CHAPTER FIVE

# Discovery and Characterization of Terpenoid Biosynthetic Pathways of Fungi

**Grayson T. Wawrzyn, Sarah E. Bloch, Claudia Schmidt-Dannert**[1]
Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, St. Paul, Minnesota, USA
[1]Corresponding author: e-mail address: schmi232@umn.edu

## Contents

1. Introduction 84
2. Identification, Cloning, and Characterization of Terpene Synthases 86
   2.1 Identification and cloning of terpene synthase genes without sequence information 86
   2.2 Identification and cloning of terpene synthase genes using sequence information 87
   2.3 Characterization of terpene synthase gene function 88
3. Structural Characterization and Engineering of Fungal Terpene Synthase Activities 90
   3.1 Site-directed mutagenesis guided by crystal structures 90
   3.2 Exploring terpene synthase structure through molecular models 91
4. Identification and Characterization of Biosynthetic Gene Clusters 92
   4.1 Characteristics of fungal biosynthetic gene clusters 92
   4.2 Obtaining a biosynthetic cluster sequence 93
   4.3 Cluster annotation 96
   4.4 Characterization of biosynthetic clusters 97
5. Genome Mining for Sesquiterpene Synthases 98
   5.1 Phylogeny-guided analysis of terpene synthase gene families 98
6. Conclusions 101
Acknowledgments 102
References 102

## Abstract

Fungi produce a myriad of terpenoids with a broad range of biological activities, many of which can be adapted to human use. This requires knowledge of the enzymes responsible for the biosynthesis of these compounds. Herein, we describe strategies for identification and characterization of putative biosynthetic genes, structural examination of important pathway enzymes with a focus on altering activity, and identification of biosynthetic clusters, and genome mining for yet-to-be-discovered pathways. Fungi are a particularly attractive class of organism for terpenoid pathway

*Methods in Enzymology*, Volume 515
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-394290-6.00005-7

discovery, as they often cluster their biosynthetic genes. The affordability of genome sequencing and the relatively small size of fungal genomes further simplify this process. While only a select few fungal strains are genetically tractable, many terpenoid biosynthetic genes are functional in *Escherichia coli* and *Saccharomyces cerevisiae*, allowing easy characterization. Identification of new terpenoid biosynthetic pathways has the potential to uncover new pharmaceutical compounds and establish new strategies for metabolic engineering.

## 1. INTRODUCTION

The human relationship with plants and microbes is in large part determined by the secondary metabolites they produce, whether harmful, like toxins and virulence factors, or helpful, such as pharmaceuticals and industrial commodities. A major class of such natural products is the terpenoids, which comprise tens of thousands of structures with a broad range of biological activities.

All terpenoids are derived from the basic five-carbon units isopentenyl diphosphate and its isomer dimethylallyl diphosphate, which are sequentially coupled via prenyltransferase enzymes to yield longer prenyl diphosphates, the direct precursors to terpene biosynthesis. Monoterpenes are derived from geranyl diphosphate (GPP, $C_{10}$), sesquiterpenes from farnesyl diphosphate (FPP, $C_{15}$), and diterpenes from geranylgeranyl diphosphate (GGPP, $C_{20}$) (Davis & Croteau, 2000). FPP and GGPP can be further coupled to give rise to tri- and tetraterpenoids ($C_{30}$ and $C_{40}$, respectively). Terpene synthases, also known as terpene cyclases, catalyze ionization of the prenyl diphosphate to give a carbocation, which is then quenched through intramolecular rearrangements and/or nucleophilic attack as guided by the active-site topology of the enzyme (Christianson, 2008; Davis & Croteau, 2000). The resulting hydrocarbon terpene skeleton can be further modified by one or more enzymes to give highly "decorated" bioactive terpenoid compounds (Fig. 5.1).

Although terpenoids have been predominantly studied as plant natural products, they are also major secondary metabolites of fungi. Several Ascomycete plant pathogens produce terpenoids as virulence factors and mycotoxins. Species of the mold genus *Fusarium*, which cause the major grain disease *Fusarium* head blight, synthesize a large family of sesquiterpenoid mycotoxins called trichothecenes, which pose a threat to humans and animals when infected crops enter into food sources (Fig. 5.1B) (Kimura et al., 2007). *Penicillium roqueforti* produces indole diterpenoids and

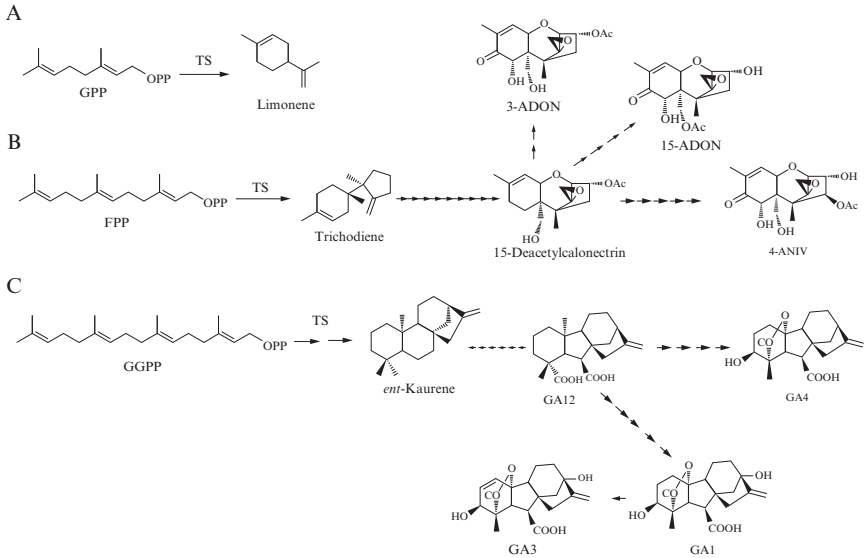



**Figure 5.1** Examples of terpenoid biosynthetic pathways. Each arrow corresponds to one chemical conversion in the pathway. (A) Single-step biosynthesis of the monoterpene limonene, catalyzed by the cyclization of geranyl diphosphate (GPP) by a terpene synthase (TS). (B) Biosynthesis of sesquiterpenoid trichothecene mycotoxins from farnesyl diphosphate (FPP). A TS catalyzes the cyclization of FPP to form the trichodiene skeleton, which is further modified to produce 3-acetyldeoxynivalenol (3-ADON), 15-acetyldeoxynivalenol (15-ADON), and 4-acetylnivalenol (4-ANIV) via the intermediate 15-deacetylcalonectrin (Kimura, Tokai, Takahashi-Ando, Ohsato, & Fujimura, 2007). (C) Biosynthesis of gibberellins (GAs) from geranylgeranyl diphosphate (GGPP). A TS cyclizes GGPP to form the *ent*-kaurene skeleton, which is further modified to form bioactive GAs (Bomke & Tudzynski, 2009).

sesquiterpenoid mycotoxins known as PR toxin, which can also pose a threat to feed stocks (Brase, Encinas, Keck, & Nising, 2009). Many *Fusarium* species also synthesize the diterpenoid plant hormones gibberellins (GAs), which act as virulence factors for grain infection (Fig. 5.1C) (Bomke & Tudzynski, 2009). Many strains of *Botrytis cinerea*, which causes gray mold disease in a broad spectrum of plant hosts, produce the sesquiterpenoid botrydial as a virulence factor (Pinedo et al., 2008). Some Ascomycetes also produce potentially useful sesquiterpenoids, such as terrecyclic acid from *Aspergillus terreus*, which has a high potential for use as an anticancer drug owing to its cytotoxicity (Brase et al., 2009).

In contrast to Ascomycetes, Basidiomycetes, including mushroom-forming higher fungi, are a veritable treasure trove of bioactive terpenoids with potential uses as pharmaceuticals. Many of these compounds exhibit

antimicrobial activity, such as the pleuromutilin sesquiterpenes (Mitterbauer & Specht, 2011), some of which have been approved for veterinary and human topical use, while others are in clinical trials as broad-spectrum antibiotics (Novak & Shlaes, 2010). Many other fungal sesquiterpenoids exhibit cytotoxic activities with high potential for use as anticancer drugs, such as the illudins produced by *Omphalotus* and *Lampteromyces* species (Zaidman, Yassin, Mahajna, & Wasser, 2005), isovelleral produced by *Lactarius vellereus* (Brase et al., 2009), and antrocin and other drimane sesquiterpenoids from *Antrodia camphorata* and related fungi (Abraham, 2001). There are also several triterpenoid products of Basidiomycota which inhibit DNA polymerase (from *Poria cocos* and *Fomitella fraxinea*) and matrix metalloproteinases essential for tumor metastasis (from *Daedalia dickinsii*), or show selective antitumor cytotoxicity (from *Ganoderma lucidum*), making them excellent candidates for cancer therapy (Zaidman et al., 2005). Recent advances in genome sequencing for Basidiomycota suggest that mushrooms have the potential to produce hundreds of terpenoid products which have so far gone unidentified.

Fungal terpenoids have proven to be a rich source of valuable bioactive natural products. Advances in the study of their biosynthesis have included a deeper understanding of terpene synthases, the first step in terpenoid biosynthesis, as well as biosynthetic gene clusters and characterization of pathway enzymes. With this understanding has come the ability to engineer terpene synthases and subsequent biosynthetic pathways, in the hopes of synthesizing novel terpenoid compounds with different bioactivities. However, despite their importance, relatively few biosynthetic pathways of fungal terpenoids have been investigated, particularly from Basidiomycota. In this chapter, we outline the major methodologies employed to define and characterize the biosynthetic pathways of fungal terpenoids, with an emphasis on sesquiterpene synthases and their emergence as a major class of biosynthetic enzymes in these organisms.

## 2. IDENTIFICATION, CLONING, AND CHARACTERIZATION OF TERPENE SYNTHASES

### 2.1. Identification and cloning of terpene synthase genes without sequence information

As knowledge of terpene biosynthesis in fungi is still quite limited, a variety of methodologies have been used in order to characterize new terpene synthases. More than two decades ago, three of the first sesquiterpene synthases cloned from fungi, aristolochene synthase (*A. terreus* and *P. roqueforti*) and trichodiene

synthase (*F. sporotrichioides*), required purification of the native protein from fungal cultures to obtain genetic information. Peptide sequencing of the resultant proteins allowed for the generation of degenerate primers aimed at specific amino acid sequences which were used for DNA hybridization or PCR amplification from cDNA libraries or genomic DNA (Cane & Kang, 2000; Hohn & Beremand, 1989a, 1989b; Proctor & Hohn, 1993). In order to amplify the *ent*-kaurene synthase (*Phaeosphaeria* sp.) and the aphidicolan-16β-ol synthase (*Phoma betae*), degenerate primers designed against the conserved aspartate/glutamate-rich (DXDD and DD(E)XXD (E)) motifs were used to amplify partial genes from fungal cDNA. Subsequent extension of genes through 5′ rapid amplification of the cDNA ends (5′-RACE) and 3′-RACE generated full-length cDNAs (Kawaide, Imai, Sassa, & Kamiya, 1997; Oikawa et al., 2001).

## 2.2. Identification and cloning of terpene synthase genes using sequence information

The approaches described above, while effective, are laborious and time consuming, resulting in a significant delay between the discovery of a biological activity and the correlation of that activity with the biosynthetic genes involved. Recent approaches have taken advantage of the fast-developing field of next-generation sequencing in order to examine cDNA libraries and even full genomes for the presence of terpene synthases. The sequencing of a cDNA library from *Armillaria gallica* allowed for the identification of a novel protoilludene sesquiterpene synthase through homology searches (Engels, Heinig, Grothe, Stadler, & Jennewein, 2011). Additionally, the mining of the published genome from the necrotrophic plant pathogen *B. cinerea* lead to the discovery of a presilphiperfolan-8β-ol sesquiterpene synthase and a nearby biosynthetic cluster (Pinedo et al., 2008). Previously, we identified six sesquiterpene synthases from the model mushroom *Coprinus cinereus* using the publicly available genome sequence (Agger, Lopez-Gallego, & Schmidt-Dannert, 2009). Finally and most recently, our group identified eleven sesquiterpene synthases from the anticancer illudin-producing fungus *Omphalotus olearius*. Subsequent homology searches of currently available Basidiomycota genomes uncovered a large complement of putative synthases, many of which have not been explored for their biosynthetic potential (Wawrzyn, Quin, Choudhary, Lopez-Gallego & Schmidt-Dannert, 2012).

Several challenges may arise during an attempt to identify, clone, and characterize fungal terpene synthases. Amplification of fungal biosynthetic genes is often complicated by poor prediction of introns, some of which can be unusually small (Misiek & Hoffmeister, 2008). Manual gene

prediction using the publicly available Web server Augustus (Stanke, Steinkamp, Waack, & Morgenstern, 2004) for gene predictions has allowed for fast confirmation using a wide variety of splicing models. Despite the usefulness of such gene model predictions, some terpene synthases still remain nonfunctional when heterologously expressed in a host such as *Escherichia coli* or *Saccharomyces cerevisiae*. For example, one of the six sesquiterpene synthases from *C. cinereus*, Cop5, was tested for all possible gene predictions using the *C. cinereus* splicing model, and still no functional protein could be produced (Agger et al., 2009). Cloning of putative sesquiterpene genes from cDNA typically yields multiple splicing isoforms, though no alternate splice variants have yet been shown to be active. Strict regulation of secondary metabolic gene clusters may result in low transcript levels under suboptimal laboratory conditions, which further complicates cloning attempts (Bayram & Braus, 2012; Merhej, Richard-Forget, & Barreau, 2011). Our strategy for amplification of sesquiterpene synthases from cDNA involves a series of nested PCR reactions using primers designed to the 5′ and 3′ ends of the gene, or using primers designed using the conserved DDxxD and NSE/DTE motifs. Amplicons obtained by these strategies typically need to be extended or combined by overlap extension PCR.

## 2.3. Characterization of terpene synthase gene function

Upon successful cloning and recombinant expression in a heterologous host, the resultant products can be characterized through analytical techniques such as HPLC, LC–MS, and GC–MS. Most fungal terpenoid biosynthetic genes are expressed well in hosts such as *E. coli* or *S. cerevisiae*, which is often not the case when expressing and characterizing plant terpene synthases, which often requires codon optimization. Analysis of terpene products typically involves extraction or capture of the desired compound(s), but in the case of small terpenes (primarily mono- and sesquiterpenes) with sufficient volatility, solid-phase microextraction (SPME) can be used to adsorb the compounds to a hydrophobic siloxane fiber.

1. Transform *E. coli* or *S. cerevisiae* with a plasmid (under the control of an inducible or constitutive promoter) containing a terpene synthase.
2. Pick single colonies and grow overnight at 37 °C in 4 mL of LB medium using the proper antibiotic selection.
3. Use the 4 mL culture to inoculate a 250-mL Erlenmeyer flask containing 50 mL of LB medium. Cover flask with aluminum foil and grow for 16–18 h at 30 °C.

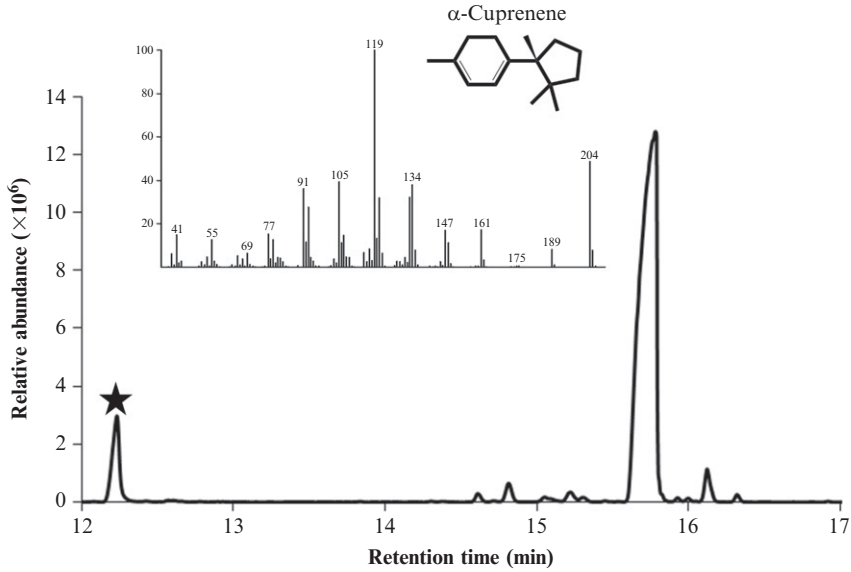



**Figure 5.2** GC–MS chromatogram for Cop6 from *Coprinus cinereus* expressed recombinantly in *E. coli* (Agger et al., 2009). The retention time and fragmentation pattern for the major product were used to identify it as α-cuprenene. Notably, the endogenous production of indole (★) by *E. coli* can be used as an internal standard to estimate relative terpene production.

4. Create a small hole in the tin foil and insert the SPME fiber (100 μm polydimethylsiloxane fiber, Supelco). Allow adsorption of compounds for 5–15 min.
5. Remove SPME fiber and load directly in the GC–MS injection port, allowing at least 10 min for desorption. GC–MS analysis is carried out on an HP GC 7890A coupled to an anion-trap mass spectrometer HP MSD triple axis detector (Agilent Technologies Inc.). Separation is carried out using a HP-5MS capillary column (30 m × 0.25 mm [inner diameter] × 1.0 μm) with an injection port temperature of 250 °C and helium as a carrier gas.
6. A program to separate sesquiterpenes typically starts near 60 °C and ramps up 8 °C $min^{-1}$ to a final oven temperature of 300 °C. Mass spectra are scanned in the range of 5–300 atomic mass units at $1^{-s}$ intervals (Fig. 5.2).
7. For product identification, calculate the retention index of each compound peak by calibrating the GC–MS first with a $C_8$–$C_{40}$ alkane mix. Use MassFinder's terpene library (Konig, Bulow, & Saritas, 1999) to carry out preliminary identification based on fragmentation pattern and retention times.

Subsequent confirmation of products can be achieved by comparison with essential oils of known composition, or synthesis of desired compounds as

standards. The compositions of many essential oils are listed in detail at The Good Scents Company Web site (http://www.thegoodscentscompany.com/), which also allows for a search of all essential oils for a specific compound. If no standard can be obtained and the compound is not present in the library, NMR analysis becomes necessary for product determination. In order to obtain enough product for NMR analysis, a number of methods can be used to increase yield. First, plasmids increasing rate-limiting steps of isoprenoid precursor pathways (Martin, Pitera, Withers, Newman, & Keasling, 2003) are available from the Addgene (www.addgene.org) plasmid repository for both *E. coli* (ID: 17815) and *S. cerevisiae* (ID: 17817). Next, once a suitable production strain of *E. coli* or *S. cerevisiae* is created, terpenoids can be produced in standard culture or a fermenter and then purified. When dealing with volatile products, the off gas can be channeled through a glass column packed with Supelpak-2SV (Sigma-Aldrich) resin for later elution of products for NMR analysis (Agger, Lopez-Gallego, Hoye, & Schmidt-Dannert, 2008).

## 3. STRUCTURAL CHARACTERIZATION AND ENGINEERING OF FUNGAL TERPENE SYNTHASE ACTIVITIES

### 3.1. Site-directed mutagenesis guided by crystal structures

While crystal structures have been solved for several microbial and plant sesquiterpene synthases, the only two available fungal sesquiterpene synthase structures are those of aristolochene synthase (*A. terreus* and *P. roqueforti*) and trichodiene synthase (*F. sporotrichioides*) (Caruthers, Kang, Rynkiewicz, Cane, & Christianson, 2000; Rynkiewicz, Cane, & Christianson, 2001; Shishova, Di Costanzo, Cane, & Christianson, 2007). All three proteins were overexpressed in *E. coli* and purified using standard techniques. These fungal enzymes share the same canonical α-helical terpene synthase fold with their plant and microbial counterparts, despite significant sequence disparity. Several helices are involved in the transition from the open to closed conformations that occurs upon binding of FPP.

An advantage of structural information is the ability to intelligently design mutants in order to change the product profiles of a terpene synthase. Site-directed mutagenesis can be employed to quickly generate a number of mutants and, in the case of sesquiterpene synthases, their activities can be quickly screened by SPME coupled with GC–MS analysis. Mutational

studies with both aristolochene synthase (Rynkiewicz, Cane, & Christianson, 2002; Vedula, Cane, & Christianson, 2005; Vedula, Jiang, Zakharian, Cane, & Christianson, 2008; Vedula, Rynkiewicz, et al., 2005; Vedula et al., 2007) and trichodiene synthase (Calvert, Taylor, & Allemann, 2002; Deligeorgopoulou & Allemann, 2003; Deligeorgopoulou, Taylor, Forcat, & Allemann, 2003; Felicetti & Cane, 2004; Forcat & Allemann, 2004; Shishova et al., 2008) have defined the specific roles of many residues involved in product formation. Interestingly, many mutants have resulted in a notable shift in product profiles, suggesting that protein engineering guided by structural information is a promising avenue for product diversification.

## 3.2. Exploring terpene synthase structure through molecular models

Our previous work with the sesquiterpene synthases Cop4 and Cop6 from *C. cinereus* aimed to elucidate the role of a conserved H-α1 loop in catalysis. To this end, we constructed structural models of both enzymes using the open, unliganded conformation and the closed conformation. These models were built using the structure of trichodiene synthase from *F. sporotrichioides* (PDB 1JFA, chain A) for Cop6 and of aristolochene synthase from *A. terreus* (PDB 2E40, chain D) for Cop4. Models were built using the Swiss Model homology-modeling server (Bordoli et al., 2009), and protein models were visualized and aligned with their template structure using PyMol (Fig. 5.3)

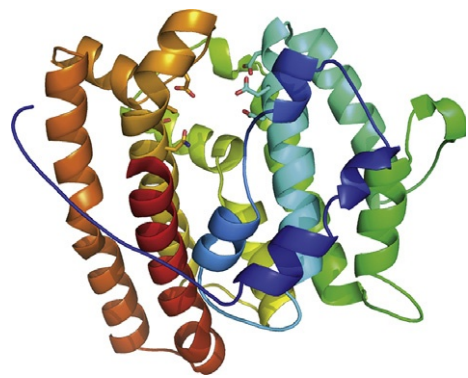


**Figure 5.3** Model of Cop4 from *Coprinus cinereus* with conserved DDxxD and NSE/DTE shown as sticks. Cop4 was modeled with aristolochene synthase from *A. terreus* (PDB 2E40, chain D) as in a previous study (Lopez-Gallego, Wawrzyn, et al., 2010), and conforms well to the classical terpene synthase fold.

(DeLano & Lam, 2005). Active-site volumes can also be calculated with CASTp (Dundas et al., 2006). In the case of Cop4 and Cop6, the size of the active site shows a strong correlation with the resultant specificity of the enzyme, with the larger active site of Cop4 resulting in many sesquiterpene products while the constricted active site of Cop6 generated only a single sesquiterpene product (Lopez-Gallego, Wawrzyn, & Schmidt-Dannert, 2010). While the product profiles of trichodiene synthase, aristolochene synthase, and Cop4 could be altered by mutagenesis, this was not the case with the highly specific α-cuprenene synthase Cop6. Mutations in the Cop6 active site either abolished activity or did not affect its product selectivity. The substrate promiscuity and reaction conditions have also been examined for their impact on product specificity. Testing both Cop4 and Cop6 at varying pH showed a dramatic shift in the fidelity of the enzyme, and this again may be attributed to the larger, more accessible active site as well as differences in the lid that covers the active site. Additionally, the introduction of non-natural substrates, such as the *cis–trans* isomer of FPP, allows for the examination and refinement of cyclization pathways leading to various products (Lopez-Gallego, Agger, Abate-Pella, Distefano, & Schmidt-Dannert, 2010).

## 4. IDENTIFICATION AND CHARACTERIZATION OF BIOSYNTHETIC GENE CLUSTERS

### 4.1. Characteristics of fungal biosynthetic gene clusters

In order to completely elucidate the biosynthetic pathway of bioactive terpenoids such as the trichothecenes and GAs made by *Fusarium* strains (Fig. 5.1), the enzymes responsible for the modification of the hydrocarbon terpene skeleton must be identified and understood. For complex terpenoids, several of these modifications must take place, corresponding to a large number of biosynthetic genes and enzymes to be identified. In fungi, this process is facilitated by a hallmark of fungal genetics, and similar to the genetic organization in bacteria: the physical clustering of secondary metabolic biosynthetic genes. A number of fungal biosynthetic pathways, including a handful for terpenoids, are carried out by coregulated genes which are clustered at a single genetic locus (Osbourn, 2010). This is believed to be advantageous to the fungus by allowing it to efficiently regulate the cluster as a whole through both transcription factors and epigenetics. The advantage to the researcher is that the identification of a single gene in a biosynthetic pathway can lead directly to identification of the remaining genes through a physical genetic linkage. However, owing to the fact that many biosynthetic clusters are tightly

controlled in response to environmental cues, their expression can be exceedingly low under laboratory conditions (Osbourn, 2010). Therefore, the use of chemical elicitors to increase expression may be necessary to detect secondary metabolite production and study the behavior of the pathway cluster *in vivo* (Chiang, Lee, Sanchez, Keller, & Wang, 2009).

A biosynthetic pathway cluster may contain as few as two genes or as many as a dozen or more, depending on the complexity of the product (Osbourn, 2010). In addition, not all genes in a biosynthetic pathway may be present in a single cluster; they may exist in more than one cluster, or some genes may be located independently throughout the genome. For example, the biosynthetic pathway for trichothecenes in *F. sporotrichioides* NRRL 3299 exists as two clusters—the *tri5* cluster containing 12 genes including the terpene synthase and a transcription factor, and another cluster containing two late-pathway genes—and a single-independent gene encodes another late-pathway enzyme (Fig. 5.4A) (Kimura et al., 2007). Interestingly, biosynthetic gene clusters also appear to function as evolutionary units. Closely related strains may contain orthologous gene clusters containing pseudogenes which have been inactivated through accumulation of mutations, leading to a difference in product profile between two strains synthesizing the same class of terpenoid (Bomke & Tudzynski, 2009; Kimura et al., 2007). In addition, a comparison of GA biosynthetic clusters across several fungal strains showed that some closely related *Fusarium* strains contain highly divergent GA gene clusters, whereas the distantly related fungi *F. fujikuroi* and *Sphaceloma manihoticola* contain highly conserved biosynthetic clusters (Fig. 5.4C) (Bomke & Tudzynski, 2009). This suggests that fungi can inherit biosynthetic clusters both vertically and horizontally, although no definitive evidence of horizontal gene transfer has been forthcoming.

## 4.2. Obtaining a biosynthetic cluster sequence

### *4.2.1 Identifying an anchor gene sequence*

In order to identify a terpenoid biosynthetic cluster, the gene sequence of at least one biosynthetic pathway enzyme must be known, hereafter referred to as the anchor sequence. In the past, the anchor sequence has often been that of the terpene synthase (Agger et al., 2009; Hohn & Beremand, 1989a, 1989b; Toyomasu et al., 2004, 2007), although not exclusively (Pinedo et al., 2008). When using genomic data, the characteristic conserved domains of terpene cyclases allow for their putative identification and use as anchor sequences.

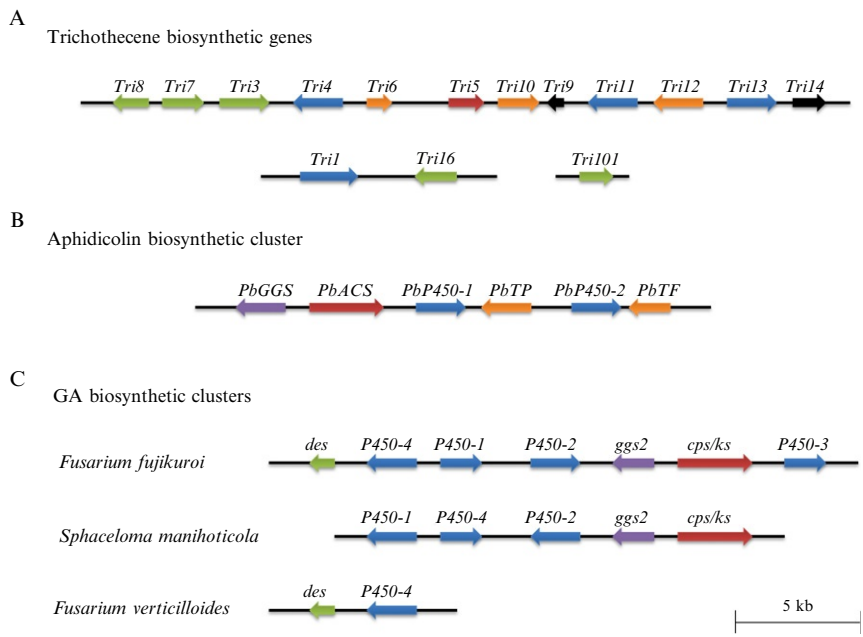



**Figure 5.4** Selected terpenoid biosynthetic gene clusters in fungi. Terpene synthase genes (*Tri5*, *PbACS*, *cps/ks*); cytochrome P540 (CYP) genes (*Tri4*, *Tri11*, *Tri13*, *Tri1*, *PbP450-x*, *P450-x*); transport or regulatory genes (*Tri6*, *Tri10*, *Tri12*, *PbTP*, *PbTF*); GGPP synthase genes (*PbGGS*, *ggs2*); other biosynthetic genes (*Tri8*, *Tri7*, *Tri3*, *Tri16*, *Tri101*, *des*); genes of unknown function (*Tri9*, *Tri14*). (A) Trichothecene biosynthetic genes of *Fusarium sporotrichioides* (Kimura et al., 2007). (B) Aphidicolin biosynthetic cluster of *Phoma betae* (Toyomasu et al., 2004). (C) Examples of the diversity of GA biosynthetic clusters (Bomke & Tudzynski, 2009). In this case, a distantly related fungus shows greater cluster similarity than a species of the same genus.

### *4.2.2 Obtaining a sequencing dataset*

#### 4.2.2.1 Small-scale sequencing approaches

If no genome sequence data is available for a fungal strain for which an anchor sequence has been identified, small-scale genomic DNA sequencing methods can be used to isolate a biosynthetic gene cluster. The first trichothecene *tri5* cluster was identified through cosmid sequencing (Kimura et al., 2003). For small (~20 kb) biosynthetic clusters predicted to contain only a handful of genes, genome walking can be used to isolate and sequence the cluster. This method was used to sequence the *P. betae* aphidicolin cluster (15.6 kb) (Toyomasu et al., 2004) (Fig. 5.4B) and the *Phomopsis amygdali* fusicoccin cluster (~20 kb) (Toyomasu et al., 2007), in both instances using

the Universal GenomeWalker Kit from Clontech. A summary of the protocol is as follows:

1. Design primary and secondary PCR primers to prime upstream and downstream walking from the anchor sequence. Primers should be as close to the anchor sequence ends as possible; for example, the primer for walking upstream of the anchor should anneal just downstream of the beginning of the anchor. Secondary PCR primers should be nested with respect to the primary primers.
2. Create four blunt-end restriction libraries of purified genomic DNA, using the four enzymes provided by the kit: *Dra*I, *Eco*RV, *Pvu*II, and *Stu*I. Purify restriction libraries via phenol–chloroform extraction as directed.
3. Perform a ligation of the purified restriction fragment libraries with the GenomeWalker Adaptor (provided), allowing for PCR priming at the ends of DNA fragments.
4. Using the primary anchor primers and Adaptor Primer 1 (provided), carry out a PCR using each of the four restriction libraries as templates.
5. If no clear bands are obtained from the primary PCR, a secondary PCR using the secondary anchor primers and Adaptor Primer 2 (provided) may be carried out using the primary PCR product as the template.
6. PCR fragments can be visualized via agarose gel electrophoresis, gel purified, and cloned into a TOPO® vector (Invitrogen) for sequencing analysis.

#### 4.2.2.2 Whole-genome sequencing

Perhaps the least labor-intensive way to identify a biosynthetic gene cluster is to use whole-genome sequencing data. Fortunately, next-generation sequencing and open-source bioinformatics software suites such as Galaxy (Giardine et al., 2005) are making it cheaper and easier to embark on a genome sequencing project. Alternatively, many draft fungal genome sequences are publicly available, and preliminary BLAST searches reveal that many of them contain terpene synthase homologs. Such an approach was used to identify a lagopodin biosynthetic cluster in *C. cinereus* (Agger et al., 2009), the botridial biosynthetic cluster in *B. cinerea* (Pinedo et al., 2008), and to identify and compare GA and other diterpene biosynthetic clusters in a number of fungal strains (Bomke & Tudzynski, 2009). Characterization of biosynthetic clusters in these genomes could lead to the discovery of new biosynthetic pathways of bioactive terpenoids without the need to obtain new sequencing data.

## 4.3. Cluster annotation

Once an anchor sequence has been located within a sequencing dataset, the surrounding regions of the genome can be searched for putative biosynthetic cluster genes. For some published genome sequences, gene prediction and annotation data may already be available. For newer datasets, gene prediction can be carried out using the freely available software Augustus (http://bioinf.uni-greifswald.de/augustus/) on the contig or scaffold containing the anchor sequence.

1. Enter the sequence to be analyzed. For sequences up to 40 kb in length, the web interface may be used; for larger sequences, Augustus should be downloaded and run locally.
2. Select the organism most closely related to the source organism of the sequence being analyzed. This informs the gene prediction model to be used. Gene prediction models for several fungal species are available.
3. Select other parameters as desired and run Augustus.
4. Augustus will output a .txt file in .gff format. The file will contain predicted gene and protein sequences and list tabular information about the features of each predicted gene sequence. Save this file to the same local folder containing the sequence file.
5. Gene predictions can be easily visualized using the freely available Argo genome browser (http://www.broadinstitute.org/annotation/argo/), which can be run using Java Webstart. Open the sequence file of the sequence you have just annotated. Argo will automatically use the Augustus output .txt file to create a feature map on the sequence.
6. Predicted gene/protein sequences can be annotated by running BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

At this point, it may become evident that the contig or scaffold containing the anchor gene does not contain the entire cluster. If this is the case, a genome-walking technique can be used to sequence beyond the end of the contig and possibly identify another contig containing the remaining biosynthetic genes.

Based on sequencing data alone, rough cluster boundaries can be predicted based on gene annotation—for instance, a housekeeping gene flanking a putative biosynthetic gene may denote the boundary of the gene cluster. However, the boundaries may not always be clear from genome sequencing data. Given that biosynthetic cluster genes are usually coregulated, expression studies can be used to define the region of coregulation and thus the boundaries of the cluster. Such studies, using qRT-PCR, were performed to determine the boundaries of the *tri5* cluster

(Brown, Dyer, McCormick, Kendra, & Plattner, 2004; Kimura et al., 2003). If microarray or RNAseq data is available for a given strain, this may also be useful in characterizing cluster transcription and determining cluster boundaries (Pinedo et al., 2008).

## 4.4. Characterization of biosynthetic clusters

### 4.4.1 Gene knockout and complementation studies

The final step to defining a biosynthetic gene cluster is to characterize the putative genes within. For a genetically tractable host, this can be carried out via gene knockout and complementation studies. If the expression of the cluster is too low under normal conditions to measure changes in the abundance of pathway intermediates, a background mutation may be made to upregulate or deregulate the cluster. For example, a high trichothecene-producing strain of *F. graminearum* F15 was made by genomic integration of an overexpression cassette containing *FgTri6*, which regulates the *tri5* cluster (Kimura et al., 2003). Conversely, a high botridial-producing strain of *B. cinerea* was engineered by knocking out the *bcg1* gene, which is responsible for downregulation of the botridial cluster (Pinedo et al., 2008). Unfortunately, many higher fungi are not genetically tractable and therefore cannot be used for knockout and complementation studies.

### 4.4.2 Pathway construction in a heterologous host

Biosynthetic cluster genes may also be characterized quite effectively in a heterologous host. Several trichothecene biosynthetic cluster genes have been characterized via overexpression in *E. coli* and purification (reviewed in Kimura et al., 2007). However, reconstitution of part or all of a biosynthetic cluster through expression in a heterologous host can lead to more rapid identification of pathway steps and enzyme function. Our laboratory was able to propose a partial biosynthetic pathway by the coexpression of the α-cuprenene synthase gene *cop6* and the two cytochrome P450 (CYP) enzymes with which it is clustered in the *C. cinereus* genome (Agger et al., 2009). The engineered yeast strain *S. cerevisiae* WAT11, which expresses the required NADPH CYP reductase to regenerate the cofactor in fungal CYPs, should be used for heterologous CYP expression and pathway construction (Proctor & Hohn, 1993). Small or partial clusters can be reconstituted in *S. cerevisiae* WAT11 as follows:

1. Clone the terpene synthase and any other cluster genes into yeast shuttle vectors downstream of a *S. cerevisiae* promoter such as the gal1-10

promoter. Shuttle vectors such as pESC (Agilent Technologies Inc.) with different auxotrophic markers can be used.
2. Cotransform the plasmids containing the terpene synthase with one or more of the other vectors containing cluster genes into *S. cerevisiae* WAT11.
3. Culture transformants containing both plasmids 48 h at 30 °C with shaking at 250 rpm in a minimal medium such as synthetic galactose minimal media.
4. Analyze the culture headspace through SPME followed by GC–MS as outlined in Section 2.3.
5. Pellet yeast cells via centrifugation and remove the media and begin stirring slowly.
6. Analyze the culture media by submerging the SPME fiber into the slowly stirring media followed by GC–MS (Section 2.3).

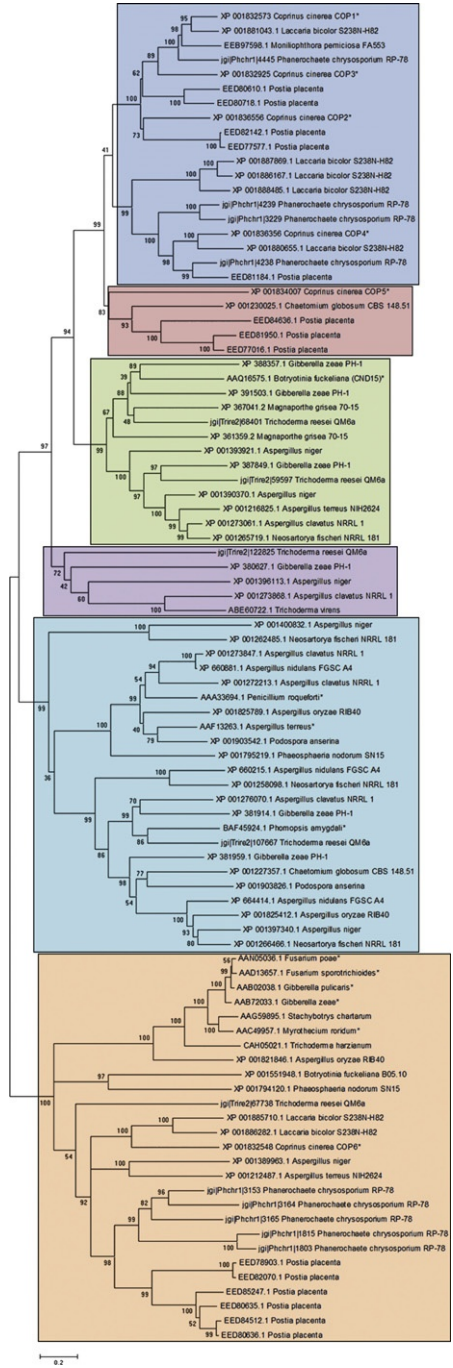
For complete heterologous expression of large clusters, methods such as DNA assembler (Shao, Zhao, & Zhao, 2009) and Gibson assembly (Gibson et al., 2009) can be used to quickly assemble full pathways onto a single yeast vector. This construct can then be transformed into *S. cerevisiae* WAT11, and analysis of pathway products and intermediates may be carried out as in steps 3–6.

## 5. GENOME MINING FOR SESQUITERPENE SYNTHASES

### 5.1. Phylogeny-guided analysis of terpene synthase gene families

The rapid influx of new fungal genomes provides an unprecedented opportunity for biosynthetic gene discovery. The ease and speed at which fungal genomes can be sequenced and assembled has generated an enormous resource in a very short period of time. The Joint Genome Institute's (JGI) Fungal Genomics Program hosts 111 Ascomycota genome sequences, and 54 Basidiomycota genomes, many of which have been added in the past year. Very few terpenoid biosynthetic enzymes have been identified in Basidiomycota in comparison to Ascomycota, and even fewer complete biosynthetic clusters have been uncovered. Homology searches can quickly uncover many terpene synthase homologs in numerous fungal species (Fig. 5.5). Phylogenetic trees can be examined and groupings of homologs can be used to infer putative functions.

In 2009, fewer than a dozen genome sequences of Basidiomycota were published. A survey of the then available genome sequences using BLAST


XP 001832573 Coprinus cinerea COP1*
XP 001881043.1 Laccaria bicolor S238N-H82
EEB97598.1 Moniliophthora perniciosa FA553
jgi|Phchr1|4445 Phanerochaete chrysosporium RP-78
XP 001832925 Coprinus cinerea COP3*
EED80610.1 Postia placenta
EED80718.1 Postia placenta
XP 001836556 Coprinus cinerea COP2*
EED82142.1 Postia placenta
EED77577.1 Postia placenta
XP 001887869.1 Laccaria bicolor S238N-H82
XP 001886167.1 Laccaria bicolor S238N-H82
XP 001888485.1 Laccaria bicolor S238N-H82
jgi|Phchr1|4239 Phanerochaete chrysosporium RP-78
jgi|Phchr1|3229 Phanerochaete chrysosporium RP-78
XP 001836356 Coprinus cinerea COP4*
XP 001880655.1 Laccaria bicolor S238N-H82
jgi|Phchr1|4238 Phanerochaete chrysosporium RP-78
EED81184.1 Postia placenta
XP 001834007 Coprinus cinerea COP5*
XP 001230025.1 Chaetomium globosum CBS 148.51
EED84636.1 Postia placenta
EED81950.1 Postia placenta
EED77016.1 Postia placenta
XP 388357.1 Gibberella zeae PH-1
AAQ16575.1 Botryotinia fuckeliana (CND15)*
XP 391503.1 Gibberella zeae PH-1
XP 367041.2 Magnaporthe grisea 70-15
jgi|Trire2|68401 Trichoderma reesei QM6a
XP 361359.2 Magnaporthe grisea 70-15
XP 001393921.1 Aspergillus niger
XP 387849.1 Gibberella zeae PH-1
jgi|Trire2|59597 Trichoderma reesei QM6a
XP 001390370.1 Aspergillus niger
XP 001216825.1 Aspergillus terreus NIH2624
XP 001273061.1 Aspergillus clavatus NRRL 1
XP 001265719.1 Neosartorya fischeri NRRL 181
jgi|Trire2|122825 Trichoderma reesei QM6a
XP 380627.1 Gibberella zeae PH-1
XP 001396113.1 Aspergillus niger
XP 001273868.1 Aspergillus clavatus NRRL 1
ABE60722.1 Trichoderma virens
XP 001400832.1 Aspergillus niger
XP 001262485.1 Neosartorya fischeri NRRL 181
XP 001273847.1 Aspergillus clavatus NRRL 1
XP 660881.1 Aspergillus nidulans FGSC A4
XP 001272213.1 Aspergillus clavatus NRRL 1
AAA33694.1 Penicillium roqueforti*
XP 001825789.1 Aspergillus oryzae RIB40
AAF13263.1 Aspergillus terreus*
XP 001903542.1 Podospora anserina
XP 001795219.1 Phaeosphaeria nodorum SN15
XP 660215.1 Aspergillus nidulans FGSC A4
XP 001258098.1 Neosartorya fischeri NRRL 181
XP 001276070.1 Aspergillus clavatus NRRL 1
XP 381914.1 Gibberella zeae PH-1
BAF45924.1 Phomopsis amygdali*
jgi|Trire2|107667 Trichoderma reesei QM6a
XP 381959.1 Gibberella zeae PH-1
XP 001227357.1 Chaetomium globosum CBS 148.51
XP 001903826.1 Podospora anserina
XP 664414.1 Aspergillus nidulans FGSC A4
XP 001825412.1 Aspergillus oryzae RIB40
XP 001397340.1 Aspergillus niger
XP 001266466.1 Neosartorya fischeri NRRL 181
AAN05036.1 Fusarium poae*
AAD13657.1 Fusarium sporotrichioides*
AAB02038.1 Gibberella pulicaris*
AAB72033.1 Gibberella zeae*
AAG59895.1 Stachybotrys chartarum
AAC49957.1 Myrothecium roridum*
CAH05021.1 Trichoderma harzianum
XP 001821846.1 Aspergillus oryzae RIB40
XP 001551948.1 Botryotinia fuckeliana B05.10
XP 001794120.1 Phaeosphaeria nodorum SN15
jgi|Trire2|67738 Trichoderma reesei QM6a
XP 001885710.1 Laccaria bicolor S238N-H82
XP 001886282.1 Laccaria bicolor S238N-H82
XP 001832548 Coprinus cinerea COP6*
XP 001389963.1 Aspergillus niger
XP 001212487.1 Aspergillus terreus NIH2624
jgi|Phchr1|3153 Phanerochaete chrysosporium RP-78
jgi|Phchr1|3164 Phanerochaete chrysosporium RP-78
jgi|Phchr1|3165 Phanerochaete chrysosporium RP-78
jgi|Phchr1|1815 Phanerochaete chrysosporium RP-78
jgi|Phchr1|1803 Phanerochaete chrysosporium RP-78
EED78903.1 Postia placenta
EED82070.1 Postia placenta
EED85247.1 Postia placenta
EED80635.1 Postia placenta
EED84512.1 Postia placenta
EED80636.1 Postia placenta
0.2

with the six sesquiterpene synthase sequences identified from *C. cinereus* (Agger et al., 2009) indicated that terpene synthases (specifically, sesquiterpene synthases) have undergone extensive gene family expansion in these fungi. A quick survey of the now available, large number of Basidiomycota genomes confirms our previous finding. In addition, many terpene synthases lie within clearly defined biosynthetic clusters. Basidiomycota thus represent a largely uncharted resource for the discovery of new terpenoid bioactive compounds.

In order to effectively find biosynthetic genes in a reasonable amount of time, several bioinformatic tools have been developed for the identification of secondary metabolite clusters. For example, programs such as antiSMASH allow not only the identification of terpenoid metabolic clusters but also the identification of other potentially relevant secondary metabolites such as polyketides and nonribosomal peptides (Medema et al., 2011). A more conventional strategy requires homology searches using the NCBI BLAST software with experimentally characterized microbial, plant, and fungal protein sequences. BLAST results can be compiled and sequence alignments computed using ClustalW (Thompson, Gibson, & Higgins, 2002) and the molecular evolutionary genetics analysis tool MEGA (Tamura et al., 2011). The resultant alignments often need to be manually curated to remove poor gene predictions and annotations that show incorrect splicing predictions, resulting in shorter or longer than expected proteins, or the absence of conserved catalytic motifs.

Phylogenetic trees can be generated using the neighbor-joining method (Saitou & Nei, 1987). The obtained alignments and phylogenetic trees yield a framework for studying other terpene synthase genes from different organisms. Once such a framework is created, the insertion of known biochemical data can facilitate the discovery of closely related enzymes with desired activities. For example, we characterized the sesquiterpene synthase Cop6 extensively, showing that it is a very product specific enzyme requiring a

**Figure 5.5** Unrooted neighbor-joining tree of fungal sesquiterpene synthase homologs. Homology searches of NCBI's nonredundant protein sequence database and in a select few fungal genome sequences obtained from the Broad Institute and JGI. Branches are labeled with their respective bootstrap values, and each branch label contains the accession number and strain name. Functionally characterized terpene synthases are indicated with an asterisk. Boxed regions identify homology groups that may possess related functions as inferred from biochemically characterized enzymes contained in these groups.

1,6 cyclization of FPP to form α-cuprenene. Cop6 is grouped with other biochemically characterized trichodiene synthases, which also proceed through a 1,6 cyclization of FPP. As more biochemical data becomes available this framework can be expanded, and cyclization mechanisms of uncharacterized genes can be inferred. Further, and perhaps most importantly, the genomic region surrounding putative biosynthetic genes can be surveyed for the presence of product-modifying enzymes as described above.

## 6. CONCLUSIONS

As technology has advanced in the fields of genomics, metabolomics, and metabolic engineering, a systematic pipeline to the identification of new bioactive terpenoids has become feasible. Terpenome profiling of fungi, followed by bioactivity screening and accompanied by genome sequencing, can lead directly from the identification of a bioactive terpenoid to its corresponding terpene synthase and potentially a cluster of biosynthetic pathway genes. A growing toolbox for synthetic biology can then allow partial or full biosynthetic pathway engineering into a heterologous host for the characterization of pathway enzymes. No longer is it necessary to rely on the pathway expression levels or genetic tractability of the fungal producer, which can take years of optimization. By comparison, drug discovery through a metabolomics, genomics, and heterologous pathway expression pipeline can be much more rapid and applied to a broader range of products.

Since the coming of the information age, and the more recent sequencing revolution, there has been an explosion in the amount of data from which novel fungal biosynthetic pathways can be mined. With the recent announcement by Illumina Inc. and Life Technologies Corp. of sequencers able to sequence the complete human genome in one day for $1000, the flow of sequencing data into current databases shows no sign of letting up. We have reached the point where, to paraphrase Robert Browning, "Our reach exceeds our grasp." Our knowledge of biosynthetic pathways is no longer limited by the amount of sequence information available to us—rather it is limited by our actual interpretation and biochemical knowledge of the sequencing data at hand. The intensive biochemical characterization of these pathways is what will finally lead to the knowledge of new natural products, new avenues for metabolic engineering, and ultimately new pharmaceuticals and medicines for the diseases of the world.

## ACKNOWLEDGMENTS

Research on terpenoid biosynthesis in the corresponding author's laboratory is supported by the National Institute of Health Grant GM080299 (to C. S.-D.). G. T. W. and S. E. B. were supported by predoctoral National Institute of Health traineeship (GM08700 (G. T. W.) and GM008347 (S. E. B.)).

## REFERENCES

Abraham, W. R. (2001). Bioactive sesquiterpenes produced by fungi: Are they useful for humans as well? *Current Medicinal Chemistry*, *8*(6), 583–606.

Agger, S. A., Lopez-Gallego, F., Hoye, T. R., & Schmidt-Dannert, C. (2008). Identification of sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120. *Journal of Bacteriology*, *190*(18), 6084–6096.

Agger, S., Lopez-Gallego, F., & Schmidt-Dannert, C. (2009). Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. *Molecular Microbiology*, *72*(5), 1181–1195.

Bayram, Ö., & Braus, G. H. (2012). Coordination of secondary metabolism and development in fungi: The velvet family of regulatory proteins. *FEMS Microbiology Reviews*, *36*(1), 1–24.

Bomke, C., & Tudzynski, B. (2009). Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry*, *70* (15–16), 1876–1893.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, *4* (1), 1–13.

Brase, S., Encinas, A., Keck, J., & Nising, C. F. (2009). Chemistry and biology of mycotoxins and related fungal metabolites. *Chemical Reviews*, *109*(9), 3903–3990.

Brown, D. W., Dyer, R. B., McCormick, S. P., Kendra, D. F., & Plattner, R. D. (2004). Functional demarcation of the *Fusarium* core trichothecene gene cluster. *Fungal Genetics and Biology*, *41*(4), 454–462.

Calvert, M. J., Taylor, S. E., & Allemann, R. K. (2002). Tyrosine 92 of aristolochene synthase directs cyclisation of farnesyl pyrophosphate. *Chemical Communications (Cambridge, England)*, (20), 2384–2385.

Cane, D. E., & Kang, I. (2000). Aristolochene synthase: Purification, molecular cloning, high-level expression in *Escherichia coli*, and characterization of the *Aspergillus terreus* cyclase. *Archives of Biochemistry and Biophysics*, *376*(2), 354–364.

Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E., & Christianson, D. W. (2000). Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium roqueforti*. *The Journal of Biological Chemistry*, *275*(33), 25533–25539.

Chiang, Y. M., Lee, K. H., Sanchez, J. F., Keller, N. P., & Wang, C. C. (2009). Unlocking fungal cryptic natural products. *Natural Product Communications*, *4*(11), 1505–1510.

Christianson, D. W. (2008). Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology*, *12*(2), 141–150.

Davis, E. M., & Croteau, R. (2000). Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. Biosynthesis: Aromatic polyketides, isoprenoids, alkaloids (Vol. 209). Berlin: Springer-Verlag Berlin, pp. 53–95.

DeLano, W. L., & Lam, J. W. (2005). PyMOL: A communications tool for computational models. *Abstracts of Papers of the American Chemical Society*, *230*, U1371–U1372.

Deligeorgopoulou, A., & Allemann, R. K. (2003). Evidence for differential folding of farnesyl pyrophosphate in the active site of aristolochene synthase: A single-point mutation

converts aristolochene synthase into an (E)-beta-farnesene synthase. *Biochemistry*, *42*(25), 7741–7747.

Deligeorgopoulou, A., Taylor, S. E., Forcat, S., & Allemann, R. K. (2003). Stabilisation of eudesmane cation by tryptophan 334 during aristolochene synthase catalysis. *Chemical Communications (Cambridge, England)*, (17), 2162–2163.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, *34*, W116–W118.

Engels, B., Heinig, U., Grothe, T., Stadler, M., & Jennewein, S. (2011). 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*, *286*(9), 6871–6878.

Felicetti, B., & Cane, D. E. (2004). Aristolochene synthase: Mechanistic analysis of active site residues by site-directed mutagenesis. *Journal of the American Chemical Society*, *126*(23), 7212–7221.

Forcat, S., & Allemann, R. K. (2004). Dual role for phenylalanine 178 during catalysis by aristolochene synthase. *Chemical Communications (Cambridge, England)*, (18), 2094–2095.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451–1455.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345.

Hohn, T. M., & Beremand, P. D. (1989a). Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene-producing fungus *Fusarium sporotrichioides*. *Gene*, *79*(1), 131–138.

Hohn, T. M., & Beremand, P. D. (1989b). Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene producing fungus *Fusarium sporotrichoides*. *Gene*, *79*(1), 131–138.

Kawaide, H., Imai, R., Sassa, T., & Kamiya, Y. (1997). *ent*-Kaurene synthase from the fungus *Phaeosphaeria* sp. L487—cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase in fungal gibberellin biosynthesis. *The Journal of Biological Chemistry*, *272*(35), 21706–21712.

Kimura, M., Tokai, T., O'Donnell, K., Ward, T. J., Fujimura, M., Hamamoto, H., et al. (2003). The trichothecene biosynthesis gene cluster of *Fusarium graminearum* F15 contains a limited number of essential pathway genes and expressed non-essential genes. *FEBS Letters*, *539*(1–3), 105–110.

Kimura, M., Tokai, T., Takahashi-Ando, N., Ohsato, S., & Fujimura, M. (2007). Molecular and genetic studies of *Fusarium* trichothecene biosynthesis: Pathways, genes, and evolution. *Bioscience, Biotechnology, and Biochemistry*, *71*(9), 2105–2123.

Konig, W. A., Bulow, N., & Saritas, Y. (1999). Identification of sesquiterpene hydrocarbons by gas phase analytical methods. *Flavour and Fragrance Journal*, *14*(6), 367–378.

Lopez-Gallego, F., Agger, S. A., Abate-Pella, D., Distefano, M. D., & Schmidt-Dannert, C. (2010). Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: Catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. *ChemBioChem*, *11*(8), 1093–1106.

Lopez-Gallego, F., Wawrzyn, G. T., & Schmidt-Dannert, C. (2010). Selectivity of fungal sesquiterpene synthases: Role of the active site's H-1 alpha loop in catalysis. *Applied and Environmental Microbiology*, *76*(23), 7723–7733.

Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology*, *21*(7), 796–802.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, *39*, W339–W346.

Merhej, J., Richard-Forget, F., & Barreau, C. (2011). Regulation of trichothecene biosynthesis in *Fusarium*: Recent advances and new insights. *Applied Microbiology and Biotechnology*, *91*(3), 519–528.

Misiek, M., & Hoffmeister, D. (2008). Processing sites involved in intron splicing of Armillaria natural product genes. *Mycological Research*, *112*, 216–224.

Mitterbauer, R., & Specht T. (2011). Cloning and sequence of *Clitopilus passeckerianus* diterpene synthase and pleuromutilin biosynthesis gene cluster, and use for producing pleuromutilin. WO/2011/110610.

Novak, R., & Shlaes, D. M. (2010). The pleuromutilin antibiotics: A new class for human use. *Current Opinion in Investigational Drugs*, *11*(2), 182–191.

Oikawa, H., Toyomasu, T., Toshima, H., Ohashi, S., Kawaide, H., Kamiya, Y., et al. (2001). 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*, *123*(21), 5154–5155.

Osbourn, A. (2010). Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. *Trends in Genetics*, *26*(10), 449–457.

Pinedo, C., Wang, C. M., Pradier, J. M., Dalmais, B., Choquer, M., Le Pecheur, P., et al. (2008). Sesquiterpene synthase from the botrydial biosynthetic gene cluster of the phytopathogen *Botrytis cinerea*. *ACS Chemical Biology*, *3*(12), 791–801.

Proctor, R. H., & Hohn, T. M. (1993). Aristolochene synthase—Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (*Ari1*) from *Penicillium roqueforti*. *The Journal of Biological Chemistry*, *268*(6), 4543–4548.

Rynkiewicz, M. J., Cane, D. E., & Christianson, D. W. (2001). Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proceedings of the National Academy of Sciences United States of America*, *98*(24), 13543–13548.

Rynkiewicz, M. J., Cane, D. E., & Christianson, D. W. (2002). X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity. *Biochemistry*, *41*(6), 1732–1741.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.

Shao, Z., Zhao, H., & Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Research*, *37*(2), e16.

Shishova, E. Y., Di Costanzo, L., Cane, D. E., & Christianson, D. W. (2007). X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry*, *46*(7), 1941–1951.

Shishova, E. Y., Yu, F., Miller, D. J., Faraldos, J. A., Zhao, Y., Coates, R. M., et al. (2008). X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *The Journal of Biological Chemistry*, *283*(22), 15431–15439.

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*, W309–W312.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739.

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, chap. 2, unit 2.3.

Toyomasu, T., Nakaminami, K., Toshima, H., Mie, T., Watanabe, K., Ito, H., et al. (2004). Cloning of a gene cluster responsible for the biosynthesis of diterpene aphidicolin, a specific inhibitor of DNA polymerase alpha. *Bioscience, Biotechnology, and Biochemistry*, *68*(1), 146–152.

Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., et al. (2007). Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *Proceedings of the National Academy of Sciences United States of America*, *104*(9), 3084–3088.

Vedula, L. S., Cane, D. E., & Christianson, D. W. (2005). Role of arginine-304 in the diphosphate-triggered active site closure mechanism of trichodiene synthase. *Biochemistry*, *44*(38), 12719–12727.

Vedula, L. S., Jiang, J., Zakharian, T., Cane, D. E., & Christianson, D. W. (2008). Structural and mechanistic analysis of trichodiene synthase using site-directed mutagenesis: Probing the catalytic function of tyrosine-295 and the asparagine-225/serine-229/glutamate-233-$Mg^{2+}$ B motif. *Archives of Biochemistry and Biophysics*, *469*(2), 184–194.

Vedula, L. S., Rynkiewicz, M. J., Pyun, H. J., Coates, R. M., Cane, D. E., & Christianson, D. W. (2005). Molecular recognition of the substrate diphosphate group governs product diversity in trichodiene synthase mutants. *Biochemistry*, *44*(16), 6153–6163.

Vedula, L. S., Zhao, Y. X., Coates, R. M., Koyama, T., Cane, D. E., & Christianson, D. W. (2007). Exploring biosynthetic diversity with trichodiene synthase. *Archives of Biochemistry and Biophysics*, *466*(2), 260–266.

Wawrzyn, G. T., Quin, M. B., Choudhary, S., Lopez-Gallego, F., & Schmidt-Dannert, C. (2012). Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. *Chemistry & Biology*, *19*(6), 772–783.

Zaidman, B. Z., Yassin, M., Mahajna, J., & Wasser, S. P. (2005). Medicinal mushroom modulators of molecular targets as cancer therapeutics. *Applied and Environmental Microbiology*, *67*(4), 453–468.