Biosynthetic Studies of Fungal Diterpene Antibiotics

Qingmei Yao

AN ABSTRACT OF THE THESIS OF

Qingmei Yao for the degree of Master of Science in Pharmacy presented on March 23, 2007.

Title: Biosynthetic Studies of Fungal Diterpene Antibiotics

Abstract approved: _____

Philip J. Proteau

The isoprenoid pathway is one of the major biosynthetic pathways of secondary metabolites. Isoprenoids with a C_{20} skeleton are known as diterpenoids and are derived from the common precursor, geranylgeranyl diphosphate (GGPP). This dissertation describes approaches to discover the diterpene synthase enzymes which provide the backbone structures for pleuromutilin and sordarin, two diterpene antibiotics produced by the fungi *Pleurotus passeckerianus* Pilat and *Sordaria araneosa* Cain, respectively. A greater understanding of these synthase

enzymes may allow us to eventually manipulate the corresponding genes, creating new scaffolds for bioactive molecules.

In this dissertation, attempts to identify the biosynthetic gene clusters, including the diterpene synthase genes, for the formation of pleuromutilin and sordarin are described. Three methods were used to locate the diterpene synthase genes.

The first approach involved the use of PCR (Polymerase Chain Reaction) with degenerate primers designed from identified and putative fungal diterpene synthase genes. A variety of conditions and degenerate primers were utilized, but no desired PCR products were found. The reason that this method did not work might be the limited overall knowledge of fungal diterpene synthases; the degenerate primers might not be suitable for *P. passeckerianus* and *S. araneosa*.

The second method attempted to locate the diterpene synthase genes through a related isoprenoid biosynthetic gene, the *ggs* genes (GGPP synthase genes). *P. passeckerianus* and *S. araneosa ggs* genes were used to try to locate the biosynthetic genes of pleuromutilin and sordarin by probing and screening a genomic DNA library and by genome walking methods. Three *ggs* gene fragments were identified from *P. passeckerianus* and two fragments were identified from *S. araneosa*. The BLAST searches strongly supported their *ggs* gene identities. About 10 kb of sequence extending from each fragment was obtained. However, no gene associated with the pleuromutilin or sordarin biosyntheses was found. This result and a recent review suggest that fungal diterpene gene clusters do not necessarily need a dedicated *ggs*. So although the *ggs* approach has worked for

several fungal clusters, it does not appear to work for all fungal diterpenes. The biosynthesis genes for the formation of pleuromutilin and sordarin appear to belong to the category that is not clustered with a dedicated *ggs* gene.

The third method utilized was Suppression Subtractive Hybridization (SSH), which is a technique used to find differentially expressed genes. This method was tried on the fungus *P. passeckerianus*. A subtracted library was formed from cDNA prepared from a 2^{nd} day culture that did not produce pleuromutilin and from cDNA from a 7^{th} day culture that produced pleuromutilin. The genes involved in the biosynthesis of pleuromutilin should be among the genes in the subtracted library. The PCR amplification results suggest that the SSH method has worked and identified a number of possible candidate genes. Several clones from the subtracted library have been sent for sequencing to find out whether or not genes associated with pleuromutilin biosynthesis are present.

In conclusion, three methods have been attempted to identify the diterpene synthase genes for the formation of pleuromutilin and sordarin. Although no diterpene biosynthetic genes have been found to date, these studies will provide a solid foundation for further studies in this area.

Biosynthetic Studies of Fungal Diterpene Antibiotics

by
Qingmei Yao

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented March 23, 2007

Commencement June 2007

Master of Science thesis of Qingmei Yao presented on March 23, 2007.

APPROVED:

_____

Major Professor, representing Pharmacy

_____

Dean of the College of Pharmacy

_____

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

_____

Qingmei Yao, Author

ACKNOWLEGEMENTS

I would like to thank my major Professor, Dr. Philip J. Proteau, for his support, guidance, and patience in this research project and dissertation. I would also like to thank Dr. T. Mark Zabriskie, Dr. Taifo Mahmud and Dr. Kerry L. McPhail for their support and guidance. They are always willing to help whenever I encounter difficulties. I would like to thank Dr. Patricia M. Flatt. She gave me so much help in constructing and screening the library and with many other experiments. I would like to thank my Graduate School Representative, Dr. Jerry R. Heidel, for his sincere advice.

I would like to thank Dr. Xihou Yin for the support and advice. He and his wife gave me a lot of help since I initially got here.

I would also like to thank all my colleagues, Dr. Takuya Ito, Dr. Jongtae Yang, Xiumei Wu and Hui Xu of the Taifo Mahmud group, Dr. Lin Zhang, David Blanchard of the Mark Zabriskie group, Dr. Michael Jackson, Rebecca Medina and Jennifer (Jean) Brown of Kerry McPhail group, Dr. Serge Fotso and Diana Ragland. I am lucky to have them as my friends and coworkers. There are laughter and happiness in the labs because of them.

Finally, I would like to give my special thanks to my family. They always stand by me no matter I am up or down. Life is meaningful because of the love.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF ABBREVIATIONS

ADP	Adenosine 5'-diphosphate
AIDS	Acquired immune deficiency syndrome
APCI	Atmospheric Pressure Chemical Ionization
ATCC	American Type Culture Collection
ATP	Adenosine 5'-triphosphate
cDNA	Complementary deoxyribonucleic acid
CDP	Copalyl diphosphate
CH_2Cl_2	Dichloromethane
CoA	Coenzyme A
d	Doublet
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DMAPP	Dimethylallyl diphosphate
DS	Diterpene synthase
DXP	1-Deoxy-D-xylulose-5-phosphate
DXR	1-Deoxy-D-xylulose-5-phosphate reductoisomerase
E. coli	*Escherichia coli*
EF	Elongation factor
ESI	Electrospray ionization
EtOAc	Ethyl acetate

LIST OF ABBREVIATIONS (Continued)

FAD	Flavin adenine dinucleotide
FPP	Farnesyl diphosphate
GAP	Glyceraldehyde 3-phosphate
GGPP	Geranylgeranyl diphosphate
GPP	Geranyl diphosphate
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
HMG-CoA	3-Hydroxy-3-methylglutaryl-CoA
HPLC	High performance liquid chromotography
IPP	Isopentenyl diphosphate
IPTG	Isopropyl-β-D-thiogalactopyranoside
J	Coupling constant
L3	Large 3
m	Multiplet
MEP	Methylerythritol phosphate
MIC	Minimum inhibition concentration
MOPS	3-[N-morpholino]propanesulfonic acid
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MVA	Mevalonic acid / Mevalonate

LIST OF ABBREVIATIONS (Continued)

NMR	Nuclear magnetic resonance
PAF	Platelet-activating factor
PCR	Polymerase chain reaction
PMA	Phosphomolybdic acid
PTC	Peptidyl transferase center
PRE	Pre-translocational
POST	Post-translocational
RF	Release factor
R_f	Retention factor
RNA	Ribonucleic acid
RP	Reversed phase
rpP0	Ribosomal protein P0
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
S	Singlet
SITL	Secondarily infected traumatic lesions
SSH	Suppression Subtractive Hybridization
TLC	Thin layer chromatography
TPP	Thiamine diphosphate
tRNA	Transfer ribonucleic acid

YM	Yeast Mold
YpSs	Yeast protein Soluble starch

Biosynthetic Studies of Fungal Diterpene Antibiotics

CHAPTER ONE

General Introduction

The kingdom of fungi

Living organisms are currently classified into five kingdoms: animals, plants, fungi, protista or protoctista, and Monera (bacteria and blue-green algae).[1] Among these five kingdoms, the kingdom of fungi often includes these major divisions: Basidiomycota, Ascomycota, Zygomycota and Chytridiomycota. Most of the interesting fungi in our neighborhoods, the common mushrooms, belong to the Basidiomycota. Basidiomycetes are also called 'Club fungi'. In this group, spores, namely basidiospores**,** are produced on club-like structures called basidia. Ascomycetes are another important group. They are also called 'Sac fungi'. Ascomycetes produce their spores, called ascospores, in sac-like structures called asci. Both Basidiomycetes and Ascomycetes may also reproduce asexually through conidia. The Chytridiomycetes produce zoospores that are capable of moving on their own through water by flagella. The Zygomycetes are called 'Conjugation fungi'. They are so called because the sexual spores, the zygospores, reproduce when two parental fungal hyphae 'conjugate', or meet each other. They can also reproduce asexually through plain spores. *Rhizopus stolonifer*, the black

bread mold, is a common species that belongs to this group. The fungus *Pleurotus passeckerianus* that will be discussed in Chapter 2 belongs to the Basidiomycetes, while the fungus *Sordaria araneosa* that will be discussed in Chapter 3 belongs to the Ascomycetes.

Isoprenoid biosynthetic pathways

Living organisms produce primary metabolites and secondary metabolites.[2] Primary metabolites include those substances such as carbohydrates, proteins, fats and nucleic acids that are essential for living organisms. Secondary metabolites are substances that are produced by certain organisms under specific conditions and do not play an essential role in the internal economy of the organisms that produce them. The pathways for the biosyntheses of primary metabolites are primarily the same except for minor variations for different organisms. In contrast, the biosynthetic pathways to the secondary metabolites are specific to certain organisms. An important biosynthetic pathway of both primary and secondary metabolites is the isoprenoid pathway. Isoprenoid compounds, also known as terpenes or terpenoids, are represented by sterols and the sidechain of ubiquinones, as well as the components of essential oils and important flavoring and fragrance agents in foods, beverages and cosmetics. Examples include limonene, menthol and camphor.[3, 4] Terpene compounds also exhibit biological activities and are used as pharmaceuticals. Artemisinin is used for treating malaria[5, 6] and the paclitaxel

(taxol®) is a strong anticancer drug (Figure 1.1).[7, 8] Ginkgolides, another diterpenoid family of compounds, have shown potent and selective antagonistic activity towards platelet-activating factor (PAF).[9] Other terpenoids, such as carotenoids, are beneficial to human health and some are essential for humans, for example, vitamins A and D.[10, 11]

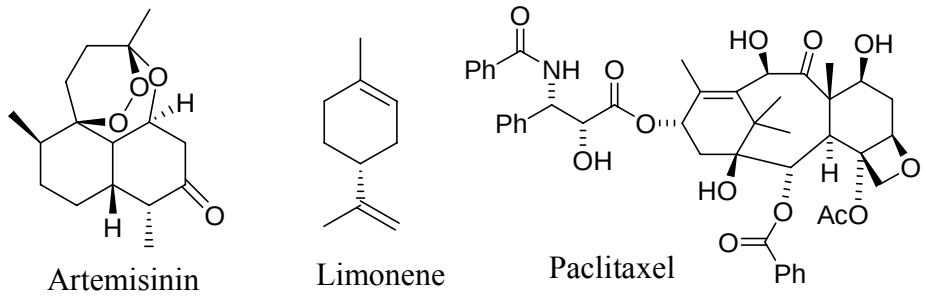


Figure 1.1. Examples of terpene compounds.

Terpenoids are classified according to the number of five carbon isoprene units in their structure: hemiterpenes (C_5, 1 isoprene unit), monoterpenes (C_{10}, 2 isoprene units), sesquiterpenes (C_{15}, 3 isoprene units), diterpenes (C_{20}, 4 isoprene units), triterpenes (C_{30}, 6 isoprene units), tetraterpenes (C_{40}, 8 isoprene units) and polyterpenes $(C_5)_n$ where 'n' may be 9–30,000.[12]

Monoterpenes, sesquiterpenes and diterpenes are typically formed through head (C1 of an allylic diphosphate, such as DMAPP) to tail (C4 of IPP) additions. Isopentenyl diphosphate (IPP) is added to its isomer dimethylallyl diphosphate (DMAPP) to synthesize geranyl diphosphate (GPP), which is the common precursor of monoterpenes. Condensation of geranyl diphosphate with additional

IPP units forms the successively larger prenyl diphosphates, farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), which are the common precursors of sesquiterpenes and diterpenes, respectively. The linear precursors undergo cyclizations, rearrangements and various modifications such as oxidations, reductions, isomerizations and hydrations to give rise to a variety of terpene compounds (Figure 1.2). The linear triterpene and tetraterpene precursors are formed through the head to head (C1 to C1) dimerizations of FPP and GGPP respectively.[3, 12]

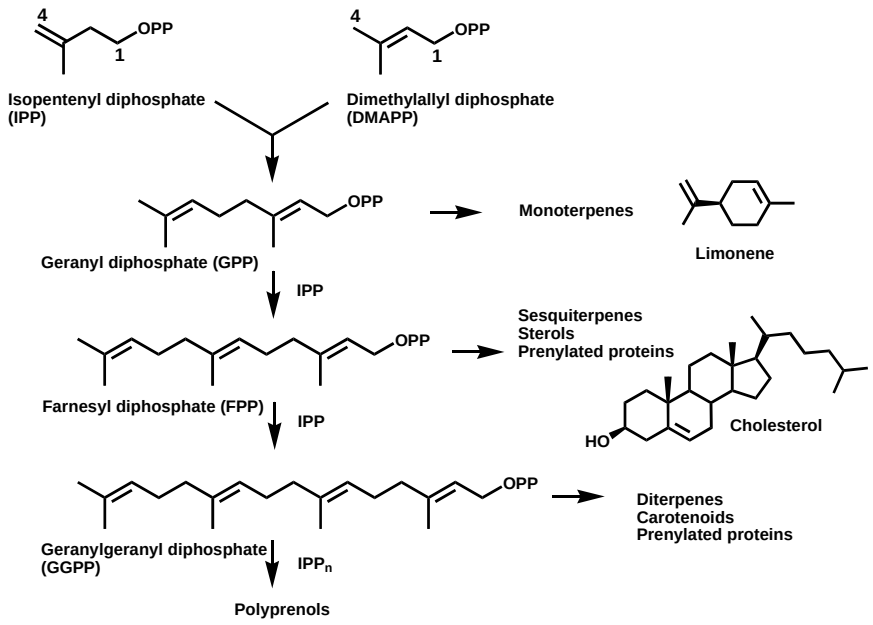


Figure 1.2. Outline of isoprenoid biosynthesis.

The diterpenoids represent a subset of isoprenoids and they arise from the C_{20} isoprenoid precursor, geranylgeranyl diphosphate (GGPP), which undergoes

cyclization reactions catalyzed by diterpene cyclases (DS) and then further modifications to give the final products. Fungal diterpenoids will be further discussed in this dissertation.

The biosynthesis of terpene compounds has been studied for more than half a century. Two independent terpene pathways, the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP) pathway, have been discovered.[13-20] Isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are the two universal precursors of terpene compounds.[21]

The mevalonate pathway was identified from early work with yeast and animal cells. Animals and fungi are found to only use this pathway.[22, 23] In this pathway, two molecules of acetyl-coenzyme A (acetyl-CoA) undergo a Claisen condensation to give acetoacetyl-CoA. Acetyl-CoA is added to the acetoacetyl-CoA to form HMG-CoA (3-hydroxy-3-methylglutary-CoA). This aldol reaction is catalyzed by HMG-CoA synthase. Mevalonic acid is then generated by HMG-CoA reductase. The six-carbon mevalonic acid is transformed into the five-carbon IPP in a series of reactions, namely the two-step phosphorylation of the primary acohol, the phosphorylation of the tertiary hydroxyl, followed by decarboxylation and loss of phosphate. Isopentenyl diphosphate is isomerized to generate DMAPP (Figure 1.3).[13-16]

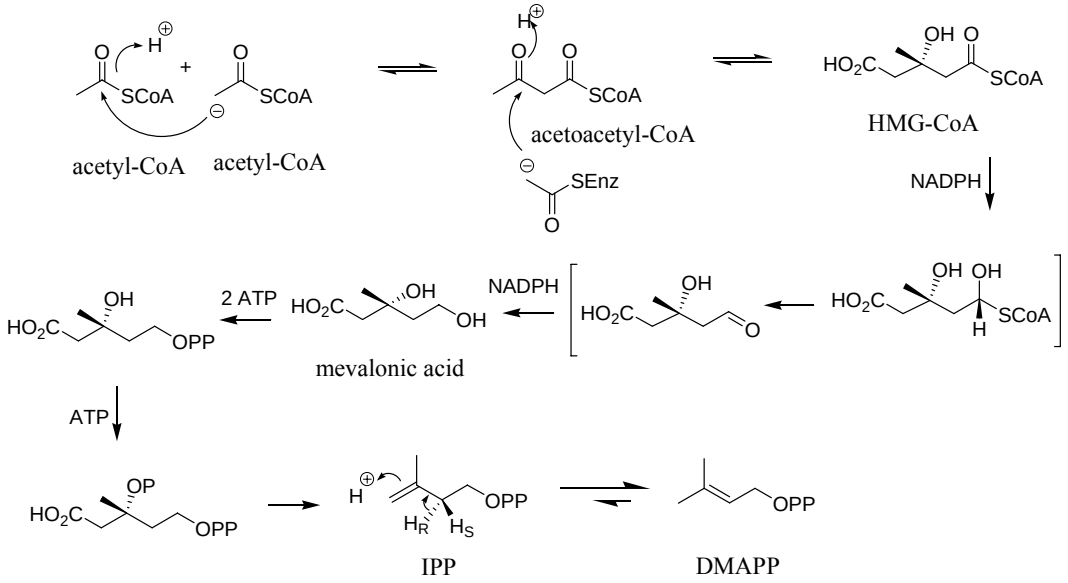


Figure 1.3. MVA pathway.[10]

Another terpene biosynthesis pathway, the methylerythritol phosphate (MEP) pathway or the mevalonate-independent pathway, was discovered in some bacteria and plants about a decade ago.[17-19] This pathway is widely used in bacteria[19, 24-26] although not in all. For examples, some eubacteria, such as *Staphylococcus* sp., were found to only use the MVA pathway.[23]

In the MEP pathway, 1-deoxy-D-xylulose 5-phosphate is formed from the glycolytic pathway intermediates pyruvic acid and glyceraldehyde 3-phosphate (GAP) (Figure 1.4). With the aid of thiamine diphosphate (TPP), pyruvic acid undergoes decarboxylation and nucleophilic addition to the aldehyde group of glyceraldehyde 3-phosphate to generate 1-deoxy-D-xylulose 5-phosphate (DXP). DXP undergoes rearrangement and subsequent reduction to form methylerythritol phosphate (MEP). This transformation is catalyzed by the enzyme 1-deoxy-D-

xylulose-5-phosphate reductoisomerase (DXR). Methylerythritol phosphate undergoes a series of reactions to give rise to IPP and DMAPP.[27-29]

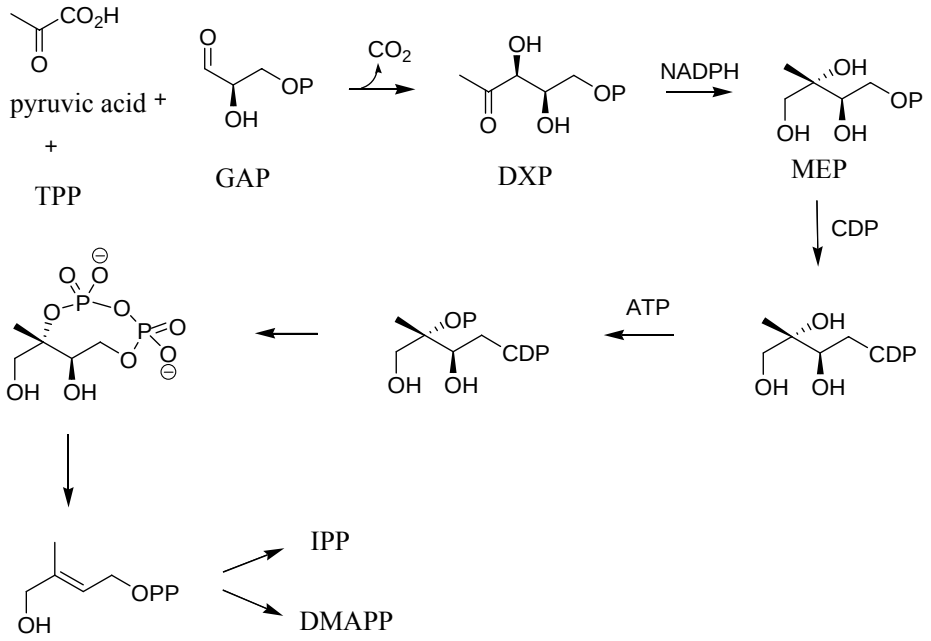
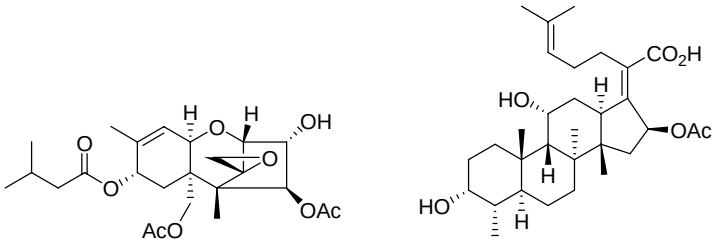


Figure 1.4. MEP pathway.[27]

Higher plants and red algae were found to use both the MVA pathway and MEP pathway,[24, 30] while green algae utilize only the MEP pathway. It was found that these two pathways occur in different compartments.[30, 31] The MVA pathway operates in the cytoplasm and gives rise to cytosolic isoprenoids such as sterols. The MEP pathway is used in chloroplasts and produces plastidic isoprenoids such as carotenoids and the phytyl sidechain of chlorophyll.

Fungal isoprenoids

Isoprenoids are among the important metabolites produced by fungi although secondary terpenoid metabolites are most abundant in plants.[32] Among the well-known isoprenoids produced by fungi are the trichothecenes, a group of sesquiterpene toxins produced by several genera of fungi, such as *Fusarium* and *Trichothecium*.[33, 34] Fewer fungal terpenoids have been developed in the medical field compared to those produced by plants, such as the antitumor agent taxol and the antimalarial agent artemisinin.[32] Fusidic acid, which is produced by the fungus *Fusidium coccineum*, is one of the known antibacterial agents marketed in Canada and Europe. Fusidic acid inhibits protein biosynthesis by preventing the translocation of the elongation factor G (EF-G) from the ribosome.[35] The structures of the T-2 toxin, which belongs to trichothecene family, and fusidic acid are shown in Figure 1.5. Pleuromutilin, an antibacterial compound, and sordarin, an antifungal compound, are fungal isoprenoids and will be discussed in Chapter 2 and Chapter 3, respectively.



T-2 toxin Fusidic acid

Figure 1.5. The structures of T-2 toxin and fusidic acid.

Prenyltransferases

The formation of geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP), the substrates for terpene cyclases, is catalyzed by enzymes called prenyltransferases. They are GPP synthase, FPP synthase and GGPP synthase, respectively.

The formation of GPP, FPP and GGPP is an electrophilic alkylation reaction initiated by the formation of a highly reactive carbocation. The carbocation is generated at C1 of DMAPP due to the departure of the allylic diphosphate leaving group. A second carbocation, a tertiary carbocation, is generated at C3 of the new GPP due to the alkylation of the C3-C4 double bond of IPP by the allylic carbocation, followed by the stereospecific elimination of a C2 H_R proton to generate a new allylic diphosphate, in which the new double bond is *trans* (*E*).[36-38] The mechanism is shown in Figure 1.6, using the formation of GPP as an example.

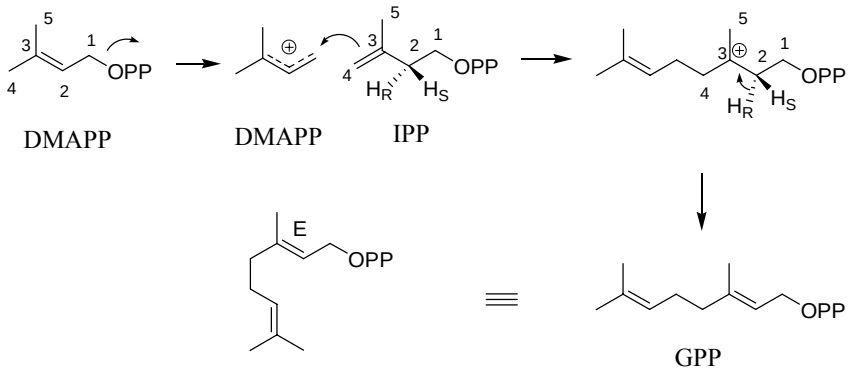


Figure 1.6. Alkylation mechanism for the formation of GPP.

The amino acid alignment for 38 FPP synthases and GGPP synthases, revealed five highly conserved motifs (Table 1.1).[39, 40] There are two aspartate rich DDXXD motifs which are essential for the prenyltransferase activity.

Table 1.1. The conserved motifs of 38 prenyltransferases. Residues are numbered according to their positions in avian FPP synthase. The conserved residues shown in upper case are present in ≥ 96% sequences. The conserved residues shown in lower case are present in ≥ 82% sequences.[40] Two aspartate-rich domains are highlighted in red.

Regions	Sequences
I	g k - - r 70 71 74
II	E - - Q/H - - - L - - D D - - D - - - - R R G 107 110 114 117 118 121 126 127 128
III	G Q/E - - d 184 185 188
IV	K t 214 215
V	G - - F/Y Q - - D D - - d - - - - - - - - -g k - - - D - - - - k 250 253 254 257 258 261 270 271 275 280

Extensive work has been done to elucidate the structures and mechanisms of prenyltransferases. The X-ray structure of avian farnesyl diphosphate synthase complexed with the substrate geranyl diphosphate (GPP) was obtained and helps to explain the mechanism of catalysis for the prenyltransferases.[41] Farnesyl diphosphate synthase catalyzes the formation of GPP from DMAPP and that of FPP from GPP. The X-ray structure shows that the enzyme catalytic site consists of a large central cavity. All five conserved motifs are located in this cavity. The two aspartate rich DDXXD motifs are located on opposite walls of the cavity.

Magnesium ions are bound to these two motifs. The aspartate residues in these two motifs make coordination interactions with the magnesium ions and facilitate the departure of the GPP diphosphate leaving group to generate the carbocation. Figure 1.7 shows that one magnesium ion forms a bridge between the diphosphate moiety of the allylic diphosphate (take DMAPP as an example) and the carboxylate oxygens in the sidechains of the two aspartates, Asp117 and Asp121, in the first DDXXD motif. This figure also shows that another magnesium ion forms a bridge between the diphosphate moiety and Ser123. The carboxylate sidechains of Asp117 and Asp121 interact weakly with this second magnesium ion.[40] The KT conserved motif may be involved in activation of the diphosphate leaving group, stabilization of the carbocation intermediates and deprotonation of the C2 of IPP.[41]

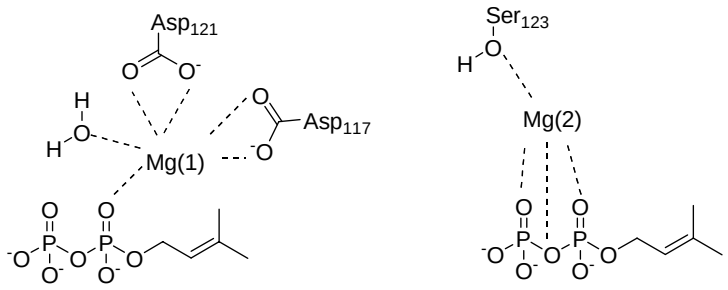


Figure 1.7. Magnesium ions coordinated with Asp117, Asp121 and Ser123.[40]

The formation of diterpenoids relies on GGPP synthase, which represent a subset of prenyltransferases. There are two classes of GGPP synthases. In bacteria and plants, the GGPP synthases that catalyze the GGPP formation for carotenoid

biosynthesis use DMAPP and three IPP molecules,[42] whereas in fungi and mammals, GGPP synthases use FPP as the allylic substrate for extension with one IPP molecule.[43, 44]

Terpene cyclases

More than 30,000 terpene compounds have been reported to date[22] and there are many more potential compounds of various structures to be discovered. They all derive from only a handful of linear isoprenoid precusors. Terpenoid cyclases catalyze the cyclizations that lead to the numerous parental ring structures. From this point of view, terpenoid cyclases do a wonderful job and thus stimulate people's interest in searching for their mechanism.

The terpene cyclase reaction is initiated by the formation of a highly reactive carbocation.[45-51] Terpene cyclases are divided into two groups according to the mechanisms of carbocation formation: ionization-dependent terpenoid cyclases and protonation-dependent terpenoid cyclases.

Ionization-dependent terpenoid cyclases include monoterpene, sesquiterpene, and some diterpene cyclases. The catalytic mechanism of this type of terpenoid cyclase is very similar to that of the prenyltransferases and the crystal structure of avian farnesyltransferase was also used to explain the mechanism of this type of terpene cyclase. It is proposed that these two enzymes derived from a common ancestor in the early evolution of terpene biosynthesis.[52] A magnesium ion promotes the

departure of a pyrophosphate (PPi) leaving group from the cyclase substrate, an allylic diphosphate molecule, to yield the carbocation,[36-38] followed by alkylation of the nucleophilic group.[53, 54] This kind of cyclase contains DDXXD/E aspartate rich motif, which is similar to that found in prenyltransferases, and a NSE/DTE non-aspartate rich motif. Both motifs together bind magnesium ions.[55] Limonene synthase belongs to this type of terpene cyclase.[56]

The second type of terpene cyclase is the protonation-dependent terpenoid cyclase. Some diterpene cyclases and triterpene cyclases are in this group. The initial carbocation is formed by the protonation of an epoxide ring or a carbon-carbon double bond. These cyclases contain a single DXDD conserved motif. Carbocation formation is initiated by the central aspartate residue as the proton donor.[51] Oxidosqualene cyclase, also known as lanosterol synthase, belongs to this type of terpene cyclase.[57, 58]

Although X-ray crystal structures have already been obtained for ionization-dependent monoterpene and sesquiterpene cyclases and protonation-dependent triterpene cyclases, no crystal structure of a diterpene cyclase has been reported. It is believed, however, that diterpene cyclases will share a common mechanism with the other terpene synthases.

Some diterpene cyclases are bifunctional cyclases. A single enzyme may contain separate ionization-dependent and proton-dependent active sites and catalyzes both the ionization-initiated and proton-initiated reactions.[59]

Methods for identification of fungal diterpene synthase genes

Overview

Information regarding biosynthetic gene clusters for terpenoids in fungi is limited. For fungal diterpenoids, only seven biosynthesis gene clusters have been described until now: *ent*-kaurene synthesis in *Phaeosphaeria* sp.,[60] *ent*-kaurene/ gibberellin synthesis in *Gibberela fujikuroi* (*Fusarium fujikuroi*),[61, 62] aphidicolan-16*β*-ol[63] and aphidicolin synthesis[64] in *Phoma betae*, paxilline synthesis in *Penicillium paxilli*,[65] alflatrem synthesis in *Aspergillus flavus*[66] and lolitrem biosynthesis in *Epichloe festucae* and *Neotyphodium lolii*.[67] The structures of the indicated terpenes are shown in Figure 1.8.

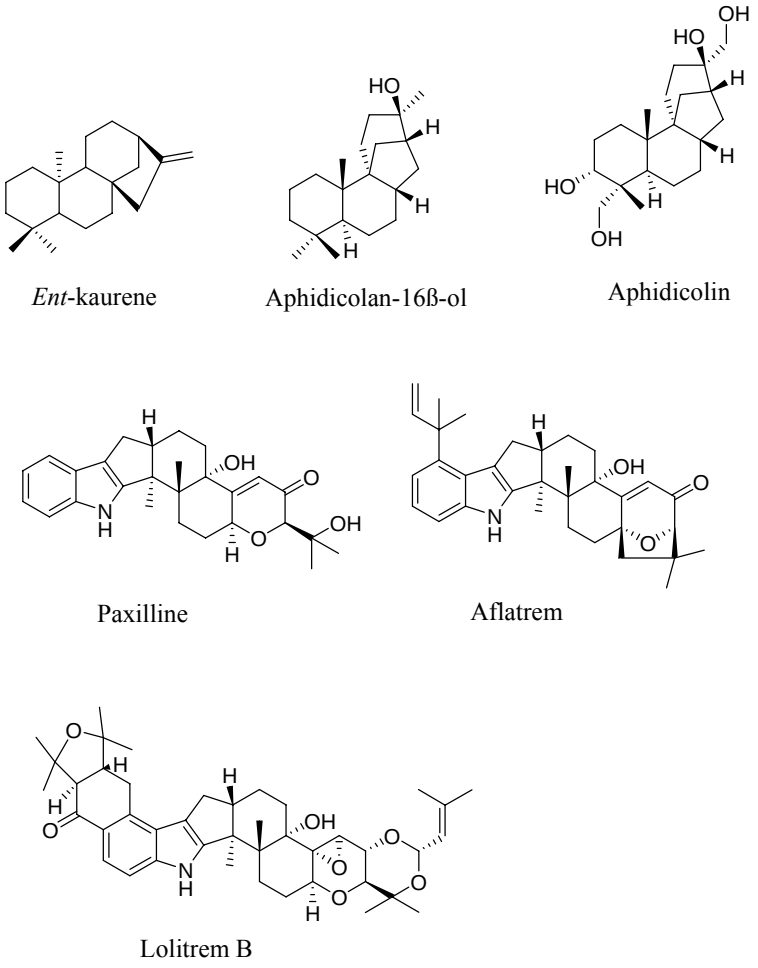


Figure 1.8. Structures of the compounds from identified fungal terpenoid biosynthetic gene clusters.

Two different approaches have been used to identify the diterpene synthase (DS) genes and/or the gene clusters: 1) Direct identification of the diterpene synthase gene using degenerate PCR primers based on the conserved motifs of plant and/or fungal DS synthases, 2) Indirect identification of the diterpene gene clusters via the GGPP synthase (*ggs*) gene.

Direct identification of the DS gene

The first fungal diterpene synthase gene, the *ent*-kaurene synthase gene, was isolated from *Phaeosphaeria* sp. L487 in 1997.[60] *Ent*-kaurene is the precursor of gibberellin which is the diterpene phytohormone found in both plants and fungi.[68] In plants, the formation of *ent*-kaurene is a two-step reaction.[69-73] Geranylgeranyl diphosphate (GGPP) is converted to copalyl diphosphate (CDP) by CDP synthase (CPS), and then copalyl diphosphate is converted to *ent*-kaurene by *ent*-kaurene synthase (KS). In order to identify the synthase gene(s) responsible for the formation of *ent*-kaurene in the fungus *Phaeosphaeria* sp. L487, degenerate primers were designed from the conserved motifs of plant CDP synthases.[71-73] Messenger RNA was isolated, cDNA was synthesized, PCR was conducted and a 873 bp fragment was obtained. The protein sequence of this fragment showed significant similarity to those of CPS,[69-71] KS,[72] and abietadiene synthase,[59] a bifunctional plant diterpene synthase. This fragment was used as a probe to screen a cDNA library and the whole open reading frame was obtained. The recombinant enzyme converted the linear geranylgeranyl diphosphate to *ent*-kaurene. This indicates that this fungal *ent*-kaurene synthase is a bifunctional enzyme and catalyzes the cyclization reaction that is catalyzed by two separate enzymes and fulfilled in two steps in plants. This bifunctional enzyme was designated as fungal-type copalyl diphosphate synthase (FCPS)/*ent*-kaurene synthase (KS) (Figure 1.9).

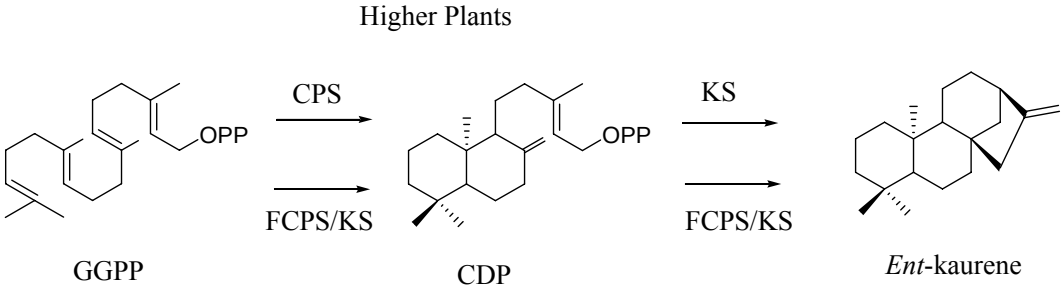


Figure 1.9. Pathway for *ent*-kaurene biosynthesis from GGPP via CDP in the plants and fungi. CPS and KS in plants catalyzes the cyclization of GGDP to CDP and that of CDP to *ent*-kaurene, respectively. Fungal bifunctional FCPS/KS catalyzes the two-step cyclization of GGDP to *ent*-kaurene via CDP.[60]

Another bifunctional copalyl diphosphate (CDP) synthase/*ent*-kaurene synthase was identified from the fungus *Gibberella fujikuroi* in 1998.[61] The same method as that used for the isolation of the FCPS/*ent*-kaurene synthase in *Phaeosphaeria* sp. was used except that the degenerate primers were slightly modified according to the sequence of *Phaeosphaeria* FCPS. In fact, the same degenerate primers as those used for the isolation of the *Phaeosphaeria* sp. FCPS were successfully used to identify the CDP synthase/*ent*-kaurene synthase in *Gibberella fujikuroi* by another research group and the paper was published in 2000.[74] This research group showed that the recombinant protein of the enzyme converted the geranylgeranyl diphosphate to copalyl diphosphate and *ent*-kaurene. This bifunctional enzyme shares 63% identity with the CDP synthase/*ent*-kaurene synthase gene from *Phaeosphaeria* sp. L487.

The gibberellin biosynthesis pathway was revealed from *Gibberella fujikuroi* after the bifunctional CDP synthase/ *ent*-kaurene synthase was identified.[62] It was very

interesting and encouraging to find out that the genes related to gibberellin biosynthesis are clustered together including three cytochrome P450 monooxygenase genes, a geranylgeranyl diphosphate synthase gene and the CDP synthase/*ent*-kaurene synthase gene that was already identified previously. Another surprise came from the *ggs* gene (*ggs2*) in this cluster because another *ggs* gene (*ggs1*) had been identified which was believed to be involved in the general isoprenoid pathway.[75] This result suggested that the *ggs2* gene was specifically involved in the gibberellin biosynthesis.

The primers that were used for the isolation of the FCPS/*ent*-kaurene synthase in *Phaeosphaeria* sp. were utilized again in 2001 to identify another fungal diterpene synthase, aphidicolan-16*β*-ol synthase.[63] This enzyme is a key enzyme in cyclization of the linear geranylgeranyl diphosphate to the diterpene compound aphidicolin in the fungus *Phoma betae* PS-13 (Figure 1.10). Reverse transcription-polymerase chain reaction (RT-PCR) was conducted using *P. betae* mRNA and the aphidicolan-16*β*-ol synthase gene was identified which shows 36-37% identity to fungal *ent*-kaurene synthases. The recombinant protein of this enzyme converted the linear geranylgeranyl diphosphate to aphidicolan-16*β*-ol.

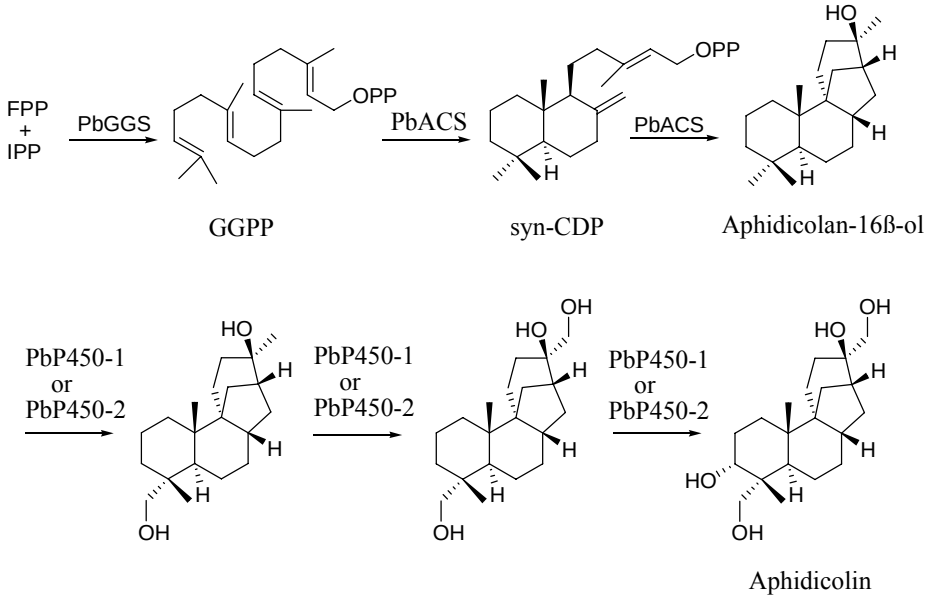


Figure 1.10. Biosynthetic scheme for aphidicolin formation from GGPP.[63]

Identification of the diterpene gene clusters via the *ggs* gene

Although three fungal diterpene synthases were identified by the homology-based method,[60, 63, 74] the fungal diterpene synthase genes in general have relatively low identity.[60, 74] All three genes encode diterpene synthases catalyzing cyclization via copalyl diphosphate (CDP) and show high identity with each other. Therefore, the homology-based method may not be a universal method for identifying other fungal diterpene synthases, especially synthases that do not utilize CDP.

Another method has been proposed that may be more universal.[64-67] The occurrence of a *ggs* gene on the gene clusters is a common feature in the biosynthesis of such diterpenoids as terpentecin (bacteria)[76], gibberellin[62] and paxilline[65], although examples of terpene biosynthesis gene clusters are limited.

Paxilline belongs to the indole-diterpene compounds which contain a diterpene moiety derived from geranylgeranyl diphosphate (GGPP) and an indole moiety derived from tryptophan (Figure 1.11).[77] Indole-diterpene compounds are commonly found in fungi of the genera of *Penicillium*, *Aspergillus* and *Claviceps*.[78] Aflatrem and the lolitrems also belong to this family.

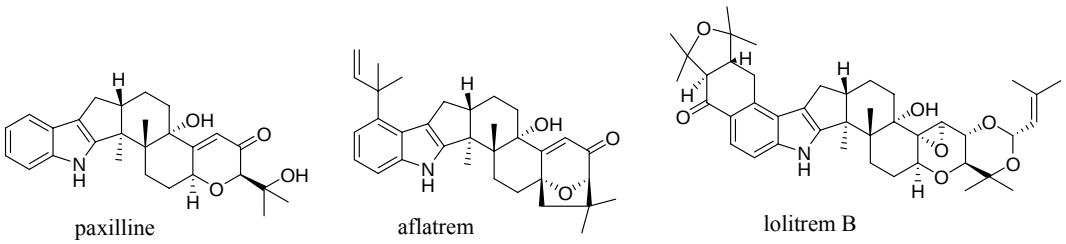


Figure 1.11. Examples of indole-diterpene compounds.

The paxilline biosynthesis pathway was identified in 2001 from the fungus *Penicillium paxilli*.[65] The locus that is involved in paxilline biosynthesis was identified on *P. paxilli* chromosome. Four mutants that did not produce paxilline were obtained by plasmid-induced chromosome deletions. Four deletions were found to be at the same locus. The paxilline biosynthesis gene cluster was revealed by DNA sequence analysis. A *ggs* gene (*paxG*) was found in the gene cluster. This *paxG* gene and two other genes, *paxM* (FAD-dependent Monooxygenase) and *paxC* (prenyl transferase) were found to be coordinately expressed and essential for paxilline biosynthesis, but not required by the fungus for growth. This suggested that there might be another GGPP synthase(s) that is/are involved in

primary metabolism. Using two degenerate primers designed from the conserved motifs of *paxG* and other fungal *ggs* genes, another *ggs* gene, *ggs1*, was identified. This gene was predicted to be involved in fungal primary metabolism. Both PaxG and *ggs*1 contain the five conserved domains found in prenyltransferases.[39, 40] These two GGPP synthases are believed to be involved in different biological functions. The *ggs1* gene was found to be constitutive expressed in low level throughout the growth cycle of *Penicillium paxilli*, while the expression of *paxG* increased dramatically just before the onset of paxilline biosynthesis.

Because GGPP is an essential precursor of carotenoids, prenylated quinones and geranylgeranylated proteins,[79] it is reasonable that two different types of *ggs* genes (*ggs*1 and *ggs*2 in *G. fujikuroi* and *ggs*1 and paxG in *P. paxilli*) are found in fungi which produce diterpenes. Identification of PbGGS in the aphidicolin gene cluster further supports the hypothesis that diterpene biosynthesis gene clusters contain a *ggs* gene.[64] This suggests that the diterpene gene cluster could be securely cloned by identification of more than one distinct *ggs* gene.[64] Sequencing the DNA flanking both *ggs* genes should lead to the cluster. This hypothesis has been used to identify gene clusters involving fungal diterpene biosynthesis since 2004.[64, 66, 67] This approach was essential for finding the gene cluster for the biosynthesis of another fungal indole-diterpene compound, aflatrem, from the fungus *Aspergillus flavus*.[66] According to the conserved domains of fungal geranylgeranyl diphosphate synthases, three degenerate primers were designed which represent the GGPP synthase conserved motifs of HRGQGMD, QIRDDYMN and

EDLTEGKF, respectively. Nested PCR was employed using *A. flavus* genomic DNA as template. Two fragments representing GGPP synthase genes, *atmG* and *ggsA* were identified which are 330 and 270 bp, respectively, from the first round PCR and 262 and 218 bp, respectively, from the second round PCR. These two fragments were used to probe a genomic DNA library and the gene cluster for aflatrem biosynthesis was identified from the GGPP synthase genes, *atmG*. Adjacent to *atmG* are two other genes, *atmC* and *atmM*. These three genes share 64 to 70% similarity to *paxG*, *paxM* and *paxC* respectively from *Penicillium paxilli*, confirming their association with indole diterpene biosynthesis.

In 2005, the same research group that worked with the aflatrem cluster, identified a cluster of three genes for another fungal indole-diterpene compound, lolitrem, that is closely related to paxilline.[67] The same three degenerate primers were used and two GGPP synthase genes, *ltmG* and *ggsA*, were identified. The *ltmG* gene was found to be involved in lolitrem biosynthesis. Adjacent to *ltmG* are *ltmM* and *ltmK* which are FAD-dependent monooxygenase and cytochrome P450 monooxygenase respectively. The success of this research further supported that diterpene gene cluster would be securely cloned by identification of more than one *ggs* gene. Based on these observations, it appears that when there are two distinct *ggs* genes in a fungal genome, one is specifically dedicated to the terpene pathway and this gene can be used to identify the other genes involved in the pathway.

It is already recognized that fungal secondary metabolic genes are organized in clusters instead of being scattered throughout the whole genome.[62, 80, 81] Therefore,

it becomes possible that identification of a single secondary metabolite biosynthetic gene would lead to finding the whole gene cluster. This method was successful in identifying the gene cluster for the aphidicolin biosynthesis by the method of genome walking.[64]

In the aphidicolin biosynthesis pathway, linear geranylgeranyl diphosphate is converted to the tetracyclic diterpene compound aphidicolan-16*β*-ol by aphidicolan-16*β*-ol synthase. Aphidicolan-16*β*-ol undergoes further oxidation reactions to generate aphidicolin (Figure 1.10). Because the aphidicolan-16*β*-ol synthase gene was already identified,[63] it was hypothesized that the whole gene cluster could be revealed by genome walking, if the genes were clustered.[64]

Compared to constructing a genomic DNA library, the PCR-based genome walking method is relatively time-saving and less laborious.[82] But it is not efficient for identifying very long sequences. Because the aphidicolin gene cluster was expected to be relatively small (less than 20 kb), it was a good candidate for genome walking. Repeated genome walking finally gave the 15.6 kb sequence of a full gene cluster. The cluster contains six open reading frames including *PbGGS*, *ACS*, *PbP450-1*, *PbP450-2*, *PbTP* and *PbTF* which encode a geranylgeranyl diphosphate synthase, the diterpene synthase (aphidicolan-16*β*-ol synthase), two P450s, the ABC transporter and a transcription factor respectively.

Summary

The isoprenoid pathway is one of the major biosynthetic pathways of secondary metabolites, which include the diterpenoids, C20 isoprenoids. The common precursor to diterpenes is geranylgeranyl diphosphate (GGPP). The amazing array of multicyclic and highly functionalized bioactive natural products formed from this linear precursor has stimulated our interest in these compounds. The long term goal of this research program is to study the diverse diterpene synthase enzymes which provide the backbone structures for a wide variety of biologically active molecules. A greater understanding of these synthase enzymes may allow us to eventually manipulate the corresponding genes, creating new scaffolds for bioactive molecules.

In this stage of the project we focused on two specific fungal diterpenoids: pleuromutilin and sordarin. Pleuromutilin and sordarin are both derived from GGPP, but GGPP undergoes different cyclization pathways to give the very different core structures.

Both of the approaches mentioned earlier were attempted to locate the pleuromutilin and sordarin biosynthetic genes: 1) Direct identification of the diterpene synthase gene using degenerate PCR primers based on the conserved motifs of plant and/or fungal DS synthases. 2) Indirect identification of the diterpene gene clusters via the GGPP synthase (*ggs*) gene. The details of these studies will be presented in Chapters 2 and 3.

References

1. Margulis, L.; Schwartz, K. V. "Five kingdoms: an illustrated guide to the phyla of life on earth," **1998**, W.H. Freeman, New York.
2. Mann, J. "Secondary metabolism," **1987**, Clarendon Press. New York: Oxford University Press.
3. Singh, N.; Luthra, R.; Sangwan, R. S.; Thakur, R. S. "Metabolism of monoterpenoids in aromatic plants," *Curr. Res. Med. Arom. Plants* **1989,** *11*, 174-97.
4. Mahmoud, S. S.; Croteau, R. "Strategies for transgenic manipulation of monoterpene biosynthesis in plants," *Trends Plant Sci.* **2002,** *7*.
5. Mann, J. "Chemical aspects of biosynthesis," **1994**, Oxford University Press.
6. Wang, D. Y.; Wu, Y.; Wu, Y. L.; Li, Y.; Shan, F. "Synthesis, iron(II)-induced cleavage and in vivo antimalarial efficacy of 10-(2-hydroxy-1-naphthyl)-deoxoqinghaosu," *J. Chem. Soc. Perkin Trans.* **1999,** *1*, 1827-31.
7. Croteau, R.; Ketchum, R. E. B.; Long, R. M.; Kaspera, R.; Wildung, M. R. "Taxol biosynthesis and molecular genetics," *Phytochem. Rev.* **2006,** *5*, 75-97.
8. Alexander, R. W. "Teasing apart the taxol pathway," *Trends Biochem. Sci.* **2001,** *26*, 152.
9. Houghton, P. "Herbal products: Ginkgo," *Pharm. J.* **1994,** *253*, 122-3.
10. Dewick, P. M. "Medicinal natural products: a biosynthetic approach," **2001**, John Wiley & Sons Ltd.
11. Sacchettini, J. C.; Poulter, C. D. "Creating isoprenoid diversity," *Science* **1997,** *277*, 1788-9.
12. McGarvey, D. J.; Croteau, R. "Terpenoid metabolism," *Plant Cell* **1995,** *7*, 1015-26.
13. Bloch, K.; Chaykin, S.; Phillps, J. W.; deWaard, A. "Mevalonic acid pyrophosphate and isopentenylpyrophosphate," *J. Biol. Chem.* **1959,** *234*, 2595-604.
14. Witting, L. A.; Porter, J. W. "Intermediates in the conversion of melaonic acid to squalene by a rat liver enzyme system," *J. Biol. Chem.* **1959,** *234*, 2841-6.
15. Bach, T. J. "Some new aspects of isoprenoid biosynthesis in plants - a review," *Lipids* **1995,** *30*, 191-202.
16. Bloch, K. "Sterol molecule: structure, biosynthesis, and function," *Steroids* **1992,** *57*, 378-82.
17. Flesch, G.; Rohmer, M. "Prokaryotic hopanoids: the biosynthesis of the bacteriohopane skeleton. Formation of isoprenic units from two distinct acetate pools and a novel type of carbon/carbon linkage between a triterpene and D-ribose," *Eur. J. Biochem.* **1988,** *175*, 405-11.

18. Rohmer, M.; Sutter, B.; Sahm, H. "Biosynthesis of the side-chain of bacteriohopanetetrol and of a carbocyclic pseudopentose from 13C-labelled glucose in *Zymomonas mobilis*," *J. Chem. Soc., Chem. Commun.* **1989,** *19*, 1471-2.
19. Eisenreich, W.; Schwarz, M.; Cartayade, A.; Arigoni, D.; Zenk, M. H.; Bacher, A. "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms," *Chem. Biol.* **1998,** *5*, R221-R33.
20. Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. "Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate," *Biochem. J.* **1993,** *295*, 517-24.
21. Ruzicka, L. "The isoprene rule and the biogenesis of terpenic compounds," *Experientia* **1953,** *9*, 357-67.
22. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R. "Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes," *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 13172-7.
23. Kuzuyama, T.; Seto, H. "Diversity of the biosynthesis of the isoprene units," *Nat. Prod. Rep.* **2003,** *20*, 171-83.
24. Rohmer, M. "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants," *Nat. Prod. Rep.* **1999,** *16*, 565-74.
25. Eisenreich, W.; Rohdich, F.; Bacher, A. "Deoxyxyluose phosphate pathway to terpenoids," *Trends Plant Sci.* **2001,** *6*, 78-84.
26. Rohmer, M. "Isoprenoid biosynthesis via the mevalonate-dependent route," *Prog. Drug Res.* **1998,** *50*, 136-54.
27. Eisenreich, W.; Bacher, A.; Arigoni, D.; Rohdich, F. "Biosynthesis of isoprenoids via the non-mevalonate pathway," *Cell Mol. Life Sci.* **2004,** *61*, 1401-26.
28. Adam, P.; Hecht, S.; Eisenreich, W.; Kaiser, J.; Grâwert, T.; Arigoni, D. "Biosynthesis of terpenes. Studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase," *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 12108-13.
29. Altincicek, B.; Duin, E. C.; Reichenberg, A.; Hedderich, R.; Kollas, A. K.; Hintz, M. "LytB protein catalyses the terminal step of the 2-C-methyl-D-ehrtol-4-phosphate pathway of isoprenoid biosynthesis," *FEBS Lett.* **2002,** *532*, 437-40.
30. Lichtenthaler, H. K. "The plants' 1-deoxyD-xylulose-5-phosphate pathway for biosynthesis of isoprenoids," *Fett-Lipid* **1998,** *100*, 128-38.
31. Kleinig, H. "The role of plastids in isoprenoid biosynthesis. Annu. Rev. Plant Physiol.," *Plant Mol. Biol.* **1989,** *40*, 39-59.
32. Inouye, S.; Abe, S.; Yamaguchi, H. "Fungal terpenoid antibiotics and enzyme inhibitors," *Mycology Series* **2004,** *Handbook of Fungal Biotechnology (2nd Edition). 20*, 379-99.
33. Grove, J. F. "Microcyclic trichothecenes," *Nat. Prod. Rep.* **1993,** *10*, 429-48.

34. Grove, J. F. "Non-macrocyclic trichothecenes, part 2," *Prog. Chem. Org. Nat. Prod.* **1996,** *69*, 1-70.
35. Verbist, L. "The antimicrobial activity of fusidic acid," *J. Antimicrob. Chemother.* **1990,** *25 (Suppl B)*, 1-5.
36. Poulter, C. D.; Satterwhite, D. M. "Mechanism of the prenyl transfer reaction. Studies with (E)- and (Z)-3-trifluoromethyl-2-buten-1-yl pyrophosphate," *Biochemistry* **1977,** *16*, 5470-8.
37. Poulter, C. D.; Argyle, J. C.; Mash, E. A. "Farnesyl pyrophosphate synthetase. Mechanistic studies of the 1'-4 coupling reaction with 2-fluorogeranyl pyrophosphate," *J. Biol. Chem.* **1978,** *253*, 7227-33.
38. Poulter, C. D.; Wiggins, P. L.; Le, A. T. "Farnesylpyrophosphate synthetase. A stepwise mechanism for the 1'-4 condensation reaction," *J. Am. Chem. Soc.* **1981,** *103*, 3926-7.
39. Chen, A. P.; Kroon, P. A.; Poulter, C. D. "Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure," *Protein Sci.* **1994,** *3*, 600-7.
40. Kellogg, B. A.; Poulter, C. D. "Chain elongation in the isoprenoid biosynthetic pathway," *Curr. Opin. Chem. Biol.* **1997,** *1*, 570-8.
41. Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D. "Regulation of product chain length by isoprenyl diphosphate synthases," *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 15018-23.
42. Dogbo, A.; Camara, B. "Purification of isopentenyl pyrophosphate isomerase and geranyl geranyl pyrophosphate synthase from Capsicum chromoplasts by affinity chromatography," *Biochim. Biophys. Acta* **1967,** *920*, 140-8.
43. Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S. "BTSl encodes a geranyl geranyl diphosphate synthase in *Saccharomyces cerevisiae*," *J. Biol. Chem.* **1995,** *270*, 21793-9.
44. Sagami, H.; Morita, Y.; Ogura, K. "Purification and properties of geranylgeranyl-diphosphate synthase from bovine brain," *J. Biol. Chem.* **1994,** *269*, 20561-6.
45. Cane, D. E. "Isoprenoid biosynthesis. Stereochemistry of the cyclization of allylic pyrophosphates," *Acc. Chem. Res.* **1985,** *18*, 220-6.
46. Cane, D. E. "Enzymic formation of sesquiterpenes," *Chem. Rev.* **1990,** *90*, 1089-103.
47. Abe, I.; Rohmer, M.; Prestwich, G. D. "Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes," *Chem. Rev.* **1993,** *93*, 2189-206.
48. Lesburg, C. A.; Caruthers, J. M.; Paschall, C. M.; Christianson, D. W. "Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism," *Curr. Opin. Struct. Biol.* **1998,** *8*, 695-703.
49. Wendt, K. U.; Schulz, G. E. "Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes," *Structure* **1998,** *6*, 127-33.

50. Davis, E. M.; Croteau, R. "Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes," *Top. Curr. Chem.* **2000,** *209*, 53-95.
51. Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. "Enzyme mechanisms for polycyclic triterpene formation," *Angew. Chem., Int. Ed.* **2000,** *39*, 2812-33.
52. Reardon, D.; Farber, G. K. "The structure and evolution of alpha/beta barrel proteins," *FASEB J.* **1995,** *9*, 497-503.
53. Davisson, V. J.; Neal, T. R.; Poulter, C. D. "Farnesyl-diphosphate synthase. Catalysis of an intramolecular prenyl transfer with bisubstrate analogs," *J. Am. Chem. Soc.* **1993,** *115*, 1235-45.
54. Davisson, V. J.; Poulter, C. D. "Farnesyl-diphosphate synthase. Interplay between substrate topology, stereochemistry, and regiochemistry in electrophilic alkylations," *J. Am. Chem. Soc.* **1993,** *115*, 1245-60.
55. Cane, D. E.; Kang, I. "Aristolochene synthase: purification, molecular cloning, high-level expression in Escherichia coli, and characterization of the *Aspergillus terreus* cyclase," *Arch. Biochem. Biophys.* **2000,** *376*, 354-64.
56. Williams, D. C.; McGarvey, D. J.; Katahira, E. J.; Croteau, R. "Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair," *Biochemistry* **1998,** *37*, 12213-20.
57. Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. "Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase," **2004,** *432*, 118-22.
58. Pandit, J.; Danley, D. E.; Schulte, G. K.; Mazzalupo, S.; Pauly, T. A.; Hayward, C. M.; Hamanaka, E. S.; Thompson, J. F.; Harwood, H. J., Jr. "Crystal structure of human squalene synthase. a key enzyme in cholesterol biosynthesis " *J. Biol. Chem.* **2000,** *275*, 30610-7.
59. Vogel, B. S.; Wildung, M. R.; Vogel, G.; R., C. "Abietadiene Synthase from Grand Fir (*Abies grandis*). cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis," *J. Biol. Chem.* **1996,** *271*, 23262-8.
60. Kawaide, H.; Imai, R.; Sassa, T.; Kamiya, Y. "*ent*-kaurene synthase from the fungus *Phaeosphaeria* sp, L487 - cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase in fungal gibberelln biosynthesis," *J. Biol. Chem.* **1997,** *272*, 21706-12.
61. Tudzynski, B.; Kawaide, H.; Kamiya, Y. "The gibberellin biosynthesis in *Gibberella fujikuroi*: cloning and characterization of the copalyl diphosphate synthase gene," *Curr. Genet.* **1998,** *34*, 234-40.
62. Tudzynski, B.; Holter, K. "Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster," *Fung. Genet. Biol.* **1998,** *25*, 157–70.

63. Oikawa, H.; Toyomasu, T.; Toshima, H.; Ohashi, S.; Kawaide, H.; Kamiya, y.; Ohtsuka, M.; Shinoda, S.; Mitsuhashi, W.; Sassa, T. "Cloning and functional expression of cDNA encoding aphidicolan-16*β*-ol synthase: a key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin," *J. Am. Chem. Soc.* **2001,** *123*, 5154-5.
64. Toyomasu, T.; Nakaminami, K. "Cloning of a gene cluster responsible for the biosynthesis of diterpene aphidicolin, a specifc inbitor of DNA polymerase," *Biosci. Biotechnol. Biochem.* **2004,** *68*, 146-52.
65. Young, C. A.; McMillan, L. "Molecular cloning and genetic analysis of an indolediterpene gene cluster from *Penicillium paxilli*," *Mol. Microbiol.* **2001,** *39*, 754-64.
66. Zhang, S.; Monahan, B. J.; Tkacz, J. S.; B., S. "Indole-diterpene gene cluster from *Aspergillus flavus*," *Appl. Environ. Microbiol.* **2004,** *70*, 6875-83.
67. Young, C. A.; Bryant, M. K.; Christensen, M. J.; Tapper, B. A.; Bryan, G. T.; Scott, B. "Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass," *Mol. Gen. Genomics* **2005,** *274*, 13–29.
68. Graebe, J. E.; Hedden, P.; Gaskin, P.; MacMillan, J. "The biosynthesis of a C19-gibberellin from mevalonic acid in a cell-free system from a higher plant," *Planta* **1974,** *120*, 307-9.
69. Sun, T. P.; Kamiya, Y. "The Arabidopsis GA1 locus encodes the cyclase *ent*-kaurene synthetase A of gibberellin biosynthesis," *Plant Cell* **1994,** *6*, 1509-18.
70. Bensen, R. J.; Johal, G. S.; Crane, V. C.; Tossberg, J. T.; Schnable, P. S.; Meeley, R. B.; Briggs, S. P. "Cloning and characterization of the maize An1 gene," *Plant Cell* **1995,** *7*, 75-84.
71. Ait-Ali, T.; Swain, S. M.; Reid, J. B.; Sun, T. P.; Kamiya, Y. "The LS locus of pea encodes the gibberellin biosynthesis enzyme *ent*-kaurene synthase A " *Plant J.* **1997,** *11*, 443-54.
72. Yamaguchi, S.; Saito, T.; Abe, H.; Yamane, H.; Murofushi, N.; Kamiya, Y. "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*, 203-13.
73. Hedden, P.; Kamiya, Y. "Gibberellin biosynthesis: enzymes, genes and their regulation," *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1997,** *48*, 431-60.
74. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Sassa, T. "Cloning of a full-length cDNA encoding *ent*-kaurene synthase from *Gibberella fujikuroi*: functional analysis of a bifunctional diterpene cyclase," *Biosci. Biotechnol. Biochem.* **2000,** *64*, 660-4.
75. Mende, K.; Homann, V.; Tudzynski, B. "The geranylgeranyl diphosphate synthase gene of *Gibberella fujikuroi*: Isolation and expression," *Mol. Gen. Genet.* **1997,** *255*, 96-105.

76. Hamano, D. T.; Kuzuyama, Y.; Itoh, T.; Funhata, N. K.; Seto, H. "Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic, terpentecin," *J. Bacteriol.* **2001,** *183*, 6085-4.
77. Parker, E. J.; Scott, D. B. "Indole-diterpene biosynthesis in Ascomycetous fungi," *Handbook of industrial mycology* **2004**, 405-26.
78. Fehr, T.; Acklin, W. "Isolation of 2 new indole derivatives from the mycelia of *Claviceps paspali*," *Helv. Chim. Acta* **1966,** *49*, 1907-10.
79. Kuzuguchi, T.; Morita, Y.; Sagam, I.; Sagam, H.; Ogua, K. "Human geraylgerayl diphosphate synthase. cDNA cloning and expression," *J. Biol. Chem.* **1999,** *274*, 5888-94.
80. Tudzynski, B.; Homann, V.; Feng, B.; Marzluf, G. A. "Isolation, characterization and disruption of the areA nitrogen regulatory gene of *Gibberella fujikuroi*," *Mol. Gen. Genet.* **1999,** *261*, 106-14.
81. Keller, N. P.; Hohn, T. M. "Metabolic pathway gene clusters in filamentous fungi," *Fung. Genet. Biol.* **1997,** *21*, 17-29.
82. Siebert, P. D.; Chenchik, A.; Kellogg, D. E.; Lukyanov, K. A.; Lukyanov, S. A. "An improved method for walking in uncloned genomic DNA," *Nucleic Acids Res.* **1995,** *23*, 1087-8.

CHAPTER TWO

Pleuromutilin Biosynthesis in *Pleurotus passeckerianus* Pilat

Antibiotic compounds

Sulfonamide drugs (known widely as "sulfa drugs") were the first known class of antibiotics.[1] The first sulfonamide drug was synthesized in 1932. However, the term 'antibiotic' was not used until 1942, when Selman Abraham Waksman, the Nobel Prize winner and famous biochemist who discovered many antibiotics including streptomycin and neomycin, coined this word.[1] It was later found that the effect of sulfonamides was exceeded by two natural products, penicillin and streptomycin, which were isolated from the fungus *Penicillium chrysogenum* and the bacterium *Streptomyces griseus*, respectively. It is believed that penicillin is the most important discovery in the history of antibiotics to date.[1]

From the 1940s to the 1960s, the term antibiotic was only applied to natural drugs produced by several fungi or bacteria. The category was later extended to include chemotherapeutic drugs and semi-synthetic antibiotics that have been developed from natural products with different modifications from the parental structures.

Natural products are still an important source of new bioactive compounds. During the period of 1981-2002, 40% of the new drugs were either natural products by themselves or analogues derived from the natural products.[2] Even for the drugs

from total synthesis and semi-synthesis, most of them were originally isolated as biologically active natural products and then the chemical syntheses were achieved later on.[2]

Most of the pharmacologically active natural products are secondary metabolites produced by certain organisms. Secondary metabolites are inexhaustible sources of new antimicrobial, antiviral, antitumor lead compounds and/or parental structures for synthetic and semisynthetic derivatives with improved pharmaceutical properties.

Today there are many different kinds of antibiotics on the market.[3] The mechanisms of known antibiotics actions include: inhibition of protein synthesis, inhibition of cell wall formation, blocking DNA or RNA synthesis and blocking cell metabolism by inhibiting metabolic enzymes. Aminoglycosides, including streptomycin and neomycin, inhibit protein synthesis by binding to the 30s subunit of bacterial ribosomes. Tetracyclines also target bacteria 30S ribosomal subunit. Macrolides, including erythromycin, inhibit protein synthesis by binding to the 50S ribosome subunit. Chloramphenicol also binds to the bacterial 50S ribosomal subunit. Beta-lactam antibiotics, including penicillins and cephalosporins, act by inhibiting the formation of the peptidoglycan layer of bacterial cell walls. Glycopeptides, including vancomycin, interfere with cell wall development by inhibiting peptidoglycan synthesis. Rifampin inhibits RNA synthesis by inhibiting one of the enzymes (DNA-dependent RNA polymerase) needed in this process. Sulfonamides block cell metabolism by inhibiting enzymes which are needed in

the biosynthesis of folic acid which is a necessary cell metabolite. Some common antibiotics are shown in Figure 2.1.

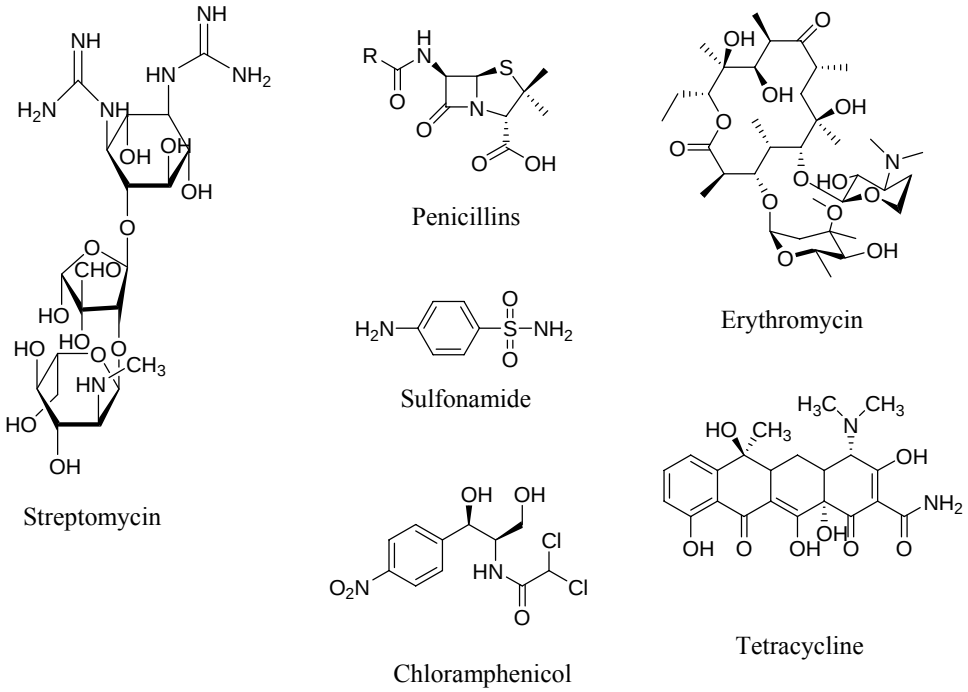


Figure 2.1. Examples of antibiotic compounds.

Although many antibiotics have been discovered and used, the war between pathogens and antibiotics never seems to end. New bacteria are discovered that are not affected by commonly used antibiotics. More importantly, use or misuse of antibiotics may result in the development of antibiotic resistance by the infecting organisms.[4, 5] Antibiotic resistance has become a serious problem in both the community and hospital environments.[6] Moreover, the occurrences of immunocompromised diseases such as AIDS and cancers pose a demand for

developing new antibiotics that can specifically fight against resistant pathogens infecting immunocompromised patients.

The pleuromutilin class of antibiotics has been used in veterinary medicine for a long time and now is being reinvestigated for use in humans. The antibacterial mechanism and the promising future of this family of compounds will be discussed. Also, the known biosynthetic pathway to pleuromutilin will be introduced. One of the goals of this project is to identify and characterize the biosynthetic genes that lead to the production of the pleuromutilin scaffold.

Introduction to pleuromutilins

Pleuromutilin (Figure 2.2, **1**) was originally isolated from the fungi *Pleurotus mutilis* (now *Clitopilus scyphoides*) and *Pleurotus passeckerianus* Pilat (now *Clitopilus passeckerianus* Pilat) in 1951.[7] The pleuromutilins are a family of antibiotic compounds that vary in the nature of the acyl sidechain. They inhibit mainly Gram-positive bacteria and mycoplasms.[8] Tiamulin (Figure 2.2, **2**) and valnemulin (Figure 2.2, **3**), two more active semisynthetic analogs, are currently used in veterinary medicine.[9, 10] Moreover, recent studies showed that valnemulin is effective for treating antibiotic resistant mycoplasma infections in immunocompromised patients.[11] Retapamulin (Altabax®) (Figure 2.2, **4**),[12] another pleuromutilin derivative, has just been approved in December 2006 by the United States FDA for the treatment of secondarily infected traumatic lesions (SITL)

which is a type of skin infection most commonly caused by the Gram-positive bacteria *Staphylococcus aureus* and *Streptococcus pyogenes*. The MIC_{90} results for retapamulin were 0.12 μg / mL for *Staphylococcus aureus* and 0.016 μg / mL for *Streptococcus pyogenes*.[13] Additional *in vitro* studies showed that resistance of *Staphylococcus aureus* and *Streptococcus pyogenes* to retapamulin develops more slowly than to many other effective agents such as vancomycin and erythromycin.[14] Another pleuromutilin derivative (GlaxoSmithKline 565154) is currently in Phase I clinical trials as an oral antibiotic.[15] The structures of pleuromutilin and its derivatives are shown in Figure 2.2.

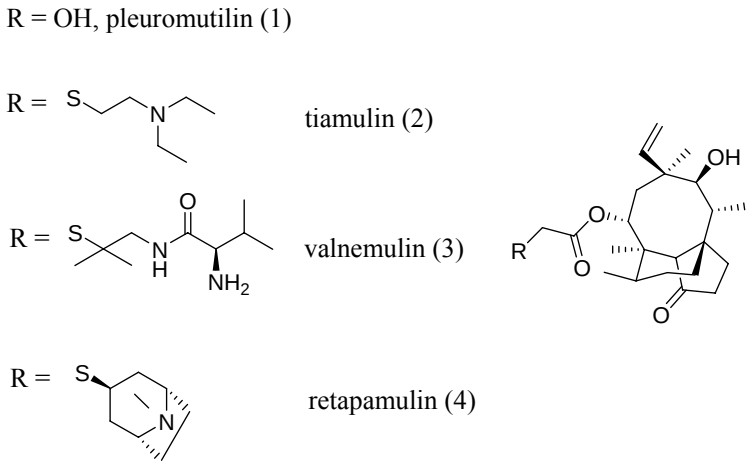


Figure 2.2. Pleuromutilin and some semisynthetic derivatives.

Antibiotic mechanism and resistance mechanism

Pleuromutilin and its derivatives function through binding to the bacterial 50S ribosomal subunit and thus inhibit protein synthesis. The antimicrobial mechanism

of the pleuromutilin class of antibiotics was revealed from the crystal structure of the *Deinococcus radiodurans* 50S ribosomal subunit in complex with tiamulin.[15, 16] Tiamulin binds to the ribosomal peptidyl transferase center of the 50S ribosomal subunit. Its tricyclic mutilin core is located inside a tight pocket confined by residues G2061, A2451, C2452, A2503, U2504, G2505, U2506 at the aminoacyl-tRNA (A-tRNA) binding site. The side chain extension of tiamulin partially overlaps with the peptidyl-tRNA (P-tRNA) binding site and enhances the interactions of tiamulin with 23S rRNA.[15] The sulfur atom in this extension forms a hydrogen bond with C2063. This side chain also offers a number of hydrophobic interactions with A2062, U2585 and C2586. These numerous binding contacts allow tiamulin to directly inhibit peptide bond formation.

This study also indicated that the binding site of pleuromutilins on 50S ribosomal subunit only partially overlaps with those of some other antibacterial agents targeting the 50S ribosome. Another paper further supports this conclusion.[17] In this paper, the binding of the pleuromutilin and three of its semisynthetic derivatives, tiamulin, valnemulin and SB-264128 which carry different side chain extensions, was investigated using chemical footprinting. This study showed that in all of the footprintings, the nucleotides A2058, A2059, G2505, and U2506 were affected which form the pocket that tiamulin binds to as shown in the crystal stucture.[15] This result indicated that all compounds are similarly anchored by the common tricyclic mutilin core in the binding pocket. It was also indicated in this paper that the side chain extensions of pleuromutilin and its derivatives adopt

distinct conformations within the peptidyl transferase cavity and thereby affect the rRNA conformation differently. The results of these studies suggest that pleuromutilins may encounter less cross-resistance with other agents, which makes them an important class of compounds to study.

Although the pleuromutilins may experience less cross-resistance, some bacteria that are resistant to pleuromutilins have been identified. The specific mechanism for tiamulin resistance has been examined further.[18] The strains that are less susceptible to tiamulin have mutations on either ribosomal protein Large 3 (L3) or the 23S rRNA at positions near to the peptidyl transferase center. These are amino acid positions 148 and 149 of the ribosomal protein L3 and nucleotide positions G2032, C2055, G2447, C2499, U2504 and A2572 of the 23S rRNA. The mutations lead to reduced binding of tiamulin to the bacteria ribosomal subunits and thus less susceptibility of the bacteria to tiamulin. Tiamulin contacts directly with the nucleotide U2504 and thus the conformation change in this position can affect tiamulin binding. All the other mutated nucleotides are either in contact with U2504 or very close to it and thus hinder tiamulin binding through alteration of position U2504.[18]

General biosynthetic pathway to pleuromutilin

The key step in the formation of the pleuromutilins **1** is the cyclization of GGPP **5** to the tricyclic intermediate **7** (Figure 2.3). The enzyme that mediates this

cyclization is a diterpene synthase (DS). Based on other fungal DS's involved in the formation of gibberellins and aphidicolin,[19-21] it is likely the *Pleurotus* DS mediates both cyclization steps, the proton-initiated step from **5** to **6** and the ionization-initiated step from **6** to **7**. The cyclization from **5** to **6** is accompanied by two 1,2-hydride shifts, a 1,2-methyl shift, and a ring contraction. The cyclization from **6** to **7** is accompanied by a 1,5-hydride shift to generate the carbocation at C14, followed by the final trapping of the carbocation with water. The general biosynthetic pathway to pleuromutilin was established by isotope labeling experiments in the 1960's.[22-27] The actions of cytochrome P-450 enzymes (functions at C3 and C11) and an acyltransferase (functions at C14 hydroxyl) are necessary to complete formation of the pleuromutilins.

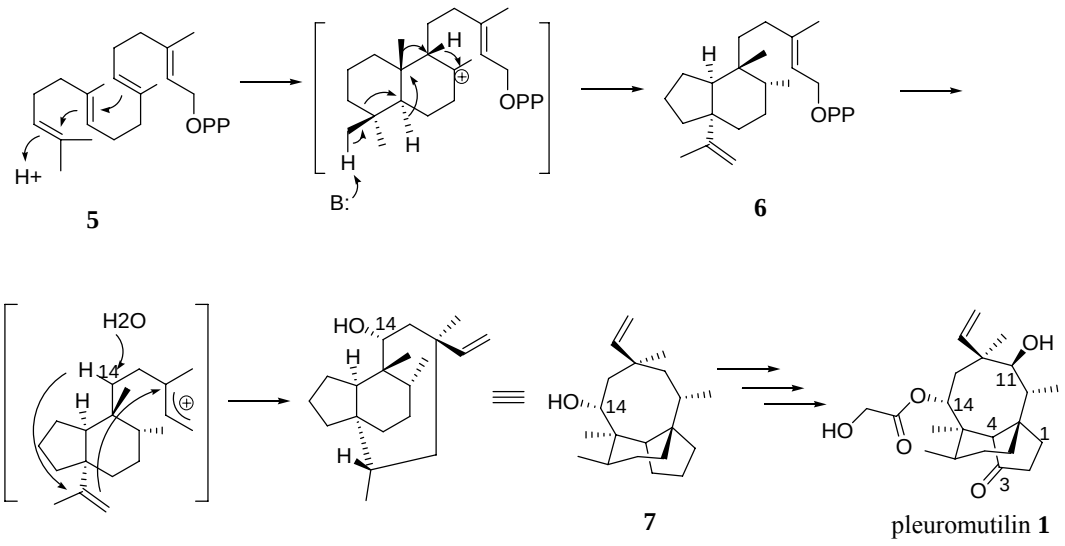


Figure 2.3. Proposed biosynthetic scheme for formation of pleuromutilin from GGPP.

Results and discussion

Culture conditions and the production of pleuromutilin compounds

Pleurotus passeckerianus Pilat (now *Clitopilus passeckerianus* Pilat) can be grown under both shaking and static conditions. Under the shaking condition, the fungus grows to numerous white, round aggregations which can be up to 1 cm diameter in seven days. Under the static condition, the mycelium covered the surface of the liquid in 500 mL Erlenmeyer flasks about three weeks after being inoculated with a 50 mL 3-day culture in the same medium (Figure 2.4). The compounds were produced under both shaking and static conditions. The growth rate of the mycelia and the yield of the compounds were not compared for these two conditions. The cultures used for this project were grown under shaking condition according to earlier work.[28] The growth medium and the conditions were also reported to produce some derivatives of pleuromutilin, mutilin, 14-acetyl-mutilin and the unsaturated fatty acid esters of pleuromutilin (Figure 2.5), but only in very small quantities, with mutilin 0.5 ~ 1%, 14-acetyl-mutilin 2 ~ 5% and mixtures of the unsaturated fatty acid esters 1 ~ 2% relative to pleuromutilin.[28]

Figure 2.4. *Pleurotus passeckerianus* Pilat grown under static condition for 20 days.

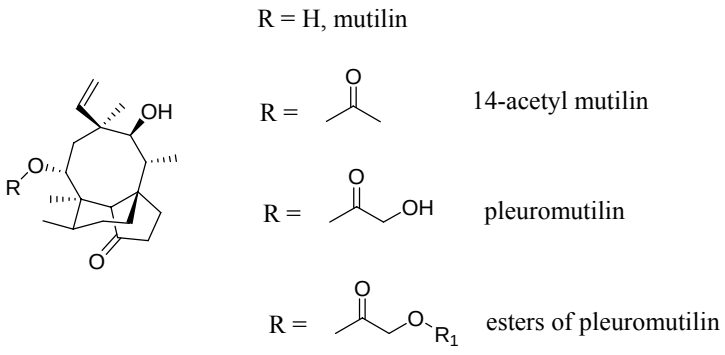


Figure 2.5. Structures of pleuromutilin and its derivatives.

Determination of the presence of the pleuromutilins

In order to determine that the fungus obtained from ATCC was the correct organism and that it does produce the secondary metabolites pleuromutilins under our laboratory conditions, pleuromutilin and its derivatives were isolated and their structures were confirmed by NMR (nuclear magnetic resonance) and MS (Mass Spectrometry). The presence of pleuromutilin and several of their derivatives was confirmed.

Nuclear Magnetic Resonance (NMR) analysis

The 1D and 2D NMR, ^{1}H, ^{13}C, COSY, DEPT, HSQC and HMBC spectra were obtained and the presence of pleuromutilin was shown (Figure 2.6 – 2.11). The ^{1}H and ^{13}C NMR chemical shifts were published in the literature, but there were no assignments for the protons and carbons.[29] Therefore, COSY, DEPT, HSQC and HMBC spectra were conducted and after analysis of these spectra, and the protons and carbon signals were assigned (Table 2.1).

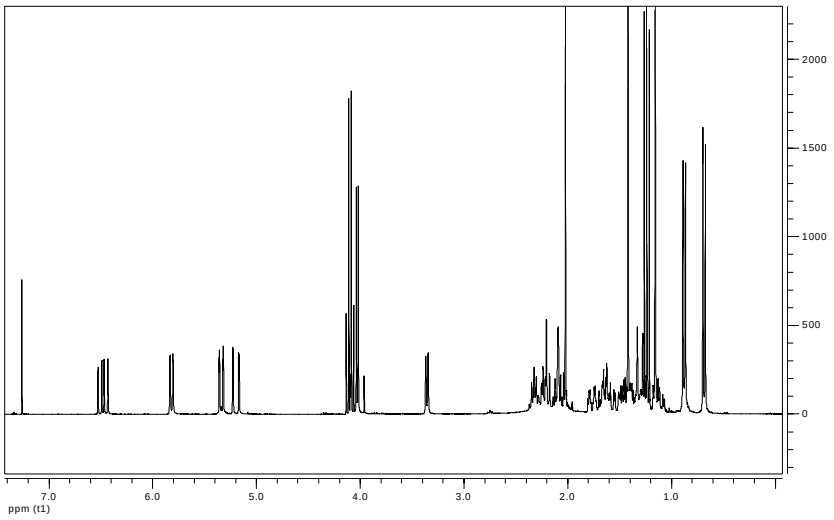


Figure 2.6. ^{1}H NMR spectrum of pleuromutilin.

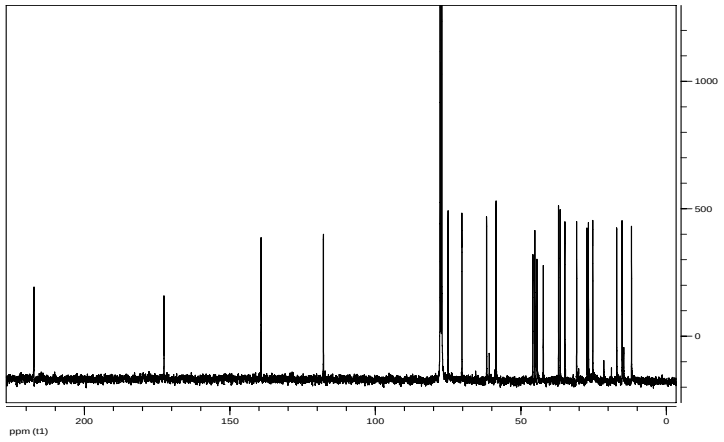


Figure 2.7. ^{13}C NMR spectrum of pleuromutilin.

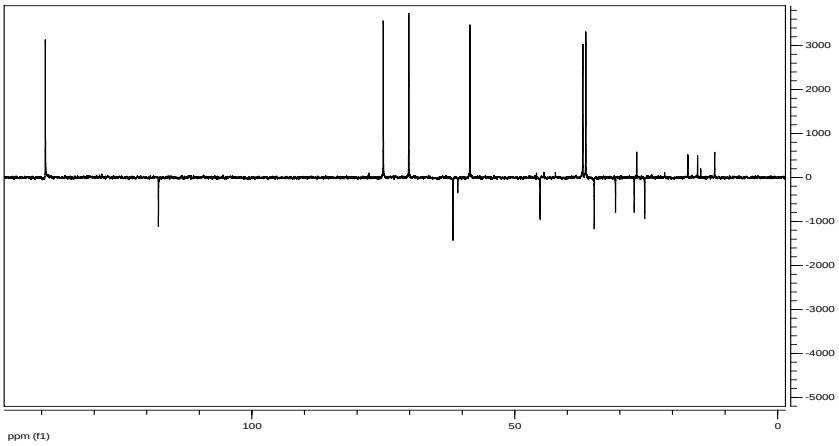


Figure 2.8. DEPT135 NMR spectrum of pleuromutilin.

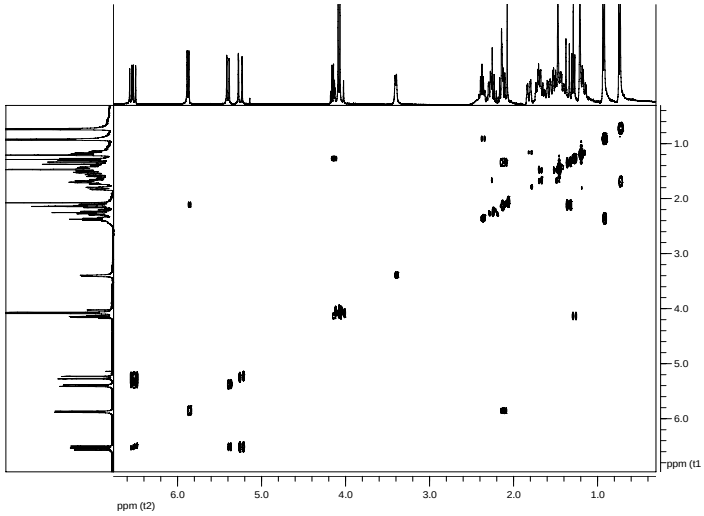


Figure 2.9. COSY NMR spectrum of pleuromutilin.

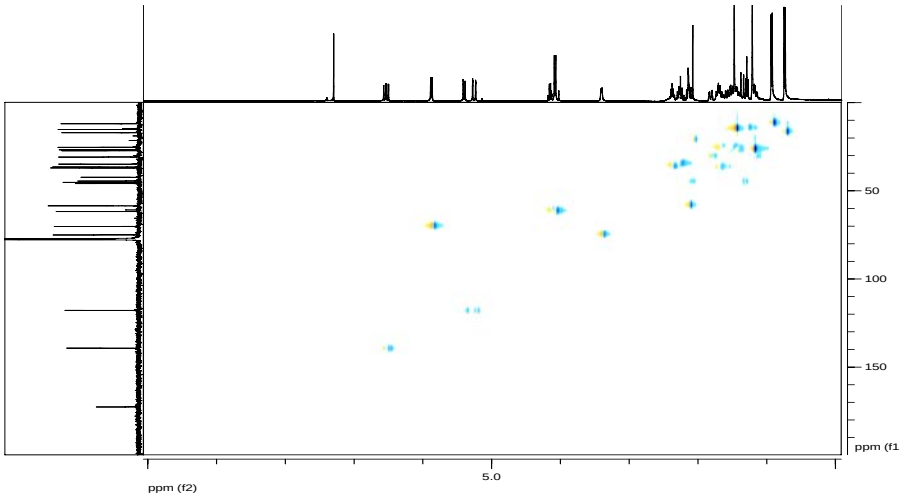


Figure 2.10. HSQC NMR spectrum of pleuromutilin.

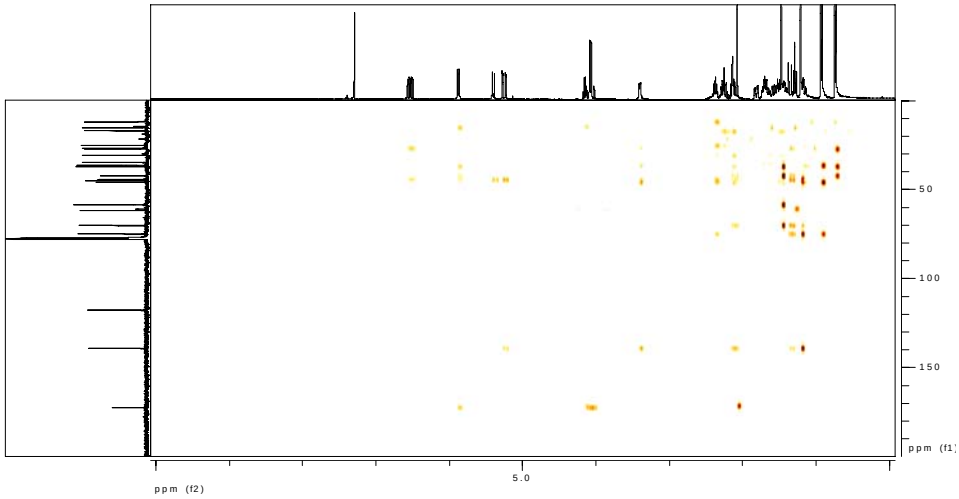


Figure 2.11. HMBC NMR spectrum of pleuromutilin.

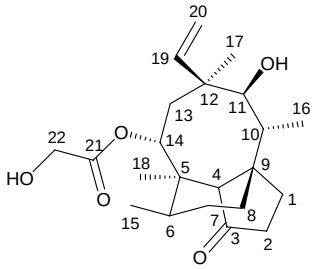


Table 2.1. The assignments of proton and carbon NMR shifts of pleuromutilin. Chemical shifts (δ) are in ppm, while coupling constants (*J*) are in Hz. The numberings of the carbons are shown in the figure on the right.

#	^{1}H (δ)	^{13}C (δ)	#	^{1}H (δ)	^{13}C (δ)
1	1.48, m	24.88	11	3.37, d, *J*=7	74.64
	1.68, m		12	---	44.06
2	2.26, m	34.47	13	1.33, m	44.85
	2.26, m			2.10, dd, *J*=9, 3	
3	---	216.83	14	5.84, d, *J*=9	69.79
4	2.09, s	58.14	15	0.70, s	14.81
5	---	41.87	16	0.89, s	16.57
6	1.68, m	36.65	17	1.18, s	11.47
7	1.40, m	26.88	18	1.44, s	26.46
	1.53, m		19	6.51, dd, *J*=18, 11	139.03
8	1.15, m	30.46	20	5.36, dd, *J*=11, 2	117.24
	1.80, m			5.22, dd, *J*=18, 2	
9	---	45.45	21	---	172.17
10	2.33, dq	36.11	22	4.11	61.35
				3.96	

Mass Spectrometry analysis

The major peak of pleuromutilin is *m/z* 303 under (+) APCI which is the loss of the glycolate ($C_2H_3O_3$; 75) sidechain from pleuromutilin (molecular formula: $C_{22}H_{34}O_5$; molecular weight: 378) (Figure 2.12). The reason that there is no *m/z* 378 peak appears may be that the molecule fragments under the high temperature of APCI probe.

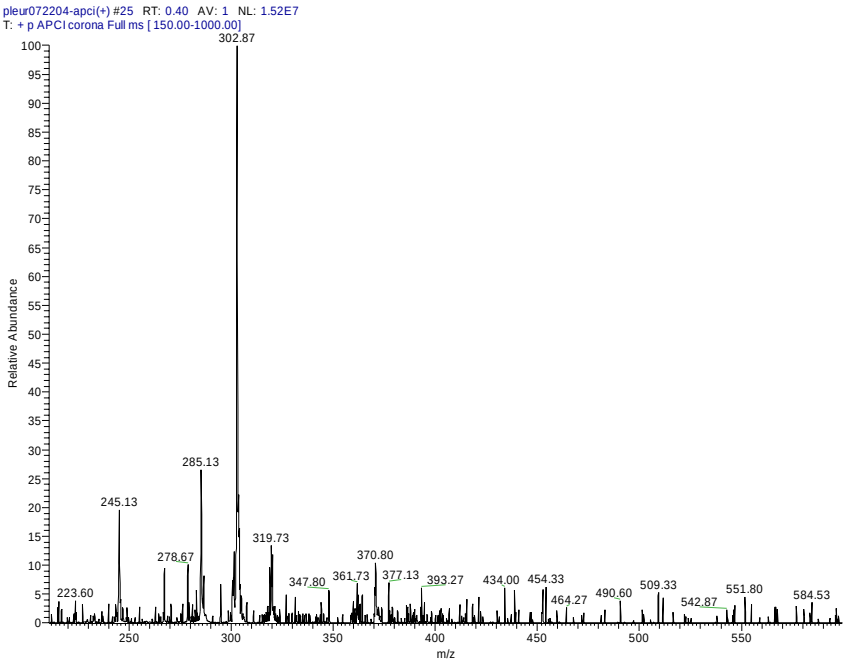


Figure 2.12. Mass spectrometry of pleuromutilin.

Time course of pleuromutilin production

In order to make sure that the isolated mRNA for subsequent experiments contains the information of pleuromutilin biosynthesis, the production curve for pleuromutilin was evaluated. The crude extracts from 1st day to 14th day cultures were used to detect the presence of pleuromutilin and its derivatives without further purification and separation. The presence of either pleuromutilin or its derivatives indicates that the genes for the biosynthesis of pleuromutilin have been turned on. Moreover, according to the literature, pleuromutilin accounts for more than 90% in the production mixture.[28] No pleuromutilin peak (*m/z* 303) was found on the 2nd day. The peaks of *m/z* 279 and *m/z* 371 belong to unknown substances as background in the carrier solvent used for LC/MS analysis. The peak of

pleuromutilins (*m/z* 303) was seen on the 5th day and was strong on the 7th day and the existence of pleuromutilins continued through at least day 14 (Figure 2.13). The identity of the peak of *m/z* 245 is unknown. It started to be seen on the 3rd day and was present throughout the remainder of the time course. According to the time course of pleuromutilin production, the 7th day *P. passeckerianus* culture was chosen for the isolation of mRNA. The presence of pleuromutilins in the 7th day culture was also confirmed by NMR.

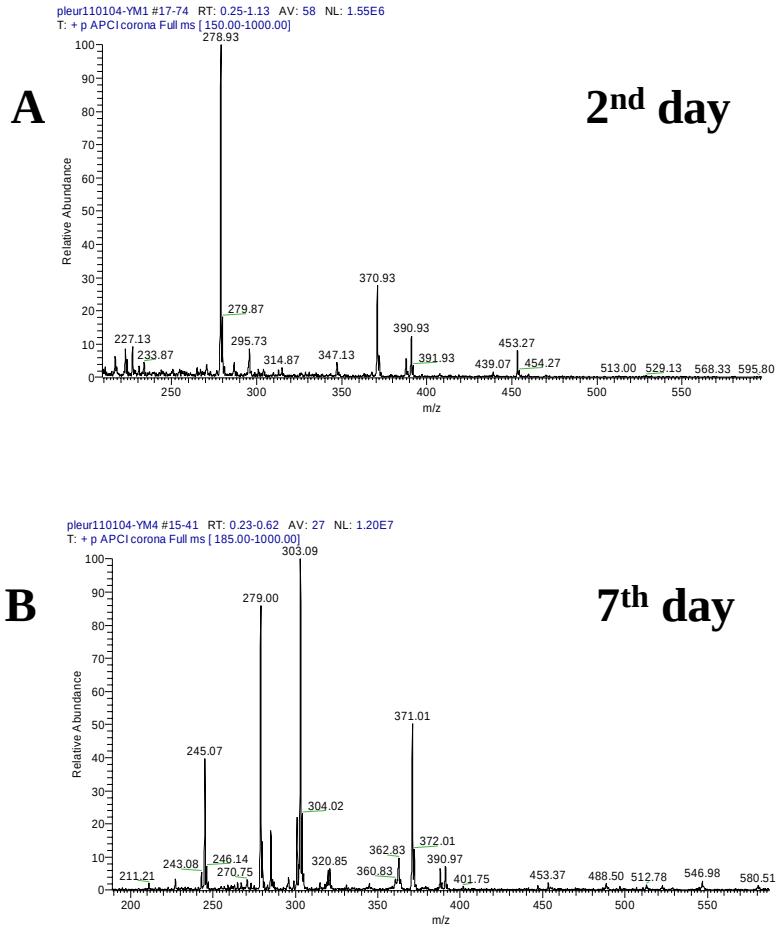


Figure 2.13. Mass Spectral data of 2nd day (A) and 7th day (B) *P. passeckerianus* cultures.

Attempted cloning of the pleuromutilin biosynthetic gene clusters

Once the laboratory culture of *P. passeckerianus* was shown to produce pleuromutilin, the biosynthetic studies could begin. In this project, most of the methods have been used on both the fungi *P. passeckerianus* and *S. araneosa* (the focus of Chapter 3). Therefore, many of the experiments were conducted together and share the same figures.

Direct identification of the diterpene synthase gene

Based on published methods,[19-21] direct identification of the diterpene synthase gene was tried. Degenerate primers were designed from the conserved motifs of fungal diterpene synthases: copalyl diphosphate synthase / *ent*-kaurene synthase (CPS/KS) from *Phaeosphaeria* sp., CPS/KS from *Gibberella fujikuroi*, the aphidicolan-16*β*-ol synthase from *Phoma betae*, and a putative diterpene synthase from *Magnaporthe grisea*. The sequences of these enzymes share 37-60% identities. The protein sequences and nucleotide sequences used to design the degenerate primers are shown in Figure 2.14 and Figure 2.15, respectively. The degenerate primers are shown in Table 2.2. Messenger RNA was isolated from the 7^{th} day mycelia, reverse transcription was conducted and PCR was performed using the degenerate primers.

```
Phoma            -------MVPISTRSEVQDGLIAQARS-------LITRIVHNSDDVYGFGTLSCTVYDTA 46
Magnaportha      MDLCRNVVRDISSHKSKMGSLTFPETANADREAALLQHLLDNYDSVHGAGFMSCTVYDTA 60
Gibberella       --MPGKIENGTPKDLKTGNDFVSAAKS-------LLDRAFKSHHSYYGLCSTSCQVYDTA 51
Phaeosphaeria    --MFAKFD-----------MLEEEARA-------LVRKVGNAVDPIYGFSTTSCQIYDTA 40
                                     :     :       *: :  .  .  :*    ** :****

Phoma            WVALVTK-HVNGIKHWLFPESFHYILASQCDDGTWCEDKTAQFDGVLNTIAGLLVLKRYR 105
Magnaportha      WLSLVKKENVDGQQRWLFPECFQYILDNQQNDGSW-EAYQSEIDGILNTAASLLALQRHS 119
Gibberella       WVAMIPK-TRDNVKQWLFPECFHYLLKTQAADGSWGSLPTTQTAGILDTASAVLALLCHA 110
Phaeosphaeria    WAAMISK-EEHGDKVWLFPESFKYLLEKQGEDGSWERHPRSKTVGVLNTAAACLALLRHV 99
                 * ::: *   .. : *****.*:*:* .*  **:*     ::  *:*:* :. *.*  :

Phoma            FRAGFQSSELDSPELTAMTDTLLKAFKAFSGAIGLD-----SGIEPDVDDSAKIVTTLNM 336
Magnaportha      LHAGFTEKQLG-PAAREVANVLNRVFETENGVVGFARLTAVPSIMADADDTAKTLSALNM 353
Gibberella       LKVGFTLKQIDGDGLRGLSTILLEALRDENGVIGFAP------RTADVDDTAKALLALSL 341
Phaeosphaeria    MKSGFTLSDLECDELSSIANTIAEGFECDHGVIGFAP------RAVDVDDTAKGLLTLTL 331
                 :: **  .::       ::  : . :.   *.:*:           *.**:** : :*.:

Phoma            HKLVGLFHRTPLLKALPKWELQASMIEASIYQGLLQDARLEVLQRP--KVDGGEYLSIIP 628
Magnaportha      ESFIKLYKATPLFSTTPEWKLRSAMAEGSLFLPLLKRRRDDVFTRE--AMTEDKYLEMIP 647
Gibberella       EKYMRLVRKTALFSPLDEWGLMASIIESSFFVPLLQAQRVEIYPRDNIKVDEDKYLSIIP 635
Phaeosphaeria    NGFQQLVGRTDLFSGVPAWELQASFLESALFVPLLRNHRLDVFDRDDIKVSKDHYLDMIP 623
                 .    *   * *:.    * * ::: *.:::  **:  * ::  *    :  ..**.:**

Phoma            FTWTSCSNSARTNASASHLWELMALSFFTYQVDEFMEAVAGPAFKGRMTHLHAIIDEAVH 688
Magnaportha      FTWTACDLRAKSGASPSFLWEMMHISMLSYQGDEYMESVAAPAFAGDMDSLRHIVHNLIP 707
Gibberella       FTWVGCNNRSRTFASNRWLYDMMYLSLLGYQTDEYMEAVAGPVFGDVS-LLHQTIDKVID 694
Phaeosphaeria    FTWVGCNNRSRTYVSTSFLFDMMIISMLGYQIDEFFEAEAAPAFAQCIGQLHQVVDKVVD 683
                 ***..*.  ::: .*   *:::* :*:: ** **::*: *.*.*      *:  :.: :

Phoma            GPFCFYYATCLLGATLTAHPPN----DCFPKSSQKYLAAATCRHLSCMCRMYNDIGSWNR 842
Magnaportha      APCYFAFVACLL-TSIVAPPSNGNAGDCFQTVEQKYFAAAANRHLATMCRMYNDYGSVSR 875
Gibberella       CAYSFAFSNCLMSANLLQ------GKDAFPSGTQKYLISSVMRHATNMCRMYNDFGSIAR 855
Phaeosphaeria    CAYSFAFACCITSATIGQ------GQSMFATVNELYLVQAAARHMTTMCRMCNDIGSVDR 846
                  .  * :  *:  :.:          . * .  : *:  :. ** : **** ** **  *
```

Figure 2.14. Protein alignments (partial) of fungal diterpene synthases. The regions used to design degenerate primers are highlighted in red.

Figure 2.15. Nucleotide alignments (partial) of some fungal diterpene synthase genes. The designed degenerate primers are highlighted in red. I: Inosine; Y: C, T; V: G, A, C; W: A, T; S: G, C; R: A, G.

Table 2.2. Degenerate primers for diterpene synthase genes.

	Primer sequences	Tm
Forward primers	F-2: GGCTITTYCCIGARWGYTTC (2I, 2^4=16 fold)	56 °C
	F-5: GAYGYIGAYGACASIGCIAAA (3I, 2^4=16 fold)	57 °C
Reverse primers	R-5: TTTIGCISTGTCRTCIRCRTC (3I, 2^4=16 fold)	57 °C
	R-6: YCCAIGTRAAIGGRATIATS (3I, 2^4=16 fold)	53 °C
	R-7: GTCGTTRYACATICGRCACA (1I, 2^3=8 fold)	57 °C

A variety of PCR conditions were tried. The primer sets of F-2 → R-5, F-2 → R-6, F-5 → R-6, F-5 → R-7, F-2 → R-7 were used, for which the expected sizes of PCR products are about 780 bp, 1680 bp, 900 bp, 1500 bp and 2285 bp, respectively. For each primer set, different PCR annealing temperatures, from 42 °C to 60 °C were tried. No desired PCR products were found.

By comparing the cDNA sequences and genomic DNA sequences of those four known or putative fungal diterpene synthases, the primers were designed from the protein conserved regions where cDNA sequences do not contain the junctions of introns and exons. Although it is still possible that these primers happen to be in the junctions of introns and exons in the genome sequences of *P. passeckerianus* and *S. araneosa*, the chance may be small. Therefore, genomic DNA was also used as a template for PCR in case that the isolated mRNA did not contain the information for pleuromutilin and sordarin biosynthesis. Many different PCR conditions were tried, but there were no positive PCR results.

During the course of this work, additional putative fungal diterpene synthase gene sequences became available. New degenerate primers were designed after including the information of two more putative fungal diterpene synthase genes from *Botrytis cinerea* and *Chaetomium globosum*. The four degenerate primers F-2, F-5, R-5, and R-7 were modified slightly and designated as F-2-new, F-5-new, R-5-new, and R-7-new, respectively; two new forward primers F-7, F-8 and two new reverse primers R-8 and R-9 were also designed (Table 2.3). The nucleotide sequences used to design the new degenerate primers are shown in Figure 2.16.

The expected sizes of PCR products for the new primers are ~1.54 Kb (F-5 → R-8), ~2 Kb (F-5 → R-9), ~0.30 Kb (F-7 → R-9), ~0.26 Kb (F-8 → R-9), ~2.5 Kb (F-2 → R-7), ~2.57 Kb (F-2 → R-8), and ~2.8 Kb (F-2 → R-9), respectively. In addition to the normal PCR, nested PCR were also performed to increase the chances of amplifying specific diterpene synthase gene fragment. Nested PCR did give cleaner results most of time, and sometimes strong bands showed up, but no correct fragments were found.

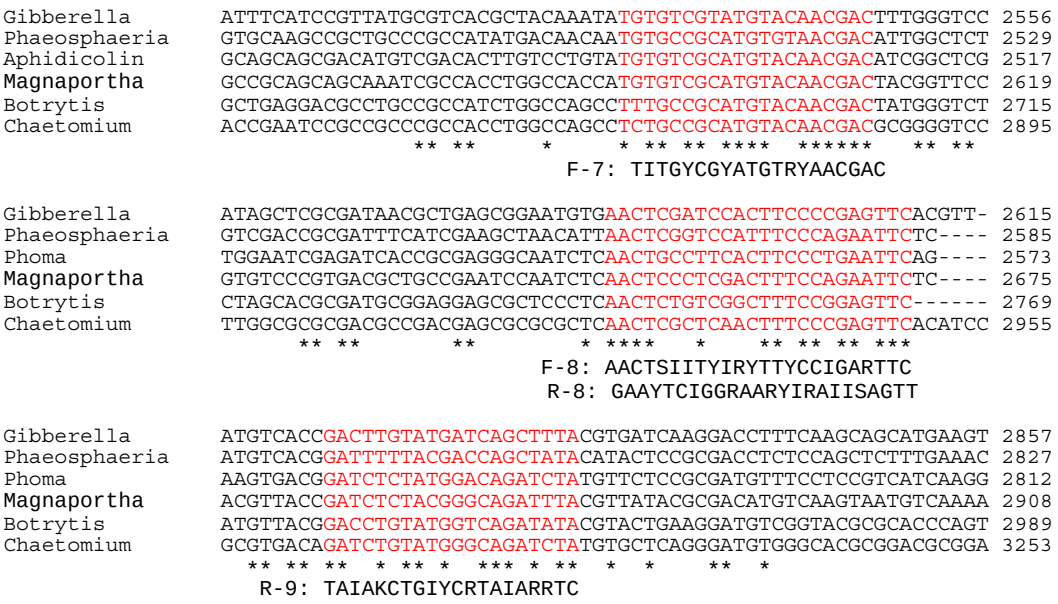


Figure 2.16. Nucleotide alignments (partial) of some fungal diterpene synthase genes for designing the new degenerate primers. The designed degenerate primers are highlighted in red. I: Inosine; Y: C, T; V: G, A, C; W: A, T; S: G, C; R: A, G.

Table 2.3. New degenerate primers for diterpene synthase genes. The bases different from the old primers are orange-highlighted. The nucleotides in blue are the ones added in new primers.

	Primer sequences	Tm
Forward primers	F-2-new: TGGYTITTYCCIGARWGYTT (2I, 2^5=32 fold)	56 °C
	F-5-new: GAYGYIGAYGACASIKCIAA (3I, 2^5=32 fold)	57 °C
	F-7: TITGYCGYATGTRYAACGAC (1I, 2^4=16 fold)	56 °C
	F-8: AACTSIITYIRYTTYCCIGARTTC (4I, 2^6=64 fold)	57 °C
Reverse primers	R-5-new: TTIGMISTGTCRTCIRCRTC	57 °C
	R-6: YCCAIGTRAAIGGRATIATS	53 °C
	R-7-new: GTCGTTRYACATRCGRCAIA	56 °C
	R-8: GAAYTCIGGRAARYIRAIISAGTT	57 °C
	R-9: TAIAKCTGIYCRTAIARRTC	51 °C

In order to test the effectiveness of the degenerate primers, the cDNA from *Botrytis cinerea* was used as a control to test these degenerate primers. A *B. cinerea* culture was kindly provided by Dr. Henrik Stotz, OSU Department of Horticulture. At the same time, gene-specific primers for the *B. cinerea* projected diterpene synthase gene were obtained and the PCR results were used as control for the degenerate primers. The F-2 → R-5 primer set and F-7 → R-9 primer set which give small sizes of PCR products and thus are relatively easier to be amplified were tried with *B. cinerea*. By comparing to the PCR products generated from the gene-specific primers for the *B. cinerea* projected diterpene synthase gene, the F-2 → R-5 degenerate primer set gave a subtle band of the desired size (~780 bp), while the F-7 → R-9 primer set gave a strong band of correct size (~300 bp). This result means that with these primers and suitable PCR conditions, the PCR could work for *P. passeckerianus* and *S. araneosa*. Therefore, these two

primer sets were tried using both the cDNA and genomic DNA of *P. passeckerianus* and *S. araneosa*. However, either multiple bands were generated, or no correct size bands were generated. The bands whose sizes were close to the correct sizes were cloned and sequenced. However, no correct fragment was found. The degenerate primer sets which should yield large sizes (larger than 2 Kb) of PCR products did not provide PCR products from *B. cinerea*. However, all primers were still tried with *P. passeckerianus* and *S. araneosa* DNA. Unfortunately, the PCR still either gave multiple bands, or incorrect size bands. One sample out of many PCR results is shown in Figure 2.17. Some PCR products showed similar patterns for both *P. passeckerianus* and *S. araneosa* although the reason for this is unclear. The bands whose sizes were close to the correct sizes were cloned and sequenced. No correct fragments were found. Some bands turned out to be products generated when a single degenerate primer happened to act as both forward and reverse primers.

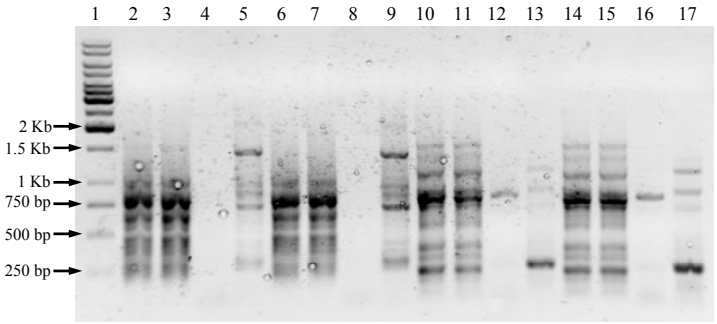


Figure 2.17. PCR products amplified from cDNAs of *P. passeckerianus* (lanes 2 to 9) and *S. araneosa* (lanes 10 to 17). Lanes 2 to 5 and lanes 10 to 13 used 46.4°C annealing temperature while lanes 6 to 9 and lanes 14 to 17 used 48.0°C annealing temperature. Lanes 2, 6, 10 and 14 used degenerate primers F-5 and R-7 (~1.50 Kb); Lanes 3, 7, 11 and 15 used degenerate primers F-5 and R-8 (~1.54 Kb);

(Figure 2.17 continued) Lanes 4, 8, 12 and 16 used degenerate primers F-7 and R-9 (~0.30 Kb); Lanes 5, 9, 13 and 17 used degenerate primers F-8 and R-9 (~0.26 Kb). Lane 1 contains a 1 Kb DNA ladder (Fermentas).

The reason that this method did not work might be the limited knowledge of fungal diterpene synthases. The fungal diterpene synthase genes identified through the homology-base method share the common intermediate (copalyl diphosphate, CDP).[19-21] Because the pleuromutilin and sordarin structures do not require the CDP intermediate, the diterpene synthases may be quite different from the known fungal diterpene synthases and the genes might also differ. Therefore, the degenerate primers might not be suitable for *P. passeckerianus* and *S. araneosa*. Another method, the indirect identification of the diterpene synthase genes via the *ggs* gene was attempted next.[30-33]

Identification of the diterpene gene cluster via the ggs gene

This method is based on the hypothesis that the terpene biosynthetic genes in sordarin and pleuromutilin-producing fungi are organized in clusters. At the onset of this project, all of the known fungal clusters all have a dedicated *ggs* gene encoding a GGPP synthase specific for diterpene production. Based on the clustering of several other fungal diterpene biosynthetic genes,[30, 32, 33] it is hypothesized that the genes for the formation of pleuromutilin and sordarin are also clustered with a dedicated *ggs* gene. If this is true, identification of the *ggs* genes dedicated to diterpene biosynthesis should lead us to the diterpene synthase genes. Therefore, the diterpene synthase genes can be found indirectly through the

ggs genes. Many fungal *ggs* gene sequences are available and they contain regions that are highly conserved. Therefore, degenerate primers can be designed that are more likely to give specific PCR products from genomic DNA. Once the *ggs* is identified, sequencing by genome walking or construction of a genomic DNA library and probing with a *ggs* gene fragment can be used to find the diterpene synthase gene.

The protein and nucleotide sequences of fungal *ggs* genes used to design the degenerate primers are shown in Figures 2.18 and 2.19, respectively. Four degenerate primers, F1 and F2 for forward primers, and R1 and R2 for reverse primers, were designed. Three degenerate primers from the previous work[33] were also used. All seven primers are shown in Table 2.4. Among these seven degenerate primers, F2 is located closely to GGPPS27, while R1 is the modified GGPPS29.

```
Phoma            QVVRMLHTASLLIDDIQDNSELRRGKPVAQNIFGTALTINSANYVYFLALEKLNSLKN-- 152
Nigrospora       RAVGMLHTASLLIDDIQDNSELRRGRPVAHKIFGPAMTINSANHMYFLALQELNALET-- 201
Gibberella       RVVGMLHESSLLIDDVQDSSELRRGFPVAHNIFGVAQTINSANYIYFVALQELHKLGN-- 226
Magnaporthe      HVVGMLHTASLLVDDVEDNSSLRRGLPVAHSIFGVPQTINSANYIYFAALQELKKLSN-- 222
Neurospora       KVISMLHTASLLVDDVEDNSVLRRGFPVAHSIFGIPQTINTSNYVYFYALQELQKLKN-- 225
Aspergillus      KVVGMLHTASLLVDDVEDNSILRRGIPVAHNIFGTAQTINSANYVYFLALQEVQKLNN-- 174
Penicillium      KVVTMLHTASLLIDDVEDNSVLRRGVPVAHNIFGTAQTINSANYVYFLALQEIQQLKN-- 172
Cryptococcus     KVVRMLHNASLLMDDVEDNSELRRGLPVAHTIYGVPQTINAANYVYFLAFQELLQLRSGV 120
                   : *** :***:**::* * **** ***: *:* . ***:.*: ** *  .:  *

Phoma            --------PNITDIFTEELLRLHRGQAMDLYWRDTLTCPTEEEYFEMVANKTGGLFRLMY 204
Nigrospora       --------PGVVDIFTDELLRLHRGQAMDLYWRETLTCPTEADYFEMTSNKTGGLFRLAY 253
Gibberella       --------PELITIFSDELVNLHRGQGMDLFWRDTLTCPTEEDYLEMVGNKTGGLFRLGI 278
Magnaporthe      --------PEAVSIFAEELIHLHRGQGMDLFWRDTLTCPTEDDYLEMVGNKTGGLFRLGI 274
Neurospora       --------PKAVSIFSEELLNLHRGQGMDLFWRDTLTCPTEDDYLEMVSNKTGGLFRLGI 277
Aspergillus      --------PTAIDIFVQELLNLHRGQGMDLFWRDTLTCPTEEDYLDMVGNKTGGLFRLAI 226
Penicillium      --------PAAIDIYVKELLNLHRGQGMDLFWRDTLTCPTEDEYLEMVGNKTGGLFRLAV 224
Cryptococcus     MGKGKDKEVDLVAAVNEELLQLHRGQGMDLFWRDSLTCPTEKEYIDMVLGKAGGLLRLAV 180
                                 **: *****.***:**: ****** :*  *   *:***:**

Phoma            RMMKAESSMPIDLLPVVELLGVIFQVVDDYKNLCSREYGKLKGFGEDLTEGKFSFPVI 262
Nigrospora       RLMKAESAVSVDLMPVVELLGVLFQVADDYKNLCSREYGDLKGVGEDLTEGKFSFPII 311
Gibberella       KLMAEEANGPTDCVPLVNLIGLIFQIRDDYMNLSSKEYSHNKGMCEDLTEGKFSFPVI 336
Magnaporthe      KLMQAESAFPTDCIPLANLIGLIFQIRDDYQNLFSREYSNNKGMCEDLTEGKFSFPVI 332
Neurospora       KLMQAESRSPVDCVPLVNIIGLIFQIADDYHNLWNREYTANKGMCEDLTEGKFSFPVI 335
Aspergillus      KLMQAESTTGIDCVSLVNVMGLIFQICDDYLNLSNKTYTQNKGLCEDLTEGKFSFPII 284
Penicillium      KLMQAESSTGKDCVSLVNVMGLIFQICDDYLNLSNTTYTHNKGLCEDLTEGKFSFPII 282
Cryptococcus     KLMMAKSDSNVDYVPLVNLISVWFQIRDDYMNLQSPAYKSNKGFCEDLTEGKFSFPVV 238
                  :*  ::    *   :.::::.: **: *** ** .  *   **. ***********::

Phoma            HSIRSNPEDLQLLHVLQQKSSNEHVKLYAIEIMES-TGSLEYTKHVVENIVSQIQEIIYS 321
Nigrospora       HTIRTNPADLQLLNILRERPTNDHVKRYAIELMEK-TGSFRYTRGVLKKLATQVHEAVHL 370
Gibberella       HSIRSNPTNLQLINILKQKTSDIQVKRYAVSYMES-TGSFEYTRKVIMVLIERARKMAEE 395
Magnaporthe      HSIRSNPANLQLLNILKQKTTDEGVKNYAVSYMEK-TGSFEYTKKVLTVLIERARKVTDE 391
Neurospora       HSIRSNPSNMQLLNILKQKTGDEEVKRYAVAYMES-TGSFEYTRKVIKVLVDRARQMTED 394
Aspergillus      HSIRSNPRNHQLISILKQRTKDEEVKLYAISYMES-TGSFAYTRKVVREFRDKALSLIDE 343
Penicillium      HSIRSNPGNHQLVSILKQKTQDEEVKRYAVQYMQS-TGSFTHTRQVVRDLRDRALSLIEA 341
Cryptococcus     HGVRADTSNRQILNVLQKRTSSIDLKQHIVDYLSNHTKSFHYTRKVLTELEVQIMEEIKA 298
                 * :*:.. . *:: :*:::. .  :*   :  : . * *: :*: *:  :  .
```

Figure 2.18. Protein alignments (partial) of fungal GGPPS. Red: regions used to
(Figure 2.18 continued) design degenerate primers; orange: primer regions from reference; blue: region used to analyze PCR results later.

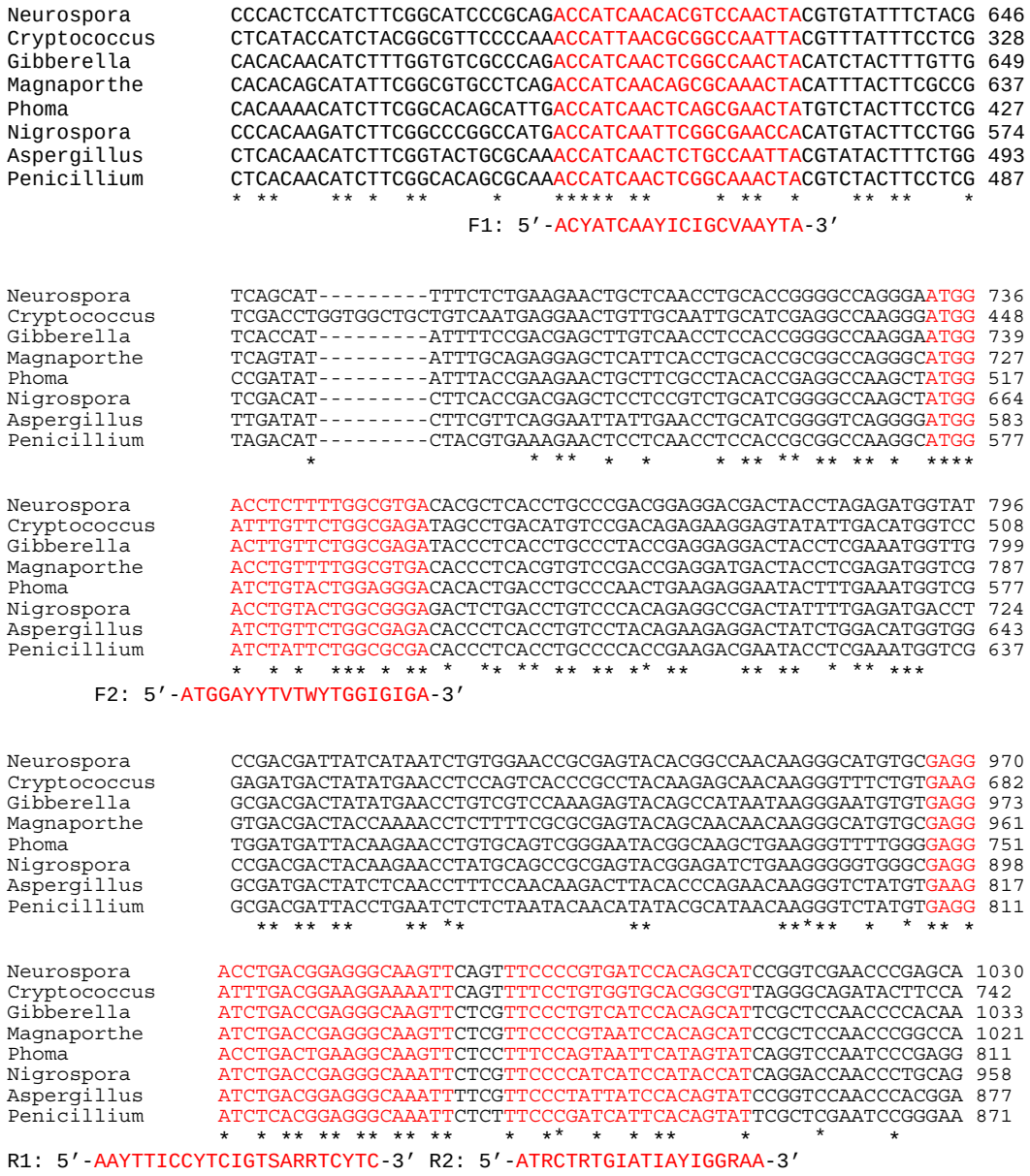


Figure 2.19. Complementary DNA (cDNA) nucleotide alignments (partial) of fungal *ggs* genes. The designed degenerate primers are highlighted in red. I: Inosine; Y: C, T; V: G, A, C; W: A, T; S: G, C; R: A, G; M: A, C.

Table 2.4. Degenerate primers for *ggs* genes. The degenerate primers GGPPS27, GGPPS28 and GGPPS29 are from the reference.[33]

	Primer sequences
Forward primers	F1: ACYATCAAYICIGCVAAYTA F2: ATGGAYYTVTWYTGGIGIGA GGPPS27: CAYMGIGGTCARGGTATTGA
Rewards primers	R1: AAYTTICCYTCIGTSARRTCYTC R2: ATRCTRTGIATIAYIGGRAA GGPPS28: TTCATRTAGTCGTCICKTATYTG GGPPS29: AACTTGCCYTCIGTSARGTCYTC

Nested PCR was conducted using these designed primers. Three different sequences representing fragments of *ggs* genes from *P. passeckerianus* were obtained, whose sizes are 319 bp, 322 bp and 345 bp (designated as p319, p322 and p345 respectively) (Figure 2.20). Two sequences representing fragments of *ggs* genes from *S. araneosa* were also obtained and they will be discussed in Chapter 3. The PCR product p319 was generated from the primers F1 and GGPPS29 for the first round PCR, and F2 and GGPPS28 for the second round PCR; p322 and p345 were generated from the primers F1 and GGPPS29 for the first round PCR, GGPPS27 and GGPPS28 for the second round PCR. The sequences of these fragments are shown below and also in the appendix.

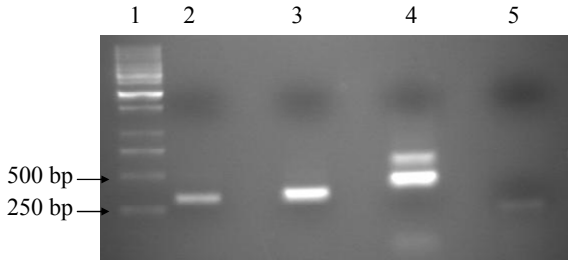


Figure 2.20. Nested PCR 2nd round products amplified from genomic DNA of *P. passeckerianus* and *S. araneosa*. Lanes 2 to 4 were from *P. passeckerianus*, while

(Figure 2.20 continued) lane 5 was from *S. araneosa*. Lane 2 was a PCR (46 °C annealing temperature) product amplified with primers (F1, GGPPS29) for the 1st round and (F2, GGPPS28) for the 2nd round. Lane 3 was a PCR (46 °C annealing temperature) product with primers (F1, GGPPS29) for the 1st round and (GGPPS27, GGPPS28) for the 2nd round. Lane 4 represented PCR (42 °C annealing temperature) products with primers (F1, GGPPS29) for the 1st round and (F2, GGPPS28) for the 2nd round. Lane 5 was a PCR (46 °C annealing temperature) product with primers (F1, GGPPS29) for the 1st round and (GGPPS27, GGPPS28) for the 2nd round. The PCR product was a faint band slightly larger than 200 bp. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

The nucleotide sequences and the translated protein sequences for the *P. passeckerianus* PCR fragments are shown below. The presence of the translated primer sequences 'HRGQGMD', 'MDLFWRD' and 'QIRDDY' in different frames indicates the presence of intron(s). The sizes of these fragments, which are larger than anticipated size of ~200 bp, also are due to the presence of intron(s). BLAST searches suggested that these three fragments from *P. passeckerianus* and the two from *S. araneosa* are potential *ggs* genes. The identification of more than one putative GGPP synthase gene may be a molecular signature for diterpene biosynthesis.[31, 32] One of them may be associated with general cell functions, while the other(s) can participate in diterpene biosynthesis.

P319:

```
CTTGATTTGTTCTGGAGAGATACCTTTACCTGCCCGACGGAGGAAGAGTATGTTTCTATGGTCAATAATAGTG
AGTTGGTTCTTGGCCCACCCGAGGCTTAGTCTTAGAACACTAGACTGATTGATACCGCTAGAAACTAGCGGCC
TTCTTCGCCTTGCTATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGACAGGTATGTGCCTGTGTTCAG
CTCCGCTTGCGCCTAAATCAACTGTCCTAATTCTTAGTGATTATGTCCCTCTTGTGAATCTCATTGGAGTATA
TTTCCAGATTCGAGACGACCTGATGAA
```

5'3' Frame 1

```
cttgatttgttctggagagataccttacctgcccgacggaggaagagtatgtttctatg
 L  D  L  F  W  R  D  T  F  T  C  P  T  E  E  E  Y  V  S  M
gtcaataatagtgagttggttcttggcccacccgaggcttagtcttagaacactagactg
 V  N  N  S  E  L  V  L  G  P  P  E  A  -  S  -  N  T  R  L
attgataccgctagaaactagcggccttcttcgccttgctatcaggctcatgatggcatg
 I  D  T  A  R  N  -  R  P  S  S  P  C  Y  Q  A  H  D  G  M
tgcgacgaccaataaggacaggtatgtgcctgtgttcagctccgcttgcgcctaaatcaa
 C  D  D  Q  -  G  Q  V  C  A  C  V  Q  L  R  L  R  L  N  Q
```

```
ctgtcctaattcttagtgattatgtccctcttgtgaatctcattggagtatatttccaga
 L  S  -  F  L  V  I  M  S  L  L  -  I  S  L  E  Y  I  S  R
ttcgagacgacctgatgaa
 F  E  T  T  -  -
```

5'3' Frame 3

```
cttgatttgttctggagagataccttacctgcccgacggaggaagagtatgtttctatggt
   -  F  V  L  E  R  Y  L  Y  L  P  D  G  G  R  V  C  F  Y  G
caataatagtgagttggttcttggcