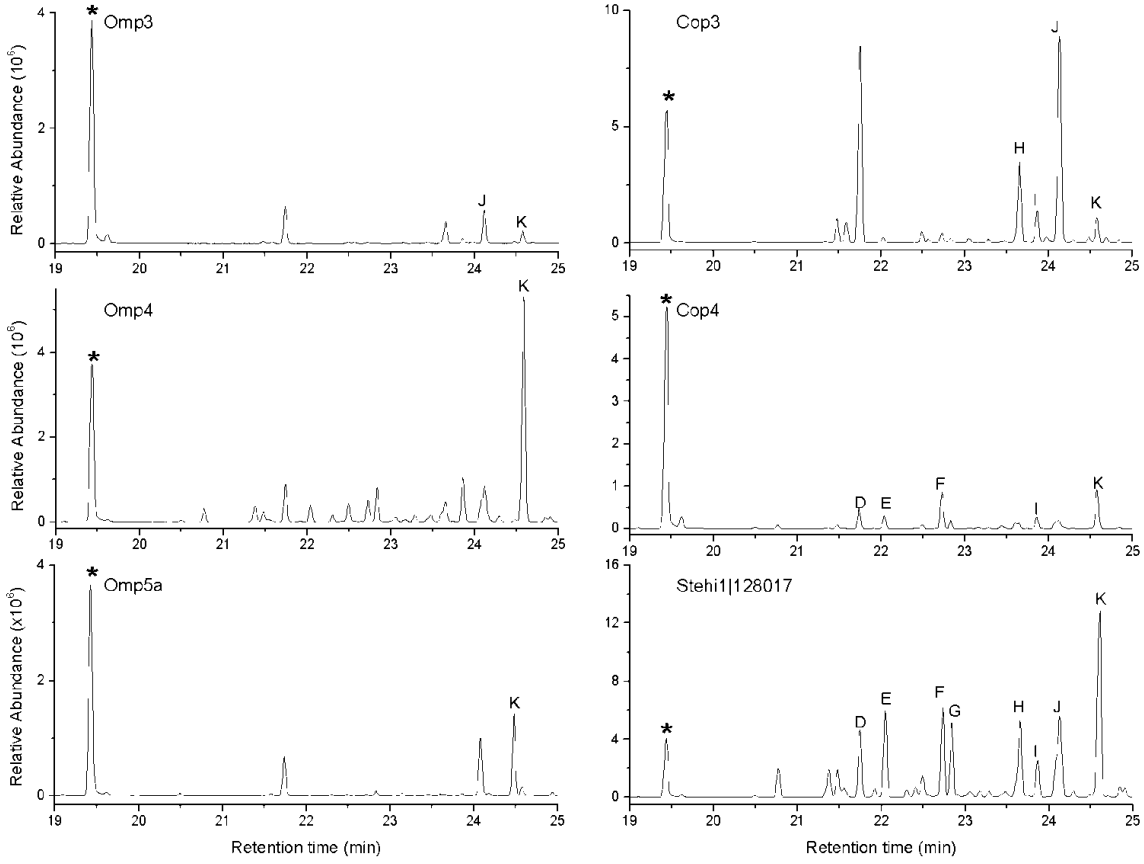


Figure S2-4: Volatile sesquiterpene production by *E. coli* cultures expressing Omp6 and Omp7.

The headspace of cultures was sampled and analyzed for production of sesquiterpenes by GC/MS. *E. coli* cells were constitutively expressing A. Omp6 and B. Omp7. Indole *****, a breakdown product of tryptophan, is naturally produced by *E. coli* and acts as an internal standard. Cells expressing Omp6 and Omp7 produce Δ-6 protoilludene **7** as their major product.

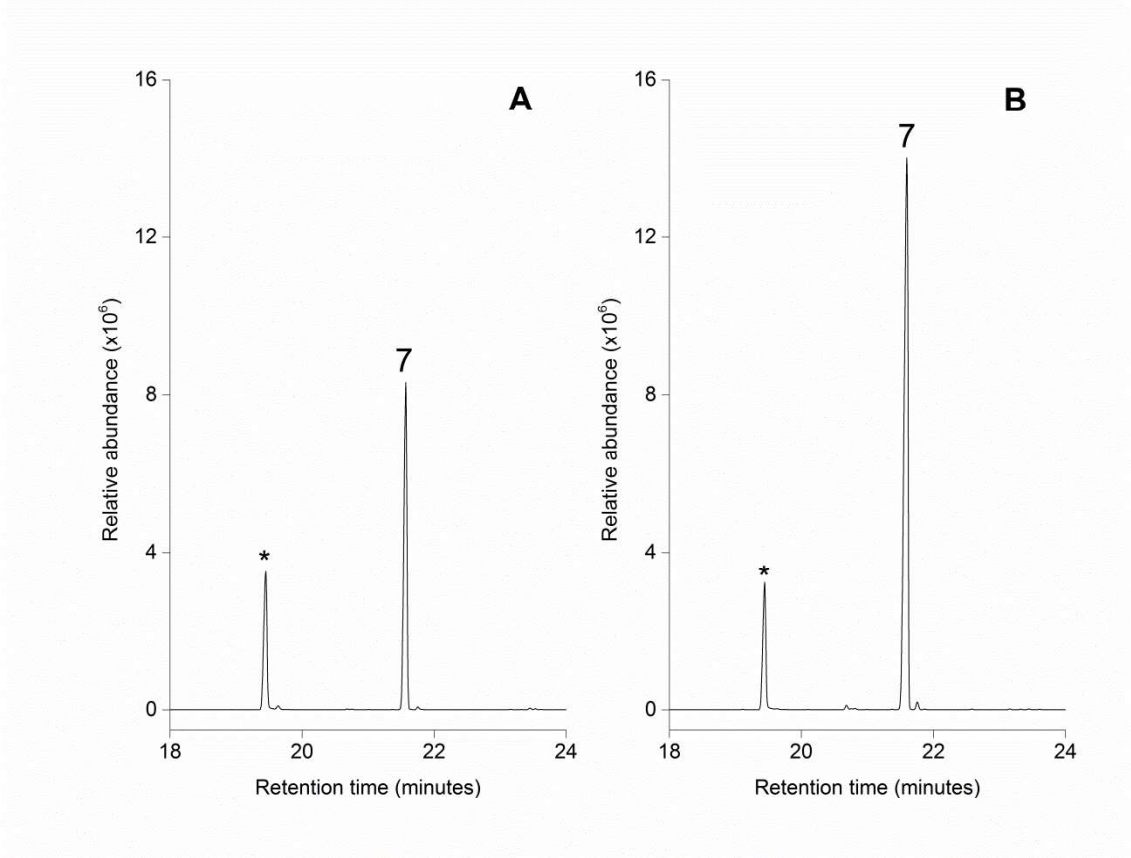


Figure S2-5: Volatile sesquiterpene production by *E. coli* cultures expressing Omp3. The headspace of cultures was sampled and analyzed for production of sesquiterpenes by GC/MS. *E. coli* cells were constitutively expressing Omp3. Indole *****, a breakdown product of tryptophan, is naturally produced by *E. coli* and acts as an internal standard. Cells expressing Omp3 produce a range of compounds, including β-elemene **8**.

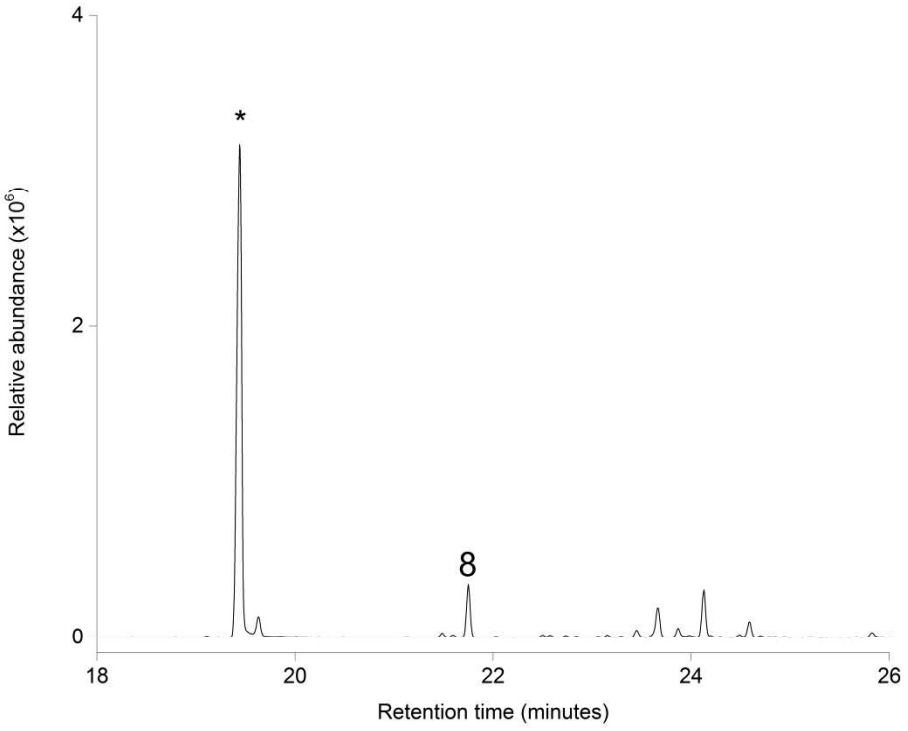
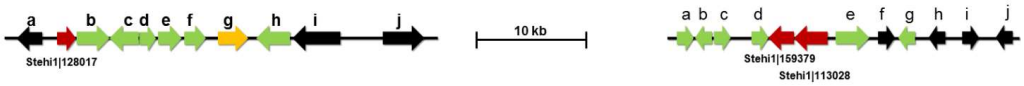


Figure S2-6: Putative biosynthetic gene clusters surrounding Stehi1|128017 and Stehi1|159379.

Manual reannotation of 10-15kb upstream and downstream of Stehi1|128017 and Stehi1|159379 revealed the presence of potential scaffold modifying enzymes such as reductases and hydrolases. However, unlike previously described biosynthetic gene clusters associated with active sesquiterpene synthases [83, 89], neither Stehi1|128017 nor Stehi1|159379 is clustered with a P450 monooxygenase. The peptide reference numbers are included for each open reading frame (Data S1). Also included in parentheses is the top conserved domain hit at NCBI. Coloring of genes according to function is consistent with the scheme used in Figure 6.



Scaffold 2: Stehi1|128017

Gene	Sup. Materials Reference	Top Conserved Domain Hit
a	scaffold2_1428k-1528k:g10.t1	WD40 domain (cd00200)
Stehi1|128017		
b	scaffold2_1428k-1528k:g12.t1	Nitrite reductase/dioxygenase (COG2146)
c	scaffold2_1428k-1528k:g13.t1	Cys/Met PLP-dependent enzyme (pfam01053)
d	scaffold2_1428k-1528k:g14.t1	HIT zinc finger (pfam04438)
e	scaffold2_1428k-1528k:g15.t1	DRE-TIM metallolyase (cd03174)
f	scaffold2_1428k-1528k:g16.t1	Crotonase/Enoyl-CoA hydratase (cd06558)
g	scaffold2_1428k-1528k:g17.t1	MFS transporter (TIGR00879)
h	scaffold2_1428k-1528k:g18.t1	Serine/Threonine kinase (cd07829)
I	scaffold2_1428k-1528k:g19.t1	PHD-finger (pfam00628)
j	scaffold2_1428k-1528k:g20.t1	Unknown (DUF1295)

Scaffold 12: Stehi1|159379

Gene	Sup. Materials Reference	Top Conserved Domain Hit
a	scaffold12_216k-316k:g11.t1	Delta-9 FA desaturase (cd03505)
b	scaffold12_216k-316k:g12.t1	Glycoside hydrolase (cd11577)
c	scaffold12_216k-316k:g13.t1	Glycosyl hydrolase (pfam01670)
d	scaffold12_216k-316k:g14.t1	Glycosyl hydrolase (pfam01670)
Stehi1|159379		
Stehi1|113028		
e	scaffold12_216k-316k:g17.t1	Glycosyl hydrolase (pfam01670)
f	scaffold12_216k-316k:g18.t1	Aldo-keto reductase (cd06660)
g	scaffold12_216k-316k:g19.t1	RING-finger (cd00162)
h	scaffold12_216k-316k:g20.t1	Acetyltransferase (COG0456)
i	scaffold12_216k-316k:g21.t1	Caspase domain (pfam00656)
j	scaffold12_216k-316k:g22.t1	BEACH domain (cd06071)

Table S2-1: Cloned sesquiterpene synthase ORFs, cloning primers, and comparison with JGI predictions.

Sequences in bold highlight differences between JGI predictions and cloned, active sesquiterpene synthases. Cloning primer binding sites are underlined. The JGI gene sequence predictions for Stehi1|25180 diverged significantly from the cloned gene. Cloned Stehi1|64702 contained an additional exon, and cloned Stehi1|159379 was shorter than the predicted sequence. Stehi1|128017 and Stehi1|73029 were predicted correctly.

Gene	JGI putative ORF	Cloned ORF
Stehi1|159379	**ATGGCACATCCAACAACCAATCCCTCTCATAGACTTG AACAA**ATGGAGTCCGTCAGGGAACACATTCCTCGCCT TCAGCACTTCCTCGGCGAGATCGGCTATCGTCACACC ACTCCTCCTGCACCGACGCTCGATTTTCTTCATGCGCA TCACCACTGGATACACCATGTACTGGGCCCCATGACT TCATGGACGGTGGCCAAACTAAACGCACTCGAAGACT CATCCTCGACCATTTTCGAGAGAGCATACCCGTTGTCC GACGCCGAGATGAAGTTCGTCCTTGCAAAGCTCACGG CCATAGCCATTTTCCTCGACGATTCTCTCGAAGATGAG GAGACCTATGACGATATCGGCAACTTCGCACATCGTG TCTACCTGGGCGAGGCCCAGCCGACCGGGGTCCTTA CCCTCTACCACCAAGGCATTCAAGAGCTTTCCAAGATG CACGAAGGCGACGCGGTCTTTAGAGGCCTCGCTGTC GCACCTTGGATTACCTTCATCGACGCCTGCATGTTAGA GAAGAGACTTCTGACCTTCGACTCGAAGCTACGCGTC AGTCCTCGCGATCTGGGATATCAGCGCCTACGAAACA GCACAGATTTTACGTCCCTCCGAGCTCCGAAGGCTAC TCCAAGTGAAGTAGAGGTTAGCTTTCCTATCTTCCTTC GCCACAAGAGTGGAATAGGAGAAGCCTATGCCGCTGC GATCTTCAAATCGTCTAGATACCAGGAGTTGCCTCTAT CAAGATTCGTCAAGTCCATGCCAGACATGATATATTAT ATCGAGCTCGTGAATGACCTGATGTCCTTCTACAAGGA GCAACTGGCCGGTGAAACAGCCAATCTGATCCATCTC CAGCATCAGTCGTGGAAGGGAGGTCAGGGAACAGGG CCATATGGATCGTGGACGTTACTCGATACTTTCAGCCG TCTGTGTGATGAGACGAGGGATGCTGCCTTTAGGGTT GACGAGCTCCTCAGACTGGATGAATGCGAAAAGATAG CGAACGGGGAGTTGAGGGGTGAAGAGGTTGGTCTCT CGCCGATGGACGTAACGATGGCGGCGCAGTGGCGTG AGTTCAGAGACGGATACGTATCTTGGCATCTGGAGTG CCAACGGTACAAGTTGGATTTCATAAAGCTTAGCACGT TTGAGTAA	ATGGAGTCCGTCAGGGAACACATTCCTCGCCTTCAGC ACTTCCTCGGCGAGATCGGCTATCGTCACACCACTCC TCCTGCACCGACGCTCGATTTTCTTCATGCGCATCAC CACTGGATACACCATGTACTGGGCCCCATGACTTCAT GGACGGTGGCCAAACTAAACGCACTCGAAGACTCATC CTCGACCATTTTCGAGAGAGCATACCCGTTGTCCGAC GCCGAGATGAAGTTCGTCCTTGCAAAGCTCACGGCCA TAGCCATTTTCCTCGACGATTCTCTCGAAGATGAGGA GACCTATGACGATATCGGCAACTTCGCACATCGTGTC TACCTGGGCGAGGCCCAGCCGACCGGGGTCCTTACC CTCTACCACCAAGGCATTCAAGAGCTTTCCAAGATGC ACGAAGGCGACGCGGTCTTTAGAGGCCTCGCTGTCG CACCTTGGATTACCTTCATCGACGCCTGCATGTTAGA GAAGAGACTTCTGACCTTCGACTCGAAGCTACGCGTC AGTCCTCGCGATCTGGGATATCAGCGCCTACGAAACA GCACAGATTTTACGTCCCTCCGAGCTCCGAAGGCTAC TCCAAGTGAAGTAGAGGTTAGCTTTCCTATCTTCCTTC GCCACAAGAGTGGAATAGGAGAAGCCTATGCCGCTG CGATCTTCAAATCGTCTAGATACCAGGAGTTGCCTCTA TCAAGATTCGTCAAGTCCATGCCAGACATGATATATTA TATCGAGCTCGTGAATGACCTGATGTCCTTCTACAAG GAGCAACTGGCCGGTGAAACAGCCAATCTGATCCATC TCCAGCATCAGTCGTGGAAGGGAGGTCAGGGAACAG GGCCATATGGATCGTGGACGTTACTCGATACTTTCAG CCGTCTGTGTGATGAGACGAGGGATGCTGCCTTTAGG GTTGACGAGCTCCTCAGACTGGATGAATGCGAAAAGA TAGCGAACGGGGAGTTGAGGGGTGAAGAGGTTGGTC TCTCGCCGATGGACGTAACGATGGCGGCGCAGTGGC GTGAGTTCAGAGACGGATACGTATCTTGGCATCTGGA GTGCCAACGGTACAAGTTGGATTTCATAAAGCTTAGC ACGTTTGAGTAA

Stehi1|128017	ATGGCTACTTCAAATCCTCCCTCCACCCCTCATCACGA GTCATTAATCCTCCCTGACCTGCTTTCTCTGTCGACAC CTTTCAATGGCTCGACAAACCCGCACTGGGCCATGGC CGCCCCAGAATCTTCAGCCTGGGTCTCTAGTTACAAC CTCTTCTCAGACCGGAAGCGGACTGACTTCATCACTG GTTCGAACGAACTCCTCGTCTCACACACATATCCTCAT GCCGATTATGATGCCTTTCGAACATGTTGTGACTTTGT AAACTTGCTGTTCGTCATTGACGAGATTAGCGACGATC AGAGCGGAAAGGCCGCGAGACGGACAGGGGAGGTGT ATCTGAACGCGATGCGCGATCCTGAATGGACAGATGG CTCTGACTTGGCGAAGATGACACAACAATTTCGCGCG CGTTTCTTGAGGTCGGTCGGACCTCAATCCTTCCGTC GCTTCCTGAGGCACAGCGAAGACTACATCGACTGCGT GGCTAAGGAGGCAGAATATCGCGAACGGGGACAAGT GCTCGACATGGATTCCTTCAAGTCCCTCAGGAGGGAG AACTCAGCGATCCGGTTGTGCTTCGGCCTGTTCGAGT TCACGTTAGGGATCGATCTTCCGGATTCAGTGTTTGAG GATGAGACGTTCATGAAGATGTACTGGGCCTCGGCGG ACATGGTGTGTTGGGCAAATGACGTATACTCGTATAAC GTGGAACAAGCTAAGGGTCACAGTGGGAATAACATCG TTACCGTCCTGATGGCTGCGAGAGACATTGACATGCA GGCTGCGAGCGACTACGTCGGTGAGTACTATGCGGA GTTAATGGAGGAGTACATGACGGCCAAGGCGGAACTG GCGTCGAAGTCGTTCGGGTCGAGGGATCTGGATGAG GACGTTTGGAAATATGTGAACGCGATGGAGAACTGGC CGATTGGGAATTTGGAGTGGTCGTTCAAAACGAACAG GTACTTCGGCACTCTACACGACGAAGTGAAACGGACG CGATTGGTCGTGATCAAGCCTCGGAAAGTCGTTGTTTA G	ATGGCTACTTCAAATCCTCCCTCCACCCCTCATCACG AGTCATTAATCCTCCCTGACCTGCTTTCTCTGTCGACA CCTTTCAATGGCTCGACAAACCCGCACTGGGCCATGG CCGCCCCAGAATCTTCAGCCTGGGTCTCTAGTTACAA CCTCTTCTCAGACCGGAAGCGGACTGACTTCATCACT GGTTCGAACGAACTCCTCGTCTCACACACATATCCTC ATGCCGATTATGATGCCTTTCGAACATGTTGTGACTTT GTAAACTTGCTGTTCGTCATTGACGAGATTAGCGACG ATCAGAGCGGAAAGGCCGCGAGACGGACAGGGGAG GTGTATCTGAACGCGATGCGCGATCCTGAATGGACAG ATGGCTCTGACTTGGCGAAGATGACACAACAATTTCG CGCGCGTTTCTTGAGGTCGGTCGGACCTCAATCCTTC CGTCGCTTCCTGAGGCACAGCGAAGACTACATCGACT GCGTGGCTAAGGAGGCAGAATATCGCGAACGGGGAC AAGTGCTCGACATGGATTCCTTCAAGTCCCTCAGGAG GGAGAACTCAGCGATCCGGTTGTGCTTCGGCCTGTTC GAGTTCACGTTAGGGATCGATCTTCCGGATTCAGTGT TTGAGGATGAGACGTTCATGAAGATGTACTGGGCCTC GGCGGACATGGTGTGTTGGGCAAATGACGTATACTCG TATAACGTGGAACAAGCTAAGGGTCACAGTGGGAATA ACATCGTTACCGTCCTGATGGCTGCGAGAGACATTGA CATGCAGGCTGCGAGCGACTACGTCGGTGAGTACTAT GCGGAGTTAATGGAGGAGTACATGACGGCCAAGGCG GAACTGGCGTCGAAGTCGTTCGGGTCGAGGGATCTG GATGAGGACGTTTGGAAATATGTGAACGCGATGGAGA ACTGGCCGATTGGGAATTTGGAGTGGTCGTTCAAAAC GAACAGGTACTTCGGCACTCTACACGACGAAGTGAAA CGGACGCGATTGGTCGTGATCAAGCCTCGGAAAGTC GTTGTTTAG
Stehi1|25180	CCTCGCCACATCAATCCTCATTATCAGGAAGTGAAGAA GGCTTCCGCTGCCTGGGCCGAAAGCTTCGGTGGTTCA ACCCAAAGGCTCAACACGCGTATAATGCCTGCGATTT CAAACGTCTGCGTACCGGGTGCGACCTCATGAATATG TTCTTCGTATTCGACGAATACTCCGACGTCTCTTCACC AAAGGACGTTATCCAACAAGCTGCCATCATAATGGATG CTCTGCGCAACCCATACGCGCCTCGTCCTGACGATGA ATGGGTCGGCGGAGAAGTTACCAGACAATTCTGGAAG CGTGCCATCAAGACCGCCACCGCAGGGGCACAAAGA CGGTTTATCGATGCGTTCGAAAGCTATACGCAGTCAGT CGTTCAACAAGCGAAGGATCGCCACCACGGGTTCATT CGTGACGTCGACAGCTATCTCGAGATGCGAAGAGAGA CGATCGGCGCGAAGCCGTCGTTCGTCGTCCTCCAGAT GGACATGACCCTCCCAGACGAGGTTCTCGCTCACCCG GTCATCCAACAGCTGTCTGCCCTGTCCACTGATATGAT ATGTCTTGGAAAC**AGACGGCTCATTCAGATACTCTGG ACCGTACAG**GACATCTGCTCTTATAATGTCGAGCAGG CTCGCGGCGACGACCTCCACAACATCATCACGATAGC GATGAACCAGTTTGATATCGATATCGCCGGTGCAATG GATTGGGTTGTGAAGTATCACGCGAAACTCGAGCGAA AGTTCCTCTACCTTTACAATAACGGTCTTCCATCATGG GGCAAGGAGCTAGACCCGCAGGTGGAGCGGTACGTC TGTGGATTA	**ATGACAGTCGTGGACAGCCC**ACAACGTTTCTATATC **CCCAATTGCCTCGAGTACTGGCCCTGG**CCTCGCCAC ATCAATCCTCATTATCAGGAAGTGAAGAAGGCTTCCG CTGCCTGGGCCGAAAGCTTCGGTGCGTTCAACCCAAA GGCTCAACACGCGTATAATGCCTGCGATTTCAA**CTTG TTGGCTTCGCTCGCATACCCCCTCGAGTCTGAAGA**A CGTCTGCGTACCGGGTGCGACCTCATGAATATGTTCT TCGTATTCGACGAATACTCCGACGTCTCTTCACCAAA GGACGTTATCCAACAAGCTGCCATCATAATGGATGCT CTGCGCAACCCATACGCGCCTCGTCCTGACGATGAAT GGGTCGGCGGAGAAGTTACCAGACAATTCTGGAAGC GTGCCATCAAGACCGCCACCGCAGGGGCACAAAGAC GGTTTATCGATGCGTTCGAAAGCTATACGCAGTCAGT CGTTCAACAAGCGAAGGATCGCCACCACGGGTTCATT CGTGACGTCGACAGCTATCTCGAGATGCGAAGAGAG ACGATCGGCGCGAAGCCGTCGTTCGTCGTCCTCCAG ATGGACATGACCCTCCCAGACGAGGTTCTCGCTCACC CGGTCATCCAACAGCTGTCTGCCCTGTCCACTGATAT GATATGTCTTGGAAACGACATCTGCTCTTATAATGTCG AGCAGGCTCGCGGCGACGACCTCCACAACATCATCA CGATAGCGATGAACCAGTTTGATATCGATATCGCCGG TGCAATGGATTGGGTTGTGAAGTATCACGCGAAACTC GAGCGAAAGTTCCTCTACCTTTACAATAACGGTCTTCC ATCATGGGGCAAGGAGCTAGACCCGCAGGTGGAGCG GTACGTCTGTGGATTA**GGAAACTGGGTTCGCGCCAG TGACCAGTGGGGGTTCGAAAGCGAGCGATACTTTGG CAAAAAGGGGAAGGAGATTTTCAAGAGGAGGTGGG TGAACTTGATGCAGCCGGAGAGAGCGCAGGACATC GGTCCGACATTGGTTGATGGGACCAGATTATGA**

Stehi1|64702	ATGGTTCGCTCTCCCGTTTCCGATAAGTTCTGCATACC GGACACGCTAGCCAGCTGGCCTTACCCTCGTATACTC AACCCCCACTATGCCGAGGAGAAAGCGGCGTCCGCT GCCTGGACGAAGGGCTTTGGCGCTTTTGGCCCGAAAG CGCAGGATGCGTTCGATCGTTGCGACTTTAAGAGATG CCGCAGCGGCTGCGATCTGATGAACCTGTTCTTTGTG ATCGACGAGCATTCAGACACACATGGCGAGGAGACAG TACGCAAGATGAAGGACGTAGTCATGGACGCCATCAG AAATCCCCACAAGCCCCGTCCAAATGACGAGTGGATT GGAGGCGAAATTGCACGACAATTCTGGGAACGTGCAA TGTGTTATGCTAGTGAGATCTCCCAGCGACGCTTCATC GACACCTTCGATGAATACTTGGAATCCGTCGTAGACCA GGCCGCAGACAGGGACAGTGCAAGGATCCGTGATAT CGAGAGCTACATTAACATTCGTCGCAACACTATCGGA GCGAAGCCCTCGTTCGTCATCATGGAGCAAGGCATGG ACATCCCGGACAACGTCTTCGAGAACGAGGTGTTTCA GAGACTTCGCATGGCCACGATAGACATGCTTTGCCTT GGAAATGACATCGTATCATACAACATTGAACAAGCTCG AGGTGATGACTCCCATAACATCGTGAGAATTGTCATGA ACGAGCTCGATACCGATGTCCCTCGCGCTATGGATTG GGTAGCTCAGAGACATACGCAACTCGAGCGCGAGTTC TTCACCGCGCTGAGCGAGCTGCCCACTTGGGGAGAG CCGATCGATGGATGGGTGAAGGAGTATGTCTATGGTC TGGGGAACTGGGTGCGCGCTAACGACCAGTGGAGTTT CGAGAGCCAAAGGTATTTTGGAACCAAGGGCATGGAG ATCATGAAGTCAAGGTGGCTTTCGGTTTTGCCCAAGGT CCGCCCTGCTGAGGTCGGTCCACAACTTGTTGACCAA TCTCTCTTATGA	ATGGTTCGCTCTCCCGTTTCCGATAAGTTCTGCATACC GGACACGCTAGCCAGCTGGCCTTACCCTCGTATACTC AACCCCCACTATGCCGAGGAGAAAGCGGCGTCCGCT GCCTGGACGAAGGGCTTTGGCGCTTTTGGCCCGAAA GCGCAGGATGCGTTCGATCGTTGCGACTTTAA**TTTGC** **TCGCCTGTTTGGCCTACCCGATCGCGACTCCAGA**GA GATGCCGCAGCGGCTGCGATCTGATGAACCTGTTCTT TGTGATCGACGAGCATTCAGACACACATGGCGAGGA GACAGTACGCAAGATGAAGGACGTAGTCATGGACGC CATCAGAAATCCCCACAAGCCCCGTCCAAATGACGAG TGGATTGGAGGCGAAATTGCACGACAATTCTGGGAAC GTGCAATGTGTTATGCTAGTGAGATCTCCCAGCGACG CTTCATCGACACCTTCGATGAATACTTGGAATCCGTC GTAGACCAGGCCGCAGACAGGGACAGTGCAAGGATC CGTGATATCGAGAGCTACATTAACATTCGTCGCAACA CTATCGGAGCGAAGCCCTCGTTCGTCATCATGGAGCA AGGCATGGACATCCCGGACAACGTCTTCGAGAACGA GGTGTTTCAGAGACTTCGCATGGCCACGATAGACATG CTTTGCCTTGGAAATGACATCGTATCATACAACATTGA ACAAGCTCGAGGTGATGACTCCCATAACATCGTGAGA ATTGTCATGAACGAGCTCGATACCGATGTCCCTCGCG CTATGGATTGGGTAGCTCAGAGACATACGCAACTCGA GCGCGAGTTCTTCACCGCGCTGAGCGAGCTGCCCAC TTGGGGAGAGCCGATCGATGGATGGGTGAAGGAGTA TGTCTATGGTCTGGGGAACTGGGTGCGCGCTAACGA CCAGTGGAGTTTCGAGAGCCAAAGGTATTTTGGAACC AAGGGCATGGAGATCATGAAGTCAAGGTGGCTTTCGG TTTTGCCCAAGGTCCGCCCTGCTGAGGTCGGTCCACA ACTTGTTGACCAATCTCTCTTATGA
Stehi1|73029	ATGGCTGTCGCTACCTCTGTTGCCACTCCTGTTCCTAC TCCCGCCTACTCTGCTGGCCGCGCTCCGGCCAAAGAG AAGAAGATCTATCTTCCCGACACACTCGCTGAGTGGCC TTGGCCTCGCGCCATCAACCCTCACTATGCTGAAGCG AAGGAAGAGTCTCAAGCATGGGCCGCAAGTTTCAATG CTTTCAGCCCGAAGGCTCAGCACGCCTTCAACCGCTG CGACTTCAACCTTCTTGCATCCCTCGCTTACCCGCTCG CCACTAAACATGGATGCCGCTCAGGCTGTGATCTCATG AACCTTTTCTTTGTCATCGACGAGTACTCTGATATCGCT CCCGTCGAGGAGGTCCGCCAGCAAAAGGACATCGTCA TGGACGCTCTCCGGAACCCACACAAACCACGTCCCGA GGGTGAATGGGTGGGCGGTGAAGTCGCACGGCAATT CTGGGCTCTGACCATCACCAACGCTAGCGCTCAGTCG CAGAAGCACTTCATCGAGACTTTCGACGAGTACCTCGA CTCCGTTGTTCAGCAGGCTGAAGACAGGTCAGAGTCA CGGATCCGCGACATTCAGAGTTATATCGACGTCCGCC GCAACACAATTGGTGCCAAGCCATCATTTGCCCTCCTC GAGCTCGATATGGACCTTCCCGACGAGGTTCTCGCCC ATCCCACTATCCAATCACTTTCCCTTGCCACCATCGAC ATGCTTTGTCTTGGCAACGACATTGTCTCGTATAACCT CGAGCAAGCTCGTGGTGATGCCAGCCACAACATCATC ACCATCGTCATGAACGAGCTCAACCTCGACGTCAACG GTGCCATGCGATGGGTAGGCGACTTCCACAAGCAGTT GGAGAAGCAATTCTTCGAGGCGTTCAACAACCTTCCCA AATGGGGCAACGCGGAGCTCGACGCTCAGATTGCAGT GTACTGCGACGGACTGGGGAACTGGGTTCGTGCAAAC GACCAGTGGAGTTTCGAGAGCGAGCGCTACTTCGGGG CTAGGGGTCTTGAGATTATGGAGACGAAGACCCTGGC GATGATGCCTATTCAGAGGACGGAGGCATTGGGCCCT CAGCTTGTTGATGACTCCATTTTGTGA	ATGGCTGTCGCTACCTCTGTTGCCACTCCTGTTCCTAC TCCCGCCTACTCTGCTGGCCGCGCTCCGGCCAAAGA GAAGAAGATCTATCTTCCCGACACACTCGCTGAGTGG CCTTGGCCTCGCGCCATCAACCCTCACTATGCTGAAG CGAAGGAAGAGTCTCAAGCATGGGCCGCAAGTTTCAA TGCTTTCAGCCCGAAGGCTCAGCACGCCTTCAACCGC TGCGACTTCAACCTTCTTGCATCCCTCGCTTACCCGCT CGCCACTAAACATGGATGCCGCTCAGGCTGTGATCTC ATGAACCTTTTCTTTGTCATCGACGAGTACTCTGATAT CGCTCCCGTCGAGGAGGTCCGCCAGCAAAAGGACAT CGTCATGGACGCTCTCCGGAACCCACACAAACCACGT CCCGAGGGTGAATGGGTGGGCGGTGAAGTCGCACGG CAATTCTGGGCTCTGACCATCACCAACGCTAGCGCTC AGTCGCAGAAGCACTTCATCGAGACTTTCGACGAGTA CCTCGACTCCGTTGTTCAGCAGGCTGAAGACAGGTCA GAGTCACGGATCCGCGACATTCAGAGTTATATCGACG TCCGCCGCAACACAATTGGTGCCAAGCCATCATTTGC CCTCCTCGAGCTCGATATGGACCTTCCCGACGAGGTT CTCGCCCATCCCACTATCCAATCACTTTCCCTTGCCAC CATCGACATGCTTTGTCTTGGCAACGACATTGTCTCGT ATAACCTCGAGCAAGCTCGTGGTGATGCCAGCCACAA CATCATCACCATCGTCATGAACGAGCTCAACCTCGAC GTCAACGGTGCCATGCGATGGGTAGGCGACTTCCACA AGCAGTTGGAGAAGCAATTCTTCGAGGCGTTCAACAA CCTTCCCAAATGGGGCAACGCGGAGCTCGACGCTCA GATTGCAGTGTACTGCGACGGACTGGGGAACTGGGTT CGTGCAAACGACCAGTGGAGTTTCGAGAGCGAGCGC TACTTCGGGGCTAGGGGTCTTGAGATTATGGAGACGA AGACCCTGGCGATGATGCCTATTCAGAGGACGGAGG CATTGGGCCCTCAGCTTGTTGATGACTCCATTTTGTGA

Table S2-2: Comparison of putative protein sequences and cloned, active sesquiterpene synthases.

Sequences in bold highlight differences between JGI predictions and of functional, cloned sesquiterpene synthases. There was no difference between the predicted and cloned Stehi1|128017 and Stehi1|73029.

Gene	JGI putative protein sequence	Cloned protein sequence
Stehi1|159379	**MAHPTTNPSHRLEQ**MESVREHIPRLQHFLGEIGYRHTTPPAPTLDFLHAHHHWIHHVLGPMTSWTVAKLNALEDSSSTIFERAYPLSDAEMKFVLAKLTAIAIFLDDSLEDEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNSTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELVNDLMSFYKEQLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDETRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE	MESVREHIPRLQHFLGEIGYRHTTPPAPTLDFLHAHHHWIHHVLGPMTSWTVAKLNALEDSSSTIFERAYPLSDAEMKFVLAKLTAIAIFLDDSLEDEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNSTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELVNDLMSFYKEQLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDETRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE	
Stehi1|128017	MATSNPPSTPHHESLILPDLLSLSTPFNGSTNPHWAMAAPESSAWVSSYNLFSDRKRTDFITGSNELLVSHTYPHADYDAFRTCCDFVNLLFVIDEISDDQSGKAARRTGEVYLNAMRDPEWTDGSDLAKMTQQFRARFLRSVGPQSFRRFLRHSEDYIDCVAKEAEYRERGQVLDMDSFKSLRRENSAIRLCFGLFEFTLGIDLPDSVFEDETFMKMYWASADMVCWANDVYSYNVEQAKGHSGNNIVTVLMAARDIDMQAASDYVGEYYAELMEEYMTAKAELASKSFGSRDLDEDVWKYVNAMENWPIGNLEWSFKTNRYFGTLHDEVKRTRLVVIKPRKVVV	MATSNPPSTPHHESLILPDLLSLSTPFNGSTNPHWAMAAPESSAWVSSYNLFSDRKRTDFITGSNELLVSHTYPHADYDAFRTCCDFVNLLFVIDEISDDQSGKAARRTGEVYLNAMRDPEWTDGSDLAKMTQQFRARFLRSVGPQSFRRFLRHSEDYIDCVAKEAEYRERGQVLDMDSFKSLRRENSAIRLCFGLFEFTLGIDLPDSVFEDETFMKMYWASADMVCWANDVYSYNVEQAKGHSGNNIVTVLMAARDIDMQAASDYVGEYYAELMEEYMTAKAELASKSFGSRDLDEDVWKYVNAMENWPIGNLEWSFKTNRYFGTLHDEVKRTRLVVIKPRKVVV
Stehi1|251780	PRHINPHYQEVKKASAAWAESFGAFNPKAQHAYNACDFKRLRTGCDLMNMFFVFDEYSDVSSPKDVIQQAAIIMDALRNPYAPRPDDEWVGGEVTRQFWKRAIKTATAGAQRRFIDAFESYTQSVVQQAKDRHHGFIRDVDSYLEMRRETIGAKPSFVVLQMDMTLPDEVLAHPVIQQLSALSTDMICLGN**RRLIQILWTVQ**DICSYNVEQARGDDLHNIITIAMNQFDIDIAGAMDWVVKYHAKLERKFLYLYNNGLPSWGKELDPQVERYVCGL	**MTVVDSPQRFYIPNCLEYWPW**PRHINPHYQEVKKASAAWAESFGAFNPKAQHAYNACDF**NLLASLAYPLESEE**RLRTGCDLMNMFFVFDEYSDVSSPKDVIQQAAIIMDALRNPYAPRPDDEWVGGEVTRQFWKRAIKTATAGAQRRFIDAFESYTQSVVQQAKDRHHGFIRDVDSYLEMRRETIGAKPSFVVLQMDMTLPDEVLAHPVIQQLSALSTDMICLGNDICSYNVEQARGDDLHNIITIAMNQFDIDIAGAMDWVVKYHAKLERKFLYLYNNGLPSWGKELDPQVERYVCGL**GNWVRASDQWGFESERYFGKKGKEIFKRRWVNLMQPERAQDIGPTLVDGTRL**
Stehi1|647022	MVRSPVSDKFCIPDTLASWPYPRILNPHYAEEKAASAAWTKGFGAFGPKAQDAFDRCDFKRCRSGCDLMNLFFVIDEHSDTHGEETVRKMKDVVMDAIRNPHKPRPNDEWIGGEIARQFWERAMCYASEISQRRFIDTFDEYLESVVDQAADRDSARIRDIESYINIRRNTIGAKPSFVIMEQGMDIPDNVFENEVFQRLRMATIDMLCLGNDIVSYNIEQARGDDSHNIVRIVMNELDTDVPRAMDWVAQRHTQLEREFFTALSELPTWGEPIDGWVKEYVYGLGNWVRANDQWSFESQRYFGTKGMEIMKSRWLSVLPKVRPAEVGPQLVDQSLL	MVRSPVSDKFCIPDTLASWPYPRILNPHYAEEKAASAAWTKGFGAFGPKAQDAFDRCDF**NLLACLAYPIATPE**RCRSGCDLMNLFFVIDEHSDTHGEETVRKMKDVVMDAIRNPHKPRPNDEWIGGEIARQFWERAMCYASEISQRRFIDTFDEYLESVVDQAADRDSARIRDIESYINIRRNTIGAKPSFVIMEQGMDIPDNVFENEVFQRLRMATIDMLCLGNDIVSYNIEQARGDDSHNIVRIVMNELDTDVPRAMDWVAQRHTQLEREFFTALSELPTWGEPIDGWVKEYVYGLGNWVRANDQWSFESQRYFGTKGMEIMKSRWLSVLPKVRPAEVGPQLVDQSLL
Stehi1|73029	MAVATSVATPVPTPAYSAGRAPAKEKKIYLPDTLAEWPWPRAINPHYAEAKEESQAWAASFNAFSPKAQHAFNRCDFNLLASLAYPLATKHGCRSGCDLMNLFFVIDEYSDIAPVEEVRQQKDIVMDALRNPHKPRPEGEWVGGEVARQFWALTITNASAQSQKHFIETFDEYLDSVVQQAEDRSESRIRDIQSYIDVRRNTIGAKPSFALLELDMDLPDEVLAHPTIQSLSLATIDMLCLGNDIVSYNLEQARGDASHNIITIVMNELNLDVNGAMRWVGDFHKQLEKQFFEAFNNLPKWGNAELDAQIAVYCDGLGNWVRANDQWSFESERYFGARGLEIMETKTLAMMPIQRTEALGPQLVDDSIL	MAVATSVATPVPTPAYSAGRAPAKEKKIYLPDTLAEWPWPRAINPHYAEAKEESQAWAASFNAFSPKAQHAFNRCDFNLLASLAYPLATKHGCRSGCDLMNLFFVIDEYSDIAPVEEVRQQKDIVMDALRNPHKPRPEGEWVGGEVARQFWALTITNASAQSQKHFIETFDEYLDSVVQQAEDRSESRIRDIQSYIDVRRNTIGAKPSFALLELDMDLPDEVLAHPTIQSLSLATIDMLCLGNDIVSYNLEQARGDASHNIITIVMNELNLDVNGAMRWVGDFHKQLEKQFFEAFNNLPKWGNAELDAQIAVYCDGLGNWVRANDQWSFESERYFGARGLEIMETKTLAMMPIQRTEALGPQLVDDSIL

Table S2-3: Peptide reference numbers and corresponding predicted peptide sequence for biosynthetic gene clusters.

Peptide sequences from Augustus are given for each predicted open reading frame in the biosynthetic gene clusters surrounding Stehi1|25180, Stehi1|64702, Stehi1|73029, Stehi1|159379 and Stehi1|128017.

scaffold4_130k-330k:g19.t1	MSNPGGPVNPPPNSSLVKLPPTPGGPGLPPPNAVPYGPPPPPGMITNSGPPPPSSVAPPPPPPPQQPQPTASIS GPPSASQGDGPGPSSAVGAGGALVPISGSTNGFPDDLDPQNVPPELKKEGSDWFAVFNPKVKRVLDVNLVHTLM HESTLIDRDANKTGDLYIRSVCFSPDGKYLATGAEDKQIRIWDIKTQRIRNIFDGHQQEIYSLDFSRDGRLIVSGSGD KTARIWDMQDGSSKTLTIHEPEAPDSGVTSVAISPDGRLVAAGSLDTIVRIWDVQTGQLVERLKGHKDSVYSVAFT PDGKGLVSGSLDKTLKYWDVRPMLAMAGKMVPGPGSTPNGKKEGGEKSNCMMNFTGHKDYVLSVAVSHDGQ WVVSGSKDRGVQFWDARSAIVQLMLQGHKNSGMPFTANRAATFVLIRVFVSVISIDLSPAGSVLATGSGDWQARI SDSFWLIA
scaffold4_130k-330k:g20.t1	MDSDLDLTLPETHLAVRWYPPSYDIRLERVPTPNGHEDVDELHVCGHEFIGYVVALGSSFTSSSPKSRPALYGTLK VGDKVVSPFTTSCGECRRCRLGFTARCESSLLFGSPRLPGAQAQYARIPHAGGTLFALPTDTQSTLSNPSLSKLTA LPDPTLILLADILPTGLSVALNALMHPKLAPILHGKAWPGEAAGLLSGSSEGIEQEGAEGTAPVRDEDRVLTVGVVG LGPVGLCAFLSLLDLLSNPANLATSALRFRLIAVDPNAARRAKLSTIIASLPTELKGSSRGEFEVCGIDEAKGVLERP GWGGGCDAVLEVGRSVKFPPIFLLLAFASDNAKNVNRELKNTSSVYRSSAIPRH
scaffold4_130k-330k:g21.t1	MSSADVLVVGAGPSGITLALALAQNGMSVRIVEKLAQMPDSQRGTGSQPRTMELYKFLGVLDDILDGGRSLVQIK QYDGEGNYIKTVSLTTGDMSPTPTNPFRQPFCVGQEVVGRIMREHLKKHNVDVEFSSEIISLQDQEKFV
scaffold4_130k-330k:g22.t1	MSSADILMVGAGPSGVTLALALAQNGVSVRIIEKLAQIPDSQRGAGSQPRTMEIYKFLGVIDDILDGGRPSLKVKQH DSEGNYIKTFDFTTGDTSPTPANPFRQPFFVGQEVVGRIMREHLKKRGVFVEFSSELVSLQDQEKYVQVGIKKAG GAIETAEFKYVVGSDGARGFVRKSIGLSLLGETRDTTEWLIGDLEIRGVDDLKEYMNVYGHAPGNMVSIAPTKRAL TENVFYTVLMGPDLDIAKALEDPRIIFNFIRNITKKPELEFGEIKSLSNYRPNNTRVVNTFGKGHVFVTGDAAHVHSP AGGQGLNSGIMDSINLAWKLALVTKGLAPVSLLDTYTEERLPVIKDMLVRTDGLLRKALSQNEERFSRPPVFNMLG VNYCFSSIVLDEFHRIEDAKGLAASAYEVDDILRAGARAPDAPKLITIPEIDHGEMALFDIFRPFVHTVLIFPASDAHV DTSSVITALAKYPQGSVSSVIVLPQYSSVSALDGISSGLVLRDEDGHAYKAYGVEVGQAKIVIVRPDGWIGAIVRGV GGVEKYFSAIMAA
scaffold4_130k-330k:g23.t1	MPGIVPHSHAEGTCPVTGATSHAYCAPQPDDKRAPCPALNTLANHGYLPRDGKQIDAAVLIRALEEGYDISAPLAY VLAYGGHFLLGQFGPFQLDDLARHNKIEHNASLIHPDARGRDEYAPIHSDPVLWKELVKNASDGKKLTTADVAEAR VRRETDEMKAGKEALDGFHAEIARGEMAIALGMFSKDGEDGIPMGTLKQWMLEERLPEGWKPTHIQGLLNTVAK AREIKGAMEKIRASEDAKEKAGESHLKLVQNFKL
scaffold4_130k-330k:g24.t1	MAIEIASTSVFDVQAERPRPKDVGILAMDMYVPKRCVSEEELEVFDGVSKGKYTIGLGQSYMAFVDDREDINSIALT VVSSLMKRFDIDPRSIGRLDVGTETIIDKSKSVKTVLMDLFKECGNHDIEGLDSKNACYGSTAAVQNATNWIESRS WDGRNAIVFTGDIAVYAKGGARAVGGAGACAILIGPNAPLVFEPIHGTYMANTYEFYKPNLHSEYPTVDGPLSVST YIAAMDAAYSAFRSKVATATKSFNPSAIPETEAGNKVDEKSVFSLADVDYPVYHSPYGKQVQKGHARLLFMDYLS APHKPKFANIDPSFLDLPYPQTINSKEIEKSFMALSKADYAARVSLSMRCASRCGNMYTASLYGGLASLLSTVDAA TLKGKRISMYAFGGGCASTFWTMRVKGDVSEIAKKMDLLKRLESMDVVSCEEYTEGLQLREDNHNLAPYQPQGD IKNIWPRAFYLDGIGSMFRRKYVQA
scaffold4_130k-330k:g26.t1	MSQATAESFAAMIEAGMSGIAYLAGILQYPAIWYRIYKYGADGISVPSQLLLALSYLTLYFLPLLGPHPSFLPYIPSLA FAFFFIPFPLYTLYLLHTRSRAIFPYTLPSSEKSQCYTISAICLGLALIFNYEWFLWGGRWGESGGGRRGTGAWGV GWLYEILWAFGTYLGAVGVWPQLAEYEREGRAELMRDRWLMCYLGLLYLSAGMTVLRSCLGAVNIHWDPLAFYA SLFTTVLFTYYLIKIFVFRSTSPAFSLPSFTSSSSSPSGRGRSWSARNWSWPWRRSEYVALSGEEEEPLYDGHGED SEEQDRIRASSAAAAAQARLAAVSDLF
scaffold4_130k-330k:g27.t1	MAIEAASTSAAPSVFDVQAERPRPKDVGILAMDMYFPKRCISEEDLEVYDGVAKGKYTIGLGQSYMAFADDREDIN SIALTVVSSLMKRFDIDPRSIGRLDVGTETIIDKSKSVKTVLMDLFKECGNHDIEGLDSKNACYGSTAAVQNATNWIE SRSWDGRNAIVFAGDIAVYAEGGARPVGGAGACAILIGPNAPLVFEPIHGTYMANTYDFYKPNLTSEYPTVDGPLS VSTYIQAMDAAYSSFRSKVASGTKALNPSAIPSSTESGKKVDEKSVFSLADVDYPVYHSPYGKQVQKGHARLLFM DYLSAPSKPEFANIDAGFLDIPYPKTLGSKEVEKAFMALSKSDYASRVAPSMRCASRVKGDVSEIAKKMDLLKRLE SMEVVSCDEYTKGLQLREENHNAAPYQPQGDIKNIWPGAFYLDNIDSMFRRKYIQA

scaffold4_130k-330k:g28.t1	MAFSLLDLLGIPLSPSRIVGFSIGALLLRWVYLRFTAISIADIPGPGPESFWLGNLRQLMGEQAVEFDKKWQDQFGG IVRFKAPLAVGLPDPLAPMPLGLASSDLSYEQEDCLLISDPKAIQYILQTSRYRFVKPYGTRFLPNMATGKGVHGAE GADHHCQRRILLVAFGSPESQALMPIFKASAQSLVQLLRDSLLLNQGSKTIDMSTILGRATLDSLGKAAFEYDFGAL HRKDNELADSLHHFTTILGAGYETITKSLVYILCELAKHSMEQGKLREELQRARADARSEGREVTTTEMDELPYFN AVIKETLRFDTVVPHLSRKATQDDVLPLSKPISATSGKVMKEIPIRKGLRIIISDVAYNFNKDIWGEDADVWRPERWL DGTFNAASVDANVGVFGNL
scaffold4_130k-330k:g29.t1	MSFDAIKIAETWLSKFSRSVQDADLPALVSTFLSNGWLRDNLIFTWDTRSLCGPEAITAYLSTTLSKAQLSNFNIVDA PYLSPTYGPLTFATVGVSFAFTFDTPIAHGRGYARLLQHNDEWQALSVFTKMSDLKGHEEQGAELGSWGNHTLS FESVMRERRAQIESEPCALIVGAGQTGLMAAARFKQMGIRTIVIEKNARVGDNWRERYPTLTLHTPKTHHAFLYHT FPSTWPTYTPRDKLSDFLEFYATAQELVVWTNSTLLPGPTYDTKTKKWFIIIDRHGTEVRINPSHIVLAAGAVGPPHI PNVPSIDSFAGDVLHASKFRGAAPYKGKRVVVIGAGNSSADICQDCCFEGASSITMVQRSSTTIMKNEAILDLIMVS FPEGVPTDVLDFKDSSIPWGLRREMLLEIKGAIAEAHREMHDGLRSRGLDVNLSDGTGTFLQFHSKYGGYWQDV GTAQLIIDGKVGVKHGVEVSRCTEKTVVFSDGSELEADVLIFATGYQSSLEAYKKIFDPDLISQTPHLWGFDDEMEV RGVWRATSHPGLWYAGGDFGICRTGSKNLALLIKATELGLK
scaffold4_130k-330k:g30.t1	MIVQVFYYYTRYPQDKLSLKLFARTMQLITLAIAHLCTQVALLLVADTANSIFDMWWIYDVLVNNWGNEDALLVAN WRNTDPAITGIIAALVQGFFCWRIWSAIAAGILKEWTLFPKGTFILVLWLAGGAVADIGITIVITWHLRRHSANFTQQS DPLNRVDHLKRIDHNTGQLG
scaffold4_130k-330k:g31.t1	MSTTADNLAAEYGPRDLKGYGQHTPDPKWPNGAKIAISIVLNYEEGSEVTGGEDGVTEMSGSEKGPGMVPMRGA RDVNTESLYEYGSRAGVWRVLRLFKEYNVHCTSYAVGQALERHPEVAIALEQDGHELASHGYRWVDRTSWTIEE EAEYARKAINAISSTSPSGTPPRGWYYGKAFTKNDARSRSLIARVFKEENIPLLYYSDYYGDDLPYWIPYPGGEKD QGLLIVPYALDTNDFRNTRHDAFISPNDFSDYLIAAFDELYREGVAGQAKMLSIGLHGRIIGRPGRIAGLRKFLEYVK EKGGVWFATREEIAKHWIETFPYEAKT
scaffold4_130k-330k:g32.t1	MPEFGMNGQFDLDRQHPFLTAANDDDDPLNDGAASMPAPYAPNQMQQPLPALYLYPLNDSFLPKHIHLPQGPRV KIGRQTNVKTAPGERNGFFDSKVLSRQHAEVWEENGKSSTVHVLRQIYIKDVKSSNGTFINGERLSPEGAESDPF ELKTDDIVEFGIDIVGEDNKTIIHHKVAARVVCVFTEQDAQSAARAEQHQNPPLANLSSQPNGSGSAFNFAGSQGP GPGQSQRRPAMQQQALAGMGGMGGNMRPPGKSGLTFDHILNRLQGELQKTRETGAELHSLTGAMSDIHETLGG NLSLPPVMPPQPSPPQATDAHPNGHNVNASSSSSALAELQAQLQETQMSLATHMDKIRSLEGMLAEHEAIKREVS TLREMMEDEKRELEMVRGRRRSNTVRADDGDGDGEEGEFVGGHHDDDDDARSISTVTPHELERVEEEDEEQLA AEQEEEERRRQQREGEESMTRPRTPEPTGMGIRMDDEYDGGRGSRSPSPPQRDRHAMSNARENDMTDRLDHL STQLESALELSRALQVSHATAQTTISLLESKVTSLESLVQASQTQVAEVQATVTETREEERKSLSEMVKEWKKGVE GQWSGVREEWREERERLNKAREEWERRVVGVEAGVTRVGDKVEKAVASLAQQQQQASQNGNARPEVGLVTP PSPRSLSSDSGKQRSRRKRSTASRGRPRSNSAGSKKEKIGDADADVDAEGYTSSSEGATLLNDGTDDKASSIHR HLQGSRTRSPWMHDDSEDEDADRRFGGKGVRKSSTGSMTTMGTEEDGLQGPAPADKDMVPMNSLARQNMQY STAVGVIVLSVAAAAVLWRVKAE
scaffold4_130k-330k:g33.t1	MVASSLLPGACILFAAVAVAGAHGLSDVELGLGYSVSDMPMSRPRSLMASLKRDKMDRLAKRDDSFARWCGKHY QVGSPADSEEVYAGLFPVPSTSDTPLLDFRCSPAIQPYLHDETEGSFIIDANVTYDVGTSFTASSNDSSSSSDVLNV MIQVTNSSTTLISNITVPVGSTGTIIPFPLSSIPPSTSPYTISCSAVLDSDATQTFAASSKVYNLPPNPYIADATNTTSV ATKIDRRSGAMLRVDKESGEWKAFLPYGFYTGFSDIIANLSVVDAMAEDGMGDDVDRCGATRFNVIHLVPGYDNT SIVETVLDRADAAGIMYIADFRDIYENLTAVAAYVEIFRTRENFLTWYTNDEPDGPADPSNYTTAAYDMFYEMDGY HPVSLVLNCLDFYWTNYTSGADIVFADVYPIAVNTSYSDRYGTVCNFTYGCCGCDDCTGELTDISNKIDLWSEYRR FVGRTRELAIWNVPQAFNGTGEFWTRIPTGAEFKMQSVLGVNHGAVGIMPWDNPYDTPDIAQGAREIAPIFSSLSS PIVALLLDPNATRTLLTAENENTTSVDIARWIVPASGNETTVLVLMTSLLYENSTVEVDVKQDGLEEKDLTLVDSVFG SGEEADAVVEDGVLTITLGSTGIAGFVFSA
scaffold4_130k-330k:g34.t1	MLLFSTIIVYPIITTAVFWSLLTSSTTLSGTVNAWTNISVHALNTALAAFEIFLTNMPPTPYLHVVFLVILLAFYLGVAYI TFATQGFYVYDFLDPSEQGALLAAYIVGIGAGCVVVFLVTRNLCLARQKIVARFLPQSQPATAEVMEVKALRRPYR TENV
scaffold4_130k-330k:g35.t1	MQLTTCLSLVLGFLVTSSSAVTVQRVQLDIVNSEISPDGFQRETVLANGTHPGPLISGNKGDIFQINVNDQLTDTTG LDTQTSIPISSDNSFLYNFTVNDQAGTFWYHSHFSNQYCDGLRGPMVIYDPEDPYANLYDIDNASTIITLADWYHYL SDDAPPGAPNPATVLINGVGRQLGDDSLALAVVSVTQGRRYRFRLVSISCDTAFNFSIDGHQFTIIEADGNNVEPVL VDSIPIYAAQRYSFVLNASQNVDNYRIRAAPDIGLTSGTFANGVNSAILRYTGAPSIEPVTNETVSLLQLNETDLHAL ENPQAPGNPGLGNVDRPYDLTVLFNGNVNTSVPANFTVNGTSWEAPGVPVLLQILSGAQRAQDLLPSGSIYPINA NDTVEIILPGVAGAGTNHPVHLHGHAFSVVKSAGSSAYNYDNPVRRDVVNIGGQEPSDITVIRFQADNPGPWFIHCHI DWHLTAGFAAVFAEDVPDVPADDPTSDAWQNLCPVYNNTYGNAPPPALKSIGPVTQVAPTPSVTLLPSSVSVYSG TPSFTVTATSSATVLPITSAAY

scaffold4_130k-330k:g36.t1	MFLGIRAGPDTGKSSTYDGGLNSAVLRYIGAPEVEPSTNQTVSLLSLNETDLHARENPQAPGNPWPGGADRIITLA PAFNGGNFTINGIAFQNPEVPVLLQIISGAKRAQDLLPHGSVYPINRGETVEIIIPGVIQAGSNHPVHLHGHSFSVVR SAGSSVYNYENPVRRDVVKVGNLQNDTTVIRFKADNPGPWIIHCHIDWHLTAGFAAVLAEDVRDTPLKDPTSRKSL PLPTIAAWRDLCPVWNQTLTFPNGTIDSIPAASLKAVAAPTTTIDLASTVTLLPSSVSVYSGTPTFSATQSSD
scaffold4_130k-330k:g37.t1	MMDYNRPPVPPRPTSSYTNSNSSGPAPPPPVPPLPPNFSFNGQDPQISLPHFEDPLVAPRPHKITPDLPADMART LDHQLSQPPQSSQRLSVVGGPQDQQYAAFRVPSRVSVLPPPQGSDWSPWNIPNSQPTPSPRPTSYHPPPPPSS SLPMNYADAPLPGAAPPPRGSSFADPTPNPGASPSLTAPLPTLSALTAALPAIQNPNGDQASKVAWCRDVLLLVD RSQAQINSSEAEPPAPGPADIQDPQLNRLAHIAVPLVVQLSNPPNPAPSPLPPHVAEAIYHRANLAASGAFPMFVQ LNPRSAFRDYEKAARAGYHAGWFKLGRDYENFGDLAHAKDCFERGVKYANESCLYRLGMANLMGQLGFAPNPD IALPLLQRAANLATIEVPQPAYVYGLLLLSEFSHVTIPPHHFTPFIPLGQTLDLEARRHLERAAFLNFAPAQYKLGHAY EFAQPPFPFDALLSVQYYSLASQQGETEADMALSKWFLCGAEGSFEKDEGLAFTFAEKAARKGLPSAEFAMGYY MEVGVGGTKDVEAAIKWYQRASQHGNTDATERLTALSQPAPQALSRTEHDTLTETTLVRKRTQAKQRSDARGPR QGQGPENETPAQQIISSVRKNSIAHRERQQGGVGGVGGMGGQMGAGQSPKYANNALPASPRMNPANKPPMQR IASGPDAFPAPQQPQSQPQPMHAASLPPGPAGAGGPGGAGAGGPGVGMGGAGAPGQQRAHPNAPRYGLSDP GPASSPAPSRGNSIPQPGSGGGSGDEGGGGAGAGGGGGAGGGGRRQAVQKGPATFEEMGFKSQKLEEKECV VILLPYEQRTYDLNTPPPPPTSSPHLLRQQRQQSKSTALILLRIIYLLILYPHFLSPLRHVPGPPTPLTSLRSILYGHFP AIIRSEAGIIQREWAKRYGGFVRAVGPFGIERVICMGRETMGRVLGEEWVDYPRVGIFVFETDFIFFTSEEKREWM KETDMYYDHIETLIEMMDFEISSQLKPSEGRIMLMYSWMSKVTLDIICDTAFGYRSNSLRDPHNELTEAYHNLLNLE NGSNITMLMALISIPGMPWFMNSKWGYKFRKIFQLSSVTAPGTVLIESMSRIRRISAQILADKIHDSGLTAADLSTKK DIMSIMVRARKGDEDAEEKGEAKHGVYKLSDEALVDQVLTFLGAGHETTASGLSWTLWLLAQNPTVQAELRAELT PLFDAEPRPGYRALKECKMLDCVVMESLRVMPPVPMTIRQAAKSNWIEGHWVRKGTYFYIPIRVINTSRSVWGED AETFRPSRWLDLPSDYVANYHLMSFIQGPHACIGKTMSILEMKAVLAAIIVKFEFLPSHEGQVAKPTAAITMSASFSA
scaffold15_10k-210k:g9.t1	MTFPNMSFSSDFQHIVDGKPRGSKTTRNVINPTTEEVLATVPVADRELLDETVVAASCAFHAWASTPMETRQAAV TDLGKLVLANLNEFAELLIKEVGKSRELANIETSISGAWLQGVAGASLPEEVLFEDDNRKAIIRYVPFGVCAAITPWN FPISLLVWKIAPALVTGNCVIVKPSPFAPLCAMKLVELAQQVLPPGVLSVLSGDDELGPWISTHPGISRISLTGSVNA GKSIMQSAASNLLSLTLELGGNDAALILPDVDPKSLAPKLFWSSFYNTGQICCAIKRAYIHEDVYEAVRDELVAFAKT VKIGNCLDPDVGMGPVQNRVQFEKVKSLIDDCREHGYTFALGGDSIQGPDGKGYFLPISIVDCPPQDSRIVQKEQF GPILPLLKWKDEDEVIKQINDSAYGLSGTIWTNDLDLAQTIARKWETVWINEAQAFHWDLPFGGFKHSGVGIEHGR QGLISWTNIQTITLNKART
scaffold15_10k-210k:g10.t1	MPSDSQLPPFDHEYKFSYTTPPNPAFTYGDPVSSTELGKVWLEGLKEGWQVYDTTKTDTSWFNQVTANPPIISIA VQNITPDGIPHLKDTGANIKALKEFTVNIASEPFIENATTTSLDAPSGVDEWELSGLTRAPSIYVKPPRILESAFSMEC TLYAAYDIPSTTCSAISTTLILGHVKYIHVRKDVLTDRERIDPAKLRAVSMLGSATYARVGAGFRLDRPRWADEKEKI ESSRKDSSIPQESTAL
scaffold15_10k-210k:g11.t1	MSHGVVLVTGITGFIATHTAKAFLEAGYTVKGTARSAGKAEEWISLFPAFKSKFQYAVVEDMIAPGAFDEAIKGCDI VAHIASPAHWNHTDNEKDILIPAIQGTRNLVEATRLEPRIKRVVFTSSFASIVDPAAPPDHVHKSEDWNPVTYEEAK ASSTPAFVYRSSKPLAERAFWDYIKEEKPAWEGSAICPWYIFVLTWPAYRLLIVSSYLAYFWNMANGTFKSGVPAV GFPVFVDVRDVALAHVRAVELDAAKGQRYLLIGGTYKPELFVDVMGRTHPELKADKLAEVDVSKVKLDTSFKYDN SKAQKDLNLTFTSLDKMVADTLDRLLLLEKELPK
scaffold15_10k-210k:g12.t1	MANELAALSSLVSKGTETIIQTYSDRGRTFPTLDDPKGPSAIEDDRQLAETINIVLAAAAQLIEMLRKPSDVIMDTVYA MHVTPALRIAVETGIAELLRESEPEGLHIDDFGKFIDVDKDKLGRCLRFLCARHIFKEVRPNVFANNRISAMLDTGKP HEAIRSRPLDKFENIADGTMAAFVSHAACDPLSRGSVELVSYLLDPASSYSTDPKITPFSRGVGEPSDWFDYIERP GNERLLKNFGAAMNVSTFLFPKDVVTTAFDWGSLEKNEVVVDVGGGIGSCTLPLARAFPHLKYVVQDRAPVVEQ GLKFWESEHPLALKEGLVRLEAHDFFQPQPIHEPRVFLLRIFAKPLALRRG

scaffold15_10k-210k:g13.t1	MHFRKQLLLPTVVLIAPSSFSALAQLDSVDPASFSSTSFDVIIVGGGTAGLVLANRLSATSLRVGVVDAGHYNTTGD PLIDVPYSPGYLINDPSASAIGNPAYDWAFSSIPQDGLLNGTVIPYPRGKVLGGSSAINNLAWQRGAREEYDSWST TFENGPRWSFDALETFFKRSENWTFPATGSDALAPNTTASQLSALAAAHGEGGPIDVRYDNFVTEIDRAFAQASV NMGFPLNPNPDGGNSSFIPLAGIPNSVDISSGKRSYAAPGYYSADIRVRPNLSVLMHATVSRIIWDDVVNNDSARA TGVEFVAGGQTYVANATQEVIVCAGTLKSPQVLELSGVGNATLLESLGIPVVLNLPQVGENLMDHILTSTDYTIKDG VFSLDWLRTNQTYLQEQQALYNSTGQGAFTYATRVTGSWPVKDIVSEDQYARMRASLDAELAALDLTPLQLAQY AKLKDMLDSGEVGWVTPIALPRGGAASIPIDDENYITVAVFQMHEFSRGSVHISTSNSSVPPQIDPKFLSFTWDLDV QVAAAQFIRRWVNTEPAASYIADAVTPPANVDTDDKWRSFVLSTLGSIAHPIGTMPMAPKAMGGLYYTRILSLRIGS AKDVLHTGVVGSDLKIYGLANVRVVDASVIPLTIGAPLQPTVYAVAENSLAAILYDPSQLPKEKVYDYIIVGAGNAGG VVASRLAEDSSINVLVVEAGGDDVTENPISQRIPVPLMGISLRNSVIDWNFTTTRQRGLHNKSMHYVRGKVLGGST TLNLMTYTRGSADDYDRLAAVTGDSGWSWSKLEPYMRKLEALTVPTDGHDTTGQILPSAHGLSGPLSISLPNHPL SVDDLIRKATQELSDRFPFNTDMNNGNPIGVGFNQATIRNGTRESVTTSYIRPALARGNNLDILLNSQVLRIVPSAHI ANIPSFRTVQFARERNASTYSLTARLEVIVSCGAISTPHLLQLSGIGDATQLKAANIRPIVPLQHVDQFIHNDTFQIQA LDEWHRTRKGIMGLSFSSMTGWLRSSALNSSLDQSAGSRSPHFAMIFTPTFVTAELKHSSNQGFLTISNIVASPAS RTHGLRVVDACIFIAAHPMGPVYIVAERGADIIKAARLAVQETIPAENAGDGVEEENRGPSPKLFVRPTIIAMTTYRR FAVAGAGNVGKAIVNELMNQKLAGQVDDIVILTRETKGSDKNSTATTATEAAPRVVVVDYASVPSLQSALANVDVV VSTLGYAAIFTEQINLASACEPAGVTLFVPSQYGLPGRTGIPTDSEFRAVLGSVGTTVFYTGVISDMLFNESPYAGL MLRTGSVIIPGDGRALISFTSRGDVARFVAHVLTTLPPSKLINREFRVEGERTSLQSIVNGYQERMGNSLAVNYLPV YALREAVEKNVGDVLSHIRLAWALGKGVVGERDELDNYEYSDWNPATALDVIV
scaffold15_10k-210k:g14.t1	MLLSSALISRPQSWNAHFRRNAASSLMAVVYGTSPVTTDNDSSVTSVNDHVSRLTRANLPGAHLVEFMPWLRHF PSSISKWKRVSEQWYEHDSKMFEDLLGSVRTRMAKGDTRLNFCVRLFEGEAQSGLSESEKAWTAGTMYAAGAE TTSAVMSWFLLAMTAFRDVQRRAQTELDTVVGRSRPPTFDDFDHLPFIRAILKEVLRWRPVDPVGLAHSTTEDDW YGGFFIPSGTIIIPNVWQMNRDPNVYGNDAAKFDPSRHLDEQGNLKATSVYAKDEGHVSYGFGRRICVGRHVANN SLFINMAMVLWAMDISEGTGGNIDALPLGVDACVEDGIVV
scaffold15_10k-210k:g15.t1	MSQFDRLKYNPNLQTIEIGAGLTWDNVYDLLVPLGYTVAGGRVRGGVGVAGFILGGGLSWITNERGLTIDTLLACDI VLPNGTVYTASDESHPDLFFGLKGGLNNFGIVTQFTLRVFPQGNIWGGEVIVPETSIQQATTVIAKYAATVSDPKAT VNIIFEYAEGQFWGSYLEQYHSGHFFDYVVEIFLPDYYSHGKPSAWPADRSVVHFPLNIAFAWFDEKEDAAVHFAL QETVAKIKSDAEAEGQVLADASLYPNLPIIGTPIEEMYGKNLEVLRAVRAQYDPHRAMDLTGGWHF
scaffold15_10k-210k:g16.t1	MGYPFNDAPDSGIMNYLPRAGIPRTLDTQSGKRAYAASAYLNEEVKMRPNLVILTNATASAIVFDDTVDRAGLMAR GVQYIVDGKTYTATATREIILSAGSLKTPQLLELSGIGNASLLRSLDIPVKLDLPGIGENLMDHPVSISDFKVKDHVET LDGLTFNETYLAKHKQLYYASHTGAFTYTPSITGGWPCKDILTSEKYARMRSQLNSALYRTPLTPLQRAQYDVMKK VLLRDDVPCFMPAVVPHGAMASLPQPNTSYITIIGFQLHQFSRGTVHINNSDPMSPPLINNAFLSIPWDLDLQVEGM RFIRKWVSAEPVASLIDELVTPPAELEDPDEWASWIKSTVWRLHYIITNSPEDWLRPSGTTAMASRELGGVVDPRL KVYGTQNVRIIDAGVIPMTVGAPIQAAVYALAEKGSEIIRQDLLGFEKSQGVEQISKRTDEL
scaffold15_10k-210k:g18.t1	MPHRPDNWDAHLRRAAASGIMSVVYDAAPISSENDSSVVAINDFVARLTRAAYPGAHFVEFFPWMKHIPSQFAKW KLQAERNYTSDSAMFEGLFGRVRERISQGDERPSLANTLVRDATRNNLSERENAWLAGTMYAAGAETTSAVMA WWMLAMVAYPETQDKARAELDRVVGRDRPPAFADIDNLPYLCAMVKEALRWRPVDPVGLPHYSMEDDWYNGH FIPAGSIIIANVWHLNRDPDIYGADAAHFNPARHLDDDGQLRVGPVDTKDEGHVTYGFGRRICVGRHVANNSLFID MAMMLWALKIEPYIDGTEPMFDLSGCIEDGLVV
scaffold15_10k-210k:g19.t1	MSRSIKRFKYRLLKNFYTARIIMQRVYGYTFENAMADPFVLVNKEASETTSTATVPGTFLVDIFPILKYVPEWVPGA GFKKTARSWRTLQESVVNGLFNVVVDQHSKGSSNPCFVSTCLEDRMNNSSDALEKNVIKSIAAVLYAAGSDTVIF HEAPSLTYKI
scaffold15_10k-210k:g20.t1	MLTPELQVIWTSTSTRAAFVVFLGVAAFSVVRYVNSPWRKLPPGPAGLPILGNVLQMGGKQWLTFSNWRDKYGD LIYLNVAGQPLVVISSQKIVGDLLNRRSKTYSDRPKNIVPNELMTGGMLITFTHYNDVWRRLRKAAHEALNKGVVH NYHATQAREAVLLAHGILSDPREWDAHLRRTAVSSIMSMLYDTAPIVSEEDPTVKRINEFVARLTRAALPGAHLVEF FPWMIHIPSGLARWKRQAQEWFLQDSLLFEGLFKAVRDRVDQGDERPSLSASLINNASRHGLSDRENAWLVGTM YGAATDTNASVMSWWIFAMVLHPEVQKRGQAEIDAVVGRERLPTFADLDKLPYIHGIVEPARFSFLRMLNDDWYE GYFIPAGSVIVANVWHLNRDERIYGPDAKEFNPSRHLDENGALRSGVFAETADEGHFTYGFGRRICVGRHVANDS LLIEIAMMLWSMNIERPTDENGKPVDMDVEGCVEDGLSV
scaffold15_10k-210k:g21.t1	MSIPSSQLGVFIKGSTPSFVVERHPVLQPGPGEALIKVRAVAVNPTDWKIIVFGGAQPGNGSGNDYAGEIVALGPG THTVEVGQRVAGWMHAAFDKTHFAFQEYLITRSDLLIKVPANIPNEEAAAIPVGATTAIIGLSRLFDFPQRAPLGKTIL VCGGASSVGMYVCQLAKISGLRVIATASRGNFEVVSSLGASAVVDRHASDVQEQIIREAGPAGVDYVFDAVSFGS TITIASSVLNAKLGGFKKAAVVLPIPPSSLDPHVEYHSVGIQTIWNKPYEFLGASIPALPRDMETAKAGYQAVSDWLE KGKFKANPVVLGDGGLHGVPDGIAAVQAGKITGSKLVYRISG

scaffold15_10k-210k:g22.t1	MYAATVILLWYIKPADFDRASSPDSSDFSPPESSSSNGSSESDFEDIASAISSVISVDIDMSKFPTRDGFTTPGSGE DDAGISAPEYPPHERFWFDDGSVTFVVEGTAYFLHRFLFKLHSTTFAQKFLAGHTDPKDPIILKDVEKDQFDAFLNV LYRTSFTSDPAQSAAAWASVLHLATEWDFPSIRNLAIKNFTPSASDVDKIVLGHRYRFEEWLLPGYTGLCERKEPL TPEEGDRLEKQDIIAIARVREELRSNASAAEPIEMTQSIRCQSCNTTHNYSFNNTQTSLPCQTNYYGNCQSIPLPVIK PRSSSSPSLPDTAAITKKVRSLIPRSTLQYVTPRFHHGHADGPEADTDVEEGHVEEGIVKVSPIEDIDAGKLEDVKD DEPKTGNCKSKGKKKKVHRPYY
scaffold15_10k-210k:g23.t1	MSSPTPSSSVNGSKDFANALAEKYTTDFPVHQPLLPQQPSPAVLDTVLLVGSRGGIGANLLSQLLRRHCVRKVYA LNRNHDDGRTSQERQRDELHLQGLDPSLAESEKLVLLEGKTSHATLGLTRDMFDEISKSVTAIIVNGWRVEFRSPL SEFEDLIDGIRNLINIALRSPQSQPPYFLFCSSLALLQDLDIPSPIPETFLEDAPPVWVSGEGYRRSKWVVESIIKDAM KKAALRATVARIGQISGSINGYWKTDEWFPIMVKSSQYLGCLPMEDRWTAWISAPLAASTIIDLIGSNVPVCHIVHS HPIPFTSILHPLSRLLNVELVSRTEWFRRLESSTSKDLTKEMKKANPAVRMPSFFLMCMTKSYQGAFNLPIVRSVEA EKGSESLRRAKERPTDGEDVRRWVDYWQKMGFLRRAGDDGSGDRLKAKL
scaffold15_10k-210k:g24.t1	MEDKYIGVILALAGAVANGFGFIIIKMGLTNASERDGTYAASSDDYAYLKNPTWWVGTITLVLAEFVNFAAYAFAPPI LVTPLGSLSVIIGAILASYLLKEELGHLGRVGCALCLLGALTIVLHAPEDKEINTVDEVLRYALQPGFMTYCFSVLVFS LVMVYGVAPRYGRSNPLIYISICSLVGSVSIMAIKGFGIAVKLTFAGSNQFIYPSTYVFGVTVIGCIAVQMNYFNKALD TFSANVVNPMYSVCFTTATVIASLLLYRGFNTDDVTNIASLLTGFAVTFIGIHVLGLSQKPGKGTKPSHEEYALVDQD ARHSEGSGVDDLEARRDSGVYTGHGGTSVDEEGMEVTNMVAPVRSSRG
scaffold2_2070k-2270k:g16.t1	MKLQFLASFLSLAASLVAAQTYPNPIAGSEDISVRDPSIWYNSDTGKYYVFATGGLITTYTSTSLSGPWTNAGTALS ACAIVDSPGNCDLWAPDINYFDGEYVLYYAASTLSSQISVIGVATSPSMEPGTWTDHGEHSVDANLIITNGGLQLSF GSWQQGIYQIGLWPDVETQASALPGTHLAGADDRPAEGGFDYKPASSDLWFFFFSDGITPLQGDTTRPAAGDEY KVRVGRGSSSQGPFYDLPGDVLTEDLDPPTGNIILPSHDNIYAPGGQSVFADPVSGRDVMVYHYVREDEVGGAS YLGINYLDFSSGWPVVVD
scaffold2_2070k-2270k:g17.t1	MQFAGLAPGPFAGLILSDFGASVTRIDRFHPSPSSSSADTKEADLLCRGKRSIAVNLKNEKGLEVVKHIVRGADVVI DPFRPGVLERMGLGPEVWLDDGQGGNGENEGLVFARLAGFRTDGPHRDMAGHDINYLALSGILSMLPGTSEKPS FPLNLLADFAGGGLTCAMGILIALFERSRSGRGQVVNTDMYQIAGLGHALPLHLSVPPLRPPHADLFNSTAASSN
scaffold2_2070k-2270k:g18.t1	MSDHSSALHGDGQDLLSHTKFAGTRADHQFIYILGTSQAALSDAFSPTELPRQYATPPVTPATANPSIPLPLSPTDP TSPSTPTSPTSSFKPSPENTPPPPPANDPPPSTASAEEYEAWKAAYEKQVEEWKQESREKREKAREERERWEKK RERERQEMERLREEMVRSQGSESLSWSNIEIGGSGVASPMPGEAELGEGEKAREELAKRAEEAMRLVQDDSGE TSVVDGRALVPGEAEGKHSIEFLERALPRVGPVHTTVAIQTTDLTPPTAPVQEGHTEEPSHGSSSPHWEDVPSDM DSSFPSMSFPDASRSHSETPSHPNNNSQPPHHQHHTAESSHAHHEHHAIDDHKHHARTAATADAHTTATLAIFDS TLSTKTRVWAFASALAINMFLPFVNGVMLGFGELVAKEIIAPWFGWRPAVTRVGLSGTAGGGVAADRSRERERER EKNATRQRTGREL
scaffold2_2070k-2270k:g19.t1	MSVNGNDSFVTLEGTPPAFGHDMLKMFGFEEGYVNLNHVAEACEKWSRKIESNPDKFMRLDMAGQLGSVRRRL ASMVGAGPDEIVIVPSALHGVDTVLRNFEWNEGDVIIGMTTTYNGVERCIQYIHDLPPHPIISEFQLTFPNTHAAIINT FRAHVRKLCAHQKSNAKAGQDLKIVAVIDSLVSNPGVLMPWKELVQICKEEGIWSVVDAAHSIGQENDINLGEAKP DFWVSNCHKWLFAKRGCAMLYIPKRNRHIIKSPYPTSYAYVTRPESATGEDIDVANLIEQFGWTGTRDYAPYLSVN AALAFRSWIGGESAISTYCHNLALEGGKRLAEILGTSLMDTTENSELTVNMVNVGIPLSDDVPFNSDTNVYFQKKLL QEWNCFAATYKHGGKWWTRCSAQIWNDISDFEYVGKAYLAACKELEAMYVKEEPKAKEWTQSDELNNEATL
scaffold2_2070k-2270k:g21.t1	MPAQLPTRKIGTTDVSSIGFGAMGIGGFSYSSTDAEKDNMKFLDEVYERGCRMWDTADCYGDSEKIIGQWFKRT GKRNEIFICTKVGYVFGDRVVNGEPEYLKSAVEKSLGRLGVDQIDLLYLHRADLQTPIEKSIEALAEFVKAGKVKYL GVSEASAAALRRAHSVHPISAYQVEYSPFALDVEDPKIGIMQTCKELNITLVAYSPLARGVLTGQYKSPKDFPEGDT RPFMPRFSEENFPNILQLVEDIDNVGKKHNATAAQVTLAWLMAQGENVIPIPGTRNVKRLDENLGAATVKLTPEEV ATLRKAAEKADLPGVPRYPPGFNDHTFIDSPSL

scaffold2_2070k-2270k:g22.t1	MVMDTLRESLFGQTMRFLTGDRVFLYPEERDPQVWYTYINSVKSSNLAIYGNINPDPVDVTPPSFIPPAQITGSHT HVDEVPLEALEIVRPPLPGNESRDSLATQVDPFEQNRKKEKIEINYEKDALLIDFAENDPANPYNWSLGKRSFITGL LCLLTVAIYIGSAIYSPGIPDVMVQFGISQVAATLGLTLFILGYGIGPMLWGPMSEIPQVGFNLLYIPTLLIFVVFQVPT GLSVNLGMLLVFRFLSGFFGSPVLANGGASIAYMFSAQKRAYPIVVWGATAVCGPVLGPVVGGFAAQAKGWKWT IWELSWLSGFTLIVIFFFLPETSHANILYRRTHRLRKLTGRPNLRCESELINEQNNHLLHDIAISTARGFALNFQEPIVF FLNIYTCLIYGLIYAWFESFAVVFGEIYGFSLPIEGLSFVGILVGVIIAVPPYCAYIYYVQEPKFNEKGELKPEHRLPPA FVGAFLVPISLFLFGWTGRASIPWIVPIIGSSIFIGGAMLLFMAVLNYLGDAYPAYAASVLAGNDFMRSTFGAAFPLF ATQMYHKLGVGWASTLLAFLAILFIPIPFVLHRSTVTPASFAATEFDVVIVGGGTAGLVLANRLSATSLNVGLIDAGH YNESGDDTIDIPTGFLTTDLLSNDDYVWPLASVPQSRLNDRSISYPRGKVLGGSSAINGLIWQRASVAEYDAWGNT FGNGDDWTWENMLPYFERMETWNAPTTTYVADDTYSTSLPSVEGESGPLDISYNTFFTPLDIDTAESTATLGFLL NPNTDGGNSTFLPIESDARCADPTTGKRSYAVAYYNETVRARSNLVVLQDTIASRIIWDNSTLNSSAIKATGVEYVG SDGVTYVVNVSKEVVLSAGSVVKSPQILELSGVGNSTILEGLGIDVVLDFPQIGENLFDHASVYTDYLVNDGIVTLDQL NWNSTYLAEQTELFNTNGTGALSWTVEADGSFPLKQIVASSNFTAFRSELDSELANKTLTPLQSAQYDLLKTWAD AGNIGWLAIRAIPSGGIVSTRESGSSYYTLGMSLYHAFSRGSVHINSTVPSASAVIDPNFLEFEWDRNALLSGLVFT RQMAAASPLVDLLNGANSPDASVQAAEDFNDFLEARLSTTNHPAGTTVMASQDVGGVVSPRLKVYGLDNVRVVD AGVFPITTSSALQQTVYVIAEKASDMIIEDLSA
scaffold2_2070k-2270k:g23.t1	MASTHLQPALDFIVENKGGFAAAALGGLAYATVQYIRSPWRKLPPGPRGIPILGNIAELGDKQWFKFEELAKQYGD LFYLNAAGQPIIVVNDHKVAADLLDRRATIYSDRPQNIVASDIMTGGLLVVFTRYNDTWRRMRKAAHEGLNKGVVH KYHHTQHVEALLLARATLALPNDWDAHLRRASASAIMSMVYNVPPITKHGDPAVTNINDFVARLTRAAYPGAYLVE FFPWMRHFPSWLAKWKKTAEDWYSMDSAMFEGLFKDVKGRLVSGDHRASLASTLINEAARHGLSDRENAWLAG TMYAAGSETTSGVMAWWMLALLAFPETQKRAQAEIDEIVGRSRPPRFSDLQSMPYMRAMVKEALRWRPVDPVG LPHRSTEDDWYDGKFIPKGSIVIANVWHLNRDPGTFGADAHLFNPARHLDANGQIGPGPADMKEEGHCTFGFGR RICVGRHVANNSLFIEMAMLLWAMNFERETGPDGKPKFLDVDGFIEDGLVVRPMPFNVKITPRFPEAAAILAQEQE FVSSE
scaffold2_2070k-2270k:g24.t1	MSSHSAAMSLLTKGAATSTLYKTFAIAGVGNVGKFIAEEFLERKAAGDATQVVILTRKEADNSALAELASRGATIKE VDYDSVDSIAAALSSIDVVICTFGYAAVFQPQFNLAAASKKAGAKLYVPSQFGLPGRAGIPSDDDLKTHLNGVPVSN FLVGTIADILVKYADYGGLVLEKGEVIIPGDGTARISFTDRRDVARYVGYVLTKLPPSKLRGRDFHIEGQRVALNDVI KEYQERSGKTLKVTYIPVEELQSNIEKNAMDILSVIRLLWASGKGVVGEEKEVDNGLFKEWNPKSVVDIALS
scaffold2_2070k-2270k:g25.t1	MDSISLNATGIPSLLDTSNVSSTTILLNLALVGLGAAVLKRIVSKRSVPLPPGPKGWPVIGSLFLMVQSANIKRWHAY AQWSQQWGSMVYMTVLGQGMLIINSRQLALEYLEKRSTVTSDRPQLFVATQLGINRFITLDKDGPRFRDQRRMM AGFMGSWAVLERFHPAFERETSTLLKTLLDEPKDFATHVHRLVGAVILPFVYGYHPNFNGEDPVMESTERGVEVF LDATKFGSFIVDSFPTLRFIPSWLPGGGWKRMLAPWQAAIDNMLDVPYVFTQKKMSEGMAEPCFVTSQIEKEGGK KLNPEREEAIKLTANAIWGGASDTSPAALKTFFLMMMLHPEIQRKAQAEIDAAIGPDGGLPNMQTRAKLPYIEAIFKE VLRMYPVTPLGFPHAAREDDVLDGYLVPKSTVIMVNVWEILHDPKYYSDPFAFKPERFLTTPTHESELDPSDVIFGF GRRYVNFVCPPSNMALPLTMRLLCFVISSVCDYLESAQLTYIKILAVFDITKPVDASGKPIEPVIDYNDGTISHPAPFE CAIKPRSAKAVELIHEAAAAAAEA
scaffold2_2070k-2270k:g26.t1	MSLSPSLRPIHAFLLFLLTFSSPSSAALLNDPSELAKGKTYDYIVVGSGNAGGVVASRLAEDPSMNVLVIEAGGNDL GDDSAAQAVQVPLHGLSLQHSRFDWNFTTTPQPALENRTLNYPRGKVFGGSTTINLMTYVRGPLDDYNRWARVT GEDSWAWDNLEPFMRKIEHITPPVDGHDITGQLTLSAHGTDGPLSLSITNYHHVLDDLTVKATSELSDRFPFIEDLN TGRPLGFAYNQNTMKDGVRHSVTSAYVRPAIARGDNLDVLLNTQVLKVSPTEWIDGVPSFRTVQFSTGPDGRCH DSLCSSRRLLTSWWKAPVQEITAKSEVVISCGSISTPHLLQLSGIGDSEHLESVGVKPIVHLPDVGQNLQDHALFSH PYAVSEPDILDAWIRDPEVGAKALEEWKENKTGLLTMGFASQTGWFRLPDDASIYDSAADPSSGPTAPHFAIIFTP MSLGSALEPDASKGFITISNVVASPASRGSVLITSSDPWAQPAINLSILTTDFDIYCMREAIKMTQELIKAPSLKDYVL GPYGGIADLTIDEELDAYIRKISTAIHHAVGTARMASRDSPTEGVVNPDLTVKGTHGLRVIDASIFPYETAAHTMAPV YIVAERGAAFIKESRAGKTVAGEAVGGQAHFEATSNPSRQAGEL
scaffold2_2070k-2270k:g27.t1	MPGKLNENLDVLVVGGGFGGLTSAIECARQGLNVTILERADAPSPLGDAIGFAPNTSRILFRWGIGDSFWEHSARG RYFNVHDGHTGKIIFREDHAGQLARFGWTSLPGSRAFYQRGLWDLAIKSGANLRNNCKVVSYTDGGLERPYVLLE SGEKISADVILASDGIHSKAKKIILGFQEQPVSSGYAIYRGYAPADSLLNDPETAHLGDGAPHTWIGPDQHTIICTFG EPGKVHSMSFATIHEDTEEAIEAWNKPVSIEDAVKSISSPDYDWDPVLKKVVSKFPDCIDWKLSWRDPLPRWISDD GRVTLLGDAAHPFLPSSAQGGSQALEDGVTLGICLALSPKTPDGCRLALKAFQKIRHKRAADVQMTGRTQIHKWHI NQTEAEKKANKELGLASAAYFDHDAEFHALNSYESVVRELTGDESFKVDPELFEKVRDVIKGDTIYANGAPEEKKA IWYGPPFEVPERDRPVAA
scaffold2_2070k-2270k:g28.t1	MASEKNSQKSQSSSHTSFLLVFGLGLLTVLVNYSPISIQELRILVPQYFGYGEELRTADACHLISQQTSNATEVHWP GSDAYDVDVHHWATSASTPSVCSVTPATTDDLGIILQIVGSSRVPIAITGGGHSLNPGFSSTPGIHIAMSRFTHIQHN EEAQTIDLGAGLRWDDVFGHLNPKGYTVAGGRVSGVGIAGFILGGGLSWITNERGLTIDTLVSCEIVLPNGTVTVAS AESNPDLFFGLKGGFNNFGGTVIVPESSIDYANAAIASFASQPRDVKATVNIIYEYAEGALNTIVIVFYNAPEPPAGLF DEILAIKGGAVTTASMPFLTLLAAGQPRNTPRRIQDEVPIQTWTPSILKQVANETRYWGEYLEKHHSGHFFDYIVEIF RSDYYSHYPLSSYSTSPTTASAWPADRSKVHCPLNIAFAWYDSAHDKAIYAALEETVRKLTVDAIAEGQDLEGASS YPNIVRPGVPVEEMYGGKERVKELQRIRDVYDPKRVMDLAGGWHF

scaffold2_2070k-2270k:g29.t1	MPAQLPTRKIGTTDVSALGFGAMGIGGFSYTSPDAEKDNMKFLDQVYERGCTMWDTADCYGESEQIIGQWPDFVIDRFKRTGKRNEIFISTKVGYVFGDRMVNGEPEYLKSAVEKSLSRLGVDQIDLLYLHRADLQTPIEKSIEALAEFVKSGKVKYLGVSEASAAALRRAHSVHPISAYQVEYSPFALDVEDPKIGIMQACKELNIALVAYSPLARGVLTGQYKSPKDFPEGDTRPYMPRFSEENFPNILQLVEDINTVGKKHSATPAQVTLAWLLAQGENVIPIPGTRNVKRLDENLGAVNVKLTEEEVVALRKAAEKADLPGVPRYPAGYDDHTFIDSPPL
scaffold2_2070k-2270k:g30.t1	MTYLSAILGAVVVAAIVLVAKSRHFRRSLPPGPKGLPIVGNINDLDGNEIHVQARKWSQEYDSDIVSITTLGHTTVFLNSAQAIQDLFVSRGAIYSDRPDMPMLIDLMGWDWTFALMRYGQKWKEHRRIFHEQFNNTISEHRQIQIPAALELLQLLNKSPDRFLDHLEHYVSRIIMKRVYGYDFEDAMKDPYVLVNKAASESTSTATVPGTFLVDTFPILKYVPEWVPGAGFQKIAKKWRKLQEAVVHGAYDMVVEQQKQGASNASFVGTCLEDRAQNLPEALSEDAIKSIAAVVYAAGSDTNLATLTCFVLAMVLHTDVQKRLQAEIDSVTDGERLPTFEDKEELPYFMDVFREVMRWLVVFPVAIPHRATTEDSYKGYYIPKDSTIIGNSWAVLHDPVTYPDPETFNPDRFVKDKSLPDPMDIAAFGYGRRSCSGKAMAIDTVWIAMATMLTVYDMEKPLNKEGKPVDPIVAFRRAAIK
scaffold12_216k-316k:g11.t1	MSSVPPSRQAKPPPSKADCPSLPLNPSKLVEKLAFGEQPIDIDIPDNYVAFTLKKERELPPLSWSNFYTRFNYVSLIVLTATPAMAIYGMCTTKLRWETALFSVLYYFITGLGITAGFHRLWAHRCYNASKPLQYFLAMAGTGAVQGSIKWWSRGHRAHHRYTDTELDPYNAHWGFWWSHVGWMLFKPIRKPGVADVSDLSRSEVVRWQHRWYLWLILAMGFGVPTVIPGLLWGDWRGGYFYAGAARLVFVHHSTFCVNSLAHWLGEAPFDDKHTPRDHFITALVTVGEGYHGFHHQFPMDYRNAIKWYQYDPTKWFIWVCYKLGLASHLKTFPENEIRKGQLTMQLKRLREKQENITWPEDVDDLPVISWQSFQDQSETRRLVLISGFIHDVGSFIDEHPGGAHLLKKYVGKDATTAFFGGVYDHSNAAHNVRPIFSPSASSPSSY
scaffold12_216k-316k:g12.t1	MVDSQATATFNPAVLQLQSDILILLQVVDAHTMILTISTVGILALGTLSMALSVKPSPRNAWSQAHIGIEDGSETTASCVIDKPVVAHHMVGYTYSYTPSDWASDIYLAATNGLDGFILNMGADPWQPARISDAFSAAEAFNANVSSTCATQSAHTFVDDIDTFDSSSSFAATGPRSTNNCTAKPFKLAISFDMSSFPCANSNDASNLRSIATAHSNSSAQLQICTADGNGGMSLQPVLSTFAGEACGFGQAGWQSVTSGFWFAPAFFDMSKGWTGWSGVQASLNWNGGWPQGNNDISFSNDQTWISETNGKGYVAPASPWFYAHYGPESYNKNFIYRPDDWLWASRWEQLIANREQVAMVEVVTWNDYTESSYITPFKGELPPYSEWVTKYDHQGWLKMLPYYIQAFKTGVYPAITQDKFYLWGRLYDKNAQPTGDSVARPANADWVHPSSTPTFLTSSSFSPTQTSNTLWAVVHLAPSLATNTSITLTCGSSTSTTSYSASGVYKLSLPFIDTCDVSAQITRNQGVLSTITRFTPSGMHYSAQGPQYWDFNAFVASA
scaffold12_216k-316k:g13.t1	MSTSIDTNAPTGAHTAGLNPTPSSATTYGATGSTNPSQSGTTNPPTTLDPTPSSATTGEPTTTSATAHDQPPPHNQPGTGVQSEDAGTGANEEYPEQKHAGAVGYGPNYGKGVGVGEKLGGLTEEVKGKVLKKPEAVQHGHDRRTGELKKKEEDAANSSPFENKSEHAPAPATSGETERAAMTAPEGTEHVWVLLKAITAVGLVAVAPKITLLPTCVPVNIGTPYLQLYKLTSNTLINIEALQPSYLSSRMLFPLLALASLSLVFALDTSSHCGQWDTIDTGIQYSILLDQWGASGASSGVSCASITSVSGSSVAWVNNWTWTGGTGVKSFTNVQLNDGINQQLSAISSMPTTWTWSQTTTGTIVADVAYDMFTSTSSGGSAANEIMIWMANINAGPISASYGSDGQPVPVASSISIAGNTWNLYQGSNGANNVFSFLPTSGQITSFSGDIMDFYDYLVSNEGLSSSQYLTTAQAGTEATSGTAVLTTYEFIPSSYSPTHPKNYVLMSSAS
scaffold12_216k-316k:g14.t1	MFFTLLAAVLPASIALAVPLKRLDTSQYCGTWDTIDSGAEYEVFLDQWGASGATSGQSCASITSYSDGTIAWVDDWTWTGGTGIKSFTNVQLNDGVNQQLSAITSMPTTWEWSQTTDGDIVADIAYDLFTSSTSGGDAENEIMIWLANINAGPISSSYGSDGSPTAIASDLSIAGQTWNLYSGTNGANAVFSFLPSSGDEITSFTGDVMDFFTYLIDNQGLSSSQFLTTAQAGTEATSGTATLTTSSYSLSISSSSSSDVASSVGLAVSSSSAASASTSAAVSAVSSAATSASVSCPAVTSVTASSAVATSVVTSKAATASSTSTQASASATSTSTSSATQSLYTDFPFP
scaffold12_216k-316k:g16.t1	MKFVLAKLTAIAIFLDDSLEDEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNSTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELVNDLMSFYKEQLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDETRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE
scaffold12_216k-316k:g17.t1	MLLSVLSLTYAAFALASPLKDRRATAVDTSVYCGQWGTISTGIEYSILLDQWGASGASSGESCAYINSVSSESVAWVNSFTWTGGTGVKSFTDIQLNDGVNQQLSAISSMPTTWDWSLSTEDGIVADIAYDMFISDTSEGSATNEIMIWMANVNAGPISATYGSDGQPTPIASSIYIASQTWNLYQGSNGANNVFSFLPTSGEFTSFSGDLMDFYTVTRCYEFFRFHACSDHVFGM
scaffold12_216k-316k:g18.t1	MVVKIPTLKLPNANNTEIPAIGTLSHLSLITIYIXLGCWGGVTKEERAKSKDWILSALKSGYRLLDTAYGYGTEQYVGEAIRESGIPREEIFVITKLPKVQESIDQSLKNLGMDYVDLYLMHFPQAVIDNDDENDLVDENGNAMPIDDGITFNETYAQMEEALAAGKARAIGVSNFSIKTLTELLKTAKVVPALNEVE
scaffold12_216k-316k:g19.t1	MLPSPTFSGPSLKFSSFALGGPASETDARRLNPAPVRISSPPPSSATASTSTPPSSHRTSLGPAGFLALLRSRRGSQSITPTPAALDIHSANNNNINNNSAAATNNNNNHSPPRQPGAARDREARRRTIAVAGISSTISSPSSSSTPPTAAFFPTSPTTADTSSPFAFSPPSTPPSLPAALPASASNTPNMPAPSVGHGFTQILRRRRSNGAIAPASPPQPSHANANAPSSNTNNSNNNTNNSQPAGSRTAPLPFASVPHGIAFASSNGNSNGGGGGGGGGGGHLFSFGHSPPNRPAVPPSANTRAQSMPGPNSSSSNATPTAANPHPTSTSASASNAANGGQGQGQGQGQGQGNRHRIRLVPHIDSRRSLRFEPIMRDVLENDAPLRIGRFTDRSGMGVAAANATGSNKLAFKSKVVSRSHAEIWCEGPPSLSALNPGPSPSPNNGNGTGVAAKFYIRDTKSSSGTFLNHVRLSAANSDSRPFELKDGDTLQLGVDYQGGTEDIYNTNALRQLRALADVTTAPLAINTTGKAPGGGTIAKKAMPKSGFADCCICLFSVTIQQALFIAPCSHAFHYKCIRPLLETHHPAFSCPLCRTFANLEEDVEVDTVSLNLKSDDDIAELGEIGTPVVELGAGGIGGGAGGIGGGVGGIGGGAGGIGGGGMGGAGINIGGQTPVAPGILALATAAATLSAREREEGGAGAETEVEEGAGATVGRRGRRTRNANTRGLPEEEMMEMEISDEHEFDGLMDEGYERDGFERDGEGLEDSAFVDADADGYEREEEDVGMLVDGDEGSGSGSADGEGSELSGGMGGGAGTNGGGSGSAVMGRRGEEGDRVVKRKR
scaffold12_216k-316k:g20.t1	MANISVKQIVDPTDEELDSIVSLVVRAFDGQNVVNMFTGNNLTLAEPLWRAHVLAGLHHGEVWVATDGTAEIMSVGVWFLPGHELYGTEELEEYFTLLSDEAINWWRNISRPKVVGFFEKYLGEKTRTDNWFAQIIATHPDYERRGLASAHIDAVYHRVFPWFFNTRIPASRSHELIQAVREESYLALGAAGKAQEAFYRSNGFKELGRVSSQTADWVCFVKGKGI

scaffold12_216k-316k:g21.t1	MEHKSPANGGIFLEIDGIRRLVETMAQLLWIVLFSDASRDNYRSKQTKSSSAGKQPTSNSIGQPLARPAKPVLAPP FYALIIGIDEYPDRAKLHGAVADAEAMKNYLRKELNVPDDQIINLYNSHATRSRIIESWKSLYMNTRIHRGDPILIYFA GHGASANAPSQWTEMAGGRNAKIQLLVPYDYDGETVQPIADVTLGYLLDELSKAKGNNITVIFDKCHAASGTRAD KNDPSLLVRSIECKNSLPGNLDEDVLGNTQRIVPRAAAPACQGQALGSHILLAACGADEYAKEQKCRGWFTKALL EVLYRVGAEKLTYSDLMKRLPDLPGQNPHCEGDHRDRLLFDGKAPSQSRTMYPVRVSTFEGKSVYTMGAGAAV GITKGSQYAIYQTPDASTDPKNRIGILVAEGKPAAFETVLTGFLEGQFDEKGWAIQTRAGEEEDLRVRIVMSNKLVP VFKKLAEDMQSRSGEKQGILLVDDDTADLDVSLDQENRVVFNVPQGACTQYGHSRLPNTLPPDVDMIYPVIRSAA RYYWHLKRSAPASESRGLSSLVRVEFHELMQENDGWDGQLRPILKPVEPLVDLNPANVVDLIVGDGGNKMYGFK VSNISATPLYVSAFYFDNSDFSILPYYQPPLGINNHIDPSIPANGSLTIGYGAGGGMPFTYYLRPEQNIDVGFLKLFIT NVPVDLSDLPQPSPFEDPARATDIAAFSKPLLVWDTVLITMVQRRYAPN
scaffold12_216k-316k:g22.t1	MLNASISHQPDVSSIPRNIHLSARYTSTFQSRNLSTFLTYDASTSMNMYLAAIASPPPSSSPSSSSPSTPAVGSPAT MHGNSFARGGPQVQALSRQNSYASQAPQSPSMSYASQTIARHGGVTNIAGSAAEAAALLRAINEGIGITDDTIVFSI TPGVFPNTDLMNRSSGSGEGAGEGAEEEEKDGVWVDEGNPFVKEGDVVPVRPRCVDQAMWEVGGAGVALRLV QLAQTPHELSRTLGILTDGLKFSWQNSEDMERIRGYDILGAMLKSKADLINMTSFETLFEFLGLRFGSPDNSTVSNP LAYRTIALDFELWSRTPQPIQRVHLQHFTTLIQGSRHKRFNIKQRIGKLSPSLVRRLLFIWQTGWYGEEAMMGSAG KRDGTLEGVIAVAMKGMFKVDETIKPIVSYLAANLHERAPAVASPVSVVSRISQDHTREKAERILVLLVSLLSNSPTH YNTFASALPLTRICLLLLGDRPTPVVAIQVLTLIDVSLTASRSFNRKFEIVSGWTLLRTVLPYAWNEGVHKVAFRILLG RTGASAHDIENGKRRSGAKKEEPAVVVCPQIMPAVFASLQRVLGVVANHVPPLNPIDESKALEPEAVAELLLEELL SLQLSCPTFRQVYKSQQTIEIFVQAFRSFVKALNSAVVIEQRTIRMLEKLMHFGVTIAMDNVVAGNQKAEILEVIKSA EVITSGDRPGSTDIDPSLVSDKRSRLNRLSSARLSVQVSERTIQKSMARMVQWRQTVITAEKKRLRKNVLDLRENH RQVSRLNDWIVPLTTERGLWEIPEERTWRLDETEGPYRVRKKLEADNDLMSRFKPTDNEHLRGLEAPALDEQSTF RVEAPWSDSYEISATDVEEDRQLAEEVVDDRHRRVRHDLEAGDVIEAVATVARISGVDSSPGLLIFGKTHLYMLDG LVENSDGEVIDVHDAPKELFFVPGSLVELDGIQRAQRWTYDQVASFSDRTFLFRDVALEIYFKDSRSLLTVFLTKRQ RSEANQRLSLTVNGQRYNELSTPGLLRSPLIGRMSARVRSGFRADQLSGAQRKWQARAREMSNFTYLSILNQISGRT PSDATQYPVMPWVLQDYSSETLDLSKQETFRDLSLPMGALTEARREAATSRYDSLASVGERPFHYGTHFSSSMIV CHFLIRLAPFTSMFKTLQGGDWDLPDRLFSNIARAYDSAAKDVRGDVRELIPEFFTCPEFLENSKNIDFGVQQANG ERIHHVKLPPWAKDDPLLFITLNRRALESDYVSERLPAWIDLIWGIKQQDPAAINVFHPLSYEGSIDLDTITDELEREA TIGIIHNFGQTPRKLFNVPHPPRFNHGLSTLPLGTLHGIEEDHNLLTQSVKTAARAPVSCVTASRAWSVIVSGSEDG SAAIWDLNKAVYVRSIWHGRADETVDGSKVHLASVNESTGYIATCSRIKLCLHTVNARPIASLDLTYSSALPYLPPQI TSLAFLEREYSYLGVLATGGPDGKITLRTWNTDKTPEGEKAKWEFVTLRILEARETGREKLMIMSLRFVGERLYHG DTNGKVYAWDLPD
scaffold2_1428k-1528k:g10.t1	MGKLKPAKIFKTAVDPGISFDDKGGHIITAGDDETFRLYSCKTAKHLKTLYSKKYGVDLPRFTHSDRNIIHASTKEDD TIRYHALHDNKYLSYFKGHTQKVISLEVSPIDDGFMSSAMDNTVRLWDLRTPTCRGLLNLPAAPIVAYDSTGVVFAV AVNRYSRILLYDQANFDKAPFLTITLEDPTLALISFPPRAIYITSMSFSTNGKYLLVGCSGDAHYVMDAFEGHLLAKL VGHASLERRRPEVPLGIEPQRGCSGEEVCWTPDSNYIVGGGLDGRICIWDVQDLGEKTGPIDLKATPRRLQPLTK MDGHPGPARCVKFNPRLAMMASAGAELAFWLPDQSPDSDDIAKDLVKKLAG
scaffold2_1428k-1528k:g12.t1	MSTKTISVLDESELKDGQMKEVDFEGGKVLLSRLGDKIHATSAFCTHYGAPLAKGVLTADGRVVCPWHGACFKVT NGDIEDAPAPNAIHSFQAHVADGKVHVTADPMKASKENMSRSPRLLATGAEVGGPGVVIVGGGSAAFHTIESLRE HGYKSPITVLSREAYAPIDRTKLSKA LITDPSKLELRSAADLKIKYGTTLRTGVTVTSIDIKAKKVILDSGKDAVSYDKLVLASGGLPKKLPVEGSDLENVYVLR GVQQAQAIDEACQEGKNLVVIGSSFISMELVVAVSKRKLKSIHVIGMEEYPFEAVLGKEVGAGLKKFHESQGVTFH PKASIKAIHKSSSEPKKASSVELESGEKLDADAVIMGVGVRPSTDYLKESGFKLEKDASIKVDEYLKVIGHDDIFAVG DIATYPEFGEYKRIEHWNVAGNHGRAVGKTLAGDPQPFVKVPIFWSAQGQQLRYCGTGAGHDDIIIKGDPGEMKFI SYYVKKGKVVAVASMQNDPVVSKCSELIRLGLMPSAEEIRAGKDVLSIDTSTTNALSAVKT
scaffold2_1428k-1528k:g13.t1	MSPSIDPKAKASSSLTGPHDLPTPEPESPHRPRQPYRFATLCATVENPDMKDQYGSSSVPIYQSSTFKGVGNEFD YSRSGNPTRSHLEHHIAKISSAAHAITVSSGMAALDIILRLLKPGDEIIAGDDLYGGTNRLLTYVKTHVGVTVHHVDT TDPTSLLKYIQAGKTAMVLLESPTNPLLKIADLATLAAEVKERAPDAIVVVDNTMMSPYLQRPLEHGADIVYDSATKY LSGHHDLMAGVIVCNRDDLHKQMAFVVNSVGNALSPFDSFLLLRGIKTLAVRMDRQQTSALAVANMLDRLGFKVH YPGLSSHPGREIHERIATGYGAVLSFTTGDKALSERIVGATRLWGISVSFGCVNSLISMPCVMSHASIDPAIRAARG LPEDLIRLCVGIEDPADLLDDLERALLEAGAITLSASETEHHFVRASLPREPVSISRAVEKLAYNDNAVTGRTKENRE WFVSAPGKIILFGEHAVVHGVTAVAASVDLRCYGLTTPRHDGKLGIEFSDLGDFSHEWVLDELPWDAVTSVNVGD AHPDVLDQRLVDAIAARALPSTINDKAKPAAIAFLYLYMSMAHGGERPSFGFTVRSTLPVGAGLGSSASFSTCVAT CLLLLHQRVNLPSLPPPTTSADASGPGHIHISHQGRRAIPSEISEEVNRWAFVAEKVLHGNPSGVDNSVAVFGGAL AYTRPGFTRKSGMEAISGFKSLRFLLIDSRVPRDTKSLVAGVGAKKAAEPEVVDKVLDAIQSITDEARRALLDPELS RPALLSALGALIEENHAHLVSLGVSHPVLEIIKARTAQSPYGLSTKLTGAGGGGCSITLVPDDMEESLLKDLMARLS EDGYVPYLTSVGGSGLGILSPYDPNNYDDHPTATAHTQTPPETPGDESALGGREPLRATLQSKPLHDLSEWASN RGRWMYV
scaffold2_1428k-1528k:g14.t1	MKLWHESVGKDGERVVDLRKPFFLLNDASQATTYGSFATPSSRHSYFPSKAARNAQATVHFVGDDVVAKRTKRH LDELERSNYTEPSSSSYPGLADDDDEPAAKSAKGRARQVISDKRLQALPGPKRKKSTMNVRSAVLYPKSLATLIEQ SGIAHLPPSVPTYLTATAAPPKVPPR ALCSVCGYWGRYKCMKCAMSYCDLNCQTTHDETRCERRVL
scaffold2_1428k-1528k:g15.t1	MCPHANGDALPNGSSDMVGIENTPNGVKSANGHTDSQRRNPYAPRASDFLNNVSNFKIIESTLREGEQFANAFFD TKTKIAIAKALDAFGVEYLELTSPAASEQSRADCEAICKLGLKAKILTHIRCHMDDARIAVETGVDGVDVVIGTSSFLR EFSHGKDMTYITKTAIEVINFVKSKGIEVRFSTEDSFRSDLVDLLSIYQTVDKIGVNRVGIADTVGCANPRQVYELIRT LRGVVGCDIEVHLHNDTGMAIANAYTALEAGATHIDTSVLGIGERVGITSLGGLVACMYAANPEYVKSKYNLKALRE IENLVAEAVEVNIPYNNPITGYCAFTHKAGIHAKAILNNPSTYEILKPEDFGLTRYVSIGHRLTGWNAVKSRVEQLGL KLTDDEIKDATAKIKELADVRTQSMDDVDSLLRVYHSGIQSGELEVGQKEALDRLLQEHRDRQRTELSAAATAAA
scaffold2_1428k-1528k:g16.t1	MPTHRKIAQEGHRFTPSEALQSGIIDEVVPGGTEAVLSSALKMAEAKASTAKGGVWGLIETEVYRDVIEGTKKEVR MVTLAMDDAAAKARL
scaffold2_1428k-1528k:g17.t1	MLGGVATQATASTVARFIGARVLIGAGLTFQTNAAPLLIAELAYPTQRAKMTSFYNASWYIGSVIFWFMPESPRWLI HTGREGKASQVLARFHANGFDEHDPLVVFEMAQIRHALKLEKEMSKSTSFKALIATPGNRRRMMIIVAIASFSQWS GNGLVSYYINLVLEGVGIQRTST KAAINGGLMIFNLFAAFGGAALVDKLGRRTLFITSNAGMLIAFSVWTLTTALFNTIHTAGAAKGKSNQNNVIACKFTH FVGRTATVPFIFIFYFFYDIAYTPMLIAYTLEILPYAIRAKGFAIMNLTVALTLAFNTFVNPWALDAIGWKYVESPLLLR DLGII

scaffold2_1428k-1528k:g18.t1	MNPNGVATPPLAARKRKRARQYTVNYSEVQEYDPDGKLREVIVIEDTPPPPTVSPATTVRTNGYSASYQPPQYSAQVRTRARAAKEAEALSTTTTSVALQAPPPKKRKRDVGDVPQATRKVAPPSSQQATVDGAATVSSWATSSAATTTNTVDPSKAPPPSCDDKEGHYIIVPDDIIYKRYRTVRLLGQGTFGKVVEAVDQETGRRVAIKIIRAIQKYRDASKIEVRVLQKLKERDPTNRHKCIHLLHWFDHRNHICLVSELLGMCVYDFLKENDFAPFPRYQIQDFARQLLGSVAFLHDLHLIHTDLKPENILLVHNGYRMVNVPVPGKRPQQTRAKRILNSTEIHLIDFGSATFNEEYHSTVVSTRHYRAPEIILGLGWSYPCDAFSLGCILVEFYTGVALFQTHDNLEHLAMMEQVMGKMPDRFARAGARSKPEYFKEGNKLDWPKPKASRQSKKEVRATRSLQDIIHPTDQINRLFLDLVRRLLCFDPAQRITVRDALNHAYFMQSISPEEL
scaffold2_1428k-1528k:g19.t1	MRGLPFPQGVDRETSYEYGTPAASTGPDDIPIDPALADLPIDPALQTPQQSHTCQELPSTPTAEYHHQSTHHRPEYSQAPQGDPFAPIPVYYPPPDDTQPVQPSPIKKTPKRKRKKVFREDECGFCQGNDDRNRDGQPERMVTCDECGRSGHPRCMQLDSAAELMYSYPWKCAECKPCEVCNDKEDDGRMLCCDSCDRGWHMDCLDPPLAEAPPGIWHCPACPQFHLEQEMQPPMFSEPEPEMHELLRMPRQSSPASSSRSQPPRGRPKSTRKKGKGRNAVEPHSLAAGDSEMEVDMNIDVETVEQPAHQAEHDEEDVSDSTPKASRRRPRKKPSLSPRKSRRASRIIQDSSPPPARASSSPKRPVLKITPPQHPSIPQRRVVLRVPQPASASTKGKEKEVESSDEEPGKGLFDDLLTAQERDISKTSITVRDKSRFDRSRQLAEDKLAVRPKLPFPEFSATPDTPIAGPSSRPLRSTIAASIHTPITPATNRSESPTPTPAPSTPAPHSLATKPPLRIRTIRFGIYDIQTWYDAPFPEEYNNIPDGRLWLCEYCLKYMKSQFMAARHRMKCKARHPPGDEIYRDGAVSIFEVDGRLNKIYCQNLCLLSKMFLDHKSLFYDVEPFLFYVMTEVDEFGAHFIGYFSKEKRSPKDYNVSCIMTLPVRQRKGWGNLLIEFSYLLSKKEQRTGSPEKPLSGLGALGYKNYWTLSLMRYLHTAPAHPRLEDISSATCMTLEDIYNTLSDLKFISVRSHFPAPKPLPGQSIKFPKGRKSGVARRHLVRSQTSQSNGGGGNGAEPAPFVPPASYEIRWNRDEVWQWLERWEQKGHLRIRPERLKWSPFLLARVKKTDGVEGRSLEEGGPGGGGLGVGLGVGFGTPATEGEASTSAMTLDHHQPEAETASVGERNDRGAQSDEEGAEEEDHEVRPVRGRKRNSGSRSPAKPRRLARAPSNKSLRSSGRLANGSATNTPIRRLRSQRGDNTPSPLSLAPSGRANIISPRKNAAVSARRTSTRGTRTRKSTTRSPRKSAAARPVRRRRRKMDSDAEYSEEEEEDEASEASEESEELPEPPDSDDDDDENTEEDAESEEELPVPTQRRRRVSGRASVKDAGIVDDDAAFAARLAAQDGPRLLRSRSATGTGSGPTSEPILKRSVSAVSAASRIVGRKRRRIESSPEPEGEPEIEAGPDEGSQQPLPNRTTEDADTPPLPNGSPSISFPPPLALPQDTDHVMEDAKSEFDSGTPFTATAHSRQSAPSDDTVIGNGDGAHVNGMGNGKGMDVVNMDVDSAVASPTQMSLLNGHGHQQGQTEMVVDAKTEFNPAQAQGLNGTPLAGLQFQYEEGSDADAEGEPDDEL
scaffold2_1428k-1528k:g20.t1	MNAFSRLIPVVGSAYGLQAALAAIFVPQANEKYYDLGGSLGFLSTAFVSLYYPQLKSSVLERRLVPLPALSSFAPRQLLLTAALGIWSARLGSFLVQRAIKAGGDSRFDEVKHQPAKFTAFWFAQATWVFLVGLPVYMVNTLPVHLHPALSYADYAAVGIYGASLLFEIIADRQKANWRNAQKNKQHEEKFITSGLWSLSRHPNYVGEVGIWTGIWALSIASLRTPYFPKLTWAVAAVSPLVTYMLLRNVRPTLSFYFPVDVLIKRKLATIQVSGVPPLEKSGDKKYGNDPKWQEYKSPPWFTSTTTTMAESSDSGEPLPRPNSFNTGTELVLHTAYSRCLDFETNVLANPAALPQTGKLSPLICARLIGRLLLEAPTDDGRANVASEIVRCQDDTNLQNMANLYKDHFIRCWLVLDDGVSMGVVTHCAHIFDRSTNENFDDSAKEPYAASVHAILDRYGGINSFQELNGANIHRLENILTLDTNIHTSFDQLKLWLKKIPDGSNSSYTVEAPSSSRFLLRILPSTQVDFTTSDPVKRPLPDPRYLKLHAACARVAHLSGAAEYIDQILREEEVTKVLAVDGGSSELLNHILMHRLGITAH

Chapter 3 Molecular cloning of hirsutene synthase, a bifunctional hirsutene synthase-3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase from *S. hirsutum*[3]

3.1: Summary

A large number of bioactive hirsutenoids have been identified from culture broths of *Stereum hirsutum*. Biosynthesis of hirsutenoids begins with cyclization of the linear precursor to all sesquiterpenoids, Farnesyl-pyrophosphate (FPP) by a sesquiterpene synthase. FPP itself is produced in fungi from the mevalonate (MVA) pathway, an essential pathway converting acetyl-CoA into isoprenyl pyrophosphates, the precursors to all isoprenoids. Here we described the cloning from cDNA and characterization of Hirsutene Synthase (HS), an unprecedented gene fusion with the second enzyme in the MVA pathway, 3-hdyroxy-3-methylglutaryl-CoA (HMG-CoA) synthase. We demonstrate *in vivo* hirsutene biosynthesis is not affected by fusion to the C-terminal HMG-CoA synthase domain. Furthermore, purification and kinetic characterization of heterologously expressed HS-HMGS confirmed production of hirsutene from FPP with typical kinetic constants for sesquiterpene synthases. By characterizing hirsutene synthase, we have now isolated the sesquiterpene synthases producing all major sesquiterpenes produced in *Stereum hirsutum* liquid culture. In addition, we demonstrated HMG-CoA synthase activity of HS-HMGS through yeast complementation. Finally, the unexpected cloning of hirsutene synthase as a bifunctional STS-HMG-CoA synthase protein led to analysis of 527 fungal genomes for MVA pathway gene duplications. Analysis revealed gene duplications involved in HMG-CoA production and depletion to be common and widespread in fungi, and are frequently found within isoprenoid biosynthetic gene clusters. This pattern suggests HMGS, HMG-CoA Reductase (HMGR), and HMG-Co lyase (HCL) gene duplications are a mechanism shared across the fungal kingdom for regulating production of isoprenoids from those biosynthetic gene clusters.

[3]The contents of this chapter are currently in preparation for publication in the Journal of Biological Chemistry. All work was planned and performed by CMF. LC-HRMS was assisted by Dr. Dana Freund

3.2: Introduction

Sesquiterpene synthases (STS) produce an amazing array of cyclic sesquiterpenes from the same linear, 15-carbon farnesyl pyrophosphate (FPP) precursor, with over 300 known scaffolds [78] with structures so varied that thereaction catalyzed by STS has repeatedly been described as "chemical wizardry".[354, 355] Class I STSs are classified according to the presence of highly conserved DD/ExxD/E and NSE/DTE motifs, which coordinate Mg^{2+} or other divalent metal ions [356] that in turn bind and abstract the terminal pyrophosphate (PP_i) group of FPP, creating a reactive farnesyl carbocation intermediate. Product specificity is then determined by the contour of the hydrophobic residues lining the active site, as well as its size and flexibility,[78] sequence of the H-α1 loop active site cap,[158] and availability of carbocation quenches, currently thought to be the PP_i removed from FPP.[139] However, the exact residues determining product formation are unknown for even the most well studied STS, rendering enzyme engineering approaches to alter product specificity of known STSs in a predictable manner elusive.[357] As a result, isolation of unique STSs from their native organisms remains essential to the biochemical study and large scale bioproduction of industrially and pharmaceutically useful sesquiterpenoids.

Hirsutene is one such precursor to antimicrobial and cytotoxic sesquiterpenoids. However, no hirsutene synthase has been characterized to date, nor have any hirsutenoid biosynthetic gene clusters been identified. Therefore, to address this gap in knowledge, as part of a larger effort to (re)create biosynthetic pathways to pharmaceutically relevant fungal sesquiterpenoids in a heterologous host as an industrial platform, I set out to use rational gene prediction to isolate the first hirsutene synthase from the brown-rot Basidiomycota *Stereum hirsutum*, a known producer of hirsutene [358] and bioactive hirsutene-derived sesquiterpenoids such as the antibiotic complicatic acid [359] and the hirsutenols (**Figure 5-1**).[168, 170] Recent studies of *Streptomycetacea* sp. have identified three cucumene synthases producing structurally similar 1- and 4-cucumene (aka isohirsut-1-ene, isohirsut-4-ene) from cryptic pathways, as bacterial biosynthesis of linear triquinanes had not been previously reported.[360-362] The structure of the

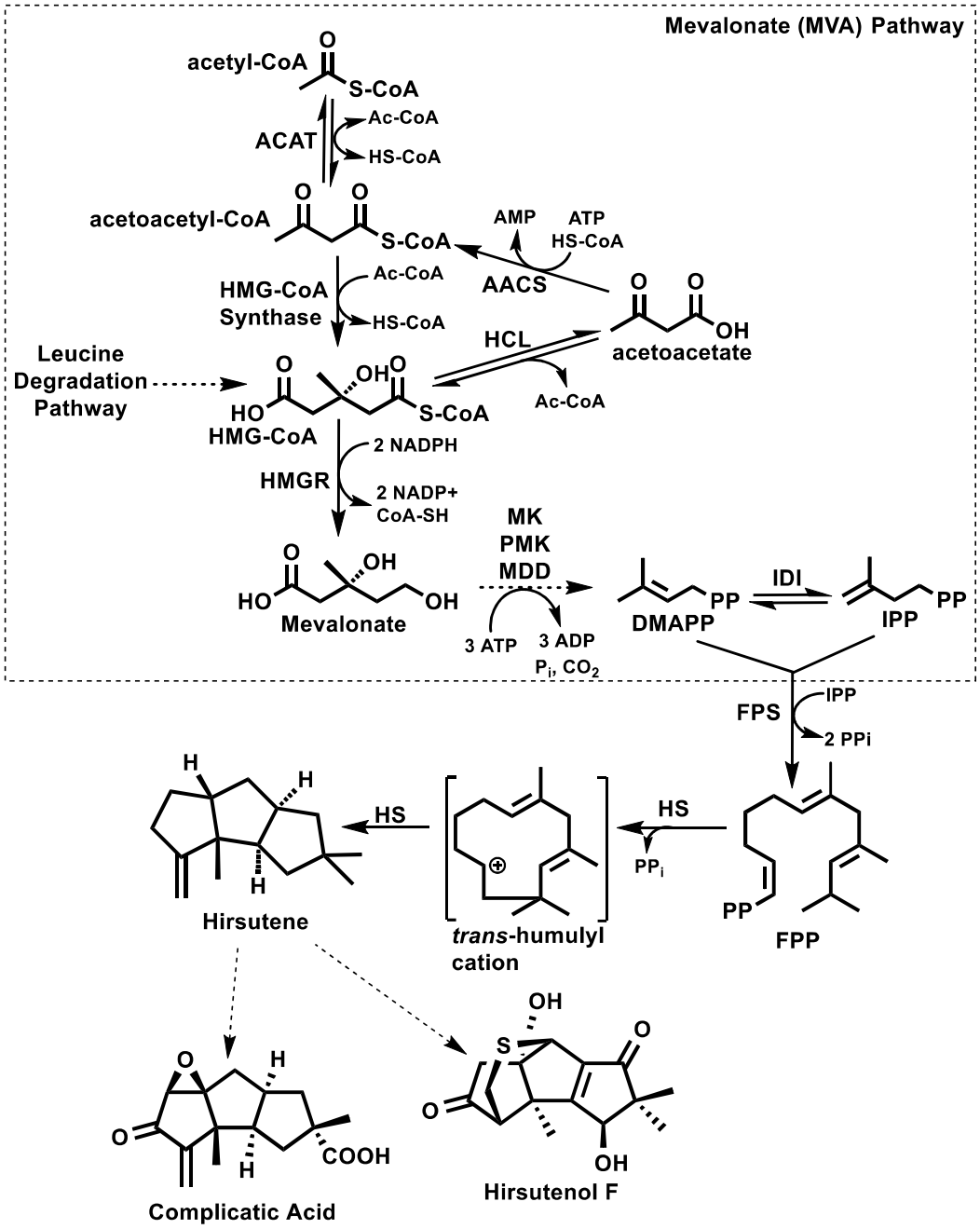


Figure 3-1: Overview of the mevalonate (MVA) pathway and hirsutenoid biosynthesis
The mevalonate pathway converts Acetyl-CoA to isopentenylpyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) via a seven enzyme pathway. HMGS and HMG-CoA reductase (HMGR) are tightly regulated, with HMGR being the rate-limiting and first committed step in isoprenoid formation. HMG-CoA is at at the intersection of the MVA pathway and leucine degradation pathways, with the reversible HMG-CoA lyase (HCL).[363] FPP synthase (FPS), the substrate for hirsutene synthase (HS). Hirsutene can then be further modified by downstream enzymes to produce bioactive sesquiterpenoids, including the bioactive complicatic acid and hirsutenol F produced by *Stereum hirsutum*.

cucumane scaffold differs from hirsutene only in the position of one methyl side group. However, this difference is significant biosynthetically, likely requiring a distinct reaction mechanism following initial cyclization of FPP to the 1,11 *trans*-humulyl cation,[364] and functionally, as no known cucumanes exhibit antimicrobial or cytotoxic activities.[14, 102] Therefore, identification of the hirsutene synthase gene is relevant to enabling biosynthesis of hirsutenoids with pharmaceutical applications and toward understanding the reaction mechanism producing hirsutene, cucumene, and related sesquiterpenes.

To guide identification of novel STS, previous research in our lab has demonstrated that Basidiomycota STS protein sesquences clade by their initial reaction mechanism, enabling rationally targeted gene cloning for the desired Basidiomycota STS based on initial cyclization mechanism. [83, 89, 358] reviewed in [364] We previously identified 18 putative STS in the *Stereum hirsutum* genome, and observed production of a diverse variety of sesquiterpenes produced by *S. hirsutum* cultures.[358] These sesquiterpenes included several derived from bisabolyl and germacrenyl intermediates, and the most abundantly produced sesquiterpenes were protoilludene and hirsutene, compounds derived from the *trans*-humulyl cation. [189, 358] We successfully predicted and cloned three protoilludene synthases, one bisabolyl-derived STS, and one germacrenyl-derived STS based on our predictive framework.[358] The goal of this earlier work was to identify protoilludene synthases in the *S. hirsutum* genome; therefore no attempt was made to clone hirsutene synthase at that time. However, application of this framework had identified three putative *trans*-humulyl cation utilizing enzymes: stehi1|16172, stehi1|122776, and stehi1|52743 which we predicted could be hirsutene synthases.[358] Here, we report the successful cloning and identification of Stehi1|52743, a bifunctional Hirsutene Synthase, 3-hydroxy-3-glutaryl-CoA (HMG-CoA) synthase (hereafter HS-HMGS). HMG-CoA synthase is the second enzyme in the isoprenoid precursor-producing mevalonate pathway (MVA). To the best of our knowledge, this is the first example of a bifunctional enzyme catalyzing non-consecutive catalytic steps in an isoprenoid pathway. With the successful molecular cloning of HS-HMGS, we have now identified all STS producing the major sesquiterpenes detected in *S. hirsutum*

cultures, as well as isolating the first known hirsutene synthase. Finally, we demonstrate that HMG-CoA synthase duplications in isoprenoid biosynthetic gene clusters are found throughout the (sequenced) fungal kingdom, suggesting such duplications are an evolutionarily conserved mechanism of increasing MVA pathway flux for the production of isoprenoids in fungi.

3.2. Materials and methods:

3.3.1 Strain growth, headspace analysis, mRNA extraction, cDNA preparation

Fungal strains, growth conditions, culture analysis, mRNA extraction and cDNA preparation were prepared as described previously.[358] Briefly, *Stereum hirsutum* FP-91666 SS-1 was grown at 22 °C in Rich medium (2% malt extract, 2% glucose, 0.5% peptone) in the dark for three weeks in 100 mL liquid cultures (500 mL flask). Tissue aliquots were ground in liquid N_2 by sterile mortar and pestle for mRNA extraction using TRIzol (Life Technologies, Grand Island, USA). cDNA was then prepared using the Superscript III First Strand Synthesis System (Life Technologies) utilizing oligo$(dT)_{20}$ primers for RT-PCR, followed by RNase H treatment.

Plasmid construction and maintenance utilized *Escherichia coli* JM109 cells, grown in LB media at 37° C, supplemented with the appropriate antibiotic(s) ampicillin (100 µg mL^{-1}), chloramphenicol (50 µg mL^{-1}) zeocin (50 µg mL^{-1}), or kanamycin (100 µg mL^{-1}). Initial *in vivo* expression of HS-HMGS and HS were performed in *E. coli* C2566 cells grown at 30° C in 50 mL LB in 250 mL shake flasks.

3.3.2 Gas Chromatography/Mass Spectrometry Analysis of volatiles:

Volatile sesquiterpenes were extracted from culture flask headspace as described previously by inserting a 100 µm polydimethylsiloxane solid phase microextration (SPME) fiber through the tin foil seal for 10 minutes.[358] Five hundred microliter *in vitro* reactions were performed overnight in sealed GC/MS vials in activity buffer (20 mM Tris pH 8, 10 mM $MgCl_2$, 500 mM NaCl, 2 uM FPP), extracted by PDMS SPME for 10 min and analyzed by GC/MS. The oven temperature started at 60 °C and was increased at 6 or 10 °C min^{-1} to a final temperature at 250 °C.

3.3.3 LC-HRMS Metabolic Fingerprinting

Fourteen-day *S. hirsutum* cultures grown in Rich medium were cultivated and confirmed to produce volatile sesquiterpenoids by SPME as described above. Ten milliliters liquid supernatant were decanted, centrifuged at 2655 g for 10 min in an Eppendorf (Hamburg, Germany) 5810R centrifuge, and clarified by vacuum filtration using qualitative filter paper (Cat. No. 28320-041, VWR. Radnor, PA). One-half milliliter clarified supernatant was overlayed with an equal volume of CHROMOSOLV ethylacetate and vortexed for 15 sec in a sealed glass GC/MS sample vial, as described previously.[365] The resulting emulsion was allowed to separate at room temperature overnight, the organic layer removed by Pasteur pipette, and placed in GC vials with glass insert liners for Liquid chromatography-high resolution mass spectroscopy (LC-HRMS) analysis.

Metabolic fingerprints were generated on an Ultimate 3000 UHPLC (ultra-high performance liquid chromatography) system with diode array detector system (Dionex/Thermo Fisher Scientific, Bremen, Germany) coupled to a Quadrupole-Orbitrap hybrid mass spectrometer. For analysis, 1 microliter of each sample was injected at 400 ul/min onto an ACQUITY UPLC C_{18} HSS T3 Column, 100Å, 1.8 µm particle size, 2.1 mm X 100 mm column (Waters, Milford, MA, USA) with column temperature of 40°C. A 20-minute gradient with mobile phase A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile) with the following gradient: initial, 10% B; 1-18min, 10-98% B; 18-20min 98% B. The UHPLC was coupled to a Q-Exactive mass spectrometer (Thermo Scientific, Bremen, Germany) with a heated electrospray ionization (HESI) source. The following MS conditions were used: full scan method acquiring positive and negative simultaneously, a scan range of 150-2000 m/z, a resolution of 35,000 (at m/z 200), target automatic gain control (AGC) of $3x10^6$, and maximum fill times of 100 ms. Data were collected and visualized in Xcalibur software version 2.2 (Thermo Scientific, Bremen, Germany). Putative compound assignments were made by comparing calculated exact masses of known *S. hirsutum* sesquiterpenoid [358] adducts to observed masses, as well as by comparison to the MetLin Library. [366]

3.3.4 Gene prediction, protein alignment, cloning of Ste5-HMGS, plasmid construction

Initial STS gene prediction, alignment, and potential transcript identification was performed as described previously [358]. Briefly, transcript predictions were made in Augustus using the region 10-15 kb up-downstream of the putative STS. Putative terpene synthases were aligned in MEGA6[315] to identify the most likely transcripts encoding functional STSs. However, this initial gene prediction failed to clone a functional STS based on the putative gene Stehi1|52743. Progressively larger portions of the Stehi1|52743-Stehi1|53380 ORF were cloned from cDNA using primers 1-4 (**Table S3-1**) until the complete ORF encoding HS-HMGS was amplified as a single transcript with primers 5-6 using touchdown PCR (68-53°C annealing temperature) and the Phusion High-Fidelity PCR Kit (New England Biolabs [NEB], Ipswich, MA). The resulting spliced PCR product was gel purified and cloned using the TOPO cloning system (Life Technologies) for sequencing.

The HS-HMGS ORF was subsequently PCR amplified using primers 7 and 10, adding the BamHI and NotI sites for subcloning into a similarly digested pUCBB plasmid [345] for expression in *E. coli*. A truncated hirsutene synthase (HS) was created by replacing the predicted HMGS start-codon methionine residue (M365) with an in-frame XhoI site to enable addition of a C-terminal His6x-tag using primers 7 and 8 upon subcloning into the pUCBB plasmid. Primers 9 and 10 were used to create a truncated HMGS (HMGS) by inserting a BamHI site upstream of the M365 residue, removing the *sh*HS subunit from the ORF. Primers 11-12 were used to generate *HS-HMGS* with flanking NdeI/NotI sites for insertion behind the NtHis6 tag of the pUCBB-Nthis6 plasmid [345] and the pET28a vector. Primers 11 and 13 were used to amplify a truncated *HS* with an artificial stop codon, which was also ligated into the pET28a vector between the NdeI and NotI restriction sites. All ligations, restriction digests followed recommended conditions described by NEB.

The yeast-expression plasmid pESC-his-HS-HMGS was made via yeast recombination cloning. pESC-his was digested with EcoRI and SpeI. The HS-HMGS ORF with YRC overhangs was amplified with primers 14 and 15 from the plasmid

pUCBB-CT5-Nthis6-HS-HMGS. Ten microliters of PCR insert was combined with 25 ul gel-purified vector digest and transformed into yeast CEN.PK2 (diploid)[56] using the LiAc/SS carrier DNA/PEG method [367], plated on SD-his selective plates and incubated for 2 days at 30° C. Yeast colonies containing pESC-his-*HS-HMGS* plasmids were then patched onto SD–his plates and grown overnight at 30° C. Plasmid isolation was performed on shake tube cells treated with 2 mg/mL zymolyse (1.2 M sorbitol, 0.1 M KPO_4,pH 7.5), incubated for 2 hours at 30°C, followed by DNA isolation using the Wizard Plus SV miniprep DNA purification system (Promega, Madison, WI). Low-concentration plasmids isolated from yeast shake tubes were then transformed into *E. coli* JM109 to obtain high concentration plasmids as necessary for sequencing and plasmid preservation.

Primers 16-17 (**Table 3-S1**) were designed using NEBuilder to generate the desired alanine replacement mutations *via* Q5 Site-Directed Mutagenesis (New England Biolabs, Ipswich, MA), using pESC-*HS-HMGS* as the initial PCR template, to produce pESC-HS-HMGS C133A.

3.3.5 Analysis of HS-HMGS gene splicing, protein sequence, and 3D structure modelling

Gene structures were visualized using Gene Structure Display Server (GSDS 2.0) [368]. Protein alignment and evolutionary history trees were performed using MEGA6 [315]. Proteins were aligned to the respective HS-HMGS subunits using MUSCLE [369]. Primary amino acid sequence alignments were exported to Jalview 2.0 [370], with known active site residues manually annotated (black boxes), and colored by the ClustalX convention for similar amino acid properties. The evolutionary analysis of the HS subunit against known fungal and bacterial 1,11-cyclizing STSs was inferred using the Maximum Likelihood method based on the JTT matrix-based model with 500 bootstrap replications. [371, 372] All positions with less than 95% site coverage were eliminated. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The analysis of 1,11 cyclization STS involved 38 amino acid sequences from *C. cinerea[83], O. olearius [89],*

S. hirsutum [358], Fusarium sp.[153, 154, 207], *A. terreus*[138], *P. roqueforti*[134], Streptomycetaceae [360-362, 373], A. gallica [162], and S. cerevisiae outgroups. There were a total of 261 positions in the final dataset.

The HMGS subunit phylogeny was similarly inferred by using the Maximum Likelihood method based on the JTT matrix-based model.[371] The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed.[372] Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model. The analysis involved 16 amino acid sequences. The protein sequences used for this alignment were obtained from several sources. *Homo Sapien h*HMGS1-2 [374], *Brassica juncea bj*HMGS [375], *Staphylococcus aureus sa*HMGS [376], *Enterococcus faecalis ef*HMGCS)[377], *Hevea brasiliensis hb*HMGS[378], *Phycomyces blakeseeanus* (*hmgS*)[379], *S. pombe* [380] and *Ganoderma lucidum* HMGS [381] sequences were obtained from GenBank. *Ustilago maydis* hcs1 sequence [382] (UMAG_05362) was obtained from EMBL, and YML126c (ERG13) from the Saccharomyces Genome Database (http://www.yeastgenome.org). *S. hirsutum* putative HMGS enzymes Stehi1|166728, Stehi1|119554 were identified by BLAST search of the *S. hirsutum* FP-91666 SS1 v1.0 genome sequence (genome.jgi.doe.gov) using yeast HMGS ERG13 (Yml126c), as was the *Phaffia rhodozyma* HMGS (jgi.doe.org). *P.rhodozyma* HMGS has not been cloned and sequenced in its entirety, but the gene prediction is consistent with RT-PCR probes described by Miao et. al. (2011)[383]. All positions with less than 95% site coverage were eliminated. There were a total of 274 positions in the final dataset.

Remote homology detection and 3D modelling was performed using the web interface for Phyre2 [384], by uploading the STS subunit of HS-HMGS (HS AA1-365). The crystal structure with the highest identity (21%) aligned in this way was *Aspergillus terreus* Aristolochene synthase (atAS) with the inhibitor FPS bound in the active site (4KUX) [139]. Seven additional structures were modelled in this way with greater than 281/365 residues aligned, confidence of 100% and ID greater than 15%. There were no

significant differences in positions of conserved amino acid residues between models, so the model with the highest sequence identity, atAS, was chosen for detailed analysis using Pymol (www.pymol.org). 3D model reliability was supported by a RMS of 0.377 (165 to 165 atoms). The resulting model was CONSURF colored using the 13 1,11 cyclizing STS created above. Complete agreement between conserved motif orientations in the primary amino acid sequence and the modeled structure between atAS and HS support the reliability of the 3D model generated using Phyre.

3.3.6 HS-HMGS purification and *in vitro* analysis

Rosetta™ (EMD Millipore, Billerica, MA) *E. coli* cells containing pET28a-SH-HMGS were cultivated in 500 mL LB medium with appropriate antibiotics (2 L flask) at 30°C and induced with 0.5 mM IPTG at OD_{600} of ~0.3. Cultures were then incubated 3 hours post-induction, monitored for hirsutene production by SPME as described above, pelleted by 10 min centrifugation at 6,000 rpm in a Beckman (Brea, CA) J2-HS floor centrifuge and immediately frozen at -80°C. Pellets were resuspended in 30 mL STS purification buffer (50 mM Tris, pH 8, 250 mM NaCl, 5 mM imidazole, 10 mM $MgCl_2$, 1 mM PMSF) supplemented with cOmplete protease inhibitor (Sigma-Aldrich, St. Louis, MO) for 6-minute sonication (1 second on, 2 seconds off), followed by 30 min centrifugation at 12,000 rpm. Soluble lysate was then purified using TALON resin (Takara Clonetech, Mountain View, CA), eluting into STS elution buffer (50 mM Tris, pH 8, 250 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$, 1 mM PMSF). Eluent was then concentrated using Amicon (EMD Millipore, Billerica, MA) Ultra -15 10,000 NMWL centrifugation filter units and desalted using PD-10 (GE Healthcare) columns. Partially purified, imidazole desalted HS-HMGS was then assayed for activity by $P_i$per pyrophosphate assay (ThermoFisher Scientific, Waltham, MA) following manufacturer instructions over a range 0-100 μM FPP, and in the presence of 0.5 mM $MgCl_2$. Fractions were analyzed for hirsutene synthase activity by *in vitro* GC/MS assay as described above, 10% SDS-PAGE, and western blot. Western blots were performed using monoclonal Mouse IgG_1 clone anti-his tag primary antibody (catalog # MAB050, Bio-Techne, Minneapolis, MN). Goat anti-Mouse IgG (H + L) human ads-HRP (cat 1031-05,

lot k1012-PM44C) was used as a secondary antibody, and visualized using Supersignal West Femto maximum sensitivity substrate detection (ThermoFisher Scientific) and exposure to GeneMate Blue Basic Autorad (BioExpress, Kaysville, UT) film. Protein concentration was determined by Bradford assay.

3.3.7 Yeast HMG-CoA Synthase complementation

The Heterozygous diploid HMGS knockout YML126c (ERG13$^{\Delta}$::KanMX) was obtained from the GE Dharmacon (Lafayette, CO) yeast knockout collection [385] and cultivated in 5 mL sporulation medium (10 g L^{-1} KoAc, 1 g L^{-1} yeast extract, 0.5 g L^{-1} dextrose) for 9 days at 30° C. Tetrad formation was confirmed by light microscopy. Tetrads were subsequently treated with 0.5 mg mL^{-1} zymolyase 20T until ascus dissolution was visually confirmed, and enriched for random spore analysis before plating on selective YPD + 200 µg mL^{-1} G418, with replica plating on non-selective (YPD) and ERG13 auxotroph selective plates (YPD + G418, + 5 mg mL^{-1} mevalonate).[386] Small colonies growing on the ERG13 auxotroph plates were replica plated on G418 + mevalonate plates and G418 plates with het/dip controls to confirm mevalonate auxotrophs. G418 resistant, mevalonate auxotroph isolates were then PCR screened for KanMX insertion at the correct genomic location using primers described by Winzeler et. al (1999), [385] as well as for mating type using Mat*a/α* locus primers described by Huxley (1990) to confirm the haploid and mating type.[387] The same Mat*a* haploid ERG13Δ::KanMX strain was then transformed with either pESC-his-HS-HMGS, pESC-his-HS-HMGS C133A and replica plated on synthetic minimal dextrose growth (SD-his) or synthetic minimal galactose (SG-his) expression plates to test for complementation by HS-HMGS and inactivated HS-HMGS C133A [381]. A minimum three isolates of each plasmid and haploid/diploid combination were tested for complementation.

3.3.8: Construction of yeast and *P. pastoris* expression vectors pPICZA-nthis6x-SH-HMGS 2µOri-URA3

All yeast expression vectors were synthesized using yeast homologous recombination cloning (YHC). pPICZA was digested with XhoI and NotI. The 2µOri-

Ura3 cassette was amplified from pESC-Ura with primers 18 & 19 to generate overhangs essential for YRC. Ten microliters of PCR insert was combined with 25 µl gel-purified vector digest and transformed into yeast CEN.PK2 (diploid)[56] using the LiAc/SS carrier DNA/PEG method,[367] plated on SD-ura selective plates and incubated for 2 days at 30° C. Yeast colonies containing pPICZa-2uOri-Ura3 plasmids were then patched onto SD –Ura plates, and used to inoculate 4 mL SD –Ura shake tubes grown overnight for plasmid isolation. All yeast growth was at 30° C. Plasmid isolation was performed on cell pellets using 2 mg/mL 20T zymolyse (200 ul in 1.2 M sorbitol, 0.1 M KPO_4,pH 7.5), incubated for 2 hours at 37°C, followed by DNA isolation using the Wizard Plus SV miniprep DNA purification system (Promega, Madison, WI). Low-concentration plasmids isolated from yeast shake tubes were then transformed into *E. coli* JM109 to obtain high concentration plasmids as necessary. pPICZA-2uOri-Ura3 was then digested with XhoI and combined with the product of primers 18 and 19 (to generate HS-HMGS containing 40nt overhangs to the host vector) and transformed into yeast as described above. pPICZa-2uOri-HS-HMGS was digested with NotI, purified by 0.7% SDS-PAGE and Wizard SV Gel and PCR Cleanup kit (Promega) to cleanly remove the 2uOri-Ura3 cassette, creating pPICza-NtHis6-HS-HMGS. The miniprepped plasmid was confirmed by PCR screen with primers 22, 23 for removal of the 3 Kb 2µOri-Ura3 fragment.

S. cerevisiae expression plasmid pESC-NtHis6x-HS-HMGS was created using primers 14 and 15 to amplify the NtHis6x-HS-HMGS ORF from pPICZa-HS-HMGS with pESC-his MCS1 overhangs flanking the EcoRI and SpeI sites.

3.3.9: Integration and expression of HS-HMGS in *Pichia Pastoris*

Pichia Pastoris X-33 was transformed using the EasyComp® method (Life Technologies), linearizing all pPIC plasmids using PmeI. Successful integration was confirmed by zeocin resistance, first screened colony PCR using a forward primer upstream of the PmeI linearization site, and a pPIC specific reverse primer (primers 23-24), testing for integration. This result was then confirmed using the upstream of PmeI F primer, and the gene-specific HMGS ORF R primer (# 6). Zeocin-resistant isolates were

then replica plated on Minimal Methanol and Minimal Dextrose plates to confirm preservation of the $MutS^+$ phenotype.

3.3.10: Extraction and analysis of sesquiterpenes produced by *S. cerevisiae* expressing HS-HMGS and HS-HMGS C133A

Four independent *S. cerevisiae* CEN.PK2 transformant colonies with either pESC-HS-HMGS or pESC-HS-HMGS C133A were each inoculated in triplicate (n = 4, 3 technical replicates) into 14 mL conical shake tubes overnight in 3 mL SD – his (1.67 g L^{-1} Yeast Nitrogen Base (Sigma-Aldrich, St. Louis, MO), 1.4 g L^{-1} Yeast DropOut Supplement (Sigma-Aldrich), 5 g L^{-1} ammonium sulfate, 20 g L^{-1} dextrose, 0.12 g L^{-1} uracil, 0.12 g L^{-1} leucine, 0.05 g L^{-1} tryptophan) and grown overnight at 30°C with 220 rpm shaking. One milliliter of each overnight culture was then injected into 50 mL SD in 250 mL non-baffled shake flasks and allowed to expand overnight at 30°C and 220 rpm. Following overnight growth, the OD_{600} of all flasks was measured using 1/10 dilution on a 96-well SpecraMax Plus 384 plate reader (Molecular Devices, Sunnyvale, CA). An appropriate volume of cells was then removed by pipette to enable resuspension in SG (same as SD, but replaces 20 g L^{-1} dextrose with 20 g L^{-1} galactose) expression medium following 2X washes in 10 mL sterile dH_2O. Each shake flask was washed 2X with dH_2O before reinoculation with the appropriate washed and SG-resuspended yeast cells. Five milliliters of dodecane (10% v:v) overlay was then added to absorb sesquiterpene products from *S. cerevisiae* CEN.PK2 cultures grown on SG medium, as described previously for other sesquiterpenes produced in liquid cultures [8, 388] with 100 µg mL^{-1}, 95% pure (-) caryophyllene oxide (Cas # 1139-30-6, Sigma-Aldrich, St. Louis, MO) as an internal control. Hirsutene concentration was quantified by comparison to a caryophyllene oxide standard curve from 9.5 µg mL^{-1} to 9.5 mg mL^{-1} in dodecane (µg mL^{-1}) = $4E^{-7}$*x(FID counts), R^2= 1, six data points from 10 µg mL^{-1} to 10 mg mL^{-1}).

Following shake flask incubation, partition to the dodecane layer was confirmed by PDMS SPME headspace extraction as described previously. *S. cerevisiae* cultures were then transferred to 60 mL test tubes and allowed to settle for approximately 10 minutes before 10 mL of the aqueous portion was extracted for OD determination, and centrifugation for later mRNA extraction. In addition, aspiration of the top ~7-10 mL of

dodecane:SG media emulsion. The resulting emulsion was centrifuged at 2655g in an Eppendorf 5810R swinging bucket centrifuge (Eppendorf, Hamburg, Germany) for 1 hour. Two hundred microliters of clarified dodecane extract was then transferred to GC vials containing glass inserts for GC/MS-FID analysis under as described above for analysis of SPME-extracted sesquiterpenoids with the following modifications. Five microliters of each dodecane extract was injected, with Methyl Tert-Butyl ether (MTBE) and ethyl acetate as the first and second wash buffers. The oven temperature started at 60 °C and was increased at 10 °C min^{-1} to a final temperature at 250 °C. The FID detector was kept at 250°C, He flow 30 mL min^{-1}, and airflow rate of 400 mL min^{-1}.

3.3.11 Identification of MVA and isoprenoid biosynthetic pathway duplications

MVA and isoprenoid biosynthetic genes were identified by BLAST search of 525 fungal genomes found in the JGI fungal genome database. To determine the extent to which HMGS gene duplications occur in fungi, and the frequency that HMGS duplications occur in isoprenoid biosynthetic gene clusters, a BLAST search was performed using the JGI: Fungi genome database for homologues to the yeast ERG13 and the two characterized Basidiomycota HMG-CoA synthases, *Ustilago maydis* Hcs1[382] and *Ganoderma lucidum* glHMGS.[381] This analysis identified 662 putative HMG-CoA synthases. The same genomes were further analyzed for all MVA pathway enzymes, as well as a subset of common isoprenyl-diphosphate utilizing enzymes including STS, GGPS-fusicoccadiene synthase gene fusions, and FPS-Squalene synthase gene fusions. E-value cutoffs were set to 10^{-3} to obtain maximum coverage, unless resulting in false-positives as described below. The terpene cyclase-PE fusion Fusicoccadiene-synthase (PrAS) has significant homology to FPP, GGPP, and STS sequences. For this reason, results were filtered to remove proteins with stronger resemblance to either FPPs, GGPPs, or STS protein subunits based on E-value. HMG-CoA lyase (HCL) enzymes share strong homology to pyruvate carboxylase (PCC) and 1-isopropyl malate synthase (LeuA). Multiple HCL sequences[389] were used as BLAST search terms, results combined, and compared E-value cutoff adjusted to 10^{-8} to remove PPC or LeuA proteins from analysis. No significant overlap was observed between other

gene types. Twelve fungal genomes auxotrophic in at least one MVA pathway gene that are known parasites or symbionts were manually removed from analysis. All genome sequences are haploid genotype or filtered bioinformatically.[390] Acetoacetyl-CoA Thiolase was omitted due to an inability to discern true MVA pathway AA-CoA thiolases from thiolases involved in other cellular processes using primary sequence alone.

3.4 Results:

3.4.1 Cloning HS-HMGS and STS gene structure

Initial attempts to clone the putative hirsutene synthase Stehi1|52743 using start/stop codon primers for a discrete gene based on the Joint Genome Initiative (JGI) Database gene prediction [113], as well as gene predictions generated using Augustus [313] were not successful in isolating spliced gene products from *S. hirsutum* cDNA. As a an alternative, primers (**Table S3-1**, 1 and 2) corresponding to the highly conserved STS DDxxD and NSE/DTE Mg^{2+} binding motifs were created. These primers produced spliced intragene DNA products 405 bases in length, easily distinguished from unspliced background amplification (760 bp) (data not shown). This intragenic amplicon was gel purified and sequenced, proving that the Stehi1|52743 gene was indeed expressed in mycelial cultures grown in YMG medium. A modified primer walking strategy was then used to gradually lengthen the target ORF to 909 BP using predicted start codons and available Expressed Sequence Tag (EST) information to amplify much of the STS subunit from its start codon to the final EST (primers 3 & 4). However, no in-frame stop codon could be obtained by PCR or using 3' RACE. With the goal of developing a better understanding of how Stehi1|52743 is spliced, the gene structures of all confirmed fungal STS were compared to aid in intron/exon prediction. Surprisingly, intron-exon patterns and phase conservation were conserved within the four previously identified initial FPP cyclization mechanisms utilized by Basidiomycota STS,(**Figure 3-2**)[89, 358, 364] as well as within the Ascomycota initial FPP cyclization groups for which STS have been characterized to date (**Table S2**). In addition, all known Basidiomycota STS except the 1,6 NPP cyclizing STS Cop1, Stehi1|159379, and HS-HMGS contain an intron between the bases coding for the N and D residues of the conserved NSE/DTE STS motif. In contrast, most Ascomycota STS lack this intron, instead largely maintaining an intron splitting the conserved DDxxD/E active site motif. Subsequent comparison of known gene structures to predicted mRNAs generated in Augustus was then used to design

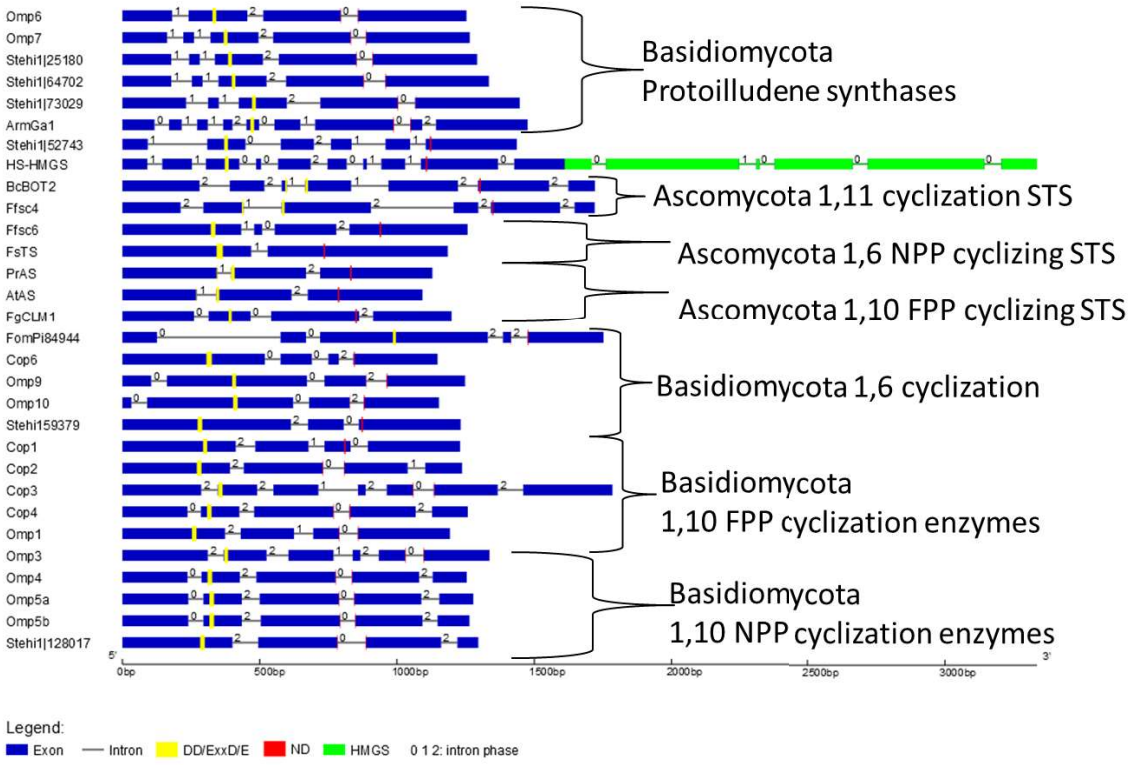


Figure 3-2: Gene structure of known fungal STS
The conserved DDxxD motif is highlighted in yellow, while the ND residues (red) of the NSE/DTE conserved motif are split by an intron in nearly all (18 of 21) Basidiomycota sesquiterpene synthase genes. The HMGS subunit of HS-HMGS, beginning in exon 10, is shown in green. Note the high gene structure and intron phase conservation within each Basidiomycota and Ascomycota cyclization group. The 5' end of the HS-HMGS gene resembles the general Basidiomycota protoilludene synthase 1:1:2 intron structure, with the DD/ExxD/E motif within exon 3. The HMGS subunit gene splicing is similar to other Basidiomycota HMGS (data not shown).

primers where Stehi1|52743 formed a gene fusion with the adjacent, downstream Stehi1|53380, a putative HMG-CoA synthase. Finally, high-fidelity PCR using forward primers corresponding to the Stehi1|52743 start codon (primer 6) with the reversed Stehi1|52743 EST F (primer 4) or a reverse primer corresponding to the HMG-CoA Synthase stop codon (primer 5) both produced discrete, spliced PCR products of the expected sizes (**Figure S3-1**), alongside amplification of the genomic-length sequence. This PCR product was gel purified, TOPO cloned, and sequenced, confirming presence of a single ORF approximately combining the Stehi1|52743-Stehi1|53380 gene annotations, hereafter referred to as Hirsutene Synthase-HMG-CoA synthase (HS-HMGS). The intron-exon pattern of the STS subunit is similar to the Basidiomycota

protoilludene synthases through the DDxxD motif on exon 3, while the remainder of the STS coding sequence generally follows the pattern of Ascomycota 1,11 cyclizing enzymes Ffsc4 (Koraiol synthase) and bcBOT2 (presilphiperfolan-8β-ol synthase). Compared to other fungal STS, HS contains an exceptionally large number of introns (9 vs. average 3.9) rare for known fungal STS, but consistent with previous reports on the intron density of another *S. hirsutum* gene, which has been attributed to a relatively recent acquisition from bacteria or another organism. [311]

3.4.2 HS protein sequence analysis

Following determination of the ORF sequence of HS-HMGS, it was necessary to analyze the protein sequence to confirm presence of conserved STS motifs and gain insights into its phyogeny. The *S. hirsutum* HS sequence diverges significantly from all previously isolated STSs, bearing the closest evolutionary relationship to bcBOT2 and FfSC4 amino acid sequences (**Figure 3-3**). which form triquinane sesquiterpenes similar in structure to hirsutene.[364] The HS phylogeny suggests a significant evolutionary distance from fungal 1,11 cyclizing enzymes, including Baisdiomycota protoilludene synthases and Ascomycota 1,11 STSs as well. The primary sequence amino acid sequence is highly divergent from protoilludene synthases, bearing only 20.8% ID to Stehi1|64702 and a similar level to (21.2% ID) to Ffsc4 (Koraiol synthase), with only slightly lower conservation to the Ascomycota 1,10 cyclizing enzyme atAS (19.2% identity). For comparison, the most divergent protoilludene synthases, Omp6 and Stehi1|64702 are 52.8% identical. However, our predictive framework still accurately predicted that HS would proceed through a 1,11 cyclization to produce a *trans*-humulyl cation intermediate *en route* to hirsutene cyclization.[89, 358] Despite this low sequence similarity to known STS, the predicted α-helical tertiary structure of HS appears to be conserved, as shown through Phyre2 protein fold recognition server (http://www.sbg.bio.ic.ac.uk/phyre2/) generated homology model of HS to the atAS

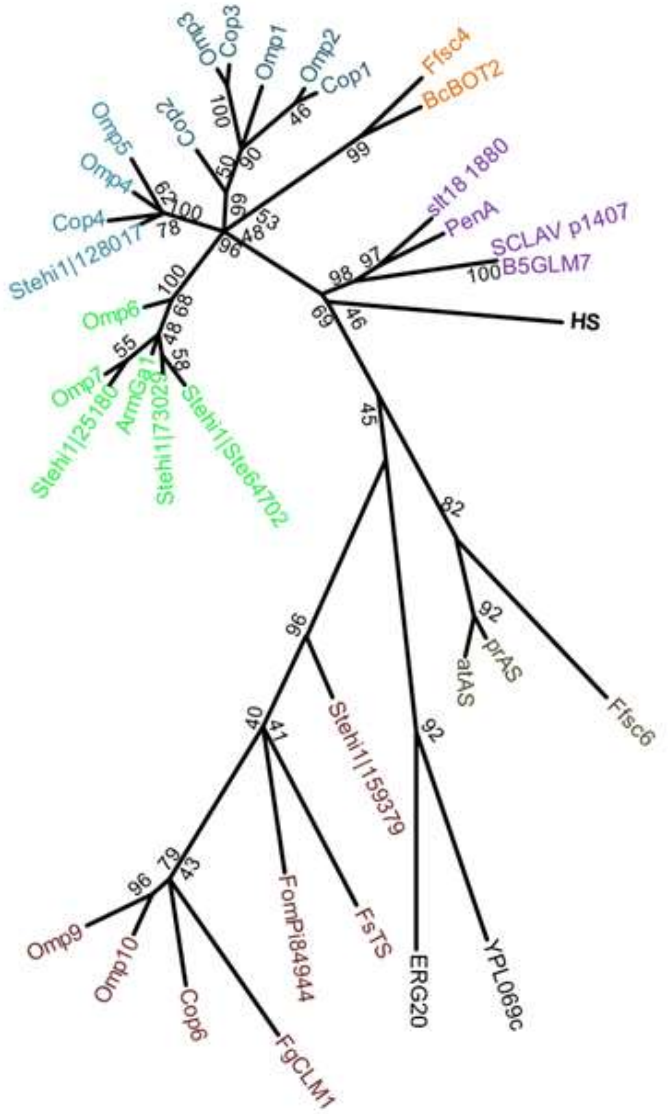


Figure 3-3: Phylogenetic analysis comparing hirsutene synthase (HS) subunit of HS-HMGS to all known fungal STS and bacterial cucumene synthases
Maximum likelihood tree shows previously described [364] clading is present among Basidiomycota employing an initial 1,10 FPP cyclization intermediate (dark blue), 1,10 NPP (blue), and 1,11 FPP cyclizing (green) protoilludene synthases. In addition, 1-6 NPP cyclizing STS from Basidiomycota and *Fusarium* sp. form a large and diverse clade (maroon). Ascomycota 1,11 FPP cyclases (orange), and bacterial 1,11 cucumene synthases (purple) form distinct clades, with HS bearing the closest phylogenetic similarity to the bacterial 1,11 FPP cyclizing STS. Ascomycota 1,10 cyclizing aristolochene synthases and 1,6 cyclzing Ffsc6 are outlyers in this analysis, while yeast FPPs and GGPPs (black) are an outgroup.

structure 4KUX (**Figure S3-4**) [138]. Low sequence identity, while retaining tertiary structure is characteristic of the terpene cyclase superfamily [357]. Combining this model with a primary protein sequence alignment to all known fungal *trans*-humulyl STSs and bacterial triquinane synthases (**Figure S3-3**) allowed prediction of active-site lining and Mg^{2+} binding residues in HS, and CONSURF color-coding of conserved residues in the protein model (**Figure S3-5a-b**). The canonical STS D^{90}DxxD and N^{216}SE/DTE Mg^{2+} binding residues are well conserved. In addition, the π-bonding residue F^{87} and pyrophosphate-H-bonding residues R^{170} and $R^{314}Y$ [140] are completely conserved across all *trans*-humulyl cyclases. In contrast, the H-α1 loop shares few residues with known *trans*-humulyl cyclizing enzymes, with only M^{226} and D^{230} bearing

similarity to other known enzymes, with the N^{233} residue completely conserved across all analyzed enzymes. The complete conservation of N^{233} reinforces its likely critical role in forming hydrogen bonds with N^{216} following FPP binding, closing the active site from water during catalysis, postulated in previous studies of the H-α1 loop.[158] Of particular interest are the hydrophobic, active site-lining residues C^{67} and W^{151}, shown in **Figure S3-5a-b** and **Figure S3-3**. There is a correlation at C^{67} between product formation and amino acid residue, where all protoilludene synthases possess an alanine, while bacterial triquinane synthases have a phenylalanine, botrydial synthase has a tryptophan residue, and koraiol synthase cysteine residues at this position. Also of note, HS W^{151} diverges from the otherwise completely conserved tyrosine residue found in *trans*-humulyl cyclases. W^{151} is analogous to atAS F^{147}, which is thought to stabilize the prenyl chain through π-bond interactions, making it important for product specificity in atAS [139]. Of the known fungal STS, only the 1,6-cyclizing Fompi1 (cuprenene synthase) [89] contains a tryptophan residue at this position, while all others possess a tyrosine or phenylalanine residue, consistent with their expected role in forming π-bonding interactions with the prenyl chain.

3.4.3 HS-HMGS is a hirsutene synthase

Both a wild-type HS-HMGS and the truncated STS subunit (HS) were subcloned into the constitutive expression plasmid pUCBB [345]. Each construct was then expressed in *E. coli* and the volatile headspace sampled by Solid-Phase MicroExtraction (SPME). HS-HMGS and HS *in vivo* expression (**Figure 3-4**) resulted in nearly identical titers and relative quantities of the *trans*-humulyl derived sesquiterpenes hirsutene (85.3%), panaginsene (7.5%), (E)-β-caryophyllene (0.6%), and α-humulene (3.1%), in similar relative concentration to those previously observed in *S. hirsutum* headspace [358]. The remaining 3.5% is divided between eight compounds with the characteristic sesquiterpene 204 m/z parent ion, but could not be identified by comparison to the Mass Finder v4 Essential oils Database [353], or any available standards. This result demonstrates that the product profile of the N-terminal hirsutene synthase subunit of HS-HMGS is independent of the C-terminal HMGS subunit. However, subtle allosteric

effects on enzyme kinetics, such as those observed in the bifunctional fusicoccadiene synthase-GGPP protein (PaFS), [391] cannot be ruled out at this time.

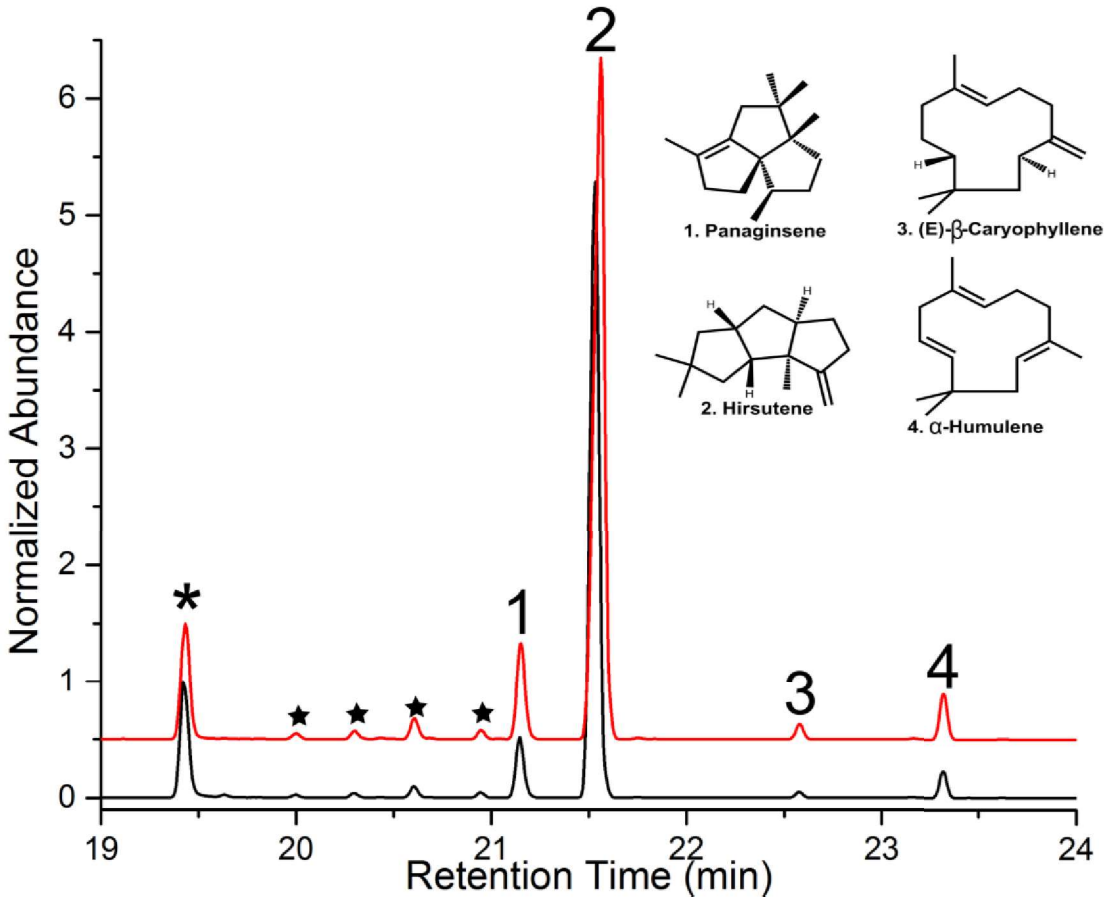


Figure 3-4: HS-HMGS GC chromatgraph comparing HS-HMGS to the HS-only artificial truncation.
HS-HMGS (black) and HS (red) produce approximately the same quantity of sesquiterpenes with the same product profile in MEP-containing *E. coli*, demonstrates that the HMGS subunit has no direct effect on the activity of HS. Indole (*), a breakdown product of tryptophan, is a natural product of *E. coli* and serves as an internal standard. Peaks marked with a star (★) possess the characteristic sesquiterpene parent ion m/z 204, but could not be identified. The chromatograph of HS-HMGS was offset by 0.5 normalized absorbance units for visibility.

Site directed mutagenesis was performed using the truncated HS subunit to test the effect on the product profile of mutations to active site residues C^{67} and W^{151}, due to their predicted role interacting with the prenyl chain during cyclization and aforementioned correlation with product formation. No mutations made to C^{67} and W^{151} residues affected the product profile of HS, while only W151G/Y retained activity (**Table S3-3**). HS W151Y exhibited 4X higher *in vivo* hirsutene production, while HS W151G hirsutene

production decreased 1/3 when expressed in *E. coli* (**Figure S3-6**). No further experiments were performed because no mutations affected product profile, only apparent *in vivo* productivity.

3.4.4 Purification and *in vitro* analysis

Extensive purification optimization revealed both HS-HMGS and its individual subunits to be poorly expressed, largely insoluble, and rapidly degraded *in vitro* when expressed in multiple standard *E. coli* protein expression strains, including BL21(DE3), Tuner, and C2566. This effect was also observed regardless of promoter and inducer concentration, including Plac*, pTrc, and T7 promoters. In all cases, typical sesquiterpene productivity was readily detected in culture headspaces, but protein purification yields remained low and unstable, losing all activity regardless of purification level, buffer, or protease inhibitor mere hours after cell lysis. Different N- or C-terminal purification tags (his6x, Trx tag) had no apparent effect on product profile, expression, or solubility. Compared to previously purified, highly soluble and highly expressed Stehi1|73029, there is no significant difference in codon usage, leading to the conclusion that the observed low accumulation of protein in lysed cells would not be improved by codon optimization, but is instead likely a property of the translated protein itself, such as inherent instability or targeted degradation by cellular proteases. Analysis of the HS model in PyMol did not identify hydrophobic surfaces expected for membrane insertion or aggregation driven by hydrophobic interactions. Controlling for degradation by *E. coli* proteases, HS-HMGS was integrated and expressed in *Pichia pastoris* X-33 using the pPIC vector system (ThermoFisher) behind the methanol-inducible promoter. This, however, failed to improve yield, solubility, or significantly reduce degradation in fractions from following Ni^{2+} or size-exclusion purification. Washes of dialysis tubing and Amicon concentrators with 1% Triton and analysis by SDS-PAGE and Western Blot did not detect entrained HS-HMGS, suggesting against aggregation or precipitation following initial Ni^{2+} purification as an explanation for the relatively rapid loss of intact HS-HMGS, as determined by GC/MS activity assays and SDS-PAGE/Western blot.

Finally, to minimize degradation or inactivation, HS-HMGS was partially purified as rapidly as possible at 4° C from Rosetta pET28a-HS-HMGS cells using TALON Resin, concentrated by Amicon filter, and immediately desalted by PD-10 column, to remove imidazole, in the presence of protease inhibitors. Following desalting, partially purified fractions were immediately assayed *in vitro* with 2 µM FPP in GC/MS tubes, total protein concentration was measured by Bradford assay, and apparent enzyme kinetics were determined by $P_i$per pyrophosphate assay. Western blot confirmed presence of HS-HMGS in concentrated supernatant, as well as detectable levels in PD-10 eluent (**Figure S3-7**). The GC/MS sesquiterpene product profile of this partially purified fraction (**Figure 3-5**) was unchanged from the *in vivo* profile previously observed in culture headspaces (**Figure 3-4**). Finally, STS kinetic parameters were determined by measuring pyrophosphate release from FPP by enzyme-coupled $P_i$per assay. HS-HMGS K_m was determined to be 6.3 ± 3.3 x10^{-6} M, with an apparent K_{cat} of 4.0 x10^{-3} ± 0.5 s^{-1} (K_{cat}/K_m = 6.30 x10^2 ± 1.40 x10^2 $M^{-1}s^{-1}$). HS-HMGS kinetic parameters are typical of previously characterized STS from *S. hirsutum*, with K_m and K_{cat} within an order of magnitude of those for Stehi1|25180.[358] Kinetic analysis for HMG-CoA synthase activity was also attempted via Ellman's reagent detection of CoA-SH release, however activity was not statistically significant above background.

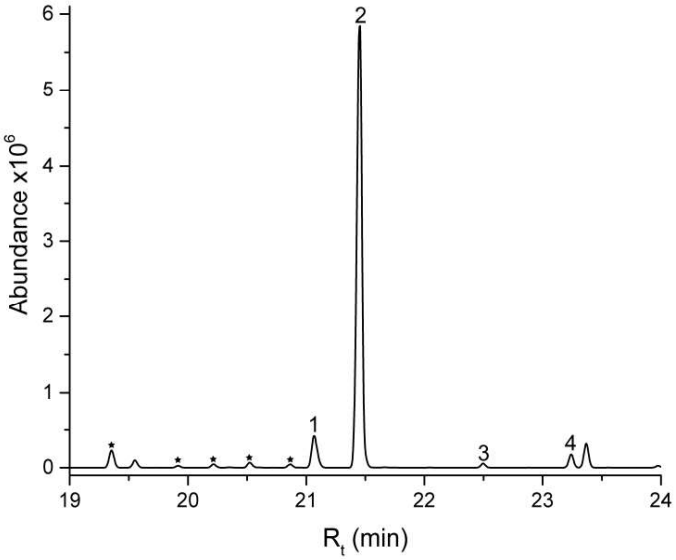


Figure 3-5: *In vitro* product profile of partially purified HS-HMGS matches *in vivo* production.
Partially purified HS-HMGS was incubated with 2 µM (2E, 6E)-FPP, and the headspace was sampled by SPME. Peaks marked with a star (★) are unidentified sesquiterpenes defined by an m/z of 204.

3.4.5 HMGS sequence analysis and predicted function

The 3' region of the HS-HMGS ORF, from nucleotide 1093 to 2526, encodes an HMG-CoA Synthase (HMGS), with all the necessary catalytic residues being conserved. The ancestral HMGS start codon is conserved as M365 of the HS-HMGS gene fusion, enabling identification and cleavage of the distinct hirsutene synthase-HMGS subunits. HMGS catalyzes a three-step ping-pong reaction consisting of an initial acylation of the catalytic cysteine, followed by acetoacetyl-coA condensation with the acyl-S-enzyme intermediate, and finally hydrolysis to release HMGS-CoA from the enzyme [28]. HMGS catalysis is dependent on a catalytic triad of conserved $C^{133}H^{280}$-E^{99} residues (HMGS numbering) [374]. Additional literature has expanded upon this conserved catalytic triad, identifying the residues Y^{134}, A^{357}, and S^{382} as being conserved in known HMGS sequences [376, 381], residues also conserved in HS-HMGS. Notably, homologous residues to HMGS C^{133} of the conserved catalytic triad have been shown in human

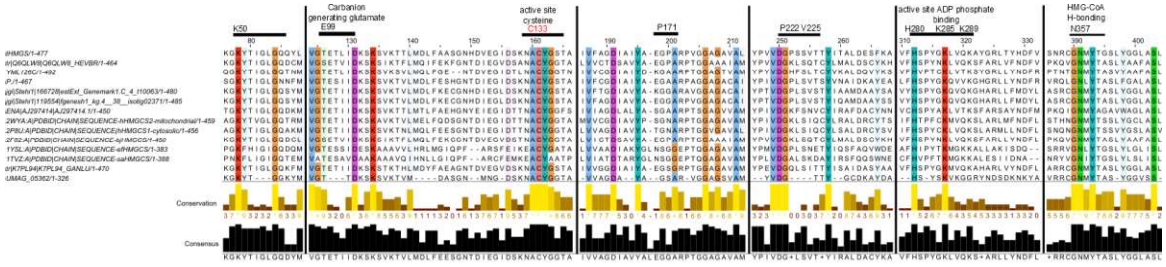


Figure 3-6: HMGS sequence alignment with functionally characterized HMG-CoA synthases.
All active site residues are conserved between *HS-HMGS* and known HMGS eukaryotic and bacterial HMGS, and no analogous mutations known to reduce hHMGS2 activity are present in *HS-HMGS*[374]. In addition, completely conserved C^{133} (red) correspond to *hb*HMGS C^{117}, where alanine mutations were previously shown to eliminate *in vitro hb*HMGS acetyl-coA binding and catalysis.[378]

HMGCS1/2 crystal structures to be acetylated by acetyl-CoA as the initial step in HMG-CoA synthesis [374]. Mutagenesis of this active site cysteine to serine or alanine is known to abolish HMG-CoA synthesis, as initial enzyme acylation is no longer possible without the reactive cysteine thiol functional group [378, 392]. To analyze HMGS for conservation of residues critical to HMG-CoA synthase activity, a primary sequence alignment was made, comparing HMGS (HS-HMGS from M365-H841) against characterized HMG-CoA Synthases from the plants *Arabidopsis thaliana* [393], *Hevea brasiliensis* [378], and plant *Brassica juncea* [375], Basidiomycota *Ustilago maydis* ([382] and *Ganoderma lucidum* [381], Zygomycota *Phycomyces blakesleeanus* [379],

Ascomycota *S. cerevisiae* and *Schizosaccaromyces pombe* [380], the human mitochondrial and cytosolic HMGS enzymes [374], and finally gram-positive bacteria *Enterococcus faecalis* [377] & *Staphylococcus aureus* [376] (**Figure 3-6**). In addition, the *S. hirsutum* genome contains three HMG-CoA synthases (Stehi1|168705, Stehi1|166728, and Stehi1|119554) in addition to HS-HMGS, the latter two genes are found flanking the previously identified protoilludene synthase Stehi1|25180, [358] while the former is not in a known gene cluster type. Protein alignment demonstrates conservation of all known substrate-binding and catalytic residues [374, 378, 379, 383], as well as absence of analogous mutations previously shown to cause clinical deficiency in *h*HMGCS2 [374] in all *S. hirsutum* HMG-CoA synthases found in STS gene clusters. Subsequent phylogenetic analysis (**Figure 3-7**) indicates the expected close evolutionary relationship within the Basidiomycota domain by these three *S. hirsutum* HMGS, and the previously characterized HMG-CoA synthases from in *Phaffia rhodozyma,* [383] *Ganoderma lucidum,* [381] and *Ustilago maydis.* [382] In addition, progressively more distant relationships between a clade formed by Ascomycota, mammalian, plant, and Zygomycota HMG-CoA synthases, the unique HMG-CoA synthase from *U. maydis*, and

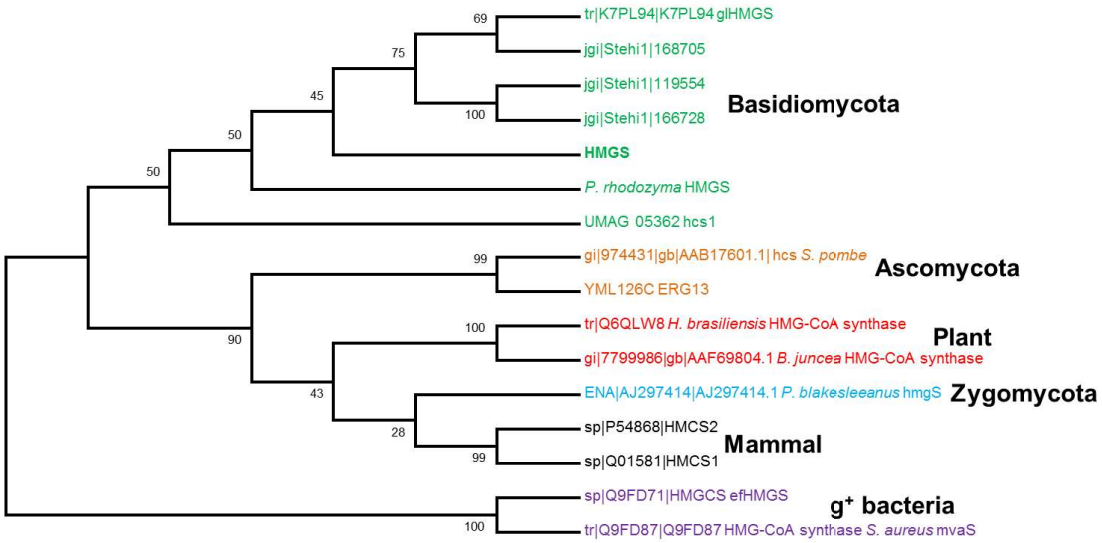


Figure 3-7: *S. hirsutum* HMGS molecular phylogeny.
Maximum likelihood, unrooted phylogenetic alignment showing HMGS shares phylogenetic conservation with Basidiomycota HMG-CoA synthases. jgi|Stehi1|168705 and jgi|Stehi1|119554 flank the protoilludene synthase Stehi1|25180 within a complex sesquiterpenoid biosynthetic gene cluster. More distal relationships between HMGS and clade formed by Ascomycota, mammalian, Zygomycota, and plant HMG-CoA synthases, while bacterial HMG-CoA synthases serve as an outgroup.

finally gram-positive bacterial HMGS enzymes are readily apparent, consistent with the evolutionary pattern described elsewhere.[379] Analysis of the amino acid sequence for localization tags using TargetP1.1 [394] did not identify localization tags in any HMG-CoA synthases except the human mitochondrial hHMGCS2, suggesting that all fungal and specifically *S. hirsutum* HMGS are cytoplasmically localized. Cytoplasmic localization is consistent with anabolic activity in the MVA pathway to the exclusion of the mitochondrially-localized ketogenesis pathway found in "higher" eukaryotes.[374] In summary, HMGS and the putative HMG-CoA synthases found in the Stehi1|25180 protoilludene gene cluster have all the characteristics of active, cytoplasmically localized HMG-CoA synthases involved in the MVA pathway.

3.4.6 Functional complementation of HS-HMGS in yeast

To determine the function of HS-HMGS, the auxotrophic HMGSΔ haploid (YML126) was obtained and transformed with pESC-his-HS-HMGS or the inactivated pESC-his-HS-HMGS C133A for expression behind a galactose inducible promoter. The resulting transformants were then cultivated on non-inducing Synthetic Dextrose (SD) medium or Synthetic Galactose (SG) induction medium for 3 days, with the het/dip parent similarly transformed as a positive control. Haploid HMGSΔ yeast expressing HS-HMGS on SG medium grew readily, while HMGSΔ haploid containing pESC-his-HS-HMGS did not grow on non-inducing SD **(Figure 3-8)**. Similarly, auxotrophs containing HS-HMGS C133A did not grow on either inducing or non-inducing conditions, demonstrating the C133A mutation eliminates HMG-CoA synthase activity in HS-HMGS, consistent with the expected role of C133 as a critical residue in the catalytic triad of a HMGS.

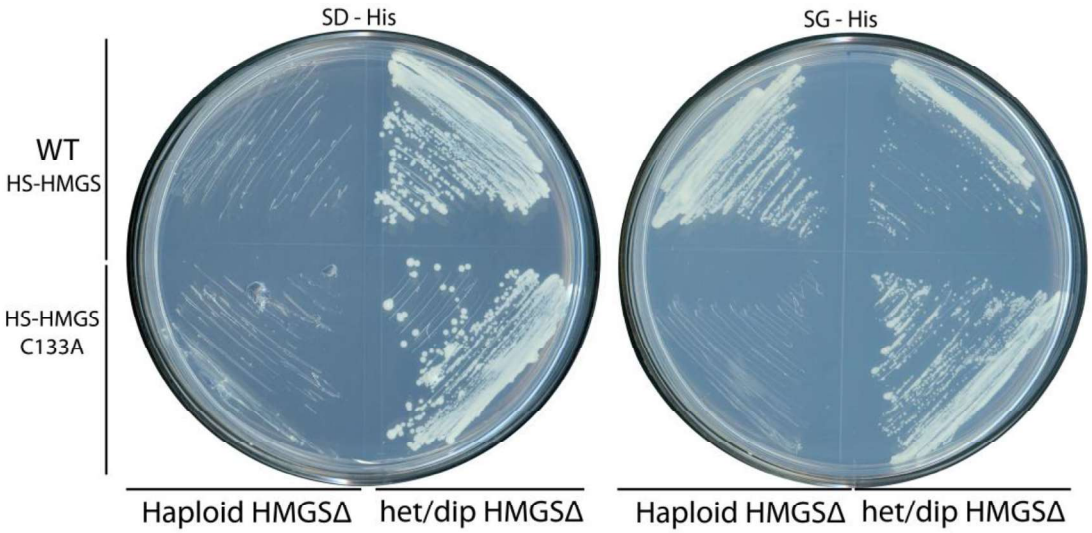


Figure 3-8: Functional complementation of Yeast ERG13 (HMGS) haploid with HS-HMGS expression.
Haploid HMGSΔ and heterozygous diploid (HMGS/HMGSΔ) parent were transformed with pESC-His-HS-HMGS or inactivated pESC-his-HS-HMGS C133A and streaked onto SD-his and SG-his and grown for 3 days at 30°C. Het/Dip positive controls grew in all conditions, while only haploids expressing WT HS-HMGS restored growth. The plates above are representative; a minimum of three isolates of each genotype were tested.

3.4.7 HS-HMGS is in a large biosynthetic gene cluster:

Most sesquiterpenoids isolated from *S. hirsutum* are derived from the hirsutene scaffold (**Figure 3-1**), requiring *S. hirsutum* to not only express at least one active hirsutene synthase, but also co-express a multitude of hirsutene modifying enzymes to produce the final, often bioactive, hirsutenoids [168, 170, 359]. LC-HRMS fingerprinting of *S. hirsutum* mycelial culture extracts enabled putative identification of several modified hirsutenoids, including hirsutic acid, and hirsutenol A, and complicatic acid (**Figure 3-9**). It is well-established that genes encoding enzymes in fungal secondary metabolic pathways and transport proteins are often co-located and in the genome, enabling study of the natural biosynthetic pathway by manipulation the neighboring genes once the cluster has been demarcated [111, 201, 216, 348, 395]. Most commonly, the first modification following sesquiterpene cyclization is oxygenation by one or more cytochrome P450 monooxygenases encoded within the gene cluster, lending their

presence particular importance to defining terpenoid biosynthetic clusters [83, 156, 192]. Manual re-prediction of the genomic region ± 50 KB on either side of HS-HMGS using Augustus [313, 314] identified a large, putative sesquiterpenoid biosynthetic cluster (**Figure 3-10**). The HS-HMGS gene cluster contains at least 19 putative biosynthetic genes, including five putative P450 monooxygenases, one putative Berberine Bridge

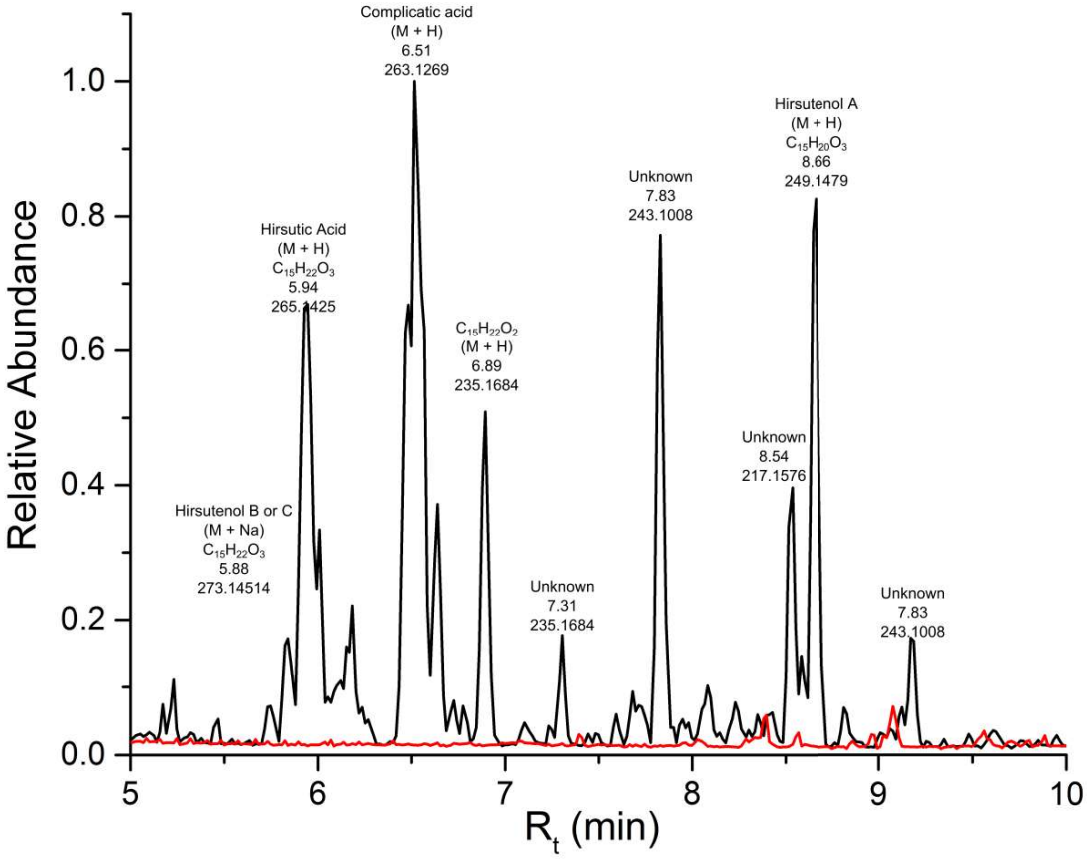
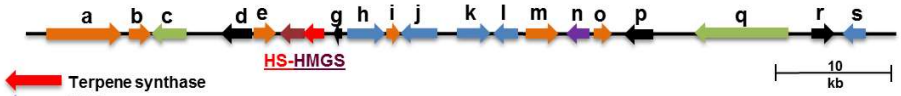


Figure 3-9: Positive ionization LC-HRMS fingerprint of *S. hirsutum* supernatant, demonstrating production and secretion of modified sesquiterpenoids liquid cultures.
Compounds were putatively identified by comparing exact mass to those predicted for known *S. hirsutum* metabolites.[358] All relevant peaks have been listed with their R_t and mass, while putative compound IDs are shown with their proposed adduct and chemical formula. The LC-HRMS spectrum of the ethyl acetate blank is shown in red.

enzyme (BBE),at least one aldo-keto reductase (predicted as a bifunctional protein with two aldo-keto reductase subunits), an MFS superfamily membrane transport protein, and five NAD(PH)- and FAD-binding oxidoreductases. Four properly spliced ORFs, encoding three of the four putative P450s and a putative BBE were cloned successfully from cDNA (*sh*HS-J, K, L and M) **(Table S5)**, suggesting co-expression with HS-HMGS

to produce modified hirsutenoids. These results suggest that at least these five genes in the HS-HMGS cluster are co-expressed and functional in *S. hirsutum*, and not a cryptic pathway or orphaned STS merely resulting from errant gene duplication. The presence of HMG-CoA synthase, a key enzyme in the isoprenoid precursor-producing MVA pathway, as a protein fusion also suggests a relatively high carbon flux toward production of hirsutene and its derivatives in the native organism.



Gene	Sup. Materials Reference	Top Conserved Domain Hit(s)
a	shHS-g34.t1	Aldo_ket_red (cd06660) (2x domains, predicted gene fusion)
b	shHS-g33.t1	NADB_Rossman Superfamily (cd05371)
c	shHS-g32.t1	Glycosyl_hydrolase_81superfamily (cl02310)
d	ShHS-g31.t1	SPC25 signal peptidase 25 kDa subunit (Pfam06703)
e	shHS-g30.t1	NAD(P)_binding_8 (pfam13450), HemY (COG1232)
HS-HMGS	shHS-g29.t1	HMG-CoA-S_euk (TIGR01833), Isoprenoid_biosyn_C1 super family (cl00210) (gene fusion)
g	shHS-g28.t1	ethD (pfam07110) [likely pseudogene]
h	shHS-g27.t1	P450 superfamily (cl12078); Rieske super family (cl00938)
i	shHS-g26.t1	PCB_SDR_a (cd05259); NAD_binding_10 (pfam13460)
J	shHS-g25.t1	P450 superfamily (cl12078)
k	shHS-1:g24.t1	P450 superfamily (cl12078)
l	shHS-1:g23.t1	P450 superfamily (cl12078)
M	shHS-1:g22.t1	FAD_binding_4 super family (cl19922); BBE superfamily (cl06869)
n	shHS-2-g21.t1	MFS superfamily (cd06174)
o	shHS-2-g20.t1	SDR_c_Rossman superfamily (cd5233)
p	shHS-2-g19.t1	Parvo_coat Superfamily (cl08286)
q	shHS-2-g18.t1	Cu_amine_oxid superfamily (cl08309)
r	shHS-2-g17.t1	PCI domain (pfam01399)
s	shHS-2-g16.t1	P450 Superfamily (cl2078)

Figure 3-10: The HS-HMGS Gene cluster.
The HS-HMGS gene is located in an unusually large biosynthetic cluster, covering approximately 100 KB and consisting of 19 putative ORFs. Predicted ORFs include catalytic activities expected for terpenoid modification and secretion, including five P450 monooxygenases, six putative NAD(P)H or FAD binding oxidoreductases, and a membrane transport protein. ORF's of HS-HMGS, three putative P450 monooxygenases, and a FAD-binding putative Berberine-bridge (BBE) enzyme successfully cloned from cDNA are shown in red.

3.5 Discussion:

3.5.1 HMG-CoA synthases are found in isoprenoid biosynthetic clusters throughout higher fungi

Yeast, and presumably every fungus, requires only one gene for each of the enzymes in the MVA pathway for free-living growth.[385, 396] Because of the well-established role of HMG-CoA reductase (HMGR) as the rate-limiting step in the MVA pathway,[48] and the apparent toxicity of HMG-CoA accumulation in engineered *E. coli*[397], one would expect that any gene duplication of MVA pathway genes to increase isoprenoid production would be favor HMGR duplication. However, genomic analysis of *Stereum hirsutum* identified two putative HMG-CoA synthase genes within the protoilludene-synthesis gene cluster of Stehi1|25180,[358] as well as HS-HMGS within its own putative STS cluster, with only a single HMGR identified in the *S. hirsutum* genome. Assuming clustered genes in fungal metabolic pathways are co-expressed, there appeared to be no flux benefit to the host organism in overexpressing HMG-CoA synthase. However, the presence of three HMG-CoA synthase genes within isoprenoid biosynthetic clusters demonstrates repeated non-random genomic localization, while conserving the residues required for activity. So how could HMGS gene duplications affect isoprenoid production in fungi?

To further understand the effect of HMG-CoA synthase duplications in fungal genomes and isoprenoid biosynthetic gene clusters, it is important to note that fungi likely lack the mitochondrial ketogenesis pathway of vertebrates [381], removing a major HMG-CoA depletion pathway utilized by mammalian cells when HMGR is blocked or repressed, e.g., by statins.[398] Indeed, none of the HMG-CoA synthases from *S. hirsutum* or other studied fungi contains a signal sequences localizing to the mitochondria, a prerequisite for HMGS-CoA synthase catabolism in the ketogenesis pathway. The absence of ketogenesis leaves only the reversible HMG-CoA lyase (HCL), [363] the final step in the leucine degradation pathway,[399] to deplete excess HMG-CoA. As such, an overproduction of HMG-CoA resulting from increased HMG-CoA synthase expression, if HMGR is held constant, would either result in a toxic accumulation of HMG-CoA, or result in an energy intensive futile cycle between HMG-

CoA, acetoacetyl-CoA, and acetoacetate (**Figure 3-1**). Finally, while the role of sterols in product feedback regulation of MVA pathway enzymes, primarily targeting HMGR is well established,[48, 400] no direct investigations have been made to determine the extent to which terpenoids affect MVA pathway flux in native organisms. Co-localization of HMGR with downstream sesquiterpenoid biosynthetic enzymes into vesicles dubbed "toxisomes" has been reported in the plant-pathogen *F. graminearum*.[92, 401] Such isoprenoid pathway compartmentalization may amplify the efficiency of co-regulated MVA pathway enzymes, although further study is needed to determine if such structures can be found throughout sesquiterpenoid producing fungi.

Engineered systems where HMG-CoA synthases have been overexpressed can provide insight into the extent to which natural HMG-CoA synthase overexpressions may affect overall isoprenoid production in "lower" eukaryotes. Toward this goal, *gl*HMGS of the triterpene-producing medicinal mushroom *Ganoderma lucidum* was overexpressed *in situ*, and found to increase production of triterpenoid ganoderic acids by 15.1-24.2% over wild type [381], demonstrating that overexpression of HMG-CoA synthase can affect flux through the isoprenoid pathway. Similar results were obtained when overexpressing *Brassica juncea* HMGS (brHMGS) in *Arabidopsis thaliana*, toward a similar increase in total sterol content of 13-26% in *A. thaliana* seedlings. [402]. Critically, qRT-PCR was performed, demonstrating increased expression of HMGR and other downstream genes in the sterol biosynthetic pathway following brHMGS overexpression.[402] These experiments suggest that the level of HMG-CoA synthase expression, and resulting increase in HMG-CoA concentration, can influence overall MVA pathway flux in "lower" eukaryotes, likely by triggering increased expression of HMGR. To test this hypothesis with HS-HMGS, *P. pastoris* and yeast were transformed with HS-HMGS or the HMG-CoA synthase deficient HS-HMGS C133A to compare total hirsutene production. However, large variations across biological isolates rendered any differences between HS-HMGS and HS-HMGS C133A statistically insignificant (**Figure S3-7**). This may be attributed the poor stability and solubility of HS-HMGS described earlier, rendering any increased MVA flux too small to be detected in these heterologous hosts. We speculate that the HS-HMGS may be stabilized in *S. hirsutum* by yet-undiscovered

chaperonins or compartmentalized in toxisomes similar to those of *F. graminearum*.[92, 401]

To determine the extent to which HMG-coA synthase gene duplications occur in fungi, and the frequency that these duplications occur in isoprenoid biosynthetic gene clusters, we performed a BLAST search of the JGI: Fungi genome database (525 genomes) for homologues to the yeast ERG13 and the two characterized Basidiomycota HMG-CoA synthases, *Ustilago maydis* Hcs1[382] and *Ganoderma lucidum* glHMGS.[381] This analysis identified 662 putative HMG-CoA synthases. Twelve fungal genomes auxotrophic in at least one MVA pathway gene that are known parasites or symbionts were manually removed from analysis. All remaining sequences are haploid or filtered bioinformatically.[390] Free living fungi require only one cytoplasmic HMGS for growth, but two copies were found to be common (103 of 513 genomes). The same genomes were further analyzed for all MVA pathway enzymes, as well as a subset of

Table 3-1: Copy number of MVA pathway and major isoprenoid biosynthetic genes in fungal genomes.
The majority of genomes contain a single copy of each MVA pathway gene, with a clear bias in duplications of early MVA pathway enzymes HMGS and HMGR, with 24 and 25% of fungal genomes possessing more than one of each respective gene. In contrast, late MVA pathway enzymes (MevK, Pmk, Pmc, and IDI) duplications are less common. Multiple copies of the reversible HMG-CoA lyase (HCL), were also identified in 37% of genomes.

Gene copy #	HCL	HMGS	HMGR	MvK	Pmk	Pmc	IDI	FPS	SQS	GGPPs	STS	PaFS
0	4%	0%	0%	1%	1%	1%	1%	0%	2%	0%	30%	78%
1	59%	76%	76%	91%	94%	95%	94%	78%	87%	3%	9%	16%
2	24%	20%	17%	8%	4%	4%	4%	17%	10%	49%	10%	4%
3	9%	3%	3%	0%	0%	0%	0%	3%	1%	18%	6%	1%
>3	3%	1%	5%	0%	0%	0%	0%	2%	0%	30%	53%	1%

common isoprenyl-diphosphate utilizing enzymes including STS, PaFS homologoues, and FPS-Squalene synthase gene fusions (**Table 3-1**). The resulting table shows a bias toward increased copies of early MVA pathway enzymes HMG-CoA Synthase (24%), HMGR (25%), and HCL (37%), while late MVA pathway enzymes (MevK, PmK, IDI) are nearly always found in single copy numbers, consistent with known MVA regulation of the early steps in the MVA pathway.[48] There was no correlation between genomes

with increased HMG-CoA synthase and increased HMGR copy numbers. That most genomes contain either no STS (29%) or more than three (53%) STS, is consistent with a trend we have observed previously in STS copy number. The terpene cyclase-PE fusion Fusicoccadiene-synthase (PrAS) has significant homology to FPP, GGPP, and STS sequences. As a result, results were filtered to remove proteins with stronger resemblance to either FPPs, GGPPs, or STS protein subunits based on E-value. No significant overlap was observed between other gene types.

Twenty of the 103 fungal genomes identified with two HMG-COA synthases were then randomly selected for manual cluster analysis, identifying 4 putative lycopene synthase clusters containing HMG-CoA Synthases out of the 40 HMG-CoA synthases manually analyzed, a 10% rate of duplication located in an isoprenoid gene cluster. Reasoning that HMGS localization in isoprenoid gene clusters is most likely the result of gene duplication(s), subsequent analysis then focused on the 19 genomes with three or more annotated HMGS encoding genes. The genomic region within 20 KB of each of the annotated HMGS was then manually analyzed for isoprenoid biosynthetic pathway enzymes. As a result, 14 of 65 (21.5%) HMG-CoA synthases analyzed from genomes with 3 or more copies were located in isoprenoid clusters, including nine in STS clusters (**Table 3-3**). Of these, three are located in putative *S. hirsutum* sesquiterpenoid biosynthetic clusters involved in protoilludene (Stehi1|25180)[358] and hirsutene biosynthesis (HS-HMGS) described above. The remaining six putative STS genes clade closely to known protoilludene synthases (aurvu1|1602279), Ascomycota aristolochene synthases (Neopa1|7973, Macph1|6367, and Conlig1|658201), and HS (Hebcy2|32419 and Hebcy2|290216) (**Figure S3-8**). In addition, three putative FPP/GGPP synthases are found in HMG-CoA synthase-containing clusters, and two putative triterpene lycopene beta-epsilon cyclases were in the vicinity of putative HMG-CoA synthase encoding genes in the genome of *Talaromyces marneffei* and *Tuber melanosporum*. Finally, a probable horizontal gene transfer is evident in the sequences of Hebcy2|32419, a putative STS-HMGS fusion with 49.9% ID to the full length HS-HMGS. Sequence alignment of the STS subunit with HS revealed maintenance of conserved STS active site motifs with high sequence conservation (81.9% ID) compared to HS. However, the predicted HMGS

subunit of Hebcy2|32419 is lacking the key catalytic triad and is predicted to have an early stop codon, suggesting Hebcy2|32419 is a functional STS fused to a vestigial C-terminal HMG-CoA synthase subunit. Another putative STS found in the *H. cylindrosporum* genome, Hebcy2|290216, and the adjacent putative HMG-CoA synthase Hebcy2|290222 also follow this same pattern, while bearing slightly less sequence conservation with HS (79.5% ID). The last common ancestor between the *Hebeloma* genus and *S. hirsutum* is currently dated to approximately 125 million years ago [403]. Taken together, these results suggest Hebcy2|32419 and Hebcy2|290216-290222 are the result of a horizontal transfer event of HS-HMGS or a relatively recent common ancestor, encode functional STS fused to vestigial HMGS subunits. To the best of my knowledge, *H. cylindrosporum* has never been systematically analyzed for sesquiterpene production.

3.6. Conclusions and Future Work:

This work has identified the first true hirsutene synthase (HS-HMGS), a surprising bifunctional protein with the second enzyme in the MVA pathway, HMG-CoA synthase. In combination with previously cloned and characterized *S. hirsutum* STS,[358] we have now identified the STS producing all the major sesquiterpenes produced by *S. hirsutum* mycelial cultures. Gene structure and protein phylogenetic analysis suggests that HS may be the result of a relatively recent horizontal gene transfer event into *S. hirsutum* from bacteria, based on its unusual intron pattern, high intron density, and similar protein structure to previously characterized bacterial cucumene synthases isolated from *Streptomyces* sp. The close similarity may also be the attributed to convergent evolution essential to 1,11 FPP cyclization reaction mechanisms, or a complex ancestry involving multiple HGT events from bacterial into fungi throughout history.[364] Furthermore, despite the significant sequence divergence from all previously characterized fungal STS, our predictive framework was successful in identifying HS as a 1,11 FPP cyclizing enzyme, [89, 358] providing further support for this predictive framework based on protein phylogeny of Basidiomycota STS. Divergent STS evolution within cyclization clades is implicit in this protein phylogeny-based approach to predicting STS function. Identification of gene structure conservation within

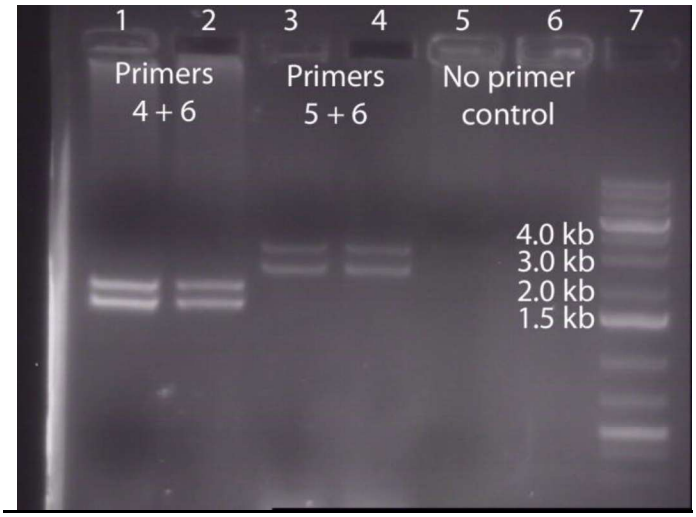
the Basidiomycota and Ascomycota phyla supports this model of STS diversification from a single common ancestor for cyclization clade, without significant convergent evolution events transforming enzymes from one cyclization mechanism to another. In addition, the cloning of HS-HMGS as an HMG-CoA synthase bifunctional protein has led to the surprising conclusion that HMG-CoA synthase duplications are not only common in fungi, but a disproportionate percentage of HMG-CoA synthases are found in isoprenoid biosynthetic gene clusters, a frequency that increases with number of HMG-CoA synthase duplications per genome.

Future work should focus on refactoring the hirsutenoid biosynthetic pathway, for which four genes in the HS-HMGS gene cluster have been cloned **(Table S3-5)**. In addition, it has been proposed that the bacterial cucumene synthase, pentalenene, and Basidiomycota protoilludene synthases share carbocation rearrangement intermediates from FPP through the 7-protoilludyl cation.[189, 364] However, identification of the bifurcation point between their respective pathways, whether before or after the 7-protoilludyl-cation intermediate, would enable refinement of the aforementioned STS predictive model, as well as greatly refine general understanding of how active site topography directs carbocation rearrangement through the central protoilludyl cation[364] in numerous proposed sesquiterpene cyclization mechanisms.[404] Finally, the discovery of early MVA pathway enzyme duplications in isoprenoid biosynthetic gene clusters suggests a common mechanism for increased production of specific fungal isoprenoids. This hypothesis could be tested through qRT-PCR paired with metabolic fluxomic analysis of natural isoprenoid synthesizing organisms containing HMG-CoA synthase duplications, e.g., S. *hirsutum*, or in an engineered species such as *G. lucidum* pGL-HMGS.[381]

3.7 Supplementary information

Table S3-1: Primers used in this study.		
No.	**Name**	**Sequence**
1	DDxxD F	ATATTTTTCCTG GCC GAC GAT TAT ATC G
2	ND R	GAA GAA ATC GTT GGT GAG GCC GAC
3	52743_EST end_R	CTC TTG CGA CCA AGC AAG GTT AC
4	HS_EST F	GTAACCTTGCTTGGTCGCAAGAG
5	HS ATG F	ATG GAT CCA TGT CTG AAA CCA AAG TTG GCA AAG TTG CTC C
6	HMGS_ORF_R	TCAATGGGTGACGGTATACGACCTCCTGTACTTCCC
7	HS_bamHI_F	CTAGTAGAAGGAGGAGATCTGGATCCATGTCTGAAACCAAAGT TGGCAAAGTTGC
8	HSt_xhoI_R	GAATCAGTGATGGTGATGGTGATGCTCGAGGAAGATGGCGTCG TCATCGATAGTGAAG
9	HMGS_F_BamHI	CTAGTAGAAGGAGGAGATCTGGATCCATGACCAACCCGCACAG CCATC
10	HMGS_R_NotI	GGTGATGGTGATGCTCGAGGCGGCCGCTCAATGGGTGACGGTA TACGACCTCC
11	HS-F-NdeI	TGGTGCCGCGCGGCAGCCATATGTCTGAAACCAAAGTTGGC
12	HMGS-R-NotI	GACTCACTCGAGGCGGCCGCTCAATGGGTGACGGTATACG
13	HS-R-NotI	GACTCACTCGAGGCGGCCGC*TCA*GAAGATGGCGTCGTCATCG
14	HR pESC-his Kozak, Nterm His6x	CATCCAAAAAAAAAGTAAGAATTTTTGAAAATTCGAATTCacc-atgggcagcagccatcatc
15	HR pESC-his HMGS R	GAT CTT ATC GTC GTC ATC CTT GTA ATC CAT CGA TAC TAG TTC AAT GGG TGA CGG TAT ACG
16	Q5 C133A F	CAAGAACGCTGCTTATGGGTCGACG
17	Q5 C133A R	CTGTCGATTCCTTCGACG
18	pPIC 2u-ura HR F	AGCCGGCCGTCTCGGATCGGTACCTCGAGGCGGCCGCcgaagcatct gtgcttcattttg
19	pPIC 2u-ura HR R	GAGATGAGTTTTTGTTCGGGCCCAAGCTGGCGGCCGCctcaaccctat ctcggtctattc
20	HR pPIC SH F	CAAAAAACAACTAATTATTCGAAACGAGGAATTCaccatgggcagcag ccatcatcatcatcatcacagcagcggcctggtgccgcgcgg
21	HR pPIC SH R	caaaatgaagcacagatgcttcgGCGGCCGCCTCGAGtcaatgggtgacggtatacgacc
22	pPICza-seq F	GCAAATGGCATTCTGACATCC
23	pPICza-seq R	caa atg gca ttc tga cat cc
24	pPIC Fseq-before PmeI	gttccccaaatggcccaaaactgac

Table S3-2: Fungal STS Intron:exon pattern and phase conservation between cyclization type and Ascomycota/Basidiomcyota Phyla.				
Gene structure group	General Exon:intron phase pattern	Intron within DDxxD motif	Intron between ND	
Basidiomycota Protoilludene synthases (FPP 1,11 cyclization)	1:(1):2:0	No (0/6)	Yes (6/6)	
Basidiomycota FPP 1,10 cyclizing enzymes	2:1:0	No (0/6)	Yes (6/6)	
Basidiomycota NPP 1,10 cyclizing enzymes	(0):2:0:2	No (0/4)	Yes (4/4)	
Basidiomycota NPP 1,6 cyclizing enzymes	0:0:2	No (0/5)	Most (3/5)	
Ascomycota FPP 1,11 cyclizing enzymes	2:1:1/2:2:2	Yes (2/2)	No (0/2)	
Ascomycota FPP 1,10 cyclizing enzymes	1:2	Most (2/3)	No (0/3)	
Ascomycota NPP 1,6 cyclizing enzymes	1:2	No (0/2)	No (0/2)	
HS-HMGS (1,11 FPP cyclization)	1:1:(0):(0):2:0:1:1:0	No	No	



Subunit	Unspliced length (BP)	Spliced RNA length (BP)
HS	1610	1092
HMGS (HS-EST-F)	1990	1640
HS-HMGS	3335	2526

Figure S3-1: PCR amplification of HS-HMGS from S. hirsutum cDNA
Phusion amplification of HMGS using a reverse primer (primer 6) corresponding to the predicted stop codon of HMGS, and forward primers corresponding to the last EST in HS (Primer 4, lanes 1-2) and the HS start codon (primer 5, lanes 3-4). The 2.5 KB PCR product in lanes 3-4 was gel purified, TOPO cloned, and sequence confirmed to produce a single ORF, HS-HMGS.

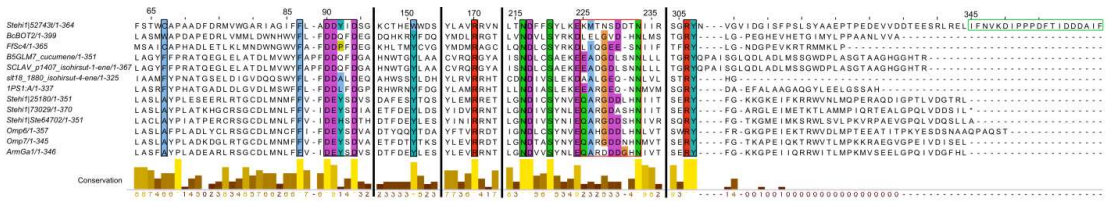


Figure S3-2: Reduced primary sequence alignment of HS compared with other known 1, 11 cyclizing terpene synthases from fungi and bacteria
This alignment shows key active site cavity residues and conserved Mg^{2+} binding motifs (black box). The H-α1 loop is boxed in red. Conservation assignments are supported by 2- and 3D alignment with atAS (**Figure S3-5**). Note the presence of C^{67} in place of aromatic or hydrophobic residues found in related *trans*-humulyl derived STSs, as well as W^{151} in place of the conserved tyrosine found in all *trans*-humulyl cyclizing enzymes. Among all known fungal STSs and bacterial *trans*-humulyl cyclizing enzymes, only Fompi1 (cuprenene synthase) possesses a tryptophan at this position (data not shown). The 21 AA linker loop between HS and HMGS subunits is shown inside a green box.

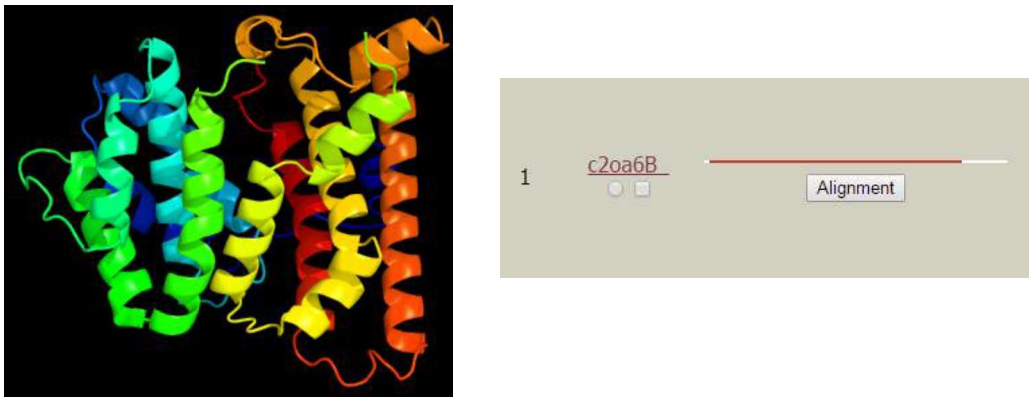


Figure S3-3: Protein structure prediction and analysis of HS.
A. Homology model of HSt, showing eight major α-helices, with 288 of 365 residues aligned to 4KUX (atAS) using the Phyre2 protein fold recognition server (http://www.sbg.bio.ic.ac.uk/phyre2/) with 100% confidence.

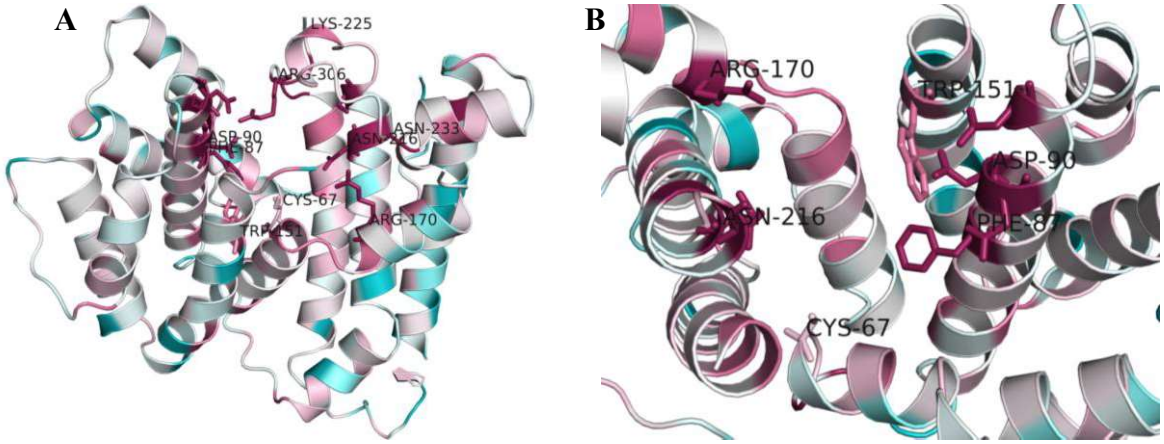


Figure S3-4: Pymol model of HS to atAS
A. (4KUX) showing all discussed active site positions are modelled facing the active site. Residue color reflects CONSURF-determined relative residue conservation between *trans*-humulyl derived STSs. Maroon colored residues are highly conserved, while white is neutral, and blue are variable. Canonical STS Mg^{2+} binding residues are highly conserved, as is π-bonding residue F^{87}. However, the H-α1 loop shows poor conservation, while C^{67} and W^{151} face the active site cavity, and are not conserved when compared to other *trans*-humulyl cyclizing STSs. **B.** Close-in of the Consurf-colored HS protein model, with H-α1 and terminal loop from E^{303}-V^{309} removed for visibility of the active site cleft, specifically showing nonpolar active site lining residues C^{67}, W^{151}, F87, Mg^{2+} binding residues N^{216} and D^{90}, and R^{170}, which likely forms hydrogen bonds with the pyrophosphate group.

Table S3-3: Site-directed mutants created to probe HS reaction specificity.
Only W151Y and W151G retained activity *in vivo*, while no significant differences were observed in product profile of these mutants. The apparent reduction in W151G hirsutene specificity is attributable to the minor products being produced near the GC/MS detection limit, not a change in product profile, as determined by the MS spectrum of integrated peaks. ND = none detected. N/A = not applicable.

Residue	Mutation	Hirsutene Production (Relative to WT)	Hirsutene fraction of total sesquiterpenes
WT	n/a	1	85.60%
C67	A	ND	n/a
C67	F	ND	n/a
C67	W	ND	n/a
W151	A	ND	n/a
W151	Y	3.98	87.80%
W151	G	0.33	68.35%

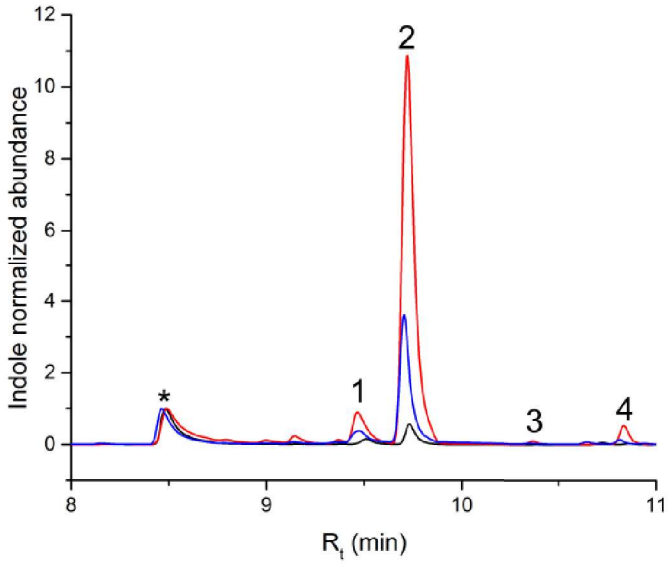


Figure S3-5: Indole (*) normalized GC/MS chromatograph comparing HS active site mutants
Indole (*) normalized GC/MS chromatograph showing the relative abundances of hirsutene (2) and other sesquiterpenes produced by wt HS (blue), HS W151Y (red) and HS W151G (black). Product distribution was unchanged, but total activity of the W151Y mutant increased, while W151G decreased.

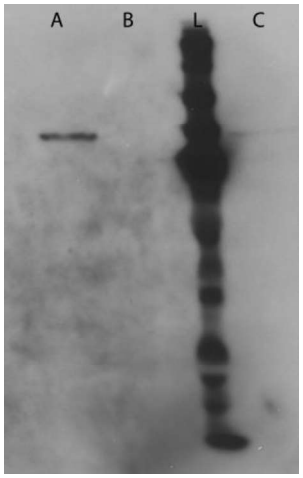


Figure S3-6: Western blot confirming presence of HS-HMGS in partial purification fractions. Lane A) Amicon 10K centrifugal concentration supernatant. Lane B) PD-10 desalting column flow through. Lane C) PD-10 desalting column elution fraction. Recombinant HS-HMGS expressed off the pET28a plasmid with an N-terminal his tag

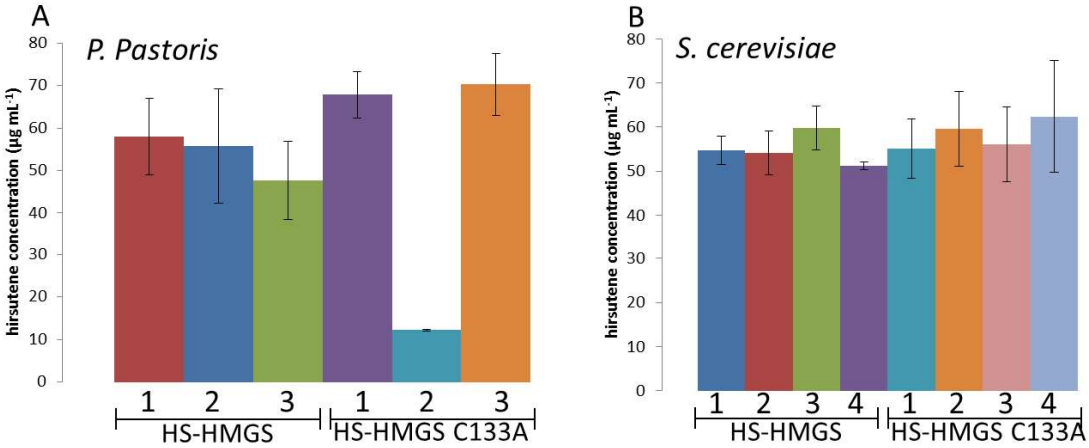


Figure S3-7: Hirsutene productivities of *S. cerevisiae* and *P. pastoris* HS-HMGS C133A. **A)** There was no significant difference between HS-HMGS and HMG-CoA synthase-deficient (C133A) HS-HMGS. *P. pastoris* with integrated pPIC-HS-HMGS (C133A) were grown overnight in 0.5% methanol expression medium. **B)** *S. cerevisiae* pESC-his-HS-HMGS (C133A) hirsutene production was not significantly affected by the C133A gene deletion.

Table S3-4: Cloned putative biosynthetic genes found in the HS-HMGS gene cluster.

Gene	Peptide reference	Scaffold number	Gene start	Gene stop	Gene length	# of introns	Transcript length	Protein length
HS-J	Stehi1|75699	4	941193	944017	2401	13	1626	541
HS-K	Stehi1|119691	4	945322	948471	2433	13	1650	549
HS-L	Stehi1|145717	4	948750	951052	2178	10	1554	517
HS-M	Stehi1|75716	4	951468	954219	2370	10	1758	585

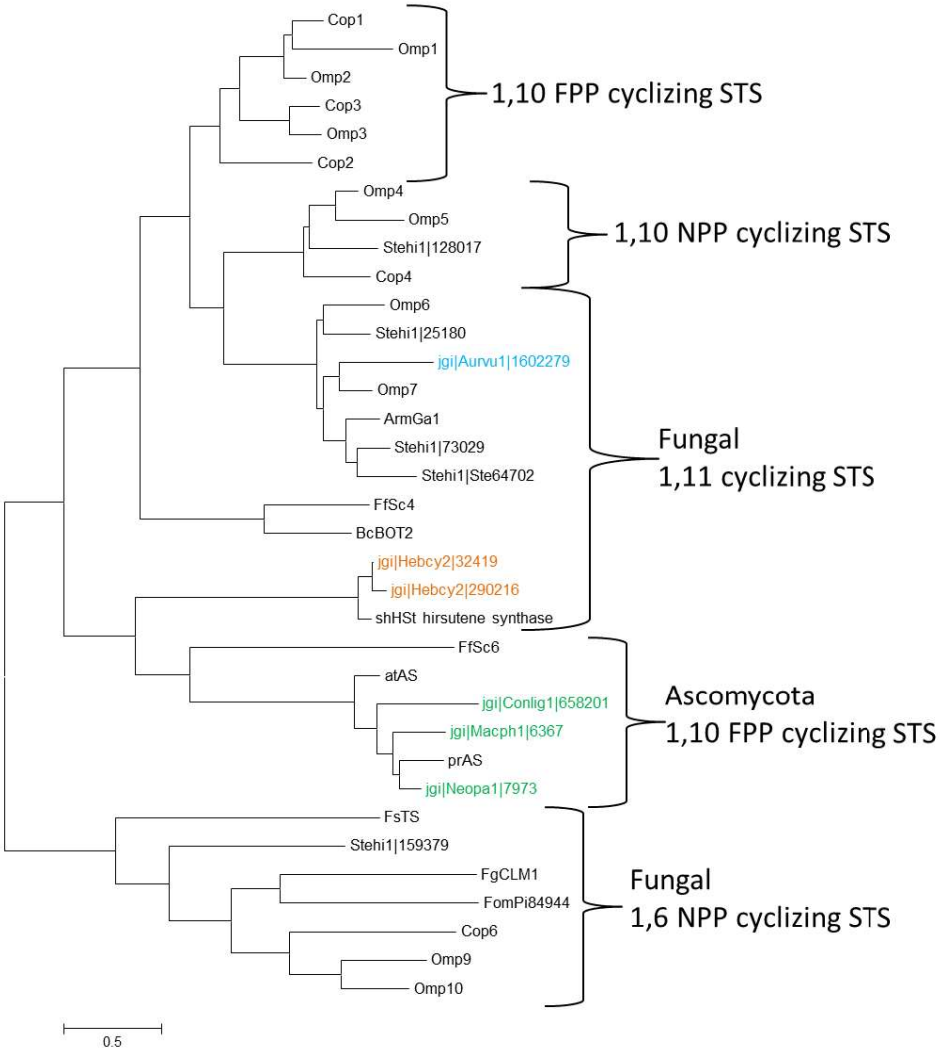


Figure S3-8: Alignment of HMGS-clustered putative STS enzymes with all known fungal STS. Putative STS found near HMGS proteins in the Ascomycota *neofusicoccum parvum, Coniochaeta ligniaria,* and *Macrophomina phaseolina* likely produce 1,10 FPP-derived sesquiterpenes similar to aristolochene (green), whereas *Auriscalpium vulgare* possesses an STS (blue) similar to protoilludene synthases, and *Hebeloma cylindrosporum* hebcy2|32419 and hebcy2|290216 (tan) are close homologues of HS-HMGS. Both genes are adjacent to 3' HMGS subunits on the same coding strand, however these HMG-CoA synthases lack all catalytic triad residues and are likely inactive.

Chapter 4 Progress toward refactoring fungal sesquiterpenoid biosynthetic pathways from *S. hirsutum* and *O. olearius*

The research in the following chapter was performed primarily by Chris Flynn, with contributions by Dr. Maureen B. Quin in developing the microsome purification assay, and Dr. Dana Freund in operating the LC-HRMS.

4.1 Summary

Biosynthesis of bioactive sesquiterpenoids requires cyclization of FPP by STSs, followed by modification by downstream enzymes, most commonly initiated by P450 monooxygenases. Specifically, the aim of this study is to demonstrate rearrangement of the protoilludene cyclobutyl ring to the highly reactive cyclopropyl ring characteristic of cytotoxic illudins, known products of *O. olearius*. We hypothesized that this conversion is catalyzed by a cytochrome P450. As a result, ten P450s cloned from protoilludene and hirsutene biosynthetic clusters were co-expressed in yeast, and screened for production of modified sesquiterpenoids by GC/MS. Six clustered oxidoreductases were also tested for activity in yeast and *E. coli*. Of all sixteen enzymes tested, only the P450 Omp7b and the VAO-type oxidase Omp7a demonstrated sesquiterpene modification, with activity dependent on co-expression with a protoilludene synthase. The reported activity by each enzyme is the first demonstration of protoilludene scaffold modification, and MS fragmentation of the major Omp7b product is consistent with cyclopropyl ring formation. Several significant non-volatile products were observed in cultures expressing Omp7a, however, they are likely breakdown products of larger compound(s) that are unstable in GC separation. Further pathway optimization is necessary, together with culture scale up and new sesquiterpenoid detection methods are required before the structures of the Omp7a and Omp7b products can be purified and definitively identified by NMR.

4.2 Introduction

The biosynthesis of bioactive sesquiterpenoids in fungi can be divided into three distinct biosynthetic steps. The first, MVA pathway production of FPP, was described in detail in Chapter 1. The second, synthesis of the unique sesquiterpene scaffolds by STSs (STS), was described in Chapters 2 and 3 with specific identification of novel protoilludene and hirsutene synthases. The final step in sesquiterpenoid biosynthesis, modification of the initial sesquiterpene scaffold to produce bioactive sesquiterpenoids, is the subject of the following chapter.

The goal of this research is to identify the genes and resulting enzymes catalyzing the conversion of protoilludene into bioactive compounds produced by *O. olearius*, illudin S and illudin M [108] (**Figure 4-1**), specifically aiming to identify the enzyme(s) catalyzing the conversion of the protoilludene cyclobutyl ring to the highly reactive cyclopropyl ring characteristic of the illudins. The mechanism of this conversion is currently unknown, and is of great chemical interest. Fungal biosynthetic pathways are often clustered in the genome. Therefore, genes near the appropriate STS are very likely to be involved in this functionalization.[111] However, there is no way to know the order of the conversions *a priori* for a newly discovered gene cluster. Comparison with previously characterized sesquiterpenoid biosynthetic pathways, such as artemisinin and trichothecene biosynthetic pathways (described in chapter 1), suggested that initial sesquiterpene functionalization is likely to be catalyzed by a cytochrome P450 monooxygenase (P450). Previous research in our lab largely refactored the lagopodin biosynthetic pathway from

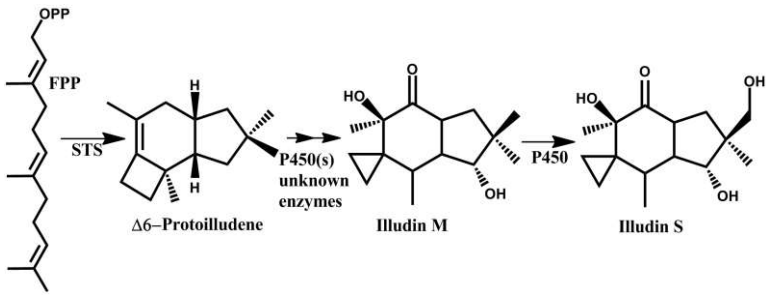


Figure 4-1: Proposed Biosynthetic pathway of illudins S and illudin M
FPP, a product of the MVA pathway in fungi, is initially cyclized by an STS to produce protoilludene.[89, 162, 358]. Subsequent oxidation and contraction of the cyclobutyl ring to the highly reactive cyclopropyl ring, characteristic of the illudins, is thought to require at least one P450.

Coprinus cinerea via co-expression of the STS Cop6 with the P450's Cox1 and Cox2.[83] This work established the methods for yeast expression, product extraction and identification, as well as providing another example of P450s catalyzing the initial functionalization of sesquiterpene scaffolds. Finally, a decrease in illudin M production, coupled to an increased accumulation of unmodified protoillene in *O. olearius* cultures exposed to the specific P450 inhibitor proadifen (0.1 mM) supports the conclusion that P450 activity is required for production of illudin M (unpublished data, CSD lab), and likely for the initial functionalization of protoilludene.

Cytochrome P450 monooxygenases are the largest and most functionally diverse enzyme superfamily found in nature.[405] P450s are heme-thiolate proteins found in all kingdoms of life, capable of catalyzing stereospecific conversions on a wide range of substrates.[406] These potential conversions include a constantly expanding list of functional group modifications, including hydroxylation, epoxidation, dealkylation, sulfoxydation, N-oxide reduction, and removal of all of the above functional groups.[406] P450s may catalyze multiple consecutive conversions with near perfect stereospecificity, for example the four consecutive oxygenations (3 hydroxylations and one epoxidation) catalyzed by Tri4, the first sesquiterpene modifying enzyme in the trichothecene pathway (**Figure 1-14**).[192]

All P450s share a reactive heme group, and require two electrons to be transferred from NAD(P)H via FAD/FMN binding cytochrome P450 reductase (CPR) or cytochrome *b5* proteins required for reduction of initial Fe^{III} to enable O_2 capture and reduction to initiate catalysis.[406] P450s are exceptionally important because of their ability to stereospecifically functionalize alkanes, represented in its simplest form in the hydroxylation of an alkane **Figure 4-2**, a common conversion in the initial activation of sesquiterpenes. Some P450's, such as the model bacterial P450 BM3, are soluble and consist of fused CPR-P450, making them easily expressed in *E. coli*, and an ideal systems

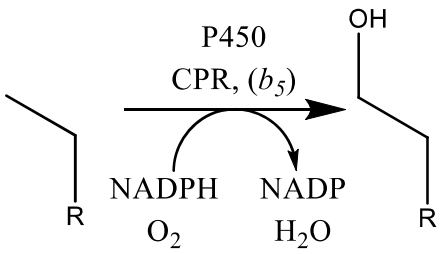


Figure 4-2: Generalized hydroxylation of alkanes by P450s and necessary CPR and potentially required cytochrome *b5* coenzymes

for engineering a wide variety of activities, given the versatility of P450 monooxygenase catalysis.[407] However, many Basidiomycota P450s, defined as Class I and Class II P450s, exist as separate proteins which must make close physical contact with reduction proteins (CPR, cyt *b5*) to enable electron transfer from NAD(P)H between their respective flavin or heme cofactors,[406] presenting a significant obstacle to heterologously expressing functional P450s. In addition, Class I and II P450s frequently contain an N-terminal membrane anchor before a conserved polyproline region, directing association into the ER membrane.[408] These characteristics require that full length, Basidiomycota P450s be expressed in eukaryotic hosts containing ER membranes, and that host possesses CPR and/or cytochrome b5 coenzymes capable of reducing the heterologous full-length P450 being studied. Heterologous expression of plant CPRs has been demonstrated to improve activity of similarly expressed plant P450s. However CPR expression also dramatically reduces growth rate in yeast,[409] further complicating functional assessment of class I and class II P450s in heterologous systems. Finally, a high-throughput survey of Basidiomycota P450s in *E. coli* recently replaced the P450 N-terminal membrane anchor with a high-expression tag, and demonstrated measurable expression in 27/304 (8.8% success rate) P450s.[408] Demonstrating the challenges of P450 functional characterization, no activity assays were reported in this study, presumably because of the difficulty in assembling functional P450-CPR-(cyt *b5*) reaction systems, much less determining the substrate and products of the subset of Basidiomycota P450s that could be functionally expressed heterologously. In summary, the catalytic diversity of P450s is unmatched by any other enzyme superfamily. However, studying the amazing chemical conversions they perform is currently severely limited by the number of coenzymes needed for catalysis, poor track record for enzyme production in heterologous hosts, required co-enzymes, and diversity of potential products making product detection particularly difficult, even when the probable substrate is known in biosynthetic terpenoid systems.

In addition to the P450s, FAD-binding enzymes of the Vanillyl Alcohol Oxidase (VAO) type can also catalyze addition of molecular oxygen to alkenes,[410] potentially catalyzing the critical cyclopropyl ring formation in illudin biosynthesis. VAO

oxidoreductases are defined by a conserved FAD-binding domain, which frequently, but not exclusively includes bicovalent binding of the FAD cofactor to cysteine and histidine residues lining the active site.[410] Site-directed mutagenesis studies of the VAO oxidoreductase Berberine Bridge Enzyme (BBE) (*S*-reticuline oxidase) that covalent protein-FAD attachment increases protein redox potential, enabling otherwise thermodynamically disfavored reaction catalysis.[411] Similar to P450s, the versatile, O_2 utilizing VAO oxidoreductases can catalyze a multitude of reactions, including hydroxylation of alkenes **(Figure 4-3)**, hydroxyl oxidation, methyl-alkene bond formation, and alkane dehydration, to name a few. Several putative VAO-type oxidases have been identified in protoilludene and hirsutene biosynthetic gene clusters. When applied to sesquiterpene modification pathways, the types of reactions catalyzed by VAO-type oxidases suggest that such enzymes likely act after P450s have initially oxygenated the sesquiterpene scaffold, or by catalyzing hydride removal near the double bond of the respective sesquiterpene scaffold. In this way, VAO-type oxidases could plausibly catalyze the cyclopropyl ring formation via at least two reaction mechanisms similar to that employed by BBE, where the bound FAD initiates a hydride transfer from an sp^3 carbon near an sp^2 hybridized carbon, triggering a concerted rearrangement of the target molecule.[412] One such ring contraction mechanisms is proposed in **Figure 4-4**. To address the possibility that a VAO-type oxidase can react with protoilludene, we decided to focus our initial cloning and pathway refactoring efforts on both VAO–type oxidoreductases and the aforementioned P450 monooxygenase enzymes.

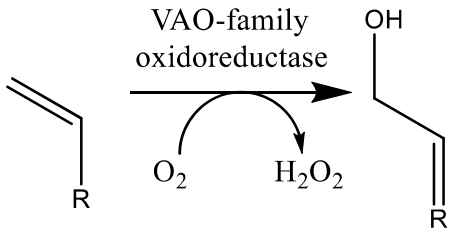


Figure 4-3: Generalized VAO enzyme alkene hydroxylation

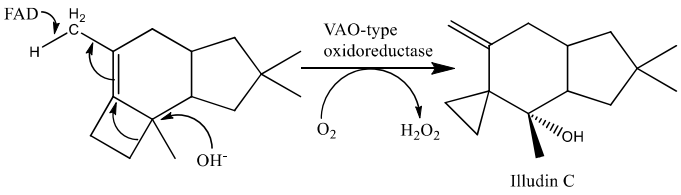


Figure 4-4: potential concerted mechanism for VAO-type oxidase catalyzed ring contractions of Δ6-protoilludene.

The aim of this research was to clone and characterize the enzymes in the protoilludene, and to a lesser extent, the hirsutene biosynthetic gene clusters in *S.*

hirsutum and *O. olearius*. Specifically, we expected the second step in sesquiterpenoid modification to be catalyzed by a clustered P450 or VAO-family oxidoreductase, and set out to test this hypothesis by co-expressing the appropriate STS with clustered oxidoreductase genes in heterologous organisms. In so doing, we intend to identify the enzyme catalyzing the critical cyclopropyl ring formation leading to illudin biosynthesis.

4.3 Materials and Methods

4.3.1 Strains and cultivation conditions

E. coli strains and methods used for DNA manipulation and gene expression were the same as those used in preceding chapters. The diploid yeast strain CEN.PK2[56] was used for all yeast homologous recombination cloning (YRC, method described in section 3.3.4) and sesquiterpenoid biosynthetic pathway refactoring and expression. The most effective conditions for P450 expression off the galactose-inducible pESC plasmid set were determined to be inoculation of 3 mL Sterile-filtered Synthetic Dextrose (SD, see section 3.2.10 for ingredients) shake tubes with single colonies from selective SD plates. Growth cultures with volumes of 50 to 500 mL in baffled shake flasks (250-2 L) covered in double-layered aluminum foil were then inoculated with approximately one to 3 mL overnight starter culture and grown overnight. Resulting high volume cultures were then pelleted by centrifugation, washed 2X in sterile dH_2O, and resuspended in their entirety in Synthetic Galactose (SG) medium, at OD's of approximately 3-6, to minimize the effect of growth inhibition caused by bioactive terpenoid production. Expression induction was typically confirmed within 2 hours by SPME extraction of the culture headspace. Typical cultures underwent 1-2 doublings in overnight SG cultivation, to reach a final OD of 6-12. All liquid yeast cultures were incubated at 30°C with 220 rpm agitation. One-half millimolar Aminolevulinic acid was supplemented (0.25 M stock) to SD and SG culture media to aid heme biosynthesis. Large-scale cultivates were performed identically, except using 2 L baffled shake flasks and 500 mL culture medium.

4.3.2 Cloning P450 and other clustered STS-clustered genes.

For the identification of putative sesquiterpene biosynthetic gene clusters, each of the scaffolds containing the protoilludene and hirsutene synthases from *S. hirsutum* and *O. olearius* were analyzed for the presence of nearby enzymes implicated in secondary metabolite biosynthesis. A previously described workflow was used for gene identification and splicing predictions, including the initial cloning of *O. olearius*

clustered enzymes Omp6h, i, m, l, and Omp7a & Omp7b.[85] *S. hirsutum* sesquiterpenoid-modifying genes were similarly identified in the genomic region surrounding the Stehi1|73029 and Stehi1|52743 (HS-HMGS) gene clusters. The results were manually scanned a minimum 20 kb up/downstream to confirm association with putative STSs using a suite of Augustus training sets, iteratively determining that the training set for *Laccaria bicolor* provided the most accurate gene predictions available for *S. hirsutum* gene predictions.[313] For prediction of the putative function of cluster genes, the most reasonable major conserved domain hit (CDD) at NCBI were used. More detailed analysis comparing conserved active site residues with similar, characterized enzymes was performed when necessary to ensure accuracy of gene structure predictions and enzyme determination.

4.3.3 Liquid chromatography-high resolution mass spectroscopy (LC-HRMS) metabolic fingerprinting

Stereum hirsutum FP-91666 SS-1 and *Omphalotus olearius* UT-643.13 (USDA Forest Products lab, Madison, WI) were grown at 25 °C on PDA plates (24 g/L Potato dextrose broth, 15 g/L agar) in the dark until sufficient mycelial front could be excised for liquid culture inoculation at a dilution rate of 1 cm^2 mycelial front per 100 mL medium. *S. hirsutum* was grown in triplicate in 500 mL liquid Rich medium (2% malt extract, 2% glucose, 0.5% peptone), contained in 2 L Erlenmeyer flasks covered with two layers of tin foil. *O. olearius* UT-643.13 was grown in triplicate in 100 mL liquid corn steep medium (1 g/L KH_2PO_4, 3 g/L $(Na)_2NO_3$ 3 g/L, 0.5 g/L KCl, 5 g/L Corn Steep Solids, 40 g/L glucose) in 500 mL Erlenmeyer flasks double layered with tin foil. All cultures were grown at room temperature (22°C) in the dark without shaking for 14 days. Metabolic fingerprint instrumentation and data analysis were collected as described in Section 3.3.3 with the help of Dr. Dana Freund of the Hegeman Lab (Department of Horticultural Sciences and Plant Biology, University of Minnesota, St. Paul, MN)

4.3.4 GC/MS detection of modified sesquiterpenoids in yeast culture

Detection of modified sesquiterpenoids in yeast cultures was performed using methods described by Agger (2009) for the detection of oxidized cuprenene

derivatives.[83] Culture headspaces were sampled for volatile sesquiterpene production by PDMS SPME as described in Chapters 2 and 3. After confirmation of volatile sesquiterpene production, culture OD was determined, and the culture was centrifuged for 5 minutes at 4000 RPM for 5 minutes in an Eppendorf swinging bucket centrifuge. Clarified supernatant was then transferred to a 125 mL Erlenmeyer flask. Non-volatile, hydrophobic compounds were then extracted by PDMS/DVB SPME for 10 minutes with gentle stirring (100 rpm), and analyzed by GC/MS. All methods were validated using the Cop6-Cox1 production of the cuprenene odixation product Compound 14 [83] prior to use in analysis of uncharacterized, refactored sesquiterpenoid modification pathways. GC/MS settings were identical to those described in earlier sections for GC/MS analysis of sesquiterpenes, with the caveat that only GC methods with temperature increases of 6 or 4°C/min (total run time 38-54 min) could be used to resolve higher molecular weight Compound 14 peak in the Cop6/Cox1 positive control.

Scale-up cultures for extraction of nonvolatile sesquiterpenoids were centrifigued at 6000 rpm (Beckman SW28 centrifuge) for 10 minutes. Clarified supernatant was then extracted with 500 mLCHROMOSOLV EtoAc, allowed to phase separate for 2 days in a capped separation funnel, and concentrated to 10 mL EtoAc by rotovap (max vacuum: 23 mmHg, waterbath temp: 40°C). Two microliters of this concentrate was then injected into GC/MS-FID for quantitative concentration determination, using caryophyllene oxide as a standard over the range of 1 to 100 μg mL^{-1}.

4.3.5 Microsome preparation of sesquiterpenoid biosynthetic compounds

Full-length fungal E-class, group I P450 monooxygenases are expected to be anchored to the ER. As a result, *in vitro* assays utilizing microsomal preparations from yeast-expressed P450's were used to test for active P450 monooxygenases. Microsomes were purified from 500 mL yeast in SG, following overnight expression of the appropriate STS and associated P450. Culture headspaces and supernatants were extracted by SPME as described above (2.5.6-7 and 4.3.4). Yeast cell pellets were then resuspended in 50 mL 100 mM Tris-HCl (pH 9.4), 10 mM DTT, incubated for 15 minutes, pelleted, and resuspended in 25 mL lyticase buffer (0.7 M Sorbitol, 7.5 g L^{-1}

yeast extract 15 g L^{-1} peptone, 0.5% glucose, 10 mM Tris-HCl, pH 7.4, 5 mM fresh DTT, 25 U lyticase), incubate at 37°C for 2 hours. Overlay sucrose cusion (0.8 M sucrose, 1.5% Figoll 400, 20 mM HEPES pH 7.4). Centrifuge at 6000 rpm for 10 minutes in Beckman SW28 centrifuge, remove top two layers with pipette, freeze (~20 mL) at -80°C. Defrost cells, centrifuge 3200 g (Eppendorf 5810 R) for 5 min. Resuspend in lysis buffer (0.1 M sorbitol, 50 mM KOAc, 2 mM EDTA, 20 mM HEPES pH 7.4, 1 mM DTT, 1 mM PMSF, 1 tablet cOmplete protease inhibitor cocktail), for sonication (30% power, 1 sec on, 2 sec off), overlay with cushion 2 (1 M Sucrose, 50 mM KOAc, 20 mM HEPES, pH 7.4, 1 mM DTT), centrifuge in SW28 centrifuge at 7200 rpm for 10 minutes. Transfer all supernatant to new tube and spin for 10 min at 21500 rpm, saving pellet. Resuspend in 25 mL wash buffer (20 mM HEPES pH 6.8, 150 mM KOAc, 250 mM sorbitol, 5 mM MgOAc), centrifuge at 21500 rpm for 10 min, resuspend pellet in 1 mL wash buffer. Sample 5 ul, add to 995 ul 2% SDS for A280 protein concentration determination (A_{280} =0.2 = 10 mg/mL). All steps performed at 4°C.

In vitro microsome activity assays with NADPH regeneration[413] in glass GC/MS vials contained contained 200 µM NADPH, 2.5 µM FMN, 2.5 µM FAD, 2 mM Glucose-6-phosphate, 2 U glucose-6-phosphate dehydrogenase, 600 µg total protein, and 2 µM FPP in 20 mM Tris pH 7.5. Reactions proceeded overnight at room temperature and were sampled by SPME for volatile and nonvolatile sesquiterpenoids by GC/MS. The extraction efficiencies of single-vial 1:1 EtOAc extraction[365] and PDMS/DVB SPME extraction were compared using the Cop6 + Cox1 α-cuprenene-oxidation pathway [83]-generating positive control. This comparison demonstrated that PDMS SPME extraction results in higher concentration of nonvolatile terpenoids, and a higher signal-to-noise ratio than solvent extraction. All subsequent GC/MS analysis for nonvolatile sesquiterpenoids utilized PDMS/DVB SPME extraction from culture supernatant or reaction buffer.

4.4 Results

4.4.1 LC-HRMS fingerprinting of *O. olearius* and *S. hirsutum*

The production of numerous hirsutenoid and protoilludene-derived compounds is well established in natural products literature, as is the production of hirsutene and protoilludene in our lab cultivations, [89, 358] from which cDNA was prepared for biosynthetic enzyme cloning. Complicating matters, it is well-known that fungi are capable of rapid epigenetic inactivation of secondary metabolite production.[321] In addition, there is little reason to assume culture conditions enabling growth, and those inducing production of desired natural products, will overlap, making the induction of cryptic pathways the subject of intense study[111, 323]. Indeed, specific chemicals are required for the plant pathogen *F. graminearum* to produce trichothecenes,[414] making any assumption that lab-grown cultures will produce the same secondary metabolites seen in nature particularly dubious. As a result, it was necessary not only to demonstrate that cultures used to prepare cDNA for use as a template for natural product gene cloning produce the desired sesquiterpene precursors (as we have done previously)[89, 358], but also to demonstrate production of modified, frequently nonvolatile, sesquiterpenoids. Liquid Chromatography-High-Resolution Mass specroscopy (LC-HRMS) was therefore performed to identify which, if any, modified sesquiterpenoids are produced by *S. hirsutum* and *O. olearius* liquid cultures used for cDNA preparation, and under conditions previously demonstrated to produce volatile sesquiterpenes.

S. hirsutum and *O. olearius* were cultivated for 14 days in liquid medium and analyzed by LC-HRMS as described in detail in section 3.3.3. Culture headspaces were confirmed by GC/MS to produce the same major sesquiterpenes identified previously under the same conditions.[89, 358] Samples of culture medium were then extracted with EtoAc and analyzed by LC-HRMS, with compounds identified by comparing the exact masses of known compounds produced by each species with observed high-resolution m/z, considering potential positive ionization adducts (**Table 4-1 and 4-2**). In addition, HRMS spectra were compared to libraries described in the mzCloud online (https://www.mzcloud.org/compound/Search) and Metlin[366] databases. Only Illudin M and Illudin S could be confirmed by comparison to these databases, as no other spectral-

based searches of the databases were strong fits, and none of the remaining sesquiterpenoids produced by *S. hirsutum* or *O. olearius* could be found in the databases at the time of search (December 2014). The LC-HRMS chromatograph of *S. hirsutum, O. olearius*, and an EtoAc extraction blank are shown in **Figure 4-5**, where the peak height of illudin M was set at 100, with all other peaks normalized relative to illudin M**.** The high concentration of illudin M compared to illudin S and their relative retention times is consistent with previous reports of illudin M and Illudin S by its close relative *O. olivascens* var. *indigo*,[415] as well as high level production of illudin M for purification by *O. olearius* in liquid culture,[416] providing further support of the identification of illudin M and Illudin S using LC-HRMS and available databases. In conclusion, LC-HRMS has been utilized to demonstrate that our lab cultivations produce and sescrete Illudin M and illudin S, and therefore the enzymes required for their production are at least transiently active in *O. olearius* cultures. Because the necessary genes must be expressed and correctly spliced for their enzymes to be active, cDNA prepared from these cultures is very likely contain functional ORF's to enable the cloning of genes involved in protoilludene modification to produce illudin S and illudin M.

Table 4-1: *O. olearius* sesquiterpenoids				
Chemical formula	Predicted Molecular mass (monoisotopic)	*O. olearius* Compounds	Observed m/z	mzCloud, MetLin library confirmed
$C_{15}H_{24}O$	220.1827	<predicted monohydroxylated intermediate>	None	-
$C_{15}H_{26}O$	222.19836	Torreyol, isoomphadione	None	-
$C_{15}H_{20}O_2$	232.14633	Illudin C Illudin E Omphadiol (Dehydrated illudaneol, illudosone hemiacetal	233.153606 [M+H]	NA
$C_{15}H_{22}O_2$	234.16198	Illudin D, Illudin E, Omphadiol	235.169256 [M + H]	NA
$C_{15}H_{24}O_2$	236.17763	<predicted hydroxylated intermediate>	None	-
$C_{14}H_{22}O_3$	238.15689	Illudiolone	None	NA
$C_{15}H_{16}O_3$	244.10994	Illudalic acid		NA
C15H20O3	248.1412	Illudin M, Illudin C2	231.138460 [M + H –H2O]	Illudin M
$C_{15}H_{22}O_3$	250.15689	Illudin A, Dihdyroilludin M, Illudaneol, Illudosone hemiacetal	None	NA
$C_{15}H_{24}O_3$	252.17254	Illudol,Illudosin	None	NA
$C_{15}H_{15}NO_3$	258.10223	illudinine	None	NA
$C_{15}H_{20}O_4$	264.136159	Illudin S	265.169256 [M+H]	Illudin S
$C_{15}H_{22}O_4$	266.15181	4a-hydroxy-dihydroilludin M, Neoilludol	None	NA
$C_{15}H_{22}O_5$	282.14672	Illudin B	none	NA

Table 4-2: *S. hirsutum* sesquiterpenoids			
Hirsutenes:	Calculated MW	Compound	Observed M/z
$C_{15}H_{22}O_2$	234.16198	<unknown compound with two double bonds, two oxygen additions>	235.169256 [M + H]
$C_{15}H_{20}O_3$	248.14124	Hirsutenol A	None
$C_{15}H_{22}O_3$	250.15689	Hirsutenol B, C	273.146108 (M + Na)
$C_{15}H_{24}O_3$	252.17254	Hirsutenol E	None
$C_{15}H_{18}O_4$	262.12051	Complicatic acid	None
$C_{15}H_{20}O_4$	264.13616	Hirsutic acid C, Compound 5 (unnamed connatusin A derivative)	265.143436 (M+H); 287.125378 ([M+ Na]
$C_{15}H_{20}O_5$	280.13107	Compound 3, connatusin A derivative	None
$C_{15}H_{17}ClO_3$	280.08662	Chlorostereone	None
$C_{15}H_{18}O_4S$	294.09258	Hirsutenol F	None
$C_{16}H_{22}O_5$	294.14672	Cpd 4 (unnamed connatusin A derive)	None

No hirsutenoids could be found in available databases to confirm *S. hirsutum* sesquiterpenoid product annotations made based on LC-HRMS. Definitive identification of *S. hirsutum* compounds would therefore require high-concentration purification followed by NMR structural identification, or chemical synthesis of standards. However, both approaches were deemed infeasible given funding constraints and local expertise. Despite being limited to putative sesquiterpenoid identifications of *S. hirsutum* sesquiterpenoids, these results demonstrate production of oxygenated sesquiterpenoids consistent with those hypothesized to be produced by actively expressed *S. hirsutum* sesquiterpenoid modifying enzymes. This was deemed sufficient to justify cloning STS associated genes for sesquiterpenoid pathway refactoring from *S. hirsutum*.

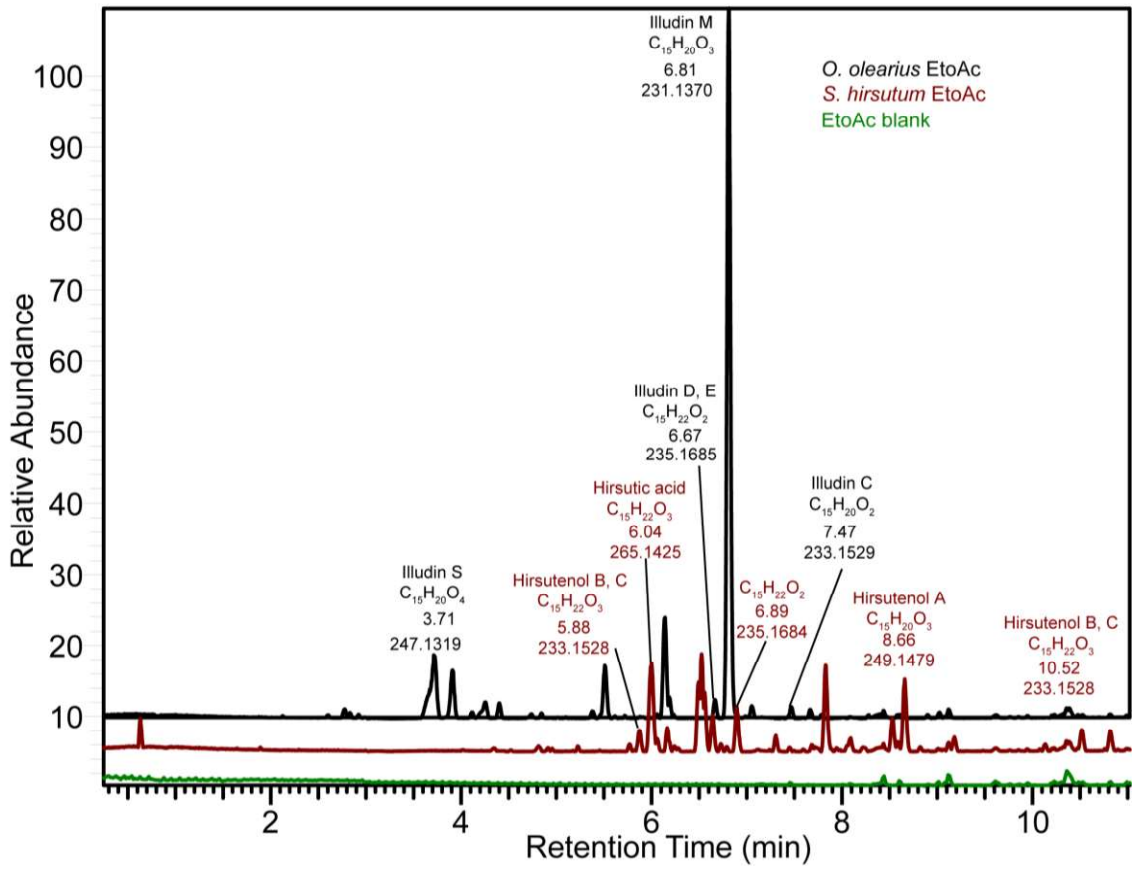


Figure 4-5: LC-HRMS chromatograph fingerprint of *O. olearius*, and *S. hirsutum* liquid cultures. The Y-axis is normalized to illudin M. *S. hirsutum O. olearius* chromatographys are offset by 5 and 10 relative abundance units for visibility. Illudin S and Illudin M assignments based on HRMW were confirmed by Metlin library comparison and consistent with literature, while all remaining compounds are putative, and determined solely by HRMW as described in **Tables 4-1 and 4-2.**

4.4.2. Summary of cloned genes and initial *in vivo* activity screens

To achieve our project goal of refactoring sesquiterpene modification pathways, we first set out to clone the enzymes in STS-containing gene clusters predicted to have oxidoreductase activity. Based on previously characterized sesquiterpenoid biosynthetic pathways, especially trichothecene (**Figure 1-14**) and artemisinic acid biosynthesis (**Figure 1-5**), we hypothesized that the initial oxygenation steps in sesquiterpene modification were likely catalyzed by cytochrome P450 monooxygenases, or NAD(P)H/FAD binding oxidoreductases whose coding genes are located in STS gene clusters. To address this hypothesis, we first

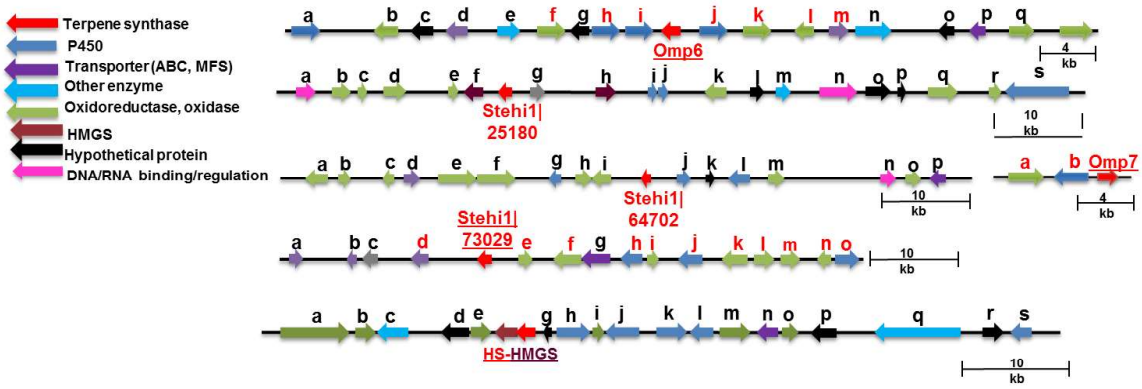


Figure 4-6: Protoilludene and hirsutene synthase gene clusters
ORFs successfully cloned are shown in red. HS-HMGS encodes a hirsutene synthase, while all remaining STS encode protoilludene synthases.

cloned these gene types from the clusters surrounding the protoilludene synthases Omp6, and Omp7,[89] Stehi1|73029, as well as the hirsutene synthase HS-HMGS gene cluster, shown in **Figure 4-6**. The genes cloned in this way are summarized in **Table 4-1**. Cloning of all P450s was attempted, however, only those whose ORFs could be successfully isolated are further described.

The sesquiterpene-dependent production of new compounds was observed only in any experiments except Omp7a and Omp7b, which will be described further in following sections. This section will therefore provide a brief overview of work performed on refactoring sesquiterpenoid-modification pathways without identifying protoilludene modifying enzymes. The remaining P450s listed above were tested for *in vivo* activity

Table 4-3: Genes cloned from fungal sesquiterpenoid biosynthetic clusters		
Clustered STS	Gene name	Putative enzyme type and sequence analysis notes
Stehi1|73029	P450-A	P450, E-class, group I
	P450-B	P450, E-class, group I
	P450-C	P450, E-class, group I
	AAT	Aminoacyl transferase
	GMC1	GMC oxidoreductase; NAD binding
	GMC2	GMC oxidoreductase; NAD binding
	FAD1	FAD binding oxidoreductase
	FAD2	FAD binding oxidoreductase
Omp6	Omp6h	P450, E-class, group I
	Omp6i	P450, E-class, group I
	Omp6j	P450, E-class, group I
	Omp6k	GMC oxidoreductase; NAD binding
	Omp6L	Aldo-keto reductase (likely inactive as cloned)
	Omp6m	Aldehyde reductase short chain reductase (SDR). NAD binding. N-terminus not conserved; key active sites not conserved (A47, W77 of 1M9H protein on pdb).
Omp7	Omp7a	FAD-binding oxidoreductase in VAO superfamily. No signal peptide, H67 residue for covalent binding to FAD. No other conserved covalent binding sites. 9 predicted N-acetyl glucosamine sites predicted, no glycosylation sites. C-terminus has BBE superfamily conservation similar to HS-M.
	Omp7b	P450, E-class, group I
HS-HMGS	HS-J	P450, E-class, group I
	HS-K	P450, E-class, group I
	HS-L	P450, E-class, group I
	HS-M	FAD-binding redox in VAO superfamily. BBE C-terminal motif. Secretion signal sequence at N-term(AA1-20). 1 predicted glycosylation site. H162 likely flavin covalent binding site; no additional conserved covalent FAD binding residues.
NA	ooCPR	*O. olearius* Cytochrome P450 reductase
NA	shCPR	*S. hirsutum* Cytochrome P450 reductase

with their associated STS in yeast strain CEN.PK2,[56] but no activity was observed. In addition, the P450s in the Stehi1|73029 cluster were also expressed in the *wat11* yeast strain, which contains an integrated, constitutively expressed *A. thaliana* cytochrome P450 reductase (CPR) ATR1.[409] The Wat11 strain has previously been used to successfully express active sesquiterpene modifying P450's cloned from plants. e.g.[191, 417, 418] However, neither expression in the Wat11 strain, nor co-expression of *shCPR*1 or *ooCPR* with respective STS clustered P450s in the CEN.PK2 yeast strain demonstrated protoilludene dependent production of modified sesquiterpenoids detected by PDMS/DVB extraction and GC/MS analysis. The genes in the hirsutene pathway were co-expressed in CEN.PK2 as well, either solely with hirsutene synthase (HS) or in combination (HS-L+ HS-M; HS-K + HS-M; HS-L + HS-M) with no observed production of modified sesquiterpenoids. In addition, the stehi1|73029 genes AAT, GMC2, FAD1, and FAD2 were also tested for activity in *E. coli* behind the constitutive LacP* promoter, but no activity was observed, and no soluble or insoluble expression was detectable by SDS-PAGE. Extensive attempts to express Omp6-clustered P450's in *P. pastoris* and *E. coli* also failed to result in P450-dependent protoilludene modifications, despite previous premature reports of success using these expression systems.[419] Finally, attempts at expression of wild-type and N-terminal membrane-anchor-truncated P450s[408] from *O. olearius* (Omp6h,i,j, and Omp7b) with their respective STS in *E. coli* BL21(DE3) were undertaken, none of which resulted in detectable protein production, determined by SDS-PAGE, and no catalytic activity was detected.

There is little reason to conclude that the low success rate observed in STS-P450 pathway refactoring can be attributed to the extraction and analysis methods utilized, unless the target pathways are far more complex than previously predicted based on the known compounds in the illudin pathway. The PDMS/DVB extraction method followed by GC/MS separation and detection was established by Agger *et. al.* (2009) in demonstrating activity of the Cop6-Cox1-Cox2 modification pathway producing oxygenated a-cuprenene derivatives, including cuparophenol and lagopodin A.[83] In addition, detection of modified sesquiterpenoids by GC/MS is well established in the literature, and is the standard technique utilized for detection of highly modified

trichothecenes,[153] [153] artemisinin precursors.[8] Finally, SPME extraction from the culture broth of *Lampteromyces japonicus* and identification of illudin S by GC/MS has been demonstrated to be highly reliable and effective, with a detection limit of 36 pg ul^{-1}. [420] Considering that detection of illudin S, related molecules, and its precursors are the primary goal of this research, utilization of SPME extraction and detection by GC/MS is strongly supported. There are examples of compounds in hybrid terpenoid pathways, such as meroterpenoid [310], and the melleolide [198, 421] pathways being detected by HPLC or TLC, presumably due to the low volatility of these higher molecular weight, polyketide-derived hybrid terpenoids. However, the sesquiterpenoids identified from *S. hirsutum* [358] and *O. olearius*,[89] lack addition of high molecular weight adducts. As a result, we jointly concluded that the SPME extraction and GC/MS detection previously employed with success to identify α-cuprenene-derived sesquiterpenoids[83] was well supported, and its use in detection of protoilludene biosynthetic pathway intermediates produced by refactored enzymes justified by extensive precedent. Despite the potential shortcoming that GC/MS detection may not identify glycosylated or polyketide-modified protoilludene derivatives, as there is no evidence that *S. hirsutum* or *O. olearius* produces such compounds in the available literature.

Proposed alternative methods for improving all facets of sesquiterpenoid modification pathway refactoring, including necessary improvements to be made in compound detection, gene expression, enzyme optimization, secretion from hosts, and improved titer of modified sesquiterpenoids are discussed in the conclusions and future work section of this chapter.

4.4.3 Microsomal preparation for in vitro analysis of P450 production

Microsomal preparation of yeast-expressed STS with clustered P450s was developed in parallel with *in vivo* activity screens, with the goal that such an in vitro analytical method would enable biosynthesis of sesquiterpenoids that may be highly toxic to the host strain, as well as developing a valuable tool for characterization of any P450s identified through tradition *in vivo* screening. As a result, a method for preparation of

yeast microsomes was developed and tested using the highly active Cop6-Cox1 system as a positive control, described in detail in section 4.3.5. Microsomal preparations of a

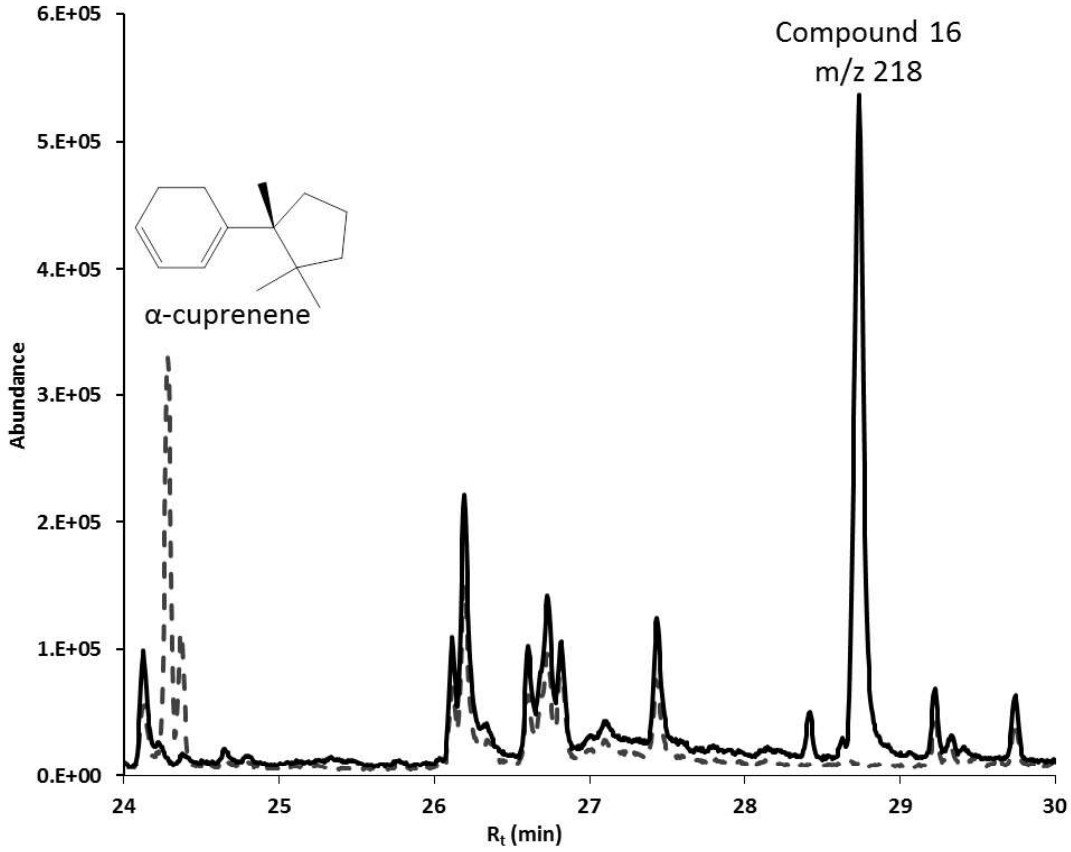


Figure 4-7: Nonvolatile PDMS-DVB SPME extraction of Cop6+Cox1-containing microsomes. Microsome reactions were incubated overnight with 2 μM FPP and NADPH. A-cuprenene accumulation is observed in the microsomes containing only Cop6 (dashed line), while the Cox1 product compound 16 is observed in Cop6+Cox1 (solid black line) containing microsomes. All other peaks are column siloxanes.

Cop6-only control and Cop6-Cox1 were incubated overnight with FPP and a NADPH regenerating pathway.[413] These microsomal preparations were then analyzed as described earlier by volatile and non-volatile SPME extraction and GC/MS (**Figure 4-7**). The Cop6 control produced solely α-cuprenene, while the Cop6+Cox1 test of P450 activity demonstrated near-complete loss of α-cuprenene from the volatile extraction (data not shown), while the nonvolatile extraction demonstrated removal of α-cuprenene and production of the 218 m/z oxygenated sesquiterpenoid previously observed as the

Cox1product Compound 16.[83] Removal of α-cuprenene suggests its complete conversion to Compound 16 when Cox1 is expressed. This result demonstrates that the microsomal preparation protocol can successfully reconstitute functional STS-P450 modification pathways from yeast expressed cells. However, subsequent microsomal preparations to analyze uncharacterized P450 listed in **Table 4-1** from the Omp6 and Omp7 gene clusters, no *in vitro* activity was observed.

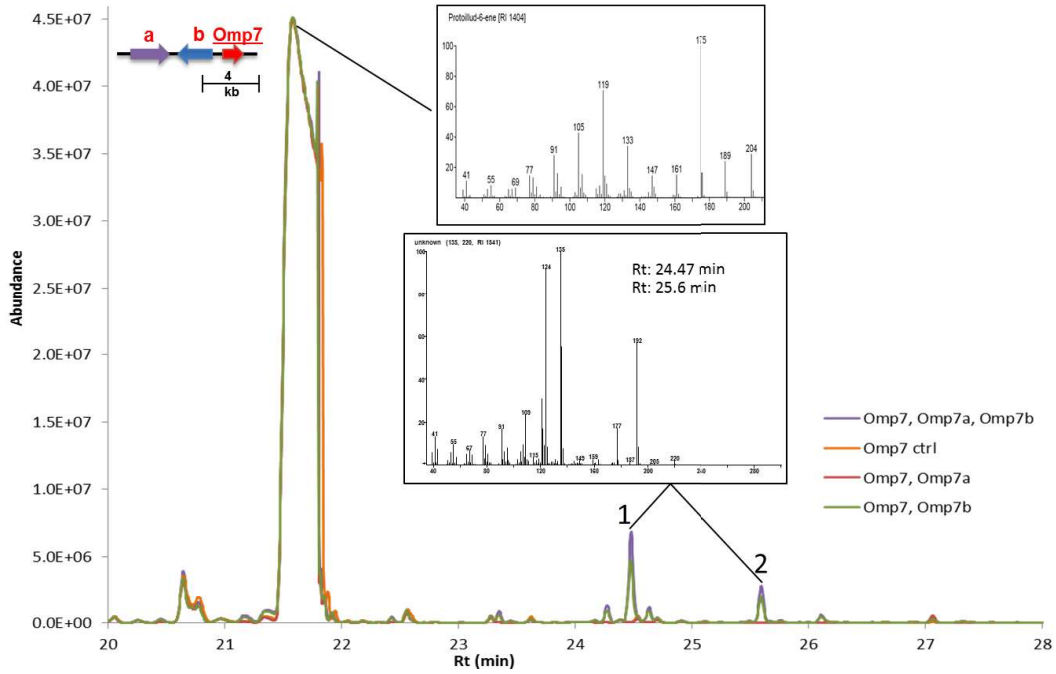


Figure 4-8: Volatile sesquiterpenoids produced by yeast expressing Omp7, Omp7a, Omp7b
Headspace extraction of terpenoids from cultures expressing Omp7 with Omp7a, Omp7b, and in combination. Strains expressing Omp7 and Omp7a either alone (red) or in combination with Omp7b (green) produced significant quantities of volatile sesquiterpenoids (parent ion m/z 220) in similar quantities, while the only volatile product of the Omp7 control (orange) and Omp7 + Omp7b (green) was protoilludene, confirming expression of Omp7.

4.4.4 Reconstructing the *O. olearius* Omp7 protoilludene modification pathway: Omp7a and Omp7b

Initial expression and activity screening identified the VAO-type oxidase Omp7a and the P450 Omp7b as the only two protoilludene modifying enzymes retaining activity during yeast expression. As a result, all further *in vivo* experimentation focused on their expression to confirm their reaction products.

Activity of Omp7a and Omp7b was demonstrated in the yeast CEN.PK2 strain,

transformed with the galactose-inducible pESC-trp-Omp7 (CSD database 1617), pESC-his-Omp7a (1618), and pESC-ura-Omp7b (1591) plasmids. All genes were under control of the GAL10 promoter, and with an ACC synthetic kozak sequence added just upstream of the start codon to aid in expression. Co-expression of *Omp7* with *Omp7b* resulted in production of several modified sesquiterpenoids, detectable in both the volatile (**Figure 4-8**) and nonvolatile extractions from overnight yeast cultures. Additional controls expressing *Omp7a* and/or *Omp7b* with an empty-vector pESC-trp produced no detectable compounds (data not shown), demonstrating that Omp7a and Omp7b activity requires presence of protoilludene. The new volatile compounds are produced only when Omp7 and Omp7b are co-expressed, have a parent ion of m/z 220, suggesting addition of a single oxygen to the protoilludene scaffold with four total degrees of unsaturation.

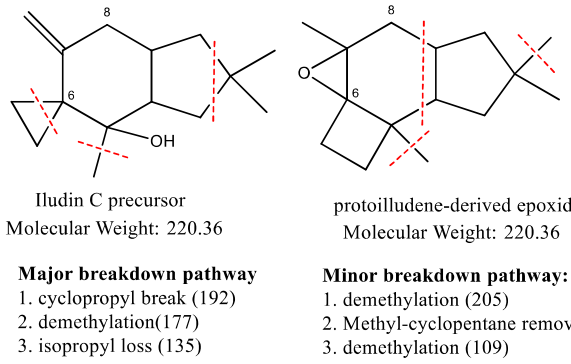


Figure 4-9: Potential Omp7b modification products and their MS breakdown
Only one breakdown pathway is shown for each molecule for clarity. However, both modified protoilludene scaffolds can follow the major and minor breakdown pathways to yield the same fragments. Dotted red lines denote plausible bond cleavages during MS fragmentation.

Intriguingly, both peaks have identical MS breakdown patterns, but their GC retention times differ by 1.1 minutes, suggesting the presence of stereoisomers. The MS breakdown patterns of these peaks consistently includes a minor 205 peak, consistent with demethylation, and major 124, 135, and 192 peaks. Hydroxyl groups are typically good leaving groups in MS, resulting in a net m/z loss of 17 from the parent ion, predicted to result in an initial 203 peak. No m/z gaps of 17 were observed, suggesting against any mechanisms involving simple hydroxylation by Omp7b. That a demethylation or 15 m/z loss is first observed is consistent with formation of a more stable epoxide or ketone group. Subsequent breakdown ion m/z values require oxygen addition to the cyclohexane-side of the protoilludene precursor, as the 135 peak can only be created from protoilludene by removal of the isobutyl group, and generation of the 124 peak requires

complete removal of a non-oxygenated cyclopenane ring, as shown in **Figure 4-9**. Ketone formation at the C8 position, a common feature of mature illudins, would also yield MS breakdown patterns consistent with observed ions; however, this possibility is not favored because it would require saturation of the double bond between C6-7, which is hypothesized to be required for contraction of the key cyclobutyl ring.

New nonvolatile compounds were also observed in yeast expressing *Omp7*+ *Omp7b* and *Omp7*+*Omp7a*. While significant and easily detectable by GC/MS, the concentration of these compounds was significantly lower than those of the aforementioned volatile sesquiterpenoids. As shown in **Figure 4-10**, several compounds with odd parent ions are present, as well as a large and irregular peak centered at Rt 23.6 minutes produced in cultures expressing Omp7a. Interestingly, while similar quantities of the same compounds were observed in cultures expressing *Omp7* with both *Omp7a* and *Omp7b* compared with two-enzyme expressing counterparts, no new compounds were observed when expressing all three enzymes together. This means that either Omp7a or Omp7b modify protoilludene as the first step in separate pathways, or that whichever enzyme is the second in the protoilludene modification pathway is promiscuous, and can accept both protoilludene, and the product of the other enzyme as substrates. This possibility would mean that the final product of Omp7, Omp7a, and Omp7b is not detectable using the current detection methods. Degradation of a larger sesquiterpenoid is supported by odd-numbered parent ions observed in the nonvolatile extraction, which for sesquiterpenoid compounds is consistent with degradation prior to MS detection, likely in the high heat of the inlet port or GC column. Therefore, I do not believe the compounds shown in **Figure 4-10** the true nonvolatile products of Omp7 + Omp7a/Omp7b, but rather breakdown products of larger modified sequiterperpenoids. Preliminary attempts to detect these nonvolatile compounds by LC-HRMS using the same conditions described in section 4.4.1 did not identify any compounds produced by heterologously expressed

Omp7 pathway genes, presumably due to product concentrations significantly lower than those observed in the native producer *O. olearius*. Because the compounds are non-volatile at room temperature, it should be possible to concentrate culture supernatants following 1:1 EtoAc extraction to enable detection using HPLC-MS or LC-HRMS, given sufficient culture medium and optimal expression. However, because the sesquiterpenoid intermediates expected are not in MS databases, simple detection the compounds produced by the Omp7 pathway would not be sufficient to identify their structures.

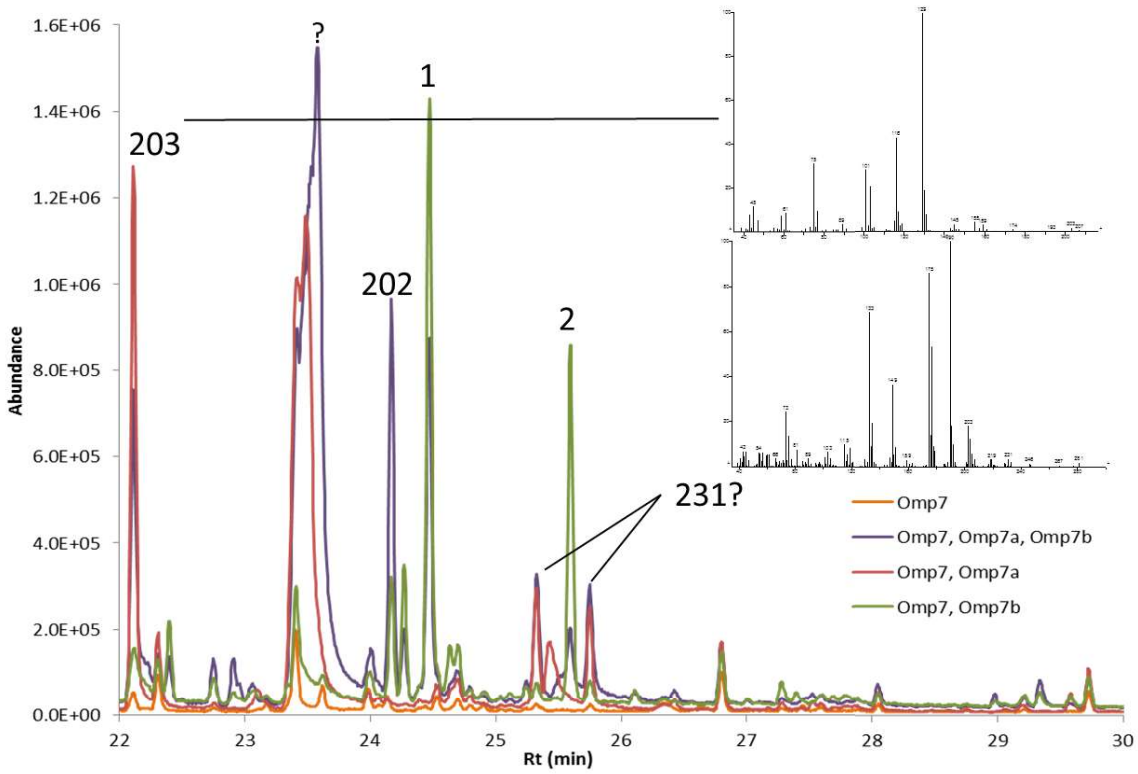


Figure 4-10: GC/MS of nonvolatile extractions from yeast expressing Omp7 with Omp7a and/or Omp7b.
The compounds with odd m/z values (203, 231) produced by Omp7+ Omp7b suggest breakdown products of larger molecules, while the large unknown peak at 23.5 minutes is weakly identified as D-allose, but is more likely to be a breakdown product of protoilludene. Compounds 1 and 2, observed in the culture headspace as well (**Figure 4-8**), are only produced by strains expressing both Omp7+Omp7b.

Therefore, the following sections focus on *in vivo* quantification of the production of Omp7 pathway products, to determine the feasibility producing sufficient sesquiterpenoids to enable structural identification using NMR of the compounds and the precursors to the degraded compounds observed by GC/MS.

4.4.5 Scale up and quantification of Omp7 pathway sesquiterpenoid production

A practical limit of one-half milligram of purified compound is required for structural identification of small molecules by NMR using available instruments.[422] While much higher concentration sesquiterpenes have been purified for NMR analysis in our lab previously,[86] no modified sesquiterpenoids have been produced for such an application using our resources. As a result, it was necessary to determine the productivity of our yeast expression system to assess the feasibility of our production system, and potential for structural identification by NMR. Scale-up of CEN.PK2 expression was performed in 500 mL culture volumes (2 L baffled shake flasks). Cultures reached final OD's of 6.27 (Omp7+Omp7a) and 8.13 (Omp7+Omp7a+Omp7b) and were extracted with EtoAc, concentrated, and analyzed by GC/MS-FID as described in section 4.3.4. Quantification of production of the major semi-volatile sesquiterpenoid (Compound 1, **Figure 4-8**), identified a total production of 50.11 µg (Omp7+Omp7a+Omp7b) and 10.77 µg (Omp7+Omp7b), respectively. Based on the volumetric productivity of Omp7+Omp7a+Omp7b, and assuming perfect purification yield, 5 L culture medium would be required for purification of Compound 1. Based on these results, production and purification of sufficient sesquiterpenoid to enable structural identification by NMR will require a combination of incremental improvements in culture density, specific productivity, and total reaction volume, or utilization of a completely different expression system. These possibilities will be discussed further in the following section.

4.5 Discussion and Future Work

The preceding chapter provided an overview of *in vivo* and *in vitro* activity screening of 16 enzymes, consisting of 10 P450s and several types of oxidoreductases found in protoilludene and hirsutene biosynethic gene clusters. This screening identified the P450 Omp7b and the VAO-type oxidoreductase Omp7a as the only enzymes with measurable protoilludene modifying activity. Further co-expression of Omp7 with Omp7b demonstrated production of new sesquiterpenoids with m/z of 220, consistent with single oxygen addition to the protoilludene scaffold. This is the first time that

protoilludene-modifying activity has been observed by a P450. In addition, the MS pattern of the resulting protoilludene-derivative is consistent with both oxygenation near the cyclobutyl ring, and consistent with rearrangement to form the critical cyclopropyl ring structure. Large-scale shake flask cultures expressing the protoilludene synthase Omp7 with Omp7b and/or Omp7a quantified the volumetric productivity of the major 220 m/z product at approximately 100 $\mu g\ L^{-1}$ in shake flask cultures. However, the production level of these modified sesquiterpenoids is currently too low to purify this and related compounds for NMR structural characterization, requiring several interconnected procedural modifications to enable effective production, purification, and structural identification of sesquiterpenoids derived from protoilludene Omp7a and Omp7b activity.

Many recently-demonstrated process improvements could result in more effective identification and characterization of active sesquiterpene-modifying enzymes. First, the success rate for expressing Basidiomycota Class II P450s, much less determining their catalytic activity, is notoriously low, with the most successful study available in the literature observing measurable protein production in only 8.8% of tested P450s.[408] Our own screening, described in section 4.4.1, similarly observed activity in only 1 of 10 P450-STS co-expression systems. This poor success rate, coupled with the difficulty in identifying P450 modification products even when active, necessitates utilization of high-throughput gene cloning strategies to make initial cloning from cDNA faster and more standardized, and should be coupled with strain development to ensure the highest possible rate of successful activity testing of P450s. Such strain development could include adaptation of the model Basidiomycota "yeast" *Ustilago maydis* for P450 expression, for which a series of plasmids have been developed.[325] Similarly, *Aspergillus oryzae* NSAR1 has been demonstrated in shake flask culture to produce P450-modified terpenoids at volumetric yields 50X higher than those obtained from our own yeast cultures (section 4.4.5), in quantities sufficient to easily enable structural characterization.[423] The use of non-standard heterologous hosts for terpenoid production may provide the critical boost to P450 expression reliability and productivity necessary to enable structural identification of Omp7a and Omp7b products without extensive strain and condition optimization.

Should expression of Omp7 pathway enzymes remain in yeast, a minimum 10X improvement in total Omp7a and Omp7b production is required to enable NMR identification of even the most highly produced protoilludene derivative. This improvement in total yield could be obtained by high-density, fed-batch yeast bioreactor cell culture. However, P450 expression in yeast is notoriously condition-dependent, and a significant change in growth conditions may negatively affect P450 activity under bioreactor growth, a possibility that remains to be tested. As a result, it is important not to underestimate the amount of time that may be required for bioreactor optimization while expressing a P450. In addition, while previous attempts at co-expression of *ooCPR* in yeast with *Omp6* and *Omp6i* had no effect on sesquiterpenoid profile, co-expression with *Omp7b* may increase its efficiency. Similarly, productivity gains may be obtained by identification, cloning, and co-expression of an *O. olearius* cytochrome *b5* to aid in electron transfer to Omp7b, in a similar manner to that previously performed in the industrial production of artemisinic acid in a highly engineered CEN.PK2 strain.[8] Finally, a revision of product detection methods is necessary to identify the true nonvolatile products of *Omp7+Omp7a* expressing cultures. The current GC/MS procedure is highly sensitive, but the soluble compounds produced by Omp7a appear to be degrading prior to entry into the MSD, resulting in odd parent ion m/z values and likely preventing identification of the true, soluble sesquiterpenoid(s) produced in cultures co-expressing *Omp7* and *Omp7a*. Replacement of GC/MS detection should employ room temperature reverse-phase purification by HPLC-MS,[421] to improve purification of heat-labile and non-volatile sesquiterpenoids that we now suspect to be intermediates in illudin biosynthesis. Such an HPLC method could be easily modified for preparative HPLC purification for subsequent structural identification by NMR.

Demonstration of protoilludene modification by Omp7b and Omp7a is a major step forward, and required extensive enzyme screening and process optimization. However, additional work is needed to improve production to enable structural identification of protoilludene derivatives by NMR. Structural confirmation of protoilludene scaffold modifications would be a major advance in biocatalysis, as no previous modifications of the protoilludene scaffold have been identified to date.

Chapter 5 Non-trichothecene sesquiterpene production requires Tri5 in *Fusarium graminearum* PH-1

The following research project developed from a collaboration with Karen Broz and Dr. Corby Kistler of the USDA-ARS Cereal Disease Lab, initially testing whether *F. graminearum* produced any non-trichothecene sesquiterpenes, then grew into the chapter below. All *F. graminearum* cultivations and media were made by Karen Broz, as were all genomically modified *F. graminearum* strains. All other work was performed by Chris Flynn.

5.1 Summary

Deoxynivalenol (DON) is the primary mycotoxin produced by the cereal disease causing ascomycete *F. graminearum*. DON is a trichothecene sesquiterpene derived *via* the trichodiene synthase (Tri5) product trichodiene. However, Tri5 is only one of eight putative STSs encoded in the *F. graminearum* genome. In addition, transcriptomic analysis revealed increased expression of four STS during infection, only two of which have been characterized. The products of these uncharacterized STS may an important role in pathogenesis. However, the products of these STS are currently unknown. In this project, we set out to identify what, if any, non-trichothecene sesquiterpenes (NTS) are produced by *F. graminearum*. Here I describe sampling of the volatile and soluble fractions of *F. graminearum* PH-1 shake flask cultures. When grown in trichothecene induction medium, PH-1 produces trichothecenes and several NTS, whereas no volatile terpenes were detected without induction. Surprisingly, a *Δtri5* deletion strain grown in inducing conditions not only ceased production of the anticipated trichothecenes, but also stopped most NTS biosynthesis. To test the possibility that *F. graminearum* Tri5 is a non-specific STS directly producing all observed sesquiterpenes, Tri5 was cloned from *F. graminearum* PH-1 cDNA, expressed in *E. coli*, and shown to produce only 8 of the 15 observed sesquiterpenes. Therefore, while Tri5 expression in *F. graminearum* PH-1 is necessary for NTS biosynthesis, direct catalysis by Tri5 is not sufficient to explain the diversity of sesquiterpenoids produced under trichothecene inducing conditions, nor can it explain the sesquiterpene deficient phenotype observed in the *Δtri5* strain.

These findings suggest that Tri5 expression, through an as-of-yet unidentified mechanism, is required for expression of NTS producing STSs. To determine the extent to which Tri5 influences NTS biosynthesis *via* either protein:protein interactions, or *via* chemical signaling by trichodiene, a catalytically inactive Tri5 was created using site-directed-mutagenesis, and verified by Circular Dichroism to retain its secondary structure. This inactive *tri5* mutant is currently being transformed into *F. graminearum* to test these hypotheses. While the role of trichothecenes in phytotoxicity is known, the biological function of NTS and their recently discovered co-regulation has not yet been determined.

5.2 Introduction

Fusarium graminearum is a fungal pathogen, and the most common cause of the major cereal crop destroying infection *Fusarium* head blight (FHB).[424] FHB is an increasingly damaging blight worldwide, recently causing billions of dollars in crop losses, with regional losses of more than 50% occurring with regularity.[425] As a result, the worldwide staple foods corn, wheat, and barley, are vulnerable to FHB, raising the cost of production, and necessitating heavy use of fungicides.[425] As a result, understanding the physiology and pathogenesis of *F. graminearum* is essential to combatting FHB infection, as well as breeding resistant cereal crops.

F. graminearum produces a suite of complex sesquiterpenoids, collectively called

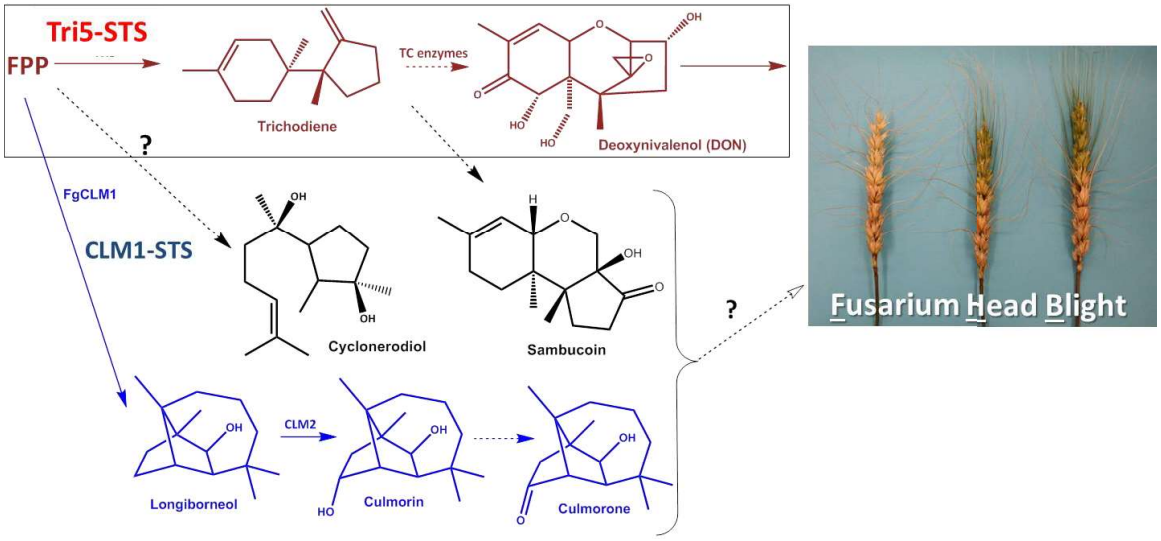


Figure 5-1: Project overview.
F. graminearum produces trichothecenes (maroon), including deoxynivalenol (DON) known to be virulence factors in wheat Fusarium head blight (FHB). *F. graminearum* also produces NTS with unknown effect on FHB, including sambucoin, and cyclonerodiol (black), and the longiborneol-derived culmorins (blue).

trichothecenes, during wheat and corn infection, but not barley infection.[426, 427] Trichothecenes encompass over 200 sesquiterpenoids all derived from trichodiene, the product of the STS Tri5. [200, 201] The primary trichothecenes are deoxynivalenol (DON, commonly known as vomitoxin), nivalenol (NIV), and T-2 toxin.[426] *Fusarium graminearum* strains are divided by chemotype, based on the predominant trichothecene product. The *F. graminearum* PH-1 chemotype primarily

produces 15-acetyldeoxynivalenol (15-ADON), and is the subject of this research. Trichothecene biosynthesis is the best characterized sesquiterpenoid biosynthetic pathway in fungi, and is described in detail in Section 1.6.2. *F. graminearum* deficient in Tri5, and resulting trichothecene production, are unable to spread across an infected wheathead.[427] Trichothecenes are not only essential to the spread of FHB, but also toxic to the human and animals who consume them from contaminated food sources, meaning FHB not only reduces crop yields, the accumulated trichothecenes can render the remaining wheat inedible.[424] In addition to trichothecenes, *Fusarium* species are known to produce the non-trichothecene sesquiterpenes (NTS) culmorin and cyclonerodiol.[153, 428] There is ample evidence that the complex cocktail of sesquiterpenes produced by an organism can have signifinant effect on chemical signaling and toxicity.[101] However, the effect of these NTS on FHB, and whether any additional NTS are produced by *F. graminearum,* were both unknown at the start of this project (**Figure 5-1**).

The initial aim of this project was to determine whether and what NTS are produced by *F. graminearum*, and to determine how that production was affected by induction. Analysis of the *F. graminearum* genome identified 8 putative STS, only two of which (CLM1, Tri5) are characterized (**Figure 5-2**).[153, 207] Trichothecene production in *F. graminearum* is activated in a tissue-specific manner in wheat endosperm, responding to production of putrescene and related polyamines that are part of the plant stress response.[429, 430] Later work demonstrated addition of either putrescene or agmatine to liquid cultures induced production of trichothecenes similar to those observed in wheat infection, providing an axenic proxy for study of trichothecene production in *F. graminearum*, with the implication that overexpressed genes are likely involved in pathogenesis.[414] This, microarray analysis of agmatine-induced *F. graminearum* showed 130 genes with greater than 10-fold difference in expression between agmatine (a polyamine similar to putrescene) and glutamine controls, most being upregulated genes from non-trichothecene producing pathway.[431] Of those 130 up-regulated genes, four are STS, including CLM1, Tri5, and the putative STS FGSG08181 and FGSG16873, demonstrating both the genomic and epigenetic potential for production

of as-of-yet undescribed NTS. Finally, a significant limitation of microarray expression analysis is that only changes in expression under different conditions are observed. If the four STS showing no change in expression are constitutively expressed, then *F. graminearum* should produce many of the same (currently unknown) NTS regardless of polyamine induction, and those NTS would be unlikely to be involved in pathogenesis. After establishing that *F. graminearum* has the genetic potential to produce NTS, I specifically set out to test the headspace of the *F. graminearum* PH-1 chemotype for sesquiterpene production with and without induction

to gain a better understanding of both what NTS are produced, with the implication that any NTS production dependent upon induction is likely to be involved in pathogenesis.

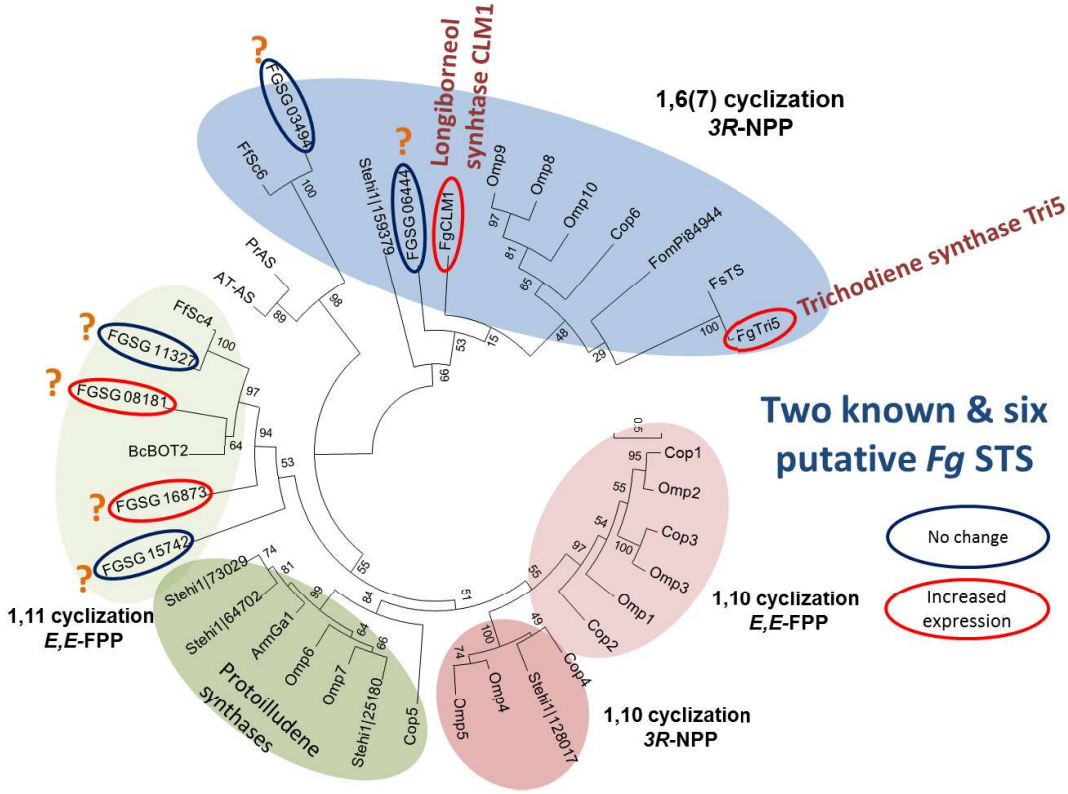


Figure 5-2: Terpenome of *F. graminearum*.
The genome of *F. graminearum* contains eight putative STSs, the activity of only two of which are currently known. Phylogenetic analysis comparing these 8 to known Basidiomycota and Ascomycota STS suggest the remaining enzymes follow 1, 6 or a 1,11 initial cyclization mechanism. Finally, microarray analysis identified four STS with increased expression upon induction with agmatine (red circles), identified two unknown STS, FGSG08181 and FGSG16873, co-expressed in inducing conditions along with the known longiborneol and trichodiene synthases. Four STS did not change expression in agmatine (blue circles).

5.3 Materials and Methods

5.3.1 Strains, media, and sesquiterpenoid profiling

Fifty milliliter *F. graminearum* cultures (250 mL shake flasks, covered in double-layered aluminum foil) were cultivated as described by Gardiner et al. (2009) [430]. Briefly, conidia are resuspended at a concentration of 10^4 mL^{-1} in either Toxin Induction Medium (TBI) supplemented with 5 mM putrescene at pH 4.5, or Non-Induction Medium (NIM) containing 10 mM sodium nitrate. All other ingredients are identical (15 g sucrose, 0.5 g KH_2PO_4, 0.25 g $MgSO_3$, 0.25 g KCl, 1 mL $FeSO_4*H_2O$ (5 mg mL^{-1} stock), and trace elements. Culture media are sterile filtered. Growth was at 25°C and 220 rpm shaking in the dark. Sesquiterpenoids were extracted from headspace as described in earlier chapters by PDMS SPME extraction. Soluble sesquiterpenoids were extracted as described in Chapter 4, except supernatant was further clarified following centrifugation by decanting and 0.45 µm syringe filtration. PDMS-DVB SPME extraction and GC/MS separation was then utilized as described in Chapter 4.

Plasmid maintenance was performed using *E. coli* strain JM109. Expression Tri5 and Tri5 N255D S229T were performed in BL21 (DE3) for initial *in vivo* activity assays in 50 mL LB (250 mL shake flask), whereas Rosetta cells were used for protein purification in 500 mL LB. All cultures utilized liquid LB supplemented with appropriate antibiotics (kanamycin, chloramphenicol). Unless otherwise stated, all *E. coli* growth was performed at 37°C, and all Tri5 expression was performed at 30°C. No difference in product profile was observed across *E. coli* cell lines.

5.3.2 Plasmid construction

Tri5 was cloned from Trizol-prepared cDNA with its single intron for insertion into the pET28a plasmid using primers F_tri5_NdeI (ggaattcc**atg**GAAAACTTTCCCACCGAGTATTTTCTCAAC) and R_tri5_NotI (aaggaaaaaagcggccgcTTACTCCACTAGCTCAATTGAACTTAGGATGGG). The intron was removed with these primers and Tri5-exon1-R (AGTCCATAGTGCTACGGATAAGGTTCAATGAGC) and Tri5-Exon2-F (CATTGAACCTTATCCGTAGCACTATGGACTTTTTTGAGGGATGTTGGATTGA

GCAGTACAAC) and reassembled with overlap-extension PCR and ligated into pET28a using the NdeI and NotI restriction sites. The resulting plasmid was sequenced and tested for trichodiene synthase activity. The intron could be identified by comparison to the well-characterized *F. sporotrichiodes* Tri5, which was also cloned with this intron, which was similarly removed prior to obtaining the active Tri5 enzyme.[207] All PCR utilized Phusion as described in earlier chapters. Tri5 N225D S229T was subsequently created via Q5 site-directed mutagenesis with competent cells, using the primers atcatcGACCCAAACCATCCAGTTC and ctcatgaCATTCTACAAGGAATTCGAC and following manufacturer conditions.

5.3.3 Trichodiene and 15-ADON feeding experiments

Trichodiene was obtained as a gift from Dr. Susan McCormick (USDA ARS), who purified from culture extracts of *F. sporotrichioides Δtri4* mutants using silica gel columns, eluted with hexane or dichloromethane and increasing amounts of acetone (TD eluted with 2%). Upon arrival, trichodiene was then analyzed by GC/MS to determine it to be ~70% trichodiene. Initially, this trichodiene-sesquiterpene mixture was then dissolved in sufficient ethanol to reach a concentration of 580 mM at 70% purity. Fifty microliters of ethanol (solvent control) or trichodiene-sesquiterpene mixture were then added at the time of induction to 50 mL *F. graminearum* shake flask cultures. Culture headspace and supernatants were analyzed for volatile and nonvolatile sesquiterpenoid production by SPME extraction and GC/MS as described above. Subsequent dose-response curve comparing trichodiene concentrations at 0, 25, 50, 125, and 250 µM where total ethanol addition was 50 ul, with varying quantities of trichodiene-acetone initial stock added depending on desired concentration. Trichodiene feeding concentration of 250 µM have been reported in literature to affect gene expression.[432]

15-ADON feeding experiments were performed similarly, except 15-ADON (Sigma-Aldrich, cas number 8833796-6) was dissolved in acetonitrile and added to induced shake flasks at concentrations of 0, 2.5, 5, 10, 25, and 50 µM at the time of induction. Reported maximum 15-ADON concentrations produced by *F. graminearum* PH-1 are approximately 50 uM.[433]

5.3.4 Tri5 and Tri5 N225D S229T expression and purification by FPLC

Tri5 and inactive Tri5 N225D S229T were purified from E. coli Rosetta cells in 500 mL LB medium (2 L shake flask). Four milliliters overnight LB shake tubes (37C) were used to inoculate large shake flasks, which were allowed to reach OD ~0.3 and induced with 0.5 mM IPTG, incubated for 10 minutes at 4°C, and incubated at 17°C with shaking at 220 rpm for 3 hours. Cells were then pelleted at 6000 rpm for 10 minutes (Beckman J2-HS) and frozen at -80°C.

Rosetta™ (EMD Millipore, Billerica, MA) *E. coli* cells containing pET28a-Tri5 and pET28a-Tri5 N225D S229T were cultivated in 500 mL LB medium with appropriate antibiotics (2 L flask) at 30°C and induced with 0.5 mM IPTG at OD_{600} of ~0.3. Induced cultures were then rested for 10 minutes at 4°C and incubated 3 additional hours at 17°C. Cultures were monitored for trichodiene production by SPME as described above, and pelleted by 10 min centrifugation at 6,000 rpm in a Beckman J2-HS (Brea, CA) floor centrifuge and immediately frozen at -80°C. Pellets were resuspended in 30 mL STS purification buffer (50 mM Tris, pH 8, 250 mM NaCl, 5 mM imidazole, 10 mM $MgCl_2$, 1 mM PMSF) supplemented with cOmplete protease inhibitor (Sigma-Aldrich) for 4-minute sonication (1 second on, 2 seconds off), followed by 30 min centrifugation at 12,000 rpm. Soluble lysate was further 0.45 um syringe filtered and purified by ÄKTA FPLC (GE Healthcare) using 5 mL His-trap^{TM} FF Ni^{2+} (GE Healthcare) affinity columns. Columns were eluted with STS elution buffer (50 mM Tris, pH 8, 250 mM NaCl, 250 mM imidazole, 10 mM $MgCl_2$, 1 mM PMSF) collecting 4 mL fractions. This sample was then concentrated using Amicon (EMD Millipore) Ultra -15 3,000 NMWL centrifugation filter units and dialyzed overnight into size exclusion buffer (5% glycerol, 20 mM Tris, pH 7.5, 50 mM NaCl, 0.22 µm filtered) for Superdex 200 10/300 (GE LifeSciences) purification with 1 mL fractionation. Protein fractions were combined and concentrated by 3000 NMWL amicon filters to a concentration of approximately 1 mg mL^{-1}. WT Tri5 fractions with increased UV_{280} absorbance were analyzed for trichodiene synthase activity in activity assay buffer (20 mM Tris, pH 8, 10 mM $MgCl_2$, 500 mM NaCl) supplemented with 2 uM FPP and 10 ul protein fraction. All fractions identified by UV_{280}

were monitored by 12% SDS-PAGE with Coomassee stain for intact Tri5 and Tri5 N225D S229T. Purified and concentrated proteins were then dialyzed overnight into CD buffer (10 mM KPO_4, pH 8.0).

5.3.4 Circular Dichroism

Tri5 and Tri5 N225D S229T purified by Ni^{2+} and size exclusion column chromatography were diluted to a concentration of 0.1 mg/mL (2.146 and 2.168 umol L^{-1} for WT and mutant, respectively) and analyzed by Circular dichroism (CD) from 260 to 185 nm with single nm steps. CD analysis was performed on a JASCO J-815 spectropolarimeter (Ishikawa-cho, Japan) following the procedure described by Greenfield (2006).{Greenfield, 2006 #456 and appropriate air and solvent-only controls. CD spectra comparing WT Tri5 and Tri5 N225D S229T molar ellipticity (Θ) from 185-260 nm were overlaid to identify any loss of α-helical structure resulting from the active site mutations performed.

5.4 Results

5.4.1 Non-trichothecene sesquiterpenoids (NTS) are produced by induced *F. graminearum* cultures

The initial goal of this research was to determine whether *F. graminearum* produces NTS with and without putrescene induction, with the implication that NTS specifically produced with induction are likely involved in FHB pathogenesis. Putrescene induction mimics wheat-head infection, providing axenic surrogate culture conditions useful for analysis of *F. graminearum* gene expression, toxin production, [414] and physiological changes[92] during pathogenesis in the absence of slow growing, living wheat heads as infection hosts. *F. graminearum* Cultures were grown in 50 mL TBI or non-inducing medium (NIM) and sampled by SPME and analyzed by GC/MS at 24, 48, 72, and 96 hours post-induction (hpi) for production of sesquiterpenes. Peak sesquiterpene production was observed at 72 hpi, and all subsequent analysis was performed at this time point. Twelve sesquiterpenes could be identified in induced WT *F. graminearum* PH-1 cultures, while none were observed in non-induced cultures **(Figure 5-4)**. These results were repeated in triplicate using independent biological cultures on

separate days. Because no sesquiterpene production was observed in non-induced cultures, the possibility that the four unidentified STSs identified in the *F. graminearum* genome (**Figure 5-1)** are constitutively expressed is precluded. In addition, tight control of NTS production alongside trichothecene biosynthesis during

infective conditions suggests an important and unexplored role in pathogenesis. Finally, most sesquiterpenes identified in putrescene induced cultures are derived from an initial

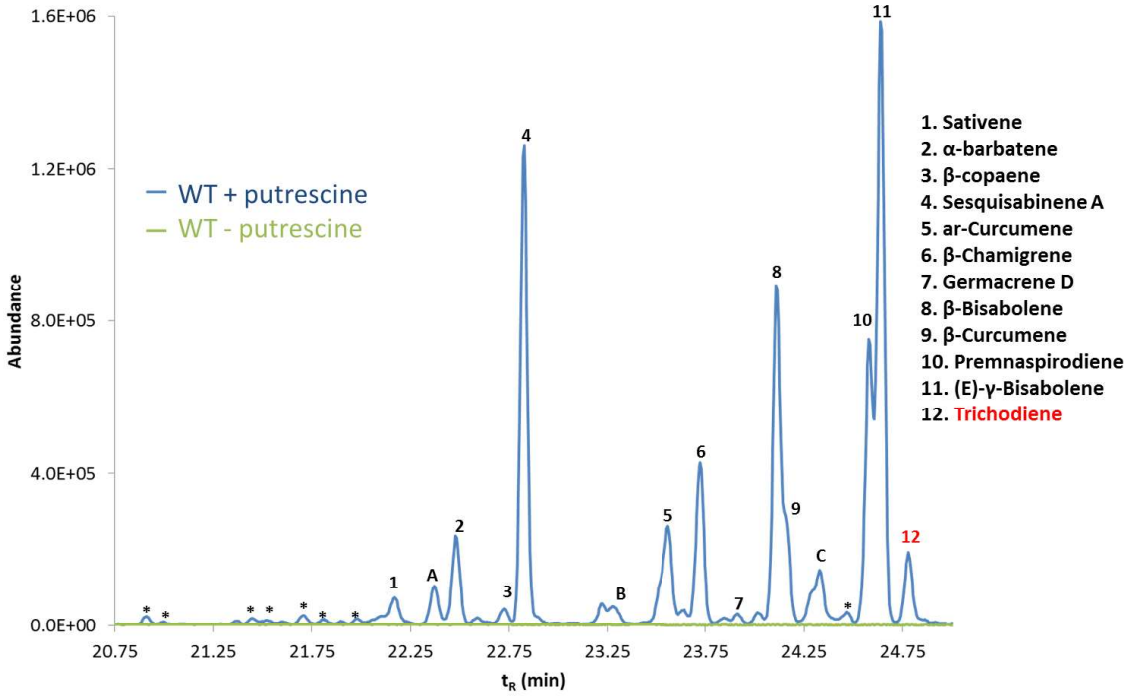


Figure 5-3: Sesquiterpene production in *F.graminearum* requires putrescene induction
Wild-type *F. graminearum* PH-1 grown on non-inducing medium (green) produces no sesquiterpenes. The same strain induced with putrescene (blue) produces 12 identifiable sesquiterpenes and three major sesquiterpenes that could not be identified (labelled A, B, C). Accumulation of trichodiene is presumably depressed because of conversion to trichothecenes by Tri4 and other trichothecene producing enzymes.

1,6 NPP cyclization mechanism (sesquiterpenes 2, 4-6, 8-12 **Figure 5-3**), they could be side products of trichodiene synthase (Tri5). To identify which sesquiterpenes are made directly by Tri5, and which are produced by other, co-expressed STSs, *tri5* was cloned from *F. graminearum* cDNA and expressed in *E. coli* BL21(DE3)using the pET28a plasmid. Overlay of the product profiles of putrescene-induced *F. graminearum* PH-1 and *E. coli* expressed *tri5* (**Figure 5-4**) show that Tri5 produces all the 1,6 NPP cyclization-derived sesquiterpenes found in *F. graminearum* cultures except ar-curcumene. In addition, the sesquiterpenes derived from 1,10 cyclization products

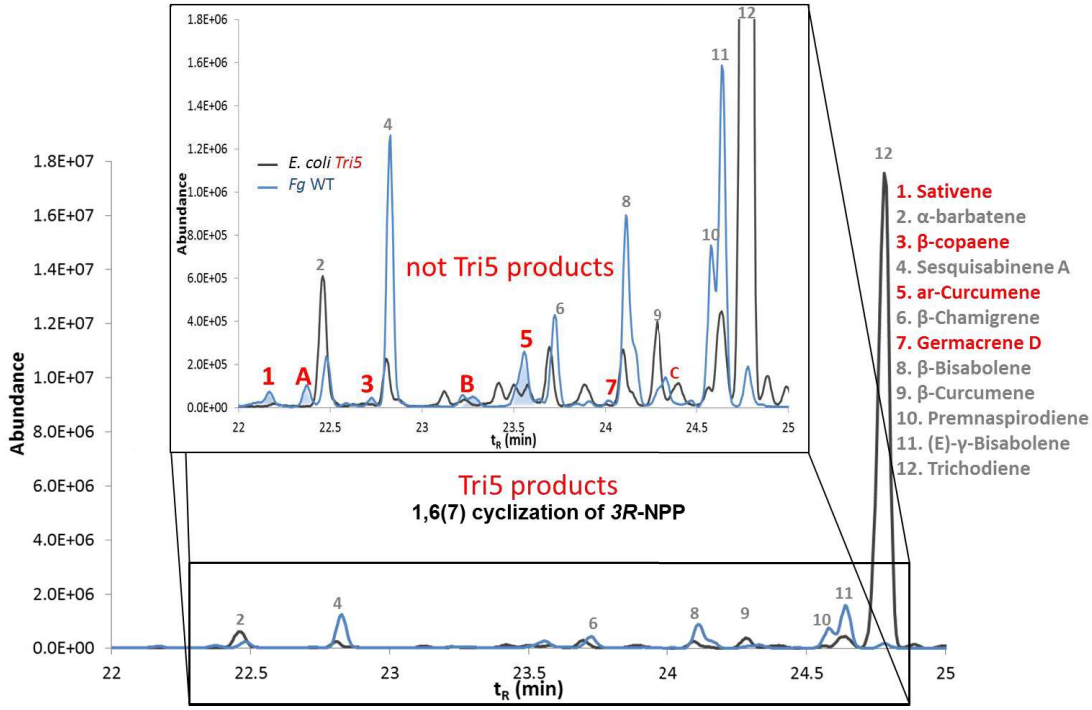


Figure 5-4: Identification of non-Tri5 derived sesquiterpenes
Overlay of GC chromatograph comparing putrescine-induced *F. graminearum* (blue) to *E. coli*-expressed Tri5 (black) to identify the sesquiterpenes not produced by Tri5. Four sesquiterpenes that could be identified, and three unidentified sesquiterpenes are not produced directly by Tri5 (inset), but nevertheless cease accumulation in *Δtri5* cultures, suggesting biosynthesis by unknown STS. Sesquiterpenes not produced by Tri5 are shown in red.

(sativene, β-copaene, germacrene D) as well as three unidentified sesquiterpenes (Compounds A-C) were observed in the *F. graminearum* headspace, but not produced by Tri5, suggesting they are produced by STSs specifically co-expressed with Tri5, most likely FGSG08181 or FGSG16873.

Attempts to clone FGSG08181 and FGSG16873 from cDNA without introns were unsuccessful, possibly due to low expression level based on sesquiterpene production. Highly expressed *tri5* could also be cloned only with its single intron which was removed to obtain the ORF encoding the active Tri5 from both *F. graminearum* as described above and previously from *F. sporotrichiodes*.[207] However, unlike Tri5, FGSG08181 contains two predicted introns, and FGSG16873 contains 8 introns, far too many to remove synthetically and obtain active enzyme, given the poor reliability of intron-exon predictions from fungi (described in detail Section 1.6.6). FGSG06444 produces a single as-of-yet unidentified sesquiterpene-alcohol with m/z 222[434] not detected in our *F. graminearum* cultures. Intron-free FGSG15742 did not produce any volatile or non-

volatile sesquiterpenes when expressed in *E. coli*. Finally, considering the possibility that non-canonical STS resembling bergomatene synthase may be responsible for production of the NTS observed in induced *F. graminearum* cultures, a bergomatene synthase[88]-like gene, FGSG03705, was identified, cloned, and expressed, but did not produce sesquiterpenes when expressed in yeast or *E. coli*. Because of the tangential nature of identifying the STS responsible for all sesquiterpene production by induced *F. graminearum* cultures to the overall project goal of identifying the sesquiterpenes produced, no further research was performed to clone and characterize unknown *F. graminearum* STS.

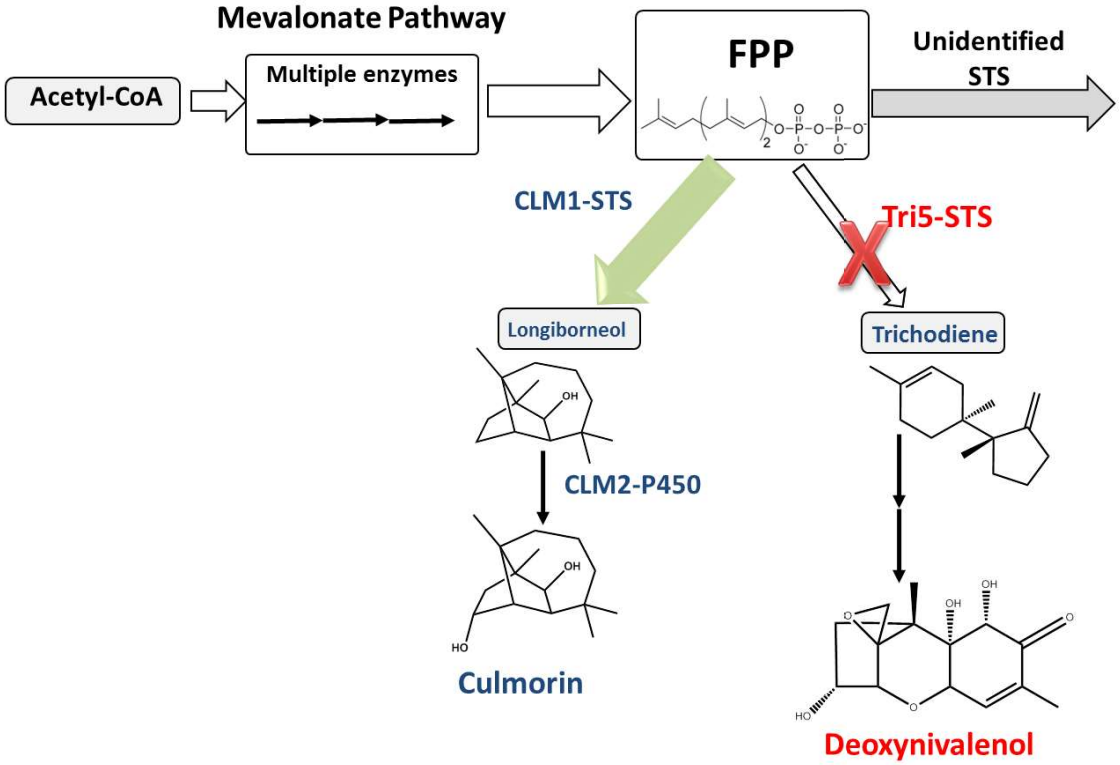


Figure 5-5 Hypothesis that *Δtri5* will increase production of non-trichothecene sesquiterpenes based on competition for FPP.
Tri5 deletion was hypothesized to remove the major depletion pathway of FPP, increasing available FPP for biosynthesis of non-trichothecenes via CLM1 and other STS. Increased production of non-trichothecene sesquiterpenoids would therefore aid detection and identification of minor sesquiterpenoids.

5.4.2 *Δtri5* eliminates trichothecene *and* most NTS production

Previous research by Gardiner et. al (2009)[431] demonstrated that deletion of the CLM1 longiborneol synthase-encoding gene[153] (previously FGSG10397) resulted in dramatic, 14-fold increase in DON production and increased virulence toward wheat. The authors of this previous work hypothesized that CLM1 competes with Tri5 for their common substrate FPP, thus reducing trichothecene production when both are expressed. Based on this FPP competition hypothesis (**Figure 5-5**), an *F. graminearum Δtri5* strain that is unable to produce trichodiene and derived trichothecenes was created to increase production of NTS. This *Δtri5* strain was then cultivated under the same conditions

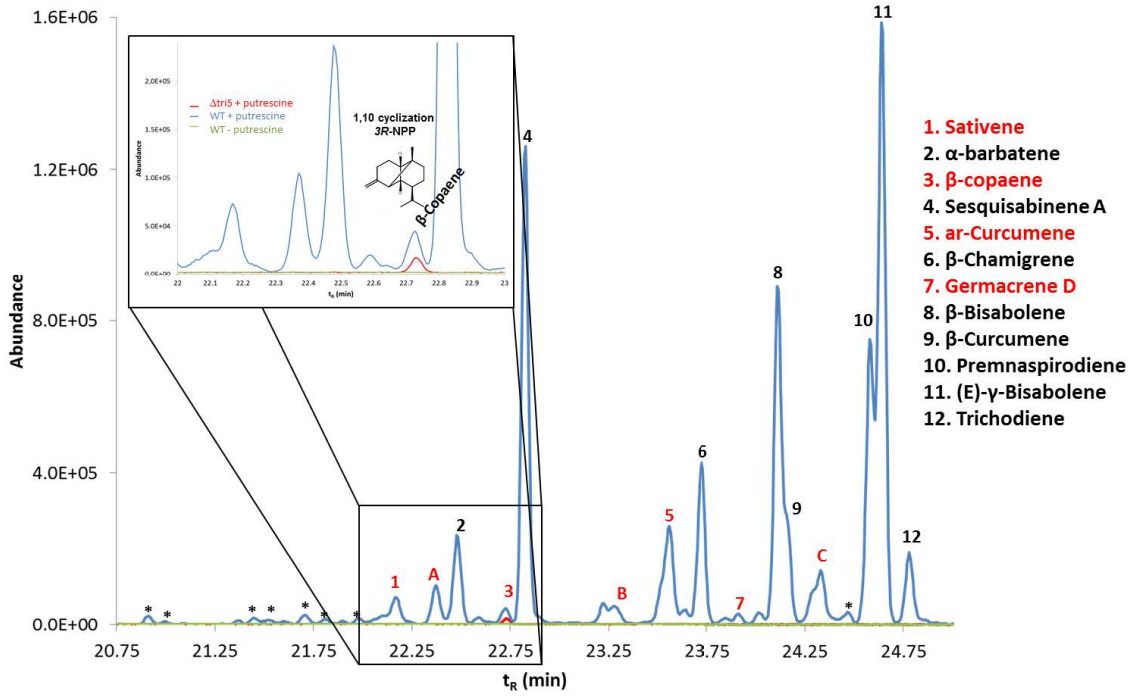


Figure 5-6: Putrescene-induced *ΔTri5* F. graminearum ceases all sesquiterpene production except β-copaene.
Complete removal of the *tri5* ORF eliminated not just the expected Tri5 products, but also nearly all volatile sesquiterpenes produced by other STSs. The *ΔTri5* mutant (red) produces only β-copaene (visible in inset), otherwise being identical to the sesquiterpene-deficient, uninduced wild-type (green). This differs greatly from the large number of sesquiterpenes, both trichothecene and NTS, produced by the putrescene-induced wild-type culture (blue). Sesquiterpenes not produced directly by Tri5 are shown in red.

described above, using induced and uninduced WT *F. graminearum* as positive and negative controls. Surprisingly, the *Δtri5* strain did not produce more NTS than wild-type induced cultures (**Figure 5-6**). Instead, it ceased production of all volatile sesquiterpenes except for the minor NTS compound β-copaene, whose production was not significantly

changed from induced WT. Neither uninduced *F. graminearum Δtri5* nor uninduced *F. graminearum* WT produced any sesquiterpenes. Representative chromatographs are shown in (**Figure 5-6**); the experiment was repeated in triplicate.

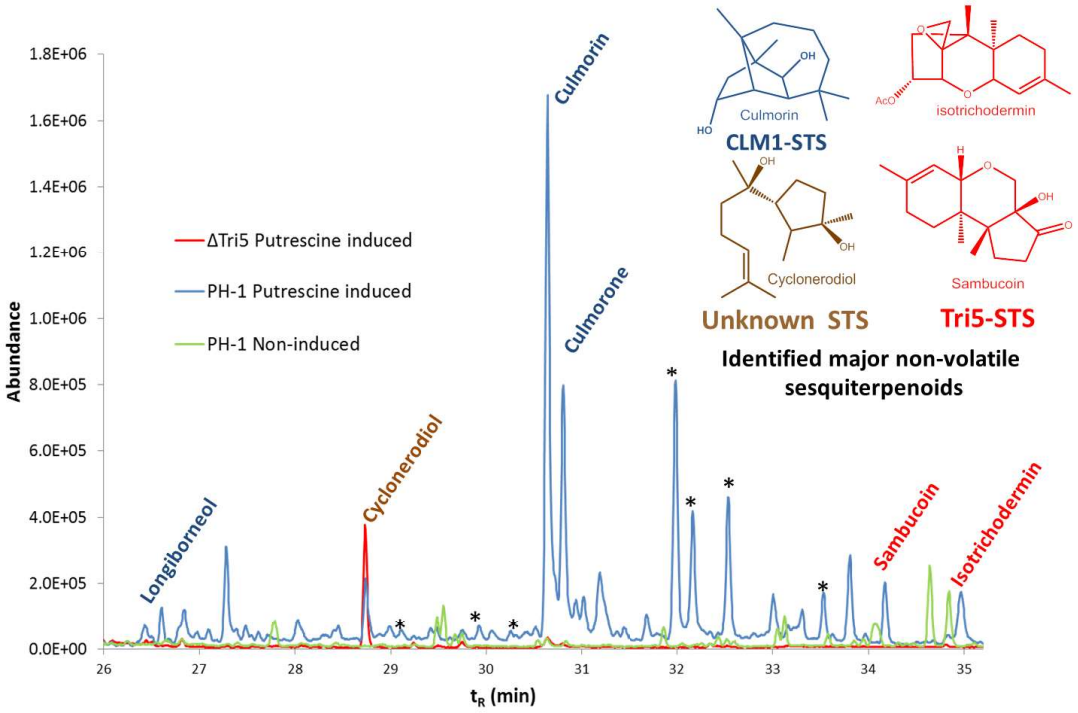


Figure 5-7: GC chromatograph of non-volatile sesquiterpenoids produced by *F. graminearum* WT (blue), *Δtri5* (red),and uninduced cultures (green)
Non-volatile sesquiterpenoids produced by WT *F. graminearum* (blue line) include the expected trichodiene-derived intermediates isotrichodermin and sambucoin, CLM1-derived longiborneol and its derivatives culmorin and culmorone (blue labels), and finally cyclonerodiol (brown). The *Δtri5* strain ceased all sesquiterpene production except cyclonerodiol (red line). The uninduced strains did not produce any sesquiterpenoids (WT uninduced: green). Unidentified peaks with characteristic sesquiterpenoid parent ions are marked with an asterisk (*).

Hydroxylated and highly-modified sesquiterpenoids are often non-volatile. To rule out the possibility that sesquiterpenoids are producted in the *F. graminearum Δtri5* strain, but not detectable due to rapid modification to less-volatile compounds, culture supernatants were then extracted and analyzed by GC/MS using previously described methods. Shown in **Figure 5-7**, induced *F. graminearum* produced large amounts of non-volatile sesquiterpenoids, including culmorin, isotrichodermin, and cyclonerodiol, as well as additional sesquiterpenoids that were strong fits to known trichothecene intermediates (between t_r 32-34 minutes), but could not be definitively identified by library comparison.

In contrast, *F. graminearum Δtri5* produced only cyclonerodiol in similar concentration to the wild-type control. Similar to the volatile extraction results, no uninduced cultures produced sesquiterpenoids. Taken together, these analyses demonstrate that both Tri5 and putrescene induction are necessary for production of trichothecene and most NTS, particularly culmorin, in *F. graminearum* cultures. β-copaene and cyclonerodiol, whose synthases are unknown, are produced independently *tri5*, but only when induced by putrescene. However, the mechanism by which Tri5 controls NTS production is unknown.

5.4.3 Trichothecene feeding does not induce production of culmorin or other non-trichothecene sesquiterpenoids (NTS)

The simplest hypothesis is that trichodiene or a trichothecene acts as chemical signal activating NTS biosynthesis. To directly test the extent to which trichothecenes affect NTS production, trichodiene and 15-ADON were added to putrescene-induced *F. graminearum* PH-1 shake flask cultures. In addition to wild-type and *ΔTri5* strains, a complete deletion of the primary trichothecene gene cluster was also tested, dubbed ΔTrichothecene Cluster or ΔTC. Seventy-percent pure trichodiene was obtained as a gift from Susan McCormick as a concentrated oil extracted from *Fusarium sporotrichioides ΔTri4* mutant treated with xanthotoxin. This trichodiene stock contained similar volatile sesquiterpenes to those observed in our *F. graminearum* cultures, obscuring any analysis of volatile NTS production. However, the trichodiene stock did not contain culmorin or trichothecenes, enabling analysis of non-volatile compounds produced by *F. graminearum* upon treatment with 580 μM trichodiene, summarized in **Figure 5-8**. With and without trichodiene treatment, all cultures produced cyclonerodiol, and *ΔTri5* strain was complemented in production of isotrichodermin, a trichothecene intermediate requiring activity of Tri4, Tri101, and at least one unknown enzyme to produce from trichodiene.

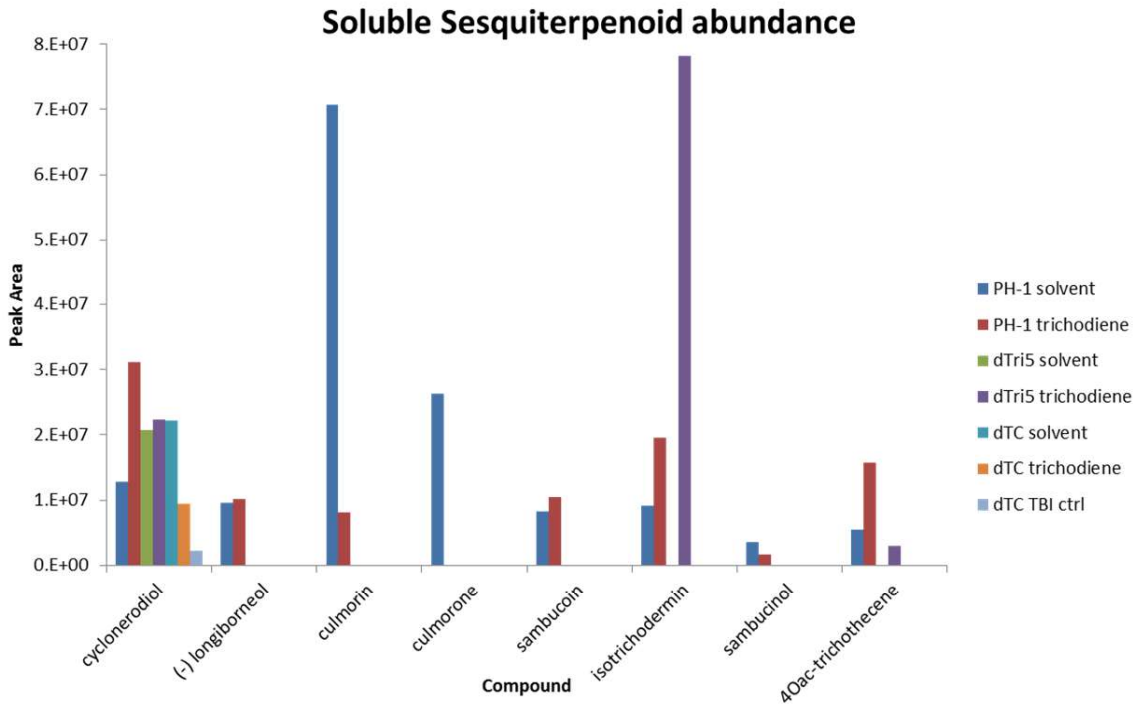


Figure 5-8: Product profile of trichodiene-fed *F. graminearum* strains
Soluble product profiles of putrescene-induced WT *F. graminearum* PH-1, *Δtri5*, and complete trichothecene cluster deletions (ΔTC) strains. All strains produced cyclonerodiol. Trichodiene supplementation (580 µM) complemented trichothecene production in the *Δtri5* strain, restoring isotrichodermin biosynthesis. In addition, trichodiene supplementation depressed culmorin production in the WT strain, suggesting inhibition at high trichodiene concentrations. No other sesquiterpenes were observed in dTC or *Δtri5* strains.

This result demonstrates expression of trichothecene pathway enzymes in the absence of Tri5, at least when complemented with trichodiene. Regarding NTS biosynthesis, only WT *F. graminearum* PH-1 produced culmorin, with 7X higher accumulation by the WT solvent (no trichodiene) control when compared to the trichodiene-fed WT strain. Five-hundred eighty micromolar trichodiene is in excess of maximum physiological concentrations of approximately 250 µM,[432] and suggests trichodiene may be inhibitory at high concentrations. To address this possibility, these same strains were then treated with trichodiene concentrations of 0, 10, 25, 50, 125, and 250 µM. However, only a weak correlation between trichodiene concentration and cyclonerodiol production was observed. A correlation was observed between trichodiene concentration and culmorin production at for the four data points between 0 and 50 uM trichodiene; however, an impurity in the trichodiene stock obscured the culmorin peak at 125 and 250 µM concentrations, making any fit inconclusive due to insufficient data points (data not

shown). Similarly, feeding 15-ADON at concentrations from 0-50 μM did not restore production of culmorin or other sesquiterpenes in the *Δtri5* strain (data not shown). Sesquiterpenoid production in *Δtri5* and ΔTC strains is dramatically reduced compared to WT, to near the limit of detection by GC/MS (~pM concentrations), severely limiting the accuracy with which such dose-response experiments can be made. I therefore decided to develop a new strategy to determine the mechanism through which Tri5 affects NTS biosynthesis.

5.4.4 Inactivation of Tri5 to discern NTS regulation by Tri5 via trichodiene production or protein-protein interactions

Based on the data presented in **Figures 5-5** through **5-7**, we know that sesquiterpenoids, most importantly culmorin, which are not directly produced by Tri5 cease production in the *tri5* strain, suggesting a mechanism of coregulation dependent on Tri5. It is therefore necessary to design an experiment to test the working model of this system (**Figure 5-9**), to determine whether co-production of NTS with trichothecenes is regulated by trichodiene (or a trichothecene derivative) as signaling molecules, or through protein:protein interaction between Tri5 with an as-yet unidentified regulatory protein. Because unmodified sesquiterpenes are volatile, lipophilic, and structurally complex compounds, they make ideal signaling molecules both for inter- and intra-species communication, with multiple examples of plant-plant and plant-insect interactions mediated by sesquiterpenes.[101] Additionally, repression of pathogenic *Fusarium* sp. has been shown to be mediated by volatile sesquiterpenes produce by an *F. oxysporum* strain containing an endocellular symbiont bacterium.[435] Whereas sesquiterpenes have been reported to be interspecies signaling molecules, there is no precedent for sesquiterpene toxin precursors acting as signaling molecules as well. There is also no precedent for STS mediating protein expression through protein:protein interactions, but it cannot be ruled out without further experimentation. To decide between these two exciting possibilities, I set out to generate a catalytically inactive Tri5 mutant, which retains is secondary structure for potential protein:protein interactions.

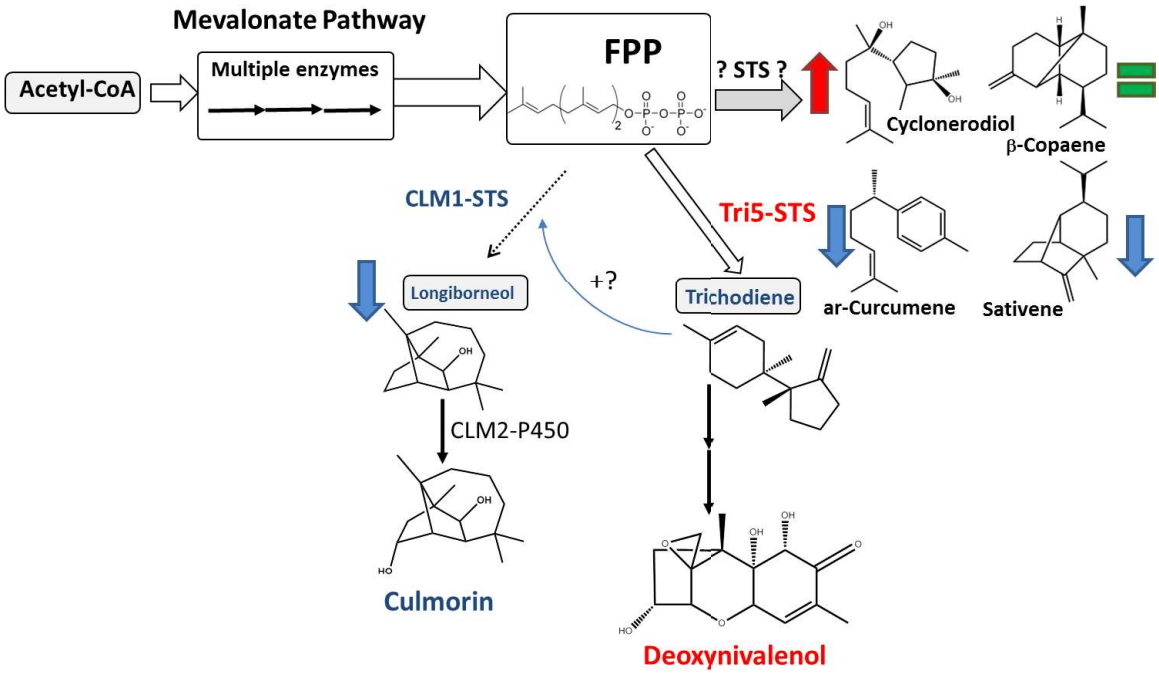


Figure 5-9: Working model of sesquiterpenoid biosynthetic regulation in *F. graminearum* by Tri5
Complete deletion of Tri5 results in increased production of cyclonerodiol, equal production of β-copaene, and complete removal of all other sesquiterpenoid production compared to induced wild-type. This suggests that Tri5 protein-protein interactions, or a product of Tri5, most likely trichodiene, are required for production of NTS in *F. graminearum*.

Complementation of the inactive *tri5* mutant into *F. graminearum* and subsequent analysis of its effect on the sesquiterpenoid product profile would directly test both hypotheses, if mutually exclusive. If culmorin and other NTS production is restored in inactive *tri5* strains, then co-regulation must be occurring *via* protein:protein interactions with unknown proteins, while lack of restoration indicates trichodiene is required for co-production of NTS. Previous mutagenesis of the *F. sporotrichioides* Tri5 (fsTri5) demonstrated complete inactivation without affecting the crystal structure by making N225D and S229T mutations in the conserved, Mg^{2+} binding NSE/DTE motif of fsTri5. [129] Before complementation in *F. graminearum*, the Tri5 N225D S229T mutant must first be confirmed to be inactive by GC/MS, and second, shown to retain secondary structure by Circular Dichroism (CD). Then, a transformation vector can be made and integrated into the *F. graminearum* genome to enable discernment between the effect of the Tri5 protein and its product trichodiene on NTS biosynthesis.

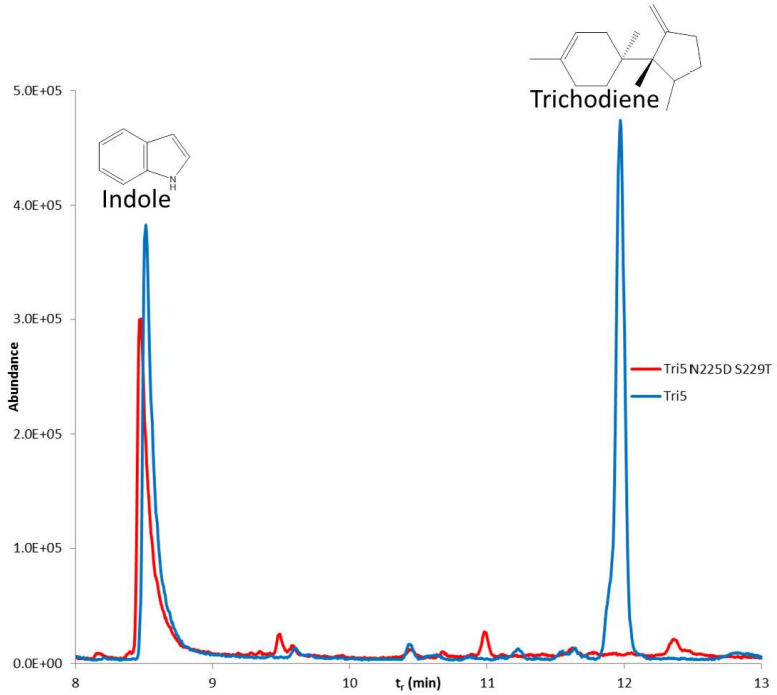


Figure 5-10: GC chromatograph showing Tri5 N225D S229T ceases production of trichodiene and other sesquiterpenes.
tri5 (blue) and *tri5* N225D S229T (red)were expressed in *E. coli* BL21 (DE3). Indole is a breakdown product of tryptophan biosynthesis produced by *E. coli* often used as an internal standard. All minor peaks in cultures expressing Tri5 N225D S229T are GC column siloxanes, and no trichodiene or other sesquiterpenes were detected.

Initial inactivation of Tri5 was performed by Q5 Site-directed mutagenesis (NEB). The Tri5 N225D S229T mutant was then expressed in *E. coli* and analyzed as before by SPME and GC/MS, and confirmed to produce no sesquiterpenes **(Figure 5-10)**. Subsequent purification and secondary structure comparison by CD (**Figure 5-11**) identified no significant difference between WT Tri5 and the Tri5 N225D S229T mutant, with both retaining the expected minima at 208 and 222 nm expected for correctly formed, primarily alpha-helical STSs.[356] Tri5 N225 S229T incubated overnight with 2 uM FPP *in vitro* also did not produce any sesquiterpenes, while similarly analyzed *in vitro* Tri5 product profile (data not shown) matched that observed previously in *in vivo* expression experiments (**Figure 5-4**).

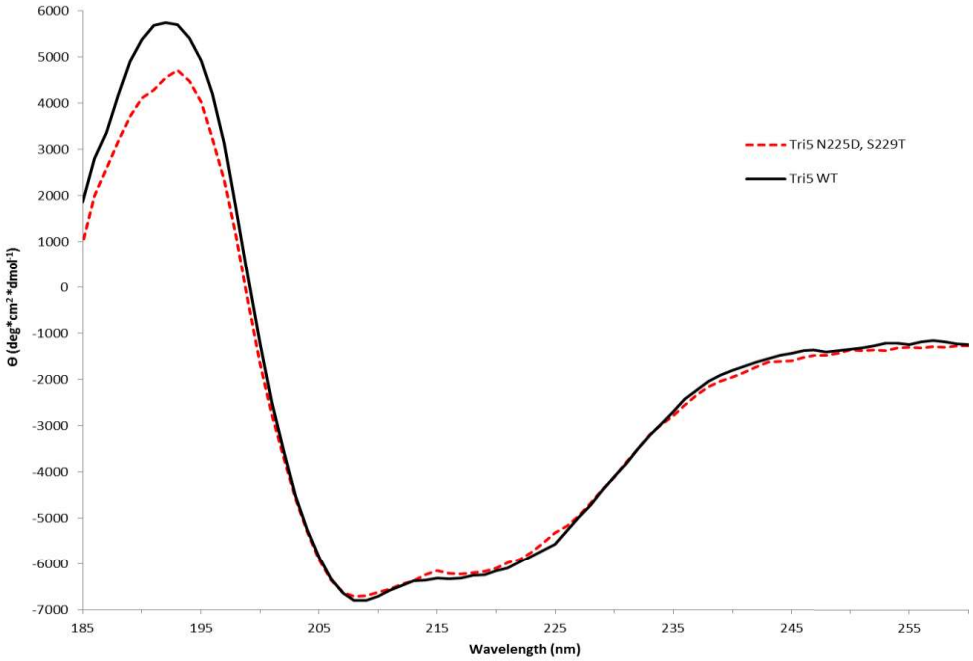


Figure 5-11: Circular dichroism reveals that Tri5 N225D, S229T active site mutations do not affect secondary structure of Tri5
Tri5 (black) and the Tri5 N225D, S229T mutant (red dash) share double minima at 208 nm and 222 nm characteristic of α-helical secondary structure conservation.

Having demonstrated complete removal of STS activity by Tri5 N225D S229T, and conservation of its secondary structure by CD, the stage is now set to create the vector for insertion into the *F. graminearum* PH-1 genome. At the time of writing, construction of this integration vector was underway, with integration and testing of *F. graminearum* Tri5 N225D S229T testing planned for the the following months.

5.5. Discussion and Future Work

The preceding chapter demonstrates that not only does *F. graminearum* produce previously unreported non-trichothecene sesquiterpenes (NTS), but also that all sesquiterpene production requires induction on putrescene. This tight regulation suggests that NTS production plays a role in pathogenesis. Furthermore, production of most NTS, primarily culmorin, requires additional expression of *tri5*. Only cyclonerodiol and β-copaene are produced by putrescene induced *Δtri5* cultures. This suggests that sesquiterpenoid regulation involves at least a two step induction pathway requiring both putrescene *and* Tri5 for NTS biosynthesis. However, the mechanism through which Tri5

affects production of NTS is currently unknown. We hypothesize that the observed regulation is mediated by either protein:protein interactions, or through chemical signaling by trichodiene or a trichothecene derivative. Should the mechanism of NTS biosynthetic regulation by Tri5 be mediated by protein:protein interactions, integration of the catalytically inactivated Tri5 N225D S229T will restore NTS biosynthesis following putrescene induction. In contrast, should trichodiene or a derivative be signaling molecules triggering biosynthesis of NTS, the inactive mutant will not restore NTS production until exogenous trichodiene is added under inducing conditions. Integrating Tri5 N225D S229T into *F. graminearum* will test both hypotheses, clarifying the mechanism of NTS regulation by Tri5.

There is no question that trichothecenes are toxic to both plants and animals, as potent inhibitors of protein synthesis.[429] However, the role NTS play in pathogenesis has not been well studied, and those studies that have occurred frequently studied either NTS in isolation, [436] neglecting potential synergistic effects, or demonstrated that *Δtri5* strains lose the ability to both produce trichothecenes and spread infection wheat heads,(e.g.[437]) a gene deletion which we now know affects both trichothecene *and* NTS biosynthesis. Nevertheless, the great unanswered questions for future work to address are these: what is the role of the complex, tightly-regulated mixture of trichothecenes and NTS during infection, and what is the mechanism controlling their co-production? Addressing these questions would clarify the epigenetic network controlling NTS biosynthesis, and the physiological effect of the complex cocktail of sesquiterpenes produced by *F. graminearum*. There is great precedent for mixtures of terpenoids eliciting synergistic signaling and toxic effects on their targets (reviewed in: [101]). A portion of the work in the preceding section consists of preliminary work to feed trichothecenes to *F. graminearum* strains to test the hypothesis that they are signaling compounds with inconclusive results for trichodiene feeding. Future feeding experiments should occur at a larger scale, use purified trichodiene as a signaling molecule, and be performed in the inactive Tri5 N225D S229T background strain to control for the possibility that protein:protein interactions are necessary but not sufficient to trigger NTS biosynthesis. Whatever the conclusion derived from testing the inactive Tri5 strain for

NTS biosynthesis, mRNA sequencing comparing its gene expression to the control strain is a logical first step in unraveling the regulatory network directing NTS biosynthesis. Furthermore, great care must be taken in analysis of total terpene production, paired with construction of strains where each pathway can be silenced or mutated without affecting the flux to other related pathways. For example, such unintended flux variations are the likely culprit for opposing conclusions regarding the effect of culmorin pathway enzymes on overall virulence in *F. graminearum*.[153, 438] Finally, identification of cyclonerodiol and β-copaene producing STS would be useful controls in understanding the mechanism through which CLM1 and other STS are dependent on Tri5 expression, given that their production requires putrescene but is independent of Tri5. The sensitivity of –omic and analytical chemistry tools such as GC/MS reveals sesquiterpenoid biosynthesis in *F. graminearum* to be far more complex and interesting than was previously imagined. Such an integrated approach combining epigenetic analysis with GC/MS analysis of the terpenome would aid in understanding the control mechanisms for cocktail of biochemicals secreted by *F. graminearum* during infection, and may provide new targets for FHB treatment and resistence as well.

Chapter 6 Concluding Remarks

The ever-expanding collective genome database represents the greatest single resource in history for the study of biochemical processes. In the field of natural product biosynthesis, this invaluable resource can be used to guide identification of genes encoding enzymes that produce unique, otherwise difficult to acquire compounds with important medical properties that can dramatically improve human health and well-being. However, the rate of genome expansion has dramatically outpaced our ability to directly test the connection between gene sequence and protein function, a situation made even more difficult when attempting to identify not just single enzymes, but entire pathways using a genome mining approach. Now more than ever, a reliable means to identify enzymes with the desired function out of the thousands of similar genes is a critical requirement for the study of natural product pathways. The work presented in this thesis demonstrates the development and application of one such method linking gene sequence to protein function, and subsequent identification of biosynthetic pathways using a genome-mining approach.

The aims of this research were 1) to employ bioinformatics to identify novel STSs (STS) in Basidiomycota genomes, streamlining cloning and characterization efforts, and 2) to use these newly-discovered STS as anchors to identify sesquiterpenoid biosynthetic gene clusters, which can be reconstituted in a heterologous host to produce sesquiterpenoids. This process would allow the characterization of (partial) pathways converting linear farnesyl-pyrophosphate (FPP) into bioactive sesquiterpenoids that are uniquely produced by Basidiomycota, and has the potential to enable biosynthesis of these complex sesquiterpenoids for medical use. While pursuing these paired goals, I have also made important insights into STS evolution, and the potential role of previously unknown sesquiterpenes in the pathogenesis of *F. graminearum*.

At the start of this work, I set out with the paired goals of identifying and characterizing protoilludene synthases and their biosynthetic gene clusters in the genome of *S. hirsutum*, and testing the reliability with which our newly developed predictive framework[89] can be applied to identify STSs based on their initial reaction mechanism.

Working with Dr. Swati Choudhary and Dr. Maureen Quin, we were able to apply this predictive framework to identify three protoilludene synthases, and two additional enzymes from other cyclization clades, demonstrating its utility and accuracy.[358] This predictive framework was then used to identify the gene encoding hirsutene synthase in the genome of *S. hirsutum*, a 1,11-cyclizing STS highly divergent from the previously identified protoilludene synthases. Comparison of the gene structures of the known STSs provides further support to the divergent evolution of STSs within each cyclization clade implicitly required for our predictive model. This differs significantly from the STSs of plants and bacteria, where no such predictions can be made reliably, suggesting a relatively high frequency of convergent evolution events between different STS cyclization mechanisms in these kingdoms.[87, 360] To date, our lab has cloned twenty-two Basidiomycota STS, with every enzyme following the predicted initial cyclization mechanism. The number of fungal STS vastly outnumbers those of bacteria and rivals that of plants. As more fungal STS are identified, this predictive framework will become more reliable and improve product accuracy, perhaps informing the evolutionary path of all fungi.

In addition to the identification of fungal STSs, my second aim was to refactor the biosynthetic pathways converting sesquiterpenes into their oxidized cytotoxic derivatives, e.g., modification of protoilludene into the illudins. Gene identification again relied on analysis of the host genome, but in this case took advantage of the fungal tendency to co-locate biosynthetic pathways in gene clusters. As a result, sixteen P450 monoxygenases, oxidoreductases, and other putative sesquiterpene modifying or accessory enzymes were cloned, and two were demonstrated to have protoilludene modifying activity. Condition and strain optimization to increase the production of these protoilludene derivatives for structural characterization by NMR is currently underway.

While my study of sesquiterpene biosynthesis began with the twin goals of identifying STSs and refactoring their biosynthetic pathways, it has led to physiological insights in the economically important plant pathogen *F. graminearum*, and to the surprising discovery of genomic characteristics affecting isoprenoid biosynthesis throughout the fungal kingdom. My analysis of the sesquiterpenes produced by *F.*

graminearum has determined that it produces several previously undescribed sesquiterpenes specifically during pathogenesis, and that production largely requires expression of the STS *tri5* through an unknown signaling mechanism. Finally, the cloning of hirsutene synthase as a bifunctional protein with HMG-CoA synthase led to identification of HMG-CoA synthase duplications throughout the fungal kingdom, with many such duplications being localized to isoprenoid biosynthetic pathways. Similar duplications were observed in all early MVA pathway enzymes across fungi, suggesting a conserved mechanism to alleviate transcriptional regulation of the MVA pathway, and thus to increase biosynthesis of specific isoprenoids, co-expressed in the same biosynthetic gene cluster.

In combination, this thesis describes a genome-driven foray into all aspects of fungal isoprenoid biosynthesis. The primary focus of this work has been the successful identification, isolation, and characterization of STSs unique to fungi,[358, 364] and refactoring the biosynthetic pathway required to produce their bioactive derivatives. This work also describes a review of engineering approaches to improve flux through the MVA pathway for isoprenoid production,[439] as well as discovery of conserved natural genomic characteristics likely to achieve the same goals. Finally, study of the *F. graminearum* terpenome has demonstrated the utility of genome-driven analysis of sesquiterpenes when applied to physiological study of economically important fungi. While many questions remain to be answered, this work has demonstrated the value of our predictive framework, and identified many enzymes, pathways, and sesquiterpenes with important biotechnological and physiological consequences that warrant further study.

References

1. Demain, A.L., *Importance of microbial natural products and the need to revitalize their discovery.* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 185-201.
2. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery.* Nat Rev Drug Discov, 2005. **4**(3): p. 206-20.
3. Kuriata-Adamusiak, R., D. Strub, and S. Lochynski, *Application of microorganisms towards synthesis of chiral terpenoid derivatives.* Applied microbiology and biotechnology, 2012. **95**(6): p. 1427-36.
4. Horwitz, S.B., *How to make taxol from scratch.* Nature, 1994. **367**(6464): p. 593-4.
5. Calveras, J., Casas, J., Parella, T., Joglar, J., Clapés, P., *Chemoenzymatic Synthesis and Inhibitory Activities of Hyacinthacines A1 and A2 Stereoisomers.* Advanced Synthesis & Catalysis, 2007. **349**(10): p. 1661-1666.
6. Cragg, G.M., *Paclitaxel (Taxol): a success story with valuable lessons for natural product drug discovery and development.* Medicinal research reviews, 1998. **18**(5): p. 315-31.
7. Gomez, L., et al., *Chemoenzymatic synthesis, structural study and biological activity of novel indolizidine and quinolizidine iminocyclitols.* Org Biomol Chem, 2012. **10**(31): p. 6309-21.
8. Paddon, C.J., et al., *High-level semi-synthetic production of the potent antimalarial artemisinin.* Nature, 2013. **496**(7446): p. 528-32.
9. Wakimoto, T., et al., *Cytotoxic tetramic acid derivative produced by a plant type-III polyketide synthase.* J Am Chem Soc, 2011. **133**(13): p. 4746-9.
10. Westfall, P.J., et al., *Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(3): p. E111-8.
11. Lopez-Gallego, F. and C. Schmidt-Dannert, *Multi-enzymatic synthesis.* Curr Opin Chem Biol, 2010. **14**(2): p. 174-83.
12. Schmidt-Dannert, C., D. Umeno, and F.H. Arnold, *Molecular breeding of carotenoid biosynthetic pathways.* Nature biotechnology, 2000. **18**(7): p. 750-3.
13. Broun, P. and C. Somerville, *Progress in plant metabolic engineering.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 8925-7.
14. Kramer, R. and W.-R. Abraham, *Volatile sesquiterpenes from fungi: what are they good for?* Phytochemistry Reviews, 2011. **11**(1): p. 15-37.
15. Sacchettini, J.C. and C.D. Poulter, *Creating isoprenoid diversity.* Science (New York, N Y), 1997. **277**(5333): p. 1788-9.
16. Lemieux, R.M. and A.I. Meyers, *Asymmetric Synthesis of (-)-Trichodiene. Generation of Vicinal Stereogenic Quaternary Centers via the Thio-Claisen Rearrangement.* J. Am. Chem. Soc. , 1998. **120**: p. 5453-5457.
17. Aungst, R.A., Jr., C. Chan, and R.L. Funk, *Total synthesis of the sesquiterpene (+/-)-illudin C via an intramolecular nitrile oxide cycloaddition.* Organic letters, 2001. **3**(16): p. 2611-3.

18. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 175-84.
19. Cane, D.E., Ikeda, H., *Exploration and mining of the bacterial terpenome.* Accounts of chemical research, 2012. **45**(3): p. 463-72.
20. Harada, H. and N. Misawa, *Novel approaches and achievements in biosynthesis of functional isoprenoids in Escherichia coli.* Appl Microbiol Biotechnol, 2009. **84**(6): p. 1021-31.
21. Kirby, J. and J.D. Keasling, *Biosynthesis of plant isoprenoids: perspectives for microbial engineering.* Annu Rev Plant Biol, 2009. **60**: p. 335-55.
22. 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 Biological and Crystallization Communications, 2013. **69**(Pt 5): p. 574-7.
23. Ajikumar, P.K., et al., *Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-4.
24. Kizer, L., et al., *Application of functional genomics to pathway optimization for increased isoprenoid production.* Appl Environ Microbiol, 2008. **74**(10): p. 3229-41.
25. Shimada, H., et al., *Increased Carotenoid Production by the Food Yeast Candida utilis through Metabolic Engineering of the Isoprenoid Pathway.* Appl Environ Microbiol, 1998. **64**(7): p. 2676-2680.
26. Haas, M.J., et al., *A process model to estimate biodiesel production costs.* Bioresour Technol, 2006. **97**(4): p. 671-8.
27. Chang, M.C. and J.D. Keasling, *Production of isoprenoid pharmaceuticals by engineered microbes.* Nat Chem Biol, 2006. **2**(12): p. 674-81.
28. Miziorko, H.M., *Enzymes of the mevalonate pathway of isoprenoid biosynthesis.* Arch Biochem Biophys, 2011. **505**(2): p. 131-43.
29. Lombard, J. and D. Moreira, *Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life.* Mol Biol Evol, 2011. **28**(1): p. 87-99.
30. Ganjewala, D., S. Kumar, and R. Luthra, *An Account of Cloned Genes of Methyl-erythritol-4-phosphate Pathway of Isoprenoid Biosynthesis in Plants.* Curr. Issues Mol. Biol. , 2009. **Suppl. 1**: p. i35-45.
31. Phillips, M.A., et al., *The plastidial MEP pathway: unified nomenclature and resources.* Trends Plant Sci, 2008. **13**(12): p. 619-23.
32. Shumskaya, M. and E.T. Wurtzel, *The carotenoid biosynthetic pathway: thinking in all dimensions.* Plant Sci, 2013. **208**: p. 58-63.
33. Vranova, E., D. Coman, and W. Gruissem, *Structure and dynamics of the isoprenoid pathway network.* Molecular plant, 2012. **5**(2): p. 318-33.
34. Enjuto, M., et al., *Arabidopsis thaliana contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(3): p. 927-31.

35. Bick, J.A., Lange, B.M., *Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane.* Archives of biochemistry and biophysics, 2003. **415**(2): p. 146-54.
36. Soliman, S.S.M., R. Tsao, and M.N. Raizada, *Chemical Inhibitors Suggest Endophytic Fungal Paclitaxel Is Derived from Both Mevalonate and Non-mevalonate-like Pathways.* Journal of Natural Products, 2011. **74**(12): p. 2497-2504.
37. Learned, R.M., *Light suppresses 3-Hydroxy-3-methylglutaryl coenzyme A reductase gene expression in Arabidopsis thaliana.* Plant physiology, 1996. **110**(2): p. 645-55.
38. Learned, R.M. and E.L. Connolly, *Light modulates the spatial patterns of 3-hydroxy-3-methylglutaryl coenzyme A reductase gene expression in Arabidopsis thaliana.* The Plant journal : for cell and molecular biology, 1997. **11**(3): p. 499-511.
39. Lee, P.C., et al., *Alteration of product specificity of Aeropyrum pernix farnesylgeranyl diphosphate synthase (Fgs) by directed evolution.* Protein Eng Des Sel, 2004. **17**(11): p. 771-7.
40. Lee, P.C., et al., *Biosynthesis of Structurally Novel Carotenoids in Escherichia coli.* Chemistry & Biology, 2003. **10**(5): p. 453-462.
41. Lee, P.C., et al., *Directed evolution of Escherichia coli farnesyl diphosphate synthase (IspA) reveals novel structural determinants of chain length specificity.* Metab Eng, 2005. **7**(1): p. 18-26.
42. Wang, K. and S.-i. Ohnuma, *Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution.* Trends in Biochemical Sciences, 1999. **24**(11): p. 445-551.
43. Wang, K.C. and S. Ohnuma, *Isoprenyl diphosphate synthases.* Biochimica et biophysica acta, 2000. **1529**(1-3): p. 33-48.
44. Kellogg, B.A. and C.D. Poulter, *Chain elongation in the isoprenoid biosynthetic pathway.* Current opinion in chemical biology, 1997. **1**(4): p. 570-8.
45. Enfissi, E.M., et al., *Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato.* Plant Biotechnol J, 2005. **3**(1): p. 17-27.
46. Chappell, J., Wolf, F., Proulx, J., Cuellar, R., Sanders, C., *Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for lsoprenoid Biosynthesis in PlanG?* Plant Physiology, 1995. **109**: p. 1337-1343.
47. Estevez, J.M., et al., *1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants.* J Biol Chem, 2001. **276**(25): p. 22901-9.
48. Goldstein, J.L. and M.S. Brown, *Regulation of the Mevalonate Pathway.* Nature, 1990. **343**(6257): p. 425-430.
49. Lee, P.C., et al., *Biosynthesis of ubiquinone compounds with conjugated prenyl side chains.* Appl Environ Microbiol, 2008. **74**(22): p. 6908-17.

50. Donald, K.A.G., Hampton, R.Y., Fritz, I.B., *Effects of Overproduction of the Catalytic Domain of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase on Squalene Synthesis in Saccharomyces cerevisiae.* Applied and environmental microbiology, 1997. **63**(9): p. 3341-3344.
51. Wilding, E.I., Kim, D-Y., Bryant, A.P., Gwynn, M.N., Lunsford, R.D., McDevitt, D., Myers, JR., D.E., Rosenberg, M., Sylvester, D., Stauffacher, C.V., Rodwell, V.W., *Essentiality, Expression, and Characterization of the Class II 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase of Staphylococcus aureus.* Journal of Bacteriology, 2000. **182**(18): p. 5147-5152.
52. Engels, B., P. Dahm, and S. Jennewein, *Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production.* Metab Eng, 2008. **10**(3-4): p. 201-6.
53. Ma, S.M., et al., *Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases.* Metab Eng, 2011. **13**(5): p. 588-97.
54. Ro, D.-K., et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**(7086): p. 940-3.
55. Tsuruta, H., et al., *High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli.* PloS one, 2009. **4**(2): p. e4489.
56. Dijken, J.P.v., et al., *An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains.* Enzyme and Microbial Technology, 2000. **26**(9-10): p. 706-714.
57. Sever, N., et al., *Accelerated degradation of HMG CoA reductase mediated by binding of insig-1 to its sterol-sensing domain.* Molecular cell, 2003. **11**(1): p. 25-33.
58. van Hoek, P., et al., *Fermentative capacity in high-cell-density fed-batch cultures of baker's yeast.* Biotechnology and bioengineering, 2000. **68**(5): p. 517-23.
59. Desmond, E. and S. Gribaldo, *Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature.* Genome Biol Evol, 2009. **1**: p. 364-81.
60. Chappell, J., *The genetics and molecular genetics of terpene and sterol origami.* Current opinion in plant biology, 2002. **5**(2): p. 151-7.
61. Misawa, N., *Pathway engineering for functional isoprenoids.* Curr Opin Biotechnol, 2011. **22**(5): p. 627-33.
62. Armstrong, G.A., Hearst, J.E., *Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis.* The Faseb Journal, 1996. **10**(2): p. 228-237.
63. Sandmann, G., *Molecular evolution of carotenoid biosynthesis from bacteria to plants.* Physiologia plantarum, 2002. **116**: p. 431-440.
64. Paine, J.A., et al., *Improving the nutritional value of Golden Rice through increased pro-vitamin A content.* Nat Biotechnol, 2005. **23**(4): p. 482-7.
65. Bhullar, N.K. and W. Gruissem, *Nutritional enhancement of rice for human health: the contribution of biotechnology.* Biotechnol Adv, 2013. **31**(1): p. 50-7.
66. Bjerkeng, B., B. Hatlen, and M. Jobling, *Astaxanthin and its metabolites idoxanthin and crustaxanthin in flesh, skin, and gonads of sexually immature and*

maturing Arctic charr (Salvelinus alpinus (L.)). Comparative biochemistry and physiology Part B, Biochemistry & molecular biology, 2000. **125**(3): p. 395-404.
67. Visser, H., A. Vanooyen, and J. Verdoes, *Metabolic engineering of the astaxanthin-biosynthetic pathway of.* FEMS Yeast Research, 2003. **4**(3): p. 221-231.
68. Schmidt, I., et al., *Biotechnological production of astaxanthin with Phaffia rhodozyma/Xanthophyllomyces dendrorhous.* Appl Microbiol Biotechnol, 2011. **89**(3): p. 555-71.
69. Song, G.H., et al., *Heterologous carotenoid-biosynthetic enzymes: functional complementation and effects on carotenoid profiles in Escherichia coli.* Applied and environmental microbiology, 2013. **79**(2): p. 610-8.
70. Umeno, D., A.V. Tobias, and F.H. Arnold, *Diversifying carotenoid biosynthetic pathways by directed evolution.* Microbiology and molecular biology reviews : MMBR, 2005. **69**(1): p. 51-78.
71. Tobias, A.V. and F.H. Arnold, *Biosynthesis of novel carotenoid families based on unnatural carbon backbones: a model for diversification of natural product pathways.* Biochimica et biophysica acta, 2006. **1761**(2): p. 235-46.
72. Jackson, H., C.L. Braun, and H. Ernst, *The chemistry of novel xanthophyll carotenoids.* Am J Cardiol, 2008. **101**(10A): p. 50D-57D.
73. Higuera-Ciapara, I., L. Felix-Valenzuela, and F.M. Goycoolea, *Astaxanthin: a review of its chemistry and applications.* Crit Rev Food Sci Nutr, 2006. **46**(2): p. 185-96.
74. Breitenbach, J., et al., *Engineering of geranylgeranyl pyrophosphate synthase levels and physiological conditions for enhanced carotenoid and astaxanthin synthesis in Xanthophyllomyces dendrorhous.* Biotechnol Lett, 2011. **33**(4): p. 755-61.
75. Li, J., et al., *An economic assessment of astaxanthin production by large scale cultivation of Haematococcus pluvialis.* Biotechnol Adv, 2011. **29**(6): p. 568-74.
76. D'Ambrosio, C., Giorio, G., Marino, I., Merendino, A., Petrozza, A., Salfi, L., Stigliani, A.L., Cellini, F., *Virtually complete conversion of lycopene into β-carotene in fruits of tomato plants transformed with the tomato lycopene β-cyclase (tlcy-b) cDNA.* Plant Science, 2004. **166**(1): p. 207-214.
77. Giuliano, G., et al., *Metabolic engineering of carotenoid biosynthesis in plants.* Trends Biotechnol, 2008. **26**(3): p. 139-45.
78. Christianson, D.W., *Structural biology and chemistry of the terpenoid cyclases.* Chem Rev, 2006. **106**(8): p. 3412-42.
79. Allemann, R.K., *Chemical wizardry? The generation of diversity in terpenoid biosynthesis.* Pure and Applied Chemistry, 2008. **80**(8): p. 1791-1798.
80. Miller, D.J. and R.K. Allemann, *Sesquiterpene synthases: passive catalysts or active players?* Nat Prod Rep, 2012. **29**(1): p. 60-71.
81. Sun, H.F., et al., *Asperolides A-C, tetranorlabdane diterpenoids from the marine alga-derived endophytic fungus Aspergillus wentii EN-48.* Journal of Natural Products, 2012. **75**(2): p. 148-52.

82. Davis, E.M., Croteau, R., *Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes.* Topics in Current Chemistry, 2000. **209**: p. 53-95.
83. Agger, S., F. Lopez-Gallego, and C. Schmidt-Dannert, *Diversity of sesquiterpene synthases in the basidiomycete Coprinus cinereus.* Mol Microbiol, 2009. **72**(5): p. 1181-95.
84. Castillo, D.A., M.D. Kolesnikova, and S.P. Matsuda, *An effective strategy for exploring unknown metabolic pathways by genome mining.* J Am Chem Soc, 2013. **135**(15): p. 5885-94.
85. Wawrzyn, G.T., S.E. Bloch, and C. Schmidt-Dannert, *Discovery and characterization of terpenoid biosynthetic pathways of fungi.* Methods Enzymol, 2012. **515**: p. 83-105.
86. Agger, S.A., et al., *Identification of sesquiterpene synthases from Nostoc punctiforme PCC 73102 and Nostoc sp. strain PCC 7120.* J Bacteriol, 2008. **190**(18): p. 6084-96.
87. Chen, F., et al., *The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom.* Plant J, 2011. **66**(1): p. 212-29.
88. Lin, H.C., et al., *The fumagillin biosynthetic gene cluster in Aspergillus fumigatus encodes a cryptic terpene cyclase involved in the formation of beta-trans-bergamotene.* J Am Chem Soc, 2013. **135**(12): p. 4616-9.
89. Wawrzyn, G.T., et al., *Draft genome of Omphalotus olearius provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota.* Chem Biol, 2012. **19**(6): p. 772-83.
90. Bartoszewska, M., Opalinski, L., Veenhuis, M., van der Klei, I.J., *The significance of peroxisomes in secondary metabolite biosynthesis in filamentous fungi.* Biotechnology letters, 2011. **33**(10): p. 1921-31.
91. Krisans, S.K., et al., *Farnesyl-diphosphate synthase is localized in peroxisomes.* The Journal of biological chemistry, 1994. **269**(19): p. 14165-9.
92. Menke, J., et al., *Cellular development associated with induced mycotoxin synthesis in the filamentous fungus Fusarium graminearum.* PLoS One, 2013. **8**(5): p. e63077.
93. Proctor, R.H., et al., *Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus Fusarium.* Molecular microbiology, 2009. **74**(5): p. 1128-42.
94. Yadav, V.G., et al., *The future of metabolic engineering and synthetic biology: towards a systematic practice.* Metabolic engineering, 2012. **14**(3): p. 233-41.
95. Galagan, J.E., et al., *Genomics of the fungal kingdom: insights into eukaryotic biology.* Genome Research, 2005. **15**(12): p. 1620-31.
96. Tedersoo, L., T.W. May, and M.E. Smith, *Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages.* Mycorrhiza, 2010. **20**(4): p. 217-63.
97. Blackwell, M., *The fungi: 1, 2, 3 ... 5.1 million species?* American Journal of Botany, 2011. **98**(3): p. 426-38.

98. Bass, D. and T.A. Richards, *Three reasons to re-evaluate fungal diversity 'on Earth and in the ocean'.* Fungal Biology Reviews, 2011. **25**(4): p. 159-164.
99. Alves, M.J., et al., *A review on antimicrobial activity of mushroom (Basidiomycetes) extracts and isolated compounds.* Planta Medica, 2012. **78**(16): p. 1707-18.
100. Fraga, B.M., *Natural sesquiterpenoids.* Natural Product Reports, 2012. **29**(11): p. 1334-66.
101. Gershenzon, J. and N. Dudareva, *The function of terpene natural products in the natural world.* Nature Chemical Biology, 2007. **3**(7): p. 408-14.
102. Abraham, W.R., *Bioactive sesquiterpenes produced by fungi - are they useful for humans as well?* Current Medicinal Chemistry, 2001. **8**: p. 583-606.
103. Elisashvili, V., *Submerged cultivation of medicinal mushrooms: bioprocesses and products (review).* International Journal of Medicinal Mushrooms, 2012. **14**(3): p. 211-39.
104. Evidente, A., et al., *Fungal metabolites with anticancer activity.* Natural Product Reports, 2014. **31**: p. 617-627.
105. Misiek, M. and D. Hoffmeister, *Fungal genetics, genomics, and secondary metabolites in pharmaceutical sciences.* Planta Medica, 2007. **73**: p. 103-115.
106. Wasser, S.P., *Current findings, future trends, and unsolved problems in studies of medicinal mushrooms.* Applied Microbiology and Biotechnology, 2011. **89**(5): p. 1323-32.
107. Zaidman, B.Z., et al., *Medicinal mushroom modulators of molecular targets as cancer therapeutics.* Applied Microbiology and Biotechnology, 2005. **67**(4): p. 453-68.
108. Schobert, R., et al., *Anticancer active illudins: recent developments of a potent alkylating compound class.* Current Medicinal Chemistry, 2011. **18**(6): p. 790-807.
109. Brase, S., et al., *Chemistry and Biology of Mycotoxins and Related Fungal Metabolites.* Chemical Reviews, 2009. **109**: p. 3903-3990.
110. Walton, J.D., *Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis.* Fungal Genetics and Biology, 2000. **30**(3): p. 167-71.
111. Wiemann, P. and N.P. Keller, *Strategies for mining fungal natural products.* Journal of Industrial Microbiology and Biotechnology, 2014. **41**(2): p. 301-13.
112. Shendure, J. and E. Lieberman Aiden, *The expanding scope of DNA sequencing.* Nature Biotechnology, 2012. **30**(11): p. 1084-1094.
113. Floudas, D., et al., *The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes.* Science, 2012. **336**(6089): p. 1715-9.
114. Chen, S., et al., *Genome sequence of the model medicinal mushroom Ganoderma lucidum.* Nature Communications, 2012. **3**: p. 913.
115. Bushley, K.E., et al., *The genome of tolypocladium inflatum: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster.* PLoS Genetics, 2013. **9**(6): p. e1003496.

116. Christianson, D.W., *Unearthing the roots of the terpenome.* Current Opinion in Chemical Biology, 2008. **12**(2): p. 141-50.
117. Wendt, K.U. and G.E. Schulz, *Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes.* Structure, 1998. **6**(2): p. 127-33.
118. Oldfield, E. and F.Y. Lin, *Terpene biosynthesis: modularity rules.* Angew Chem Int Ed Engl, 2012. **51**(5): p. 1124-37.
119. Dickschat, J.S., *Isoprenoids in three-dimensional space: the stereochemistry of terpene biosynthesis.* Nat Prod Rep, 2011. **28**(12): p. 1917-36.
120. Tantillo, D.J., *Biosynthesis via carbocations: Theoretical studies on terpene formation* Natural Product Reports, 2011. **28**(12): p. 1956-1956.
121. 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.
122. Starks, C.M., et al., *Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase.* Science, 1997. **277**(5333): p. 1815-1820.
123. Cane, D.E., Q. Xue, and B.C. Fitzsimmons, *Trichodiene Synthase. Probing the Role of the Highly Conserved Aspartate-Rich Region by Site-Directed Mutagenesis.* Biochemistry, 1996. **35**: p. 12369-12376.
124. Cane, D.E., et al., *Trichodiene Biosynthesis and the Stereochemistry of the Enzymatic Cyclization of Farnesyl Pyrophosphate.* Bioorganic Chemistry, 1985. **13**(3): p. 246-265.
125. Cane, D.E., S. Swanson, and P.P.N. Murthy, *Trichodiene biosynthesis and the enzymatic cyclization of farnesyl pyrophosphate.* Journal of the American Chemical Society, 1981. **103**: p. 2136-2138.
126. Baer, P., et al., *Hedycaryol synthase in complex with nerolidol reveals terpene cyclase mechanism.* Chembiochem, 2014. **15**(2): p. 213-6.
127. Rabe, P., C.A. Citron, and J.S. Dickschat, *Volatile terpenes from actinomycetes: a biosynthetic study correlating chemical analyses to genome data.* Chembiochem, 2013. **14**(17): p. 2345-54.
128. Garms, S., T.G. Kollner, and W. Boland, *A multiproduct terpene synthase from Medicago truncatula generates cadalane sesquiterpenes via two different mechanisms.* J Org Chem, 2010. **75**(16): p. 5590-600.
129. 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-Mg2+B motif.* Archives of Biochemistry and Biophysics, 2008. **469**(2): p. 184-94.
130. Felicetti, B. and D.E. Cane, *Aristolochene synthase: mechanistic analysis of active site residues by site-directed mutagenesis.* Journal of the American Chemical Society, 2004. **126**(23): p. 7212-21.
131. Seemann, M., et al., *Pentalenene synthase. Analysis of active site residues by site-directed mutagenesis.* Journal of the American Chemical Society, 2002. **124**(26): p. 7681-9.

132. Faraldos, J.A., V. Gonzalez, and R.K. Allemann, *The role of aristolochene synthase in diphosphate activation.* Chemical Communications, 2012. **48**(26): p. 3230-2.
133. 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.
134. 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.
135. Gennadios, H.A., et al., *Crystal structure of (+)-delta-cadinene synthase from Gossypium arboreum and evolutionary divergence of metal binding motifs for catalysis.* Biochemistry, 2009. **48**(26): p. 6175-83.
136. 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.
137. 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.
138. 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.
139. Chen, M., et al., *Mechanistic insights from the binding of substrate and carbocation intermediate analogues to aristolochene synthase.* Biochemistry, 2013. **52**(32): p. 5441-53.
140. 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.
141. Rynkiewicz, M.J., D.E. Cane, and D.W. Christianson, *X-ray Crystal Structures of D100E Trichodiene Synthase and Its Pyrophosphate Complex Reveal the Basis for Terpene Product Diversity.* Biochemistry, 2002. **41**: p. 1731-1741.
142. Vedula, L.S., D.E. Cane, and D.W. Christianson, *Role of arginine-304 in the diphosphate-triggered active site closure mechanism of trichodiene synthase.* Biochemistry, 2005. **44**(38): p. 12719-27.
143. Hohn, T.M. and F. Vanmiddlesworth, *Purification and characterization of the sesquiterpene cyclase trichodiene synthetase from Fusarium sporotrichioides.* Archives of Biochemistry and Biophysics, 1986. **251**(2): p. 756-61.
144. Osborne, L.E. and J.M. Stein, *Epidemiology of Fusarium head blight on small-grain cereals.* International Journal of Food Microbiology, 2007. **119**(1-2): p. 103-8.
145. Cane, D.E. and H.J. Ha, *Trichodiene Biosynthesis and the Role of Nerolidyl Pyrophosphate in the Enzymatic Cyclization of Farnesyl Pyrophosphate.* Journal of the American Chemical Society, 1988. **110**(20): p. 6865-6870.
146. Vedula, L.S., et al., *Exploring biosynthetic diversity with trichodiene synthase.* Archives of Biochemistry and Biophysics, 2007. **466**(2): p. 260-6.

147. Dickschat, J.S., et al., *Biosynthesis of sesquiterpenes by the fungus Fusarium verticillioides.* ChemBiochem, 2011. **12**(13): p. 2088-95.
148. Hong, Y.J. and D.J. Tantillo, *Which is more likely in trichodiene biosynthesis: hydride or proton transfer?* Org Lett, 2006. **8**(20): p. 4601-4.
149. Miller, D.J., et al., *Stereochemistry of eudesmane cation formation during catalysis by aristolochene synthase from Penicillium roqueforti.* Organic and Biomolecular Chemistry, 2008. **6**(13): p. 2346-54.
150. Faraldos, J.A., et al., *A 1,6-ring closure mechanism for (+)-delta-cadinene synthase?* J Am Chem Soc, 2012. **134**(13): p. 5900-8.
151. Faraldos, J.A., et al., *Templating effects in aristolochene synthase catalysis: elimination versus cyclisation.* Organic and Biomolecular Chemistry, 2011. **9**(20): p. 6920-3.
152. Motohashi, K., et al., *New sesquiterpenes, JBIR-27 and -28, isolated from a tunicate-derived fungus, Penicillium sp. SS080624SCf1.* J Antibiot (Tokyo), 2009. **62**(5): p. 247-50.
153. McCormick, S.P., N.J. Alexander, and L.J. Harris, *CLM1 of Fusarium graminearum encodes a longiborneol synthase required for culmorin production.* Applied and Environmental Microbiology, 2010. **76**(1): p. 136-41.
154. Brock, N.L., et al., *Genetic dissection of sesquiterpene biosynthesis by Fusarium fujikuroi.* Chembiochem, 2013. **14**(3): p. 311-5.
155. Pinedo, C., et al., *Sesquiterpene synthase from the botrydial biosynthetic gene cluster of the phytopathogen Botrytis cinerea.* ACS Chemical Biology, 2008. **3**(12): p. 791-801.
156. Wang, C.M., et al., *Biosynthesis of the sesquiterpene botrydial in Botrytis cinerea. Mechanism and stereochemistry of the enzymatic formation of presilphiperfolan-8beta-ol.* J Am Chem Soc, 2009. **131**(24): p. 8360-1.
157. 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. 1093-106.
158. 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. 7723-33.
159. Kelner, M.J., et al., *Preclinical evaluation of illudins as anticancer agents.* Cancer Research, 1987. **47**: p. 3186-3189.
160. McMorris, T.C. and M. Anchel, *FUNGAL METABOLITES. THE STRUCTURES OF THE NOVEL SESQUITERPENOIDS ILLUDIN-S AND -M.* Journal of the American Chemical Society, 1965. **87**: p. 1594-600.
161. Nair, M.S. and M. Anchel, *Metabolic products of clitocybe illudens. IX. Structure of illudacetalic acid and its conversion to illudinine.* Tetrahedron Letters, 1972. **13**(27): p. 2753-2754.
162. Engels, B., et al., *Cloning and characterization of an Armillaria gallica cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides.* The Journal of Biological Chemistry, 2011. **286**(9): p. 6871-8.

163. Martin, F., et al., *Sequencing the fungal tree of life.* New Phytologist, 2011. **190**(4): p. 818-821.
164. Cuomo, C.A. and B.W. Birren, *The Fungal Genome Initiative and Lessons Learned from Genome Sequencing.* Methods in Enzymology, 2010. **470**: p. 833-855.
165. Ayer, W.A. and M.H. Saeedighomi, *The Sterepolides - New Isolactaranes from Stereum-Purpureum.* Tetrahedron Letters, 1981. **22**(22): p. 2071-2074.
166. Abell, C. and A.P. Leech, *The Absolute Stereochemistry of the Enzymic Cyclization to Form the Sterpurene Sesquiterpenes.* Tetrahedron Letters, 1988. **29**(34): p. 4337-4340.
167. Ainsworth, A.M., et al., *Production and Properties of the Sesquiterpene, (+)-Torreyol, in Degenerative Mycelial Interactions between Strains of Stereum.* Mycological Research, 1990. **94**: p. 799-809.
168. Yun, B.S., et al., *New tricyclic sesquiterpenes from the fermentation broth of Stereum hirsutum.* Journal of Natural Products, 2002. **65**(5): p. 786-8.
169. Kim, Y.H., et al., *Methoxylaricinolic acid, a new sesquiterpene from the fruiting bodies of Stereum ostrea.* The Journal of Antibiotics, 2006. **59**(7): p. 432-4.
170. Yoo, N.H., et al., *Hirsutenols D, E and F, new sesquiterpenes from the culture broth of Stereum hirsutum.* J Antibiot (Tokyo), 2006. **59**(2): p. 110-3.
171. Li, G.H., et al., *Stereumin A-E, sesquiterpenoids from the fungus Stereum sp. CCTCC AF 207024.* Phytochemistry, 2008. **69**(6): p. 1439-45.
172. Opatz, T., H. Kolshorn, and H. Anke, *Sterelactones: new isolactarane type sesquiterpenoids with antifungal activity from Stereum sp. IBWF 01060.* The Journal of Antibiotics, 2008. **61**(9): p. 563-7.
173. Liermann, J.C., et al., *Hirsutane-Type Sesquiterpenes with Uncommon Modifications from Three Basidiomycetes.* Journal of Organic Chemistry, 2010. **75**(9): p. 2955-2961.
174. Isaka, M., et al., *Sterostreins A-E, new terpenoids from cultures of the basidiomycete Stereum ostrea BCC 22955.* Organic Letters, 2011. **13**(18): p. 4886-9.
175. Isaka, M., et al., *Sterostreins F-O, illudalanes and norilludalanes from cultures of the basidiomycete Stereum ostrea BCC 22955.* Phytochemistry, 2012. **79**: p. 116-20.
176. Zheng, X., et al., *Stereumins K-P, sesquiterpenes from the fungus Stereum sp. CCTCC AF 2012007.* Phytochemistry, 2013. **86**: p. 144-50.
177. Ma, K., et al., *New benzoate derivatives and hirsutane type sesquiterpenoids with antimicrobial activity and cytotoxicity from the solid-state fermented rice by the medicinal mushroom Stereum hirsutum.* Food Chemistry, 2014. **143**: p. 239-45.
178. Zu, L., et al., *Effect of isotopically sensitive branching on product distribution for pentalenene synthase: support for a mechanism predicted by quantum chemistry.* Journal of the American Chemical Society, 2012. **134**(28): p. 11369-71.
179. Fraga, B.M., *Natural sesquiterpenoids.* Natural Product Reports, 2013. **30**(9): p. 1226-64.

180. Erkel, G. and T. Anke, *Products from Basidiomycetes*, in *Biotechnology: Products of Secondary Metabolism*, H.-J. Rehm and G. Reed, Editors. 1997, Wiley. p. 490-533.
181. Jansen, B.J.M. and A. de Groot, *Occurrence, biological activity and synthesis of drimane sesquiterpenoids.* Natural Product Reports, 2004. **21**: p. 449-477.
182. 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.
183. Luo, D.-Q., et al., *Humulane-Type Sesquiterpenoids from the Mushroom Lactarius mitissimus.* Journal of Natural Products, 2006. **69**(9): p. 1354-1357.
184. Simon, B., et al., *Collybial, a New Antibiotic Sesquiterpenoid from Collybia Confluens (Basidiomycetes).* Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1995. **50**(3-4): p. 173-180.
185. Coates, R.M., et al., *Carbocationic rearrangements of silphinane derivatives.* Journal of Organic Chemistry, 1998. **63**(25): p. 9166-9176.
186. Ayer, W.A. and E.R. Cruz, *The Tremulanes, a New Group of Sesquiterpenes from the Aspen Rotting Fungus Phellinus-Tremulae.* Journal of Organic Chemistry, 1993. **58**(26): p. 7529-7534.
187. Cruz, E.R., *The biosynthesis of the new tremulane sesquiterpenes isolated from Phellinus tremulae.* Canadian Journal of Chemistry, 1997. **75**(6): p. 834-839.
188. Abraham, W.R., et al., *New Trans-Fused Africanols from Leptographium-Lundbergii.* Tetrahedron, 1986. **42**(16): p. 4475-4480.
189. Feline, T.C., et al., *Biosynthesis of hirsutic acid C using 13C nuclear magnetic resonance spectroscopy.* Journal of the Chemical Society, Chemical Communications, 1974. **2**(2): p. 63-64.
190. Tanasova, M. and S.J. Sturla, *Chemistry and biology of acylfulvenes: sesquiterpene-derived antitumor agents.* Chem Rev, 2012. **112**(6): p. 3578-610.
191. Covello, P.S., et al., *Functional genomics and the biosynthesis of artemisinin.* Phytochemistry, 2007. **68**(14): p. 1864-71.
192. Tokai, T., et al., *Fusarium Tri4 encodes a key multifunctional cytochrome P450 monooxygenase for four consecutive oxygenation steps in trichothecene biosynthesis.* Biochemical and Biophysical Research Communications, 2007. **353**(2): p. 412-7.
193. Daniewski, W.M. and G. Vidari, *Constituents of Lactarius (mushrooms)*, in *Constituents of Lactarius (mushrooms)*. 1999, Springer-Verlag Wien. p. 69-171.
194. Ayer, W.A. and M.H. Saeedighomi, *1-Sterpurene-3,12,14-Triol and 1-Sterpurene, metabolites of silver-leaf disease fungus Stereum purpureum.* Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1981. **59**(16): p. 2536-2538.
195. Schuffler, A., et al., *Isolactarane and Sterpurane Sesquiterpenoids from the Basidiomycete Phlebia uda.* Journal of Natural Products, 2012. **75**(7): p. 1405-1408.

196. Wang, Y., et al., *Two new sesquiterpenes and six norsesquiterpenes from the solid culture of the edible mushroom Flammulina velutipes.* Tetrahedron, 2012. **68**: p. 3012-3018.
197. Sterner, O., et al., *New Sterpurane and Isolactarane Sesquiterpenes from the Fungus Merulius-Tremellosus.* Tetrahedron, 1990. **46**(7): p. 2389-2400.
198. Lackner, G., et al., *Assembly of melleolide antibiotics involves a polyketide synthase with cross-coupling activity.* Chem Biol, 2013. **20**(9): p. 1101-6.
199. Opatz, T., et al., *The creolophins: A family of linear triquinanes from Creolophus cirrhatus (Basidiomycete).* European Journal of Organic Chemistry, 2007(33): p. 5546-5550.
200. Grove, J.F., *The trichothecenes and their biosynthesis.* Progress in the Chemistry of Organic Natural Products, 2007. **88**: p. 63-130.
201. McCormick, S.P., et al., *Trichothecenes: from simple to complex mycotoxins.* Toxins (Basel), 2011. **3**(7): p. 802-14.
202. Desjardins, A.E., *From yellow rain to green wheat: 25 years of trichothecene biosynthesis research.* Journal of Agricultural and Food Chemistry, 2009. **57**(11): p. 4478-84.
203. Kimura, M., et al., *Molecular and Genetic Studies of Fusarium Trichothecene Biosynthesis: Pathways, Genes, and Evolution.* Bioscience Biotechnology and Biochemistry, 2007. **71**(9): p. 2105-2123.
204. Kazan, K., D.M. Gardiner, and J.M. Manners, *On the trail of a cereal killer: recent advances in Fusarium graminearum pathogenomics and host resistance.* Molecular Plant Pathology, 2012. **13**(4): p. 399-413.
205. Audenaert, K., et al., *Deoxynivalenol: a major player in the multifaceted response of Fusarium to its environment.* Toxins (Basel), 2014. **6**(1): p. 1-19.
206. Gu, Q., et al., *A transcription factor FgSte12 is required for pathogenicity in Fusarium graminearum.* Molecular Plant Pathology, 2014.
207. Hohn, T.M. and P.D. Beremand, *Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene-producing fungus Fusarium sporotrichioides.* Gene, 1989. **79**(1): p. 131-8.
208. Brown, D.W., et al., *Inactivation of a cytochrome P-450 is a determinant of trichothecene diversity in Fusarium species.* Fungal Genetics and Biology, 2002. **36**: p. 224-233.
209. Nasmith, C.G., et al., *Tri6 is a global transcription regulator in the phytopathogen Fusarium graminearum.* PLoS Pathogens, 2011. **7**(9): p. e1002266.
210. McCormick, S.P., N.J. Alexander, and R.H. Proctor, *Fusarium Tri4 encodes a multifunctional oxygenase required for trichothecene biosynthesis.* Canadian Journal of Microbiology, 2006. **52**(7): p. 636-42.
211. Hesketh, A.R., et al., *Revision of the stereochemistry in trichodiol, trichotriol and related compounds, and concerning their role in the biosynthesis of trichothecene mycotoxins.* Phytochemistry, 1993. **32**(1): p. 105-116.

212. McCormick, S.P., et al., *Disruption of TRI101, the gene encoding trichothecene 3-O-acteryltransferase from Fusarium sporotrichioides.* Applied and Environmental Microbiology, 1999. **65**(12): p. 5252-5256.
213. Ohsato, S., et al., *Transgenic rice plants expressing trichothecene 3-O-acetyltransferase show resistance to the Fusarium phytotoxin deoxynivalenol.* Plant Cell Reports, 2007. **26**(4): p. 531-8.
214. Alexander, N.J., T.M. Hohn, and S.P. McCormick, *The TRI11 gene of Fusarium sporotrichioides encodes a cytochrome P450 monoxygenase required for C15 hydroxylation in trichothecene biosynthesis.* Applied and Environmental Microbiology, 1998. **64**(1): p. 221-225.
215. McCormick, S.P., T.M. Hohn, and A.E. Desjardins, *Isolation and characterization of Tri3, a gene encoding 15-O-acetyltransferase from Fusarium sporotrichioides.* Applied and Environmental Microbiology, 1996. **62**(2): p. 353-359.
216. Brown, D.W., et al., *Functional demarcation of the Fusarium core trichothecene gene cluster.* Fungal Genetics and Biology, 2004. **41**(4): p. 454-62.
217. Alexander, N.J., et al., *The genetic basis for 3-ADON and 15-ADON trichothecene chemotypes in Fusarium.* Fungal Genetics and Biology, 2011. **48**(5): p. 485-95.
218. Alexander, N.J., S.P. McCormick, and T.M. Hohn, *TRI12, a trichothecene efflux pump from Fusarium sporotrichioides: gene isolation and expression in yeast.* Molecular Genetics and Genomics, 1999. **261**(6): p. 977-84.
219. Menke, J., Y. Dong, and H.C. Kistler, *Fusarium graminearum Tri12p influences virulence to wheat and trichothecene accumulation.* Molecular Plant-Microbe Interactions, 2012. **25**(11): p. 1408-18.
220. Chanda, A., et al., *A key role for vesicles in fungal secondary metabolism.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19533-8.
221. Chanda, A., L.V. Roze, and J.E. Linz, *A possible role for exocytosis in aflatoxin export in Aspergillus parasiticus.* Eukaryot Cell, 2010. **9**(11): p. 1724-7.
222. Hidalgo, P.I., et al., *Molecular characterization of the PR-toxin gene cluster in Penicillium roqueforti and Penicillium chrysogenum: cross talk of secondary metabolite pathways.* Fungal Genet Biol, 2014. **62**: p. 11-24.
223. Zi, J., S. Mafu, and R.J. Peters, *To Gibberellins and Beyond! Surveying the Evolution of (Di)Terpenoid Metabolism.* Annual Review of Plant Biology, 2014. **65**(259-286).
224. Peters, R.J., *Two rings in them all: the labdane-related diterpenoids.* Nat Prod Rep, 2010. **27**(11): p. 1521-30.
225. Abe, I., *Enzymatic synthesis of cyclic triterpenes.* Natural Product Reports, 2007. **24**(6): p. 1311-31.
226. Kawaide, H., et al., *ent-Kaurene synthase from the fungus Phaeosphaeria.* The Journal of Biological Chemistry, 1997. **272**: p. 21706-21712.
227. Kawaide, H., T. Sassa, and Y. Kamiya, *Functional analysis of the two interacting cyclase domains in ent-kaurene synthase from the fungus Phaeosphaeria.* The Journal of Biological Chemistry, 2000. **275**: p. 2276-2280.

228. Sun, T.P. and Y. Kamiya, *The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis.* Plant Cell, 1994. **6**(10): p. 1509-18.
229. Yamaguchi, S., et al., *Molecular cloning and characterization of a cDNA encoding the gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin (Cucurbita maxima L.).* Plant J., 1996. **10**(2): p. 203-13.
230. Koksal, M., et al., *Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase.* Nat Chem Biol, 2011. **7**(7): p. 431-3.
231. Koksal, M., et al., *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis.* Nature, 2011. **469**(7328): p. 116-20.
232. Zhou, K., et al., *Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from Abies grandis.* J Biol Chem, 2012. **287**(9): p. 6840-50.
233. Tudzynski, B., H. Kawaide, and Y. Kamiya, *Gibberellin biosynthesis in Gibberella fujikuroi: cloning and characterization of the copalyl diphosphate synthase gene.* Curr Genet, 1998. **34**(3): p. 234-40.
234. Bromann, K., et al., *Identification and characterization of a novel diterpene gene cluster in Aspergillus nidulans.* PLoS One, 2012. **7**(4): p. e35450.
235. Toyomasu, T., et al., *Identification of diterpene biosynthetic gene clusters and functional analysis of labdane-related diterpene cyclases in Phomopsis amygdali.* Biosci Biotechnol Biochem, 2008. **72**(4): p. 1038-47.
236. Oikawa, H., et al., *Cloning and functional expression of cDNA encoding Aphidicolan-16 beta-ol synthase: A key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin.* Journal of the American Chemical Society, 2001. **123**(21): p. 5154-5155.
237. Mitterbauer, R. and T. Specht, *Cloning and sequence of Clitopilus passeckerianus diterpene synthase and pleuromutilin biosynthesis gene cluster, and use for producing pleuromutilin.* WO/2011/110610, 2011. **EP2011/053571**: p. WO2011110610A1.
238. Mende, K., V. Homann, and B. Tudzynski, *The geranylgeranyl diphosphate synthase gene of Gibberella fujikuroi: isolation and expression.* Molecular Genetics and Genomics, 1997. **255**(1): p. 96-105.
239. Tudzynski, B. and K. Holter, *Gibberellin biosynthetic pathway in Gibberella fujikuroi: evidence for a gene cluster.* Fungal Genet Biol, 1998. **25**(3): p. 157-70.
240. Bomke, C. and B. Tudzynski, *Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria.* Phytochemistry, 2009. **70**(15-16): p. 1876-93.
241. Toyomasu, T., et al., *Biosynthetic Gene-Based Secondary Metabolite Screening: A New Diterpene, Methyl Phomopsenonate, from the Fungus Phomopsis amygdali.* The Journal of Organic Chemistry, 2009. **74**: p. 1541-1548.
242. Toyomasu, T., et al., *Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi.* Proceedings of the National Academy of Sciences, 2007. **104**(9): p. 3084-8.

243. Minami, A., et al., *Identification and functional analysis of brassicicene C biosynthetic gene cluster in Alternaria brassicicola.* Bioorg Med Chem Lett, 2009. **19**(3): p. 870-4.
244. Wendt, K.U., K. Poralla, and G.E. Schulz, *Structure and Function of a Squalene Cyclase.* Science, 1997. **277**(5333): p. 1811-1815.
245. Tarshis, L.C., et al., *Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution.* Biochemistry, 1994. **33**(36): p. 10871-7.
246. Chiba, R., et al., *Identification of ophiobolin F synthase by a genome mining approach: a sesterterpene synthase from Aspergillus clavatus.* Org Lett, 2013. **15**(3): p. 594-7.
247. Hashimoto, M., et al., *Functional analyses of cytochrome P450 genes responsible for the early steps of brassicicene C biosynthesis.* Bioorg Med Chem Lett, 2009. **19**(19): p. 5640-3.
248. Ono, Y., et al., *Dioxygenases, key enzymes to determine the aglycon structures of fusicoccin and brassicicene, diterpene compounds produced by fungi.* J Am Chem Soc, 2011. **133**(8): p. 2548-55.
249. Asahi, K., et al., *Cotylenin A, a plant-growth regulator, induces the differentiation in murine and human myeloid leukemia cells.* Biochemical and Biophysical Research Communications, 1997. **238**(3): p. 758-763.
250. Sassa, T., et al., *Structural confirmation of cotylenin A, a novel fusicoccane-diterpene glycoside with potent plant growth-regulating activity from Cladosporium fungus sp. 501-7W.* Bioscience Biotechnology and Biochemistry, 1998. **62**(9): p. 1815-1818.
251. Sassa, T., M. Togashi, and T. Kitaguchi, *Cotylenins, Leaf Growth-Substances Produced by a Fungus .4. Structures of Cotylenins-a, Cotylenin-B, Cotylenin-C, Cotylenin-D and Cotylenin-E.* Agricultural and Biological Chemistry, 1975. **39**(9): p. 1735-1744.
252. Toyomasu, T., et al., *Cloning of a Gene Cluster Responsible for the Biosynthesis of Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase □.* Bioscience Biotechnology and Biochemistry, 2004. **68**(1): p. 146-152.
253. Wang, Q.-X., et al., *Coicenals A-D, Four New Diterpenoids with New Chemical Skeletons from the Plant Pathogenic Fungus Bipolaris coicis.* Organic Letters, 2013. **15**(15): p. 3982-3985.
254. Afiyatullov, S., et al., *New Diterpenic Altrosides of the Fungus Acremonium striatisporum Isolated from a Sea Cucumber.* Journal of Natural Products, 2000. **63**: p. 848-850.
255. Feng, Y., et al., *Guanacastane Diterpenoids from the Plant Endophytic Fungus Cercospora sp.* Journal of Natural Products, 2014. **77**(4): p. 873-881.
256. Wijeratne, E.M., et al., *Geopyxins A-E, ent-kaurane diterpenoids from endolichenic fungal strains Geopyxis aff. majalis and Geopyxis sp. AZ0066: structure-activity relationships of geopyxins and their analogues.* Journal of Natural Products, 2012. **75**(3): p. 361-9.

257. Wang, X.N., et al., *Smardaesidins A-G, isopimarane and 20-nor-isopimarane diterpenoids from Smardaea sp., a fungal endophyte of the moss Ceratodon purpureus.* Journal of Natural Products, 2011. **74**(10): p. 2052-61.
258. Kawaide, H., T. Sassa, and Y. Kamiya, *Plant-like biosynthesis of Gibberellin A1 in the fungus Phaeosphaeria sp. L487.* Phytochemistry, 1995. **39**(2): p. 305-310.
259. Bomke, C., et al., *Isolation and characterization of the gibberellin biosynthetic gene cluster in Sphaceloma manihoticola.* Applied and Environmental Microbiology, 2008. **74**(17): p. 5325-39.
260. Hartley, A.J., et al., *Investigating pleuromutilin-producing Clitopilus species and related basidiomycetes.* FEMS Microbiology Letters, 2009. **297**(1): p. 24-30.
261. Shibata, H., A. Irie, and Y. Morita, *New antibacterial diterpenoids from the Sarcodon scabrosus fungus.* Bioscience Biotechnology and Biochemistry, 1998. **62**(12): p. 2450-2452.
262. Valdivia, C., et al., *Diterpenoids from Coprinus heptemerus.* Tetrahedron, 2005. **61**(40): p. 9527-9532.
263. Liu, Y., et al., *Guanacastane-type diterpenoids with cytotoxic activity from Coprinus plicatilis.* Bioorganic and Medicinal Chemistry Letters, 2012. **22**(15): p. 5059-62.
264. Ou, Y.X., et al., *Guanacastane-type diterpenoids from Coprinus radians.* Phytochemistry, 2012. **78**: p. 190-6.
265. Li, Y.-Y. and Y.-M. Shen, *Four Novel Diterpenoids from Crinipellis sp. 113.* Helvetica Chimica Acta, 2010. **93**: p. 2151-2157.
266. Mazur, X., et al., *Two new bioactive diterpenes from Lepista sordida.* Phytochemistry, 1996. **43**(2): p. 405-7.
267. Smanski, M.J., et al., *Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis.* Current Opinion in Chemical Biology, 2012. **16**(1-2): p. 132-41.
268. Shen, J.W., Y. Ruan, and B.J. Ma, *Diterpenoids of macromycetes.* Journal of Basic Microbiology, 2009. **49**(3): p. 242-55.
269. Tudzynski, B., *Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology.* Applied Microbiology and Biotechnology, 2005. **66**(6): p. 597-611.
270. Hedden, P., et al., *Gibberellin Biosynthesis in Plants and Fungi: A Case of Convergent Evolution?* Journal of Plant Growth Regulation, 2001. **20**(4): p. 319-331.
271. Tudzynski, B., et al., *The gibberellin 20-oxidase of Gibberella fujikuroi is a multifunctional monooxygenase.* The Journal of Biological Chemistry, 2002. **277**(24): p. 21246-53.
272. Tudzynski, B., et al., *Characterization of the final two genes of the gibberellin biosynthesis gene cluster of Gibberella fujikuroi: des and P450-3 encode GA4 desaturase and the 13-hydroxylase, respectively.* The Journal of Biological Chemistry, 2003. **278**(31): p. 28635-43.

273. Malonek, S., et al., *The NADPH-cytochrome P450 reductase gene from Gibberella fujikuroi is essential for gibberellin biosynthesis.* The Journal of Biological Chemistry, 2004. **279**(24): p. 25075-84.
274. Kawaide, H., *Biochemical and Molecular Analyses of Gibberellin Biosynthesis in Fungi.* Bioscience Biotechnology and Biochemistry, 2006. **70**(3): p. 583-590.
275. Fujii, R., et al., *Total biosynthesis of diterpene aphidicolin, a specific inhibitor of DNA polymerase alpha: heterologous expression of four biosynthetic genes in Aspergillus oryzae.* Biosci Biotechnol Biochem, 2011. **75**(9): p. 1813-7.
276. Noike, M., et al., *An enzyme catalyzing O-prenylation of the glucose moiety of fusicoccin A, a diterpene glucoside produced by the fungus Phomopsis amygdali.* Chembiochem, 2012. **13**(4): p. 566-73.
277. Noike, M., et al., *Molecular breeding of a fungus producing a precursor diterpene suitable for semi-synthesis by dissection of the biosynthetic machinery.* PLoS One, 2012. **7**(8): p. e42090.
278. Hill, R.A. and J.D. Connolly, *Triterpenoids.* Natural Product Reports, 2013. **30**(7): p. 1028-65.
279. Kristan, K. and T.L. Rizner, *Steroid-transforming enzymes in fungi.* Journal of Steroidal Biochemistry and Molecular Biology, 2012. **129**(1-2): p. 79-91.
280. Garaiova, M., et al., *Squalene epoxidase as a target for manipulation of squalene levels in the yeast Saccharomyces cerevisiae.* FEMS Yeast Research, 2014. **14**(2): p. 310-23.
281. Thoma, R., et al., *Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase.* Nature, 2004. **432**: p. 118-122.
282. Corey, E.J. and S.C. Virgil, *An Experimental Demonstration of the Stereochemistry of Enzymic Cyclization of 2,3-Oxidosqualene to the Protosterol System, Forerunner of Lanosterol and Cholesterol.* Journal of the American Chemical Society, 1991. **113**: p. 4025-4026.
283. Siedenburg, G. and D. Jendrossek, *Squalene-hopene cyclases.* Applied and Environmental Microbiology, 2011. **77**(12): p. 3905-15.
284. Poralla, K., et al., *A specific amino acid repeat in squalene and oxidosqualene cyclases.* Trends in Biochemical Sciences, 1994. **19**(4): p. 157-8.
285. Frickey, T. and E. Kannenberg, *Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer.* Environmental Microbiology, 2009. **11**(5): p. 1224-41.
286. Corey, E.J., S.P.T. Matsuda, and B. Bartel, *Molecular cloning, characterization, and overexpression of ERG7, the Saccharomyces cerevisiae gene encoding lanosterol synthase.* Proceedings of the National Academy of Sciences, 1994. **91**: p. 2211-2215.
287. Corey, E.J., et al., *Studies on the substrate binding segments and catalytic action of lanosterol synthase.* Journal of the American Chemical Society, 1997. **119**: p. 1289-1296.
288. Chang, C.H., et al., *Protein engineering of oxidosqualene-lanosterol cyclase into triterpene monocyclase.* Organic and Biomolecular Chemistry, 2013. **11**(25): p. 4214-9.

289. Cheng, C.R., et al., *Cytotoxic triterpenoids from Ganoderma lucidum.* Phytochemistry, 2010. **71**(13): p. 1579-85.
290. Godio, R.P. and J.F. Martin, *Modified oxidosqualene cyclases in the formation of bioactive secondary metabolites: biosynthesis of the antitumor clavaric acid.* Fungal Genetics and Biology, 2009. **46**(3): p. 232-42.
291. Racolta, S., et al., *The triterpene cyclase protein family: a systematic analysis.* Proteins, 2012. **80**(8): p. 2009-19.
292. Phillips, D.R., et al., *Biosynthetic diversity in plant triterpene cyclization.* Current Opinion in Plant Biology, 2006. **9**(3): p. 305-14.
293. Moses, T., et al., *Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro.* New Phytol, 2013. **200**(1): p. 27-43.
294. Sawai, S. and K. Saito, *Triterpenoid biosynthesis and engineering in plants.* Front Plant Sci, 2011. **2**: p. 25.
295. Xue, Z., et al., *Divergent evolution of oxidosqualene cyclases in plants.* New Phytol, 2012. **193**(4): p. 1022-38.
296. Rios, J.L., et al., *Lanostanoids from fungi: a group of potential anticancer compounds.* J Nat Prod, 2012. **75**(11): p. 2016-44.
297. Zhao, M., et al., *Protostane and fusidane triterpenes: a mini-review.* Molecules, 2013. **18**(4): p. 4054-80.
298. Shu, S., et al., *De Novo Sequencing and Transcriptome Analysis of Wolfiporia cocos to Reveal Genes Related to Biosynthesis of Triterpenoids.* PLoS One, 2013. **8**(8): p. e71350.
299. Liu, D., et al., *The genome of Ganoderma lucidum provides insights into triterpenes biosynthesis and wood degradation [corrected].* PLoS One, 2012. **7**(5): p. e36146.
300. Yu, G.J., et al., *Deep insight into the Ganoderma lucidum by comprehensive analysis of its transcriptome.* PLoS One, 2012. **7**(8): p. e44031.
301. Mitsuguchi, H., et al., *Biosynthesis of Steroidal Antibiotic Fusidanes: Functional Analysis of Oxidosqualene Cyclase and Subsequent Tailoring Enzymes from Aspergillus fumigatus.* Journal of the American Chemical Society, 2009. **131**: p. 6402-6411.
302. Nordberg, H., et al., *The genome portal of the Department of Energy Joint Genome Institute: 2014 updates.* Nucleic Acids Research, 2014. **42**(Database issue): p. D26-31.
303. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of Molecular Biology, 1990. **215**(3): p. 403-10.
304. Fedorova, N., V. Moktali, and M.H. Medema, *Fungal Secondary Metabolism: Methods and Protocols*. Methods in Molecular Biology, ed. N.P. Keller and G. Turner. Vol. 944. 2012: Springer Science + Buisiness media.
305. Weber, T., *In silico tools for the analysis of antibiotic biosynthetic pathways.* International Journal of Medical Microbiology, 2014. **304**(3-4): p. 230-5.

306. Blin, K., et al., *antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers.* Nucleic Acids Research, 2013. **41**(Web Server issue): p. W204-12.
307. Wiemann, P., et al., *Prototype of an intertwined secondary-metabolite supercluster.* Proceedings of the National Academy of Sciences, 2013. **110**(42): p. 17065-70.
308. Pahirulzaman, K.A., K. Williams, and C.M. Lazarus, *A toolkit for heterologous expression of metabolic pathways in Aspergillus oryzae.* Methods Enzymol, 2012. **517**: p. 241-60.
309. Tagami, K., et al., *Rapid Reconstitution of Biosynthetic Machinery for Fungal Metabolites in Aspergillus oryzae: Total Biosynthesis of Aflatrem.* Chembiochem, 2014.
310. Itoh, T., et al., *Reconstitution of a fungal meroterpenoid biosynthesis reveals the involvement of a novel family of terpene cyclases.* Nat Chem, 2010. **2**(10): p. 858-64.
311. Da Lage, J.L., et al., *Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial alpha-amylase gene to Basidiomycetes.* BMC Evol Biol, 2013. **13**: p. 40.
312. Flynn, C.M. 2014.
313. Hoff, K.J. and M. Stanke, *WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W123-8.
314. Stanke, M., et al., *AUGUSTUS: a web server for gene finding in eukaryotes.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W309-12.
315. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Molecular biology and evolution, 2013. **30**(12): p. 2725-9.
316. Solovyev, V., et al., *Automatic annotation of eukaryotic genes, pseudogenes and promoters.* Genome Biology, 2006. **7 Suppl 1**: p. S10.1-12.
317. Nakayashiki, H., *RNA silencing in fungi: mechanisms and applications.* FEBS letters, 2005. **579**(26): p. 5950-7.
318. Noble, L.M. and A. Andrianopoulos, *Fungal genes in context: genome architecture reflects regulatory complexity and function.* Genome Biology and Evolution, 2013. **5**(7): p. 1336-52.
319. Bertrand, S., et al., *Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery.* Biotechnology Advances, 2014. **32**(6): p. 1180-1204.
320. Chiang, Y.M., et al., *Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms.* Current Opinion in Chemical Biology, 2011. **15**(1): p. 137-43.
321. Cichewicz, R.H., *Epigenome manipulation as a pathway to new natural product scaffolds and their congeners.* Natural Product Reports, 2010. **27**(1): p. 11-22.
322. Du, L., et al., *Diarylcyclopentendione metabolite obtained from a Preussia typharum isolate procured using an unconventional cultivation approach.* Journal of Natural Products, 2012. **75**(10): p. 1819-23.

323. Brakhage, A.A. and V. Schroeckh, *Fungal secondary metabolites - strategies to activate silent gene clusters.* Fungal Genetics and Biology, 2011. **48**(1): p. 15-22.
324. Steinberg, G. and J. Perez-Martin, *Ustilago maydis, a new fungal model system for cell biology.* Trends in Cell Biology, 2008. **18**(2): p. 61-7.
325. Feldbrugge, M., R. Kellner, and K. Schipper, *The biotechnological use and potential of plant pathogenic smut fungi.* Applied Microbiology and Biotechnology, 2013. **97**(8): p. 3253-65.
326. Yaegashi, J., B.R. Oakley, and C.C. Wang, *Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in Aspergillus nidulans.* Journal of Industrial Microbiology and Biotechnology, 2014. **41**(2): p. 433-42.
327. O'Brien, H.E., et al., *Fungal community analysis by large-scale sequencing of environmental samples.* Appl Environ Microbiol, 2005. **71**(9): p. 5544-50.
328. Zhong, J.J. and J.H. Xiao, *Secondary metabolites from higher fungi: discovery, bioactivity, and bioproduction.* Adv Biochem Eng Biotechnol, 2009. **113**: p. 79-150.
329. Fraga, B.M., *Natural sesquiterpenoids.* Nat Prod Rep, 2011. **28**(9): p. 1580-610.
330. Lesburg, C.A., et al., *Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism.* Curr Opin Struct Biol, 1998. **8**(6): p. 695-703.
331. Yamada, Y., D.E. Cane, and H. Ikeda, *Diversity and analysis of bacterial terpene synthases.* Methods Enzymol, 2012. **515**: p. 123-62.
332. Bhatti, H.N., et al., *Microbial transformation of sesquiterpenoids.* Nat Prod Commun, 2009. **4**(8): p. 1155-68.
333. Wilkinson, B. and J. Micklefield, *Mining and engineering natural-product biosynthetic pathways.* Nat Chem Biol, 2007. **3**(7): p. 379-86.
334. Cane, D.E. and I. Kang, *Aristolochene synthase: purification, molecular cloning, high-level expression in Escherichia coli, and characterization of the Aspergillus terreus cyclase.* Arch Biochem Biophys, 2000. **376**(2): p. 354-64.
335. Cane, D.E., et al., *Overexpression in Escherichia coli of soluble aristolochene synthase from Penicillium roqueforti.* Arch Biochem Biophys, 1993. **304**(2): p. 415-9.
336. Proctor, R.H. and T.M. Hohn, *Aristolochene synthase. Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Ari1) from Penicillium roqueforti.* J Biol Chem, 1993. **268**(6): p. 4543-8.
337. Citron, C.A., et al., *Terpenoids are widespread in actinomycetes: a correlation of secondary metabolism and genome data.* Chembiochem, 2012. **13**(2): p. 202-14.
338. Keeling, C.I., et al., *Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (Picea spp.).* BMC Plant Biol, 2011. **11**: p. 43.
339. Li, G., et al., *Stereumins H-J, stereumane-type sesquiterpenes from the fungus Stereum sp.* J Nat Prod, 2011. **74**(2): p. 296-9.

340. Faraldos, J.A., et al., *Conformational analysis of (+)-Germacrene A by variable temperature NMR and NOE spectroscopy.* Tetrahedron, 2007. **63**(32): p. 7733-7742.
341. Ayer, W.A., T.T. Nakashima, and M.H. Saeedighomi, *Studies on the Biosynthesis of the Sterpurenes.* Canadian Journal of Chemistry-Revue Canadienne De Chimie, 1984. **62**(3): p. 531-533.
342. Faraldos, J.A., et al., *Bisabolyl-derived sesquiterpenes from tobacco 5-epi-aristolochene synthase-catalyzed cyclization of (2Z,6E)-farnesyl diphosphate.* J Am Chem Soc, 2010. **132**(12): p. 4281-9.
343. Hertweck, C., *Hidden biosynthetic treasures brought to light.* Nat Chem Biol, 2009. **5**(7): p. 450-2.
344. Martin, F. and M.A. Selosse, *The Laccaria genome: a symbiont blueprint decoded.* New Phytol, 2008. **180**(2): p. 296-310.
345. Vick, J.E., et al., *Optimized compatible set of BioBrick vectors for metabolic pathway engineering.* Appl Microbiol Biotechnol, 2011. **92**(6): p. 1275-86.
346. Kollner, T.G., et al., *Two pockets in the active site of maize sesquiterpene synthase TPS4 carry out sequential parts of the reaction scheme resulting in multiple products.* Arch Biochem Biophys, 2006. **448**(1-2): p. 83-92.
347. Gutta, P. and D.J. Tantillo, *Theoretical studies on farnesyl cation cyclization: pathways to pentalenene.* J Am Chem Soc, 2006. **128**(18): p. 6172-9.
348. Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism - from biochemistry to genomics.* Nat Rev Microbiol, 2005. **3**(12): p. 937-47.
349. Westfelt, L., *(-)-Torreyol (d-Cadinol).* Acta Chemica Scandinavica, 1970. **24**: p. 1618-1622.
350. Thompson, J.D., T.J. Gibson, and D.G. Higgins, *Multiple sequence alignment using ClustalW and ClustalX.* Curr Protoc Bioinformatics, 2002. **Chapter 2**: p. Unit 2 3.
351. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.* Mol Biol Evol, 2011. **28**(10): p. 2731-9.
352. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.
353. Joulain, D., and Konig, W.A., *The atlas of spectral data of sesquiterpene hydrocarbons*, in *Hamburg, Germany: E.B. Verlag*. 1998.
354. Major, D.T., Y. Freud, and M. Weitman, *Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects.* Current Opinion in Chemical Biology, 2014. **21C**: p. 25-33.
355. Greenhagen, B. and J. Chappell, *Molecular scaffolds for chemical wizardry: Learning nature's rules for terpene cyclases.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(24): p. 13479-13481.
356. Quin, M.B., S.N. Michel, and C. Schmidt-Dannert, *Moonlighting Metals: Insights into Regulation of Cyclization Pathways in Fungal Delta(6) -Protoilludene Sesquiterpene Synthases.* Chembiochem, 2015. **16**(15): p. 2191-9.

357. 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. 1153-75.
358. 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. 2480-91.
359. Comer, F.W., et al., *The structure and chemistry of hirsutic acid.* Tetrahedron, 1967. **23**(12): p. 4761-4768.
360. Yamada, Y., et al., *Terpene synthases are widely distributed in bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(3): p. 857-62.
361. Yamada, Y., et al., *Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered Streptomyces host.* The Journal of antibiotics, 2015. **68**(6): p. 385-94.
362. Chow, J.-Y., et al., *Computational-guided discovery and characterization of a sesquiterpene synthase from Streptomyces clavuligerus.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(18): p. 5661-6.
363. Bischoff, K.M. and V.W. Rodwell, *Biosynthesis and characterization of (S)-and (R)-3-hydroxy-3-methylglutaryl coenzyme A.* Biochem Med Metab Biol, 1992. **48**(2): p. 149-58.
364. Quin, M.B., C.M. Flynn, and C. Schmidt-Dannert, *Traversing the fungal terpenome.* Natural product reports, 2014. **31**(10): p. 1449-73.
365. O'Maille, P.E., J. Chappell, and J.P. Noel, *A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases.* Anal Biochem, 2004. **335**(2): p. 210-7.
366. Smith, C.A., et al., *METLIN: a metabolite mass spectral database.* Ther Drug Monit, 2005. **27**(6): p. 747-51.
367. Gietz, R.D. and R.H. Schiestl, *High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method.* Nature protocols, 2007. **2**(1): p. 31-4.
368. Hu, B., et al., *GSDS 2.0: an upgraded gene feature visualization server.* Bioinformatics (Oxford, England), 2015. **31**(8): p. 1296-7.
369. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic acids research, 2004. **32**(5): p. 1792-7.
370. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics (Oxford, England), 2009. **25**(9): p. 1189-91.
371. Jones, D.T., W.R. Taylor, and J.M. Thornton, *The Rapid Generation of Mutation Data Matrices from Protein Sequences.* Computer Applications in the Biosciences, 1992. **8**(3): p. 275-282.
372. Felsenstein, *Confidence limits on phylogenies: An approach using the bootstrap.* Evolution, 1985. **39**(4): p. 783-791.

373. Cane, D.E., et al., *Pentalenene synthase. Purification, molecular cloning, sequencing, and high-level expression in Escherichia coli of a terpenoid cyclase from Streptomyces UC5319.* Biochemistry, 1994. **33**(19): p. 5846-57.
374. Shafqat, N., et al., *Crystal structures of human HMG-CoA synthase isoforms provide insights into inherited ketogenesis disorders and inhibitor design.* J Mol Biol, 2010. **398**(4): p. 497-506.
375. Pojer, F., et al., *Structural basis for the design of potent and species-specific inhibitors of 3-hydroxy-3-methylglutaryl CoA synthases.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11491-6.
376. Campobasso, N., et al., *Staphylococcus aureus 3-hydroxy-3-methylglutaryl-CoA synthase - Crystal structure and mechanism.* Journal of Biological Chemistry, 2004. **279**(43): p. 44883-44888.
377. Steussy, C.N., et al., *X-ray crystal structures of HMG-CoA synthase from Enterococcus faecalis and a complex with its second substrate/inhibitor acetoacetyl-CoA.* Biochemistry, 2005. **44**(43): p. 14256-14267.
378. Sirinupong, N., et al., *Molecular cloning of a new cDNA and expression of 3-hydroxy-3-methylglutaryl-CoA synthase gene from Hevea brasiliensis.* Planta, 2005. **221**(4): p. 502-12.
379. Ruiz-Albert, J., E. Cerda-Olmedo, and L.M. Corrochano, *Genes for mevalonate biosynthesis in Phycomyces.* Mol Genet Genomics, 2002. **266**(5): p. 768-77.
380. Katayama, S., et al., *Molecular cloning and sequencing of the hcs gene, which encodes 3-hydroxy-3-methylglutaryl coenzyme A synthase of Schizosaccharomyces pombe.* Yeast, 1995. **11**(15): p. 1533-1537.
381. Ren, A., et al., *Molecular characterization and expression analysis of GlHMGS, a gene encoding hydroxymethylglutaryl-CoA synthase from Ganoderma lucidum (Ling-zhi) in ganoderic acid biosynthesis pathway.* World journal of microbiology & biotechnology, 2013. **29**(3): p. 523-31.
382. Winterberg, B., et al., *Elucidation of the complete ferrichrome A biosynthetic pathway in Ustilago maydis.* Mol Microbiol, 2010. **75**(5): p. 1260-71.
383. Miao, L., et al., *Astaxanthin biosynthesis is enhanced by high carotenogenic gene expression and decrease of fatty acids and ergosterol in a Phaffia rhodozyma mutant strain.* FEMS Yeast Res, 2011. **11**(2): p. 192-201.
384. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature protocols, 2015. **10**(6): p. 845-58.
385. Winzeler, E.A., et al., *Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis.* Science, 1999. **285**(5429): p. 901-6.
386. Lum, P.Y., S. Edwards, and R. Wright, *Molecular, functional and evolutionary characterization of the gene encoding HMG-CoA reductase in the fission yeast, Schizosaccharomyces pombe.* Yeast, 1996. **12**(11): p. 1107-24.
387. Huxley, C., E.D. Green, and I. Dunham, *Rapid assessment of S. cerevisiae mating type by PCR.* Trends Genet, 1990. **6**(8): p. 236.
388. Martin, V.J., Y. Yoshikuni, and J.D. Keasling, *The in vivo synthesis of plant sesquiterpenes by Escherichia coli.* Biotechnology and bioengineering, 2001. **75**(5): p. 497-503.

389. Tuinstra, R.L., et al., *Evaluation of 3-hydroxy-3-methylglutaryl-coenzyme A lyase arginine-41 as a catalytic residue: use of acetyldithio-coenzyme A to monitor product enolization.* Biochemistry, 2004. **43**(18): p. 5287-95.
390. Grigoriev, I.V., *(GP-1503) Genome sequence list of heterokaryotic or homokaryotic*, C.M. Flynn, Editor. 2015.
391. Chen, M., et al., *Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase.* ACS Chem Biol, 2016.
392. Misra, I., C. Narasimhan, and H.M. Miziorko, *Avian 3-hydroxy-3-methylglutaryl-CoA synthase. Characterization of a recombinant cholesterogenic isozyme and demonstration of the requirement for a sulfhydryl functionality in formation of the acetyl-enzyme reaction intermediate.* J Biol Chem, 1993. **268**(16): p. 12129-35.
393. Montamat, F., et al., *Isolation and characterization of a cDNA encoding Arabidopsis thaliana 3-hydroxy-3-methylglutaryl-coenzyme A synthase.* Gene, 1995. **167**(1-2): p. 197-201.
394. Emanuelsson, O., et al., *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* Journal of Molecular Biology, 2000. **300**(4): p. 1005-1016.
395. Hohn, T., S.P. McCormick, and A.E. Desjardins, *Evidence for a gene cluster involving trichothecene-pathway biosynthetic genes in Fusarium sporotrichioides.* Current Genetics, 1993. **24**: p. 291-295.
396. Basson, M.E., et al., *Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting enzyme of sterol biosynthesis.* Mol Cell Biol, 1988. **8**(9): p. 3797-808.
397. Pitera, D.J., et al., *Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli.* Metab Eng, 2007. **9**(2): p. 193-207.
398. Tobert, J.A., *Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors.* Nature reviews Drug discovery, 2003. **2**(7): p. 517-26.
399. Isaac, D.T., et al., *The 3-hydroxy-methylglutaryl coenzyme A lyase HCL1 is required for macrophage colonization by human fungal pathogen Histoplasma capsulatum.* Infect Immun, 2013. **81**(2): p. 411-20.
400. Vik, A. and J. Rine, *Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 2001. **21**(19): p. 6395-6405.
401. Kistler, H.C. and K. Broz, *Cellular compartmentalization of secondary metabolism.* Front Microbiol, 2015. **6**: p. 68.
402. Wang, H., et al., *Overexpression of Brassica juncea wild-type and mutant HMG-CoA synthase 1 in Arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance.* Plant biotechnology journal, 2012. **10**(1): p. 31-42.
403. Skrede, I., et al., *Evolutionary history of Serpulaceae (Basidiomycota): molecular phylogeny, historical biogeography and evidence for a single transition of nutritional mode.* Bmc Evolutionary Biology, 2011. **11**.
404. Lodewyk, M.W., D. Willenbring, and D.J. Tantillo, *Pentalenene formation mechanisms redux.* Org Biomol Chem, 2014. **12**(6): p. 887-94.

405. Pateraki, I., A.M. Heskes, and B. Hamberger, *Cytochromes P450 for terpene functionalisation and metabolic engineering.* Adv Biochem Eng Biotechnol, 2015. **148**: p. 107-39.
406. Cresnar, B. and S. Petric, *Cytochrome P450 enzymes in the fungal kingdom.* Biochim Biophys Acta, 2011. **1814**(1): p. 29-35.
407. Jung, S.T., R. Lauchli, and F.H. Arnold, *Cytochrome P450: taming a wild type enzyme.* Curr Opin Biotechnol, 2011. **22**(6): p. 809-17.
408. Ichinose, H. and H. Wariishi, *High-level heterologous expression of fungal cytochrome P450s in Escherichia coli.* Biochem Biophys Res Commun, 2013. **438**(2): p. 289-94.
409. Urban, P., et al., *Cloning, yeast expression, and characterization of the coupling of two distantly related Arabidopsis thaliana NADPH-cytochrome P450 reductases with P450 CYP73A5.* J Biol Chem, 1997. **272**(31): p. 19176-86.
410. Leferink, N.G., et al., *The growing VAO flavoprotein family.* Arch Biochem Biophys, 2008. **474**(2): p. 292-301.
411. Winkler, A., T.M. Kutchan, and P. Macheroux, *6-S-cysteinylation of bi-covalently attached FAD in berberine bridge enzyme tunes the redox potential for optimal activity.* J Biol Chem, 2007. **282**(33): p. 24437-43.
412. Winkler, A., et al., *A concerted mechanism for berberine bridge enzyme.* Nat Chem Biol, 2008. **4**(12): p. 739-41.
413. Jennewein, S., et al., *Taxol biosynthesis: taxane 13 alpha-hydroxylase is a cytochrome P450-dependent monooxygenase.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13595-600.
414. Gardiner, D.M., et al., *Early activation of wheat polyamine biosynthesis during Fusarium head blight implicates putrescine as an inducer of trichothecene mycotoxin production.* BMC Plant Biol, 2010. **10**: p. 289.
415. Kirchmair, M., R. Poder, and C.G. Huber, *Identification of illudins in Omphalotus nidiformis and Omphalotus olivascens var. indigo by column liquid chromatography-atmospheric pressure chemical ionization tandem mass spectrometry.* J Chromatogr A, 1999. **832**(1-2): p. 247-52.
416. McCloud, T.G., et al., *ISOLATION OF ILLUDIN S FROM THE MATURE BASIDIOCARP OF OMPHALOTUS ILLUDENS, AND ISOLATION OF ILLUDIN S AND M FROM THE FERMENTATION BROTH OF OMPHALOTUS OLEARIS*, in *American Society of Pharmacognosy, 37th Annual Meeting*. 1996, National Cancer Institute: Natural Products Support Group: Santa Cruz, CA.
417. Schoendorf, A., et al., *Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from Taxus and functional expression in yeast.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1501-6.
418. Takahashi, S., et al., *Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates.* J Biol Chem, 2007. **282**(43): p. 31744-54.
419. Wawrzyn, G., *Discovery and characterization of sesquiterpenoid biosynthetic pathways from Basidiomycota*, in *Biochemistry, Molecular Biology, and*

Biophysics Department. 2014, University of Minnesota: Minneapolis, MN. p. 187.
420. Kanamori-Kataoka, M., Y. Seto, and M. Kuramoto, *Development of a method for determining illudin S in food by gas chromatography-mass spectrometry.* Journal of Health Science, 2006. **52**(3): p. 237-242.
421. Wick, J., et al., *A Fivefold Parallelized Biosynthetic Process Secures Chlorination of Armillaria mellea (Honey Mushroom) Toxins.* Appl Environ Microbiol, 2015. **82**(4): p. 1196-204.
422. Hoye, T.R., *[natural product concentration target for NMR identification]*. 2013.
423. Matsuda, Y., et al., *Molecular Basis for Stellatic Acid Biosynthesis: A Genome Mining Approach for Discovery of Sesterterpene Synthases.* Org Lett, 2015. **17**(18): p. 4644-7.
424. Goswami, R.S. and H.C. Kistler, *Heading for disaster: Fusarium graminearum on cereal crops.* Molecular Plant Pathology, 2004. **5**(6): p. 515-525.
425. McMullen, M., et al., *A Unified Effort to Fight an Enemy of Wheat and Barley: Fusarium Head Blight.* Plant Disease, 2012. **96**(12): p. 1712-1728.
426. Desjardins, A.E., *Trichothecenes: from yellow rain to green wheat.* Asm News, 2003. **69**(4): p. 182-185.
427. Jansen, C., et al., *Infection patterns in barley and wheat spikes inoculated with wild-type and trichodiene synthase gene disrupted Fusarium graminearum.* Proc Natl Acad Sci U S A, 2005. **102**(46): p. 16892-7.
428. Lauren, D.R., S.T. Sayer, and M. di Menna, *Trichothecene production by fusarium species isolated from grain and pasture throughout New Zealand.* Mycopathologia, 1992. **120**: p. 167-176.
429. Woloshuk, C.P. and W.B. Shim, *Aflatoxins, fumonisins, and trichothecenes: a convergence of knowledge.* Fems Microbiology Reviews, 2013. **37**(1): p. 94-109.
430. Gardiner, D.M., K. Kazan, and J.M. Manners, *Nutrient profiling reveals potent inducers of trichothecene biosynthesis in Fusarium graminearum.* Fungal Genet Biol, 2009. **46**(8): p. 604-13.
431. Gardiner, D.M., K. Kazan, and J.M. Manners, *Novel genes of Fusarium graminearum that negatively regulate deoxynivalenol production and virulence.* Mol Plant Microbe Interact, 2009. **22**(12): p. 1588-600.
432. McCormick, S.P., et al., *Tri1 in Fusarium graminearum encodes a P450 oxygenase.* Appl Environ Microbiol, 2004. **70**(4): p. 2044-51.
433. Broz, K., *15-ADON concentration range is 0-50 uM*. 2014.
434. McCormick, S.P., *[personal communication FGSG06444y produces m/z 222 product in yeast*. 2015.
435. Minerdi, D., et al., *Volatile organic compounds: a potential direct long-distance mechanism for antagonistic action of Fusarium oxysporum strain MSA 35.* Environ Microbiol, 2009. **11**(4): p. 844-54.
436. Wang, Y.Z. and J.D. Miller, *Effects of Fusarium-Graminearum Metabolites on Wheat Tissue in Relation to Fusarium Head Blight Resistance.* Journal of Phytopathology-Phytopathologische Zeitschrift, 1988. **122**(2): p. 118-125.

437. Bai, G.H., A.E. Desjardins, and R.D. Plattner, *Deoxynivalenol-nonproducing Fusarium graminearum causes initial infection, but does not cause disease spread in wheat spikes.* Mycopathologia, 2002. **153**(2): p. 91-98.
438. Bahadoor, A., et al., *Hydroxylation of Longiborneol by a Clm2-Encoded CYP450 Monooxygenase to Produce Culmorin in Fusarium graminearum.* Journal of Natural Products, 2016. **79**(1): p. 81-88.
439. Quin, M.B., et al., *Complex Natural Products Synthesis: Isoprenoids, Polyketides, Phenylpropanoids, Alkaloids, and Non/-Ribosomal Peptides.*, in *Science of Synthesis: Biocatalysis in Organic Synthesis*, K. Faber, W.D. Fessner, and N.J. Turner, Editors. 2015, Georg Theime Verlag KG: Stuttgart, New York. p. 361-402.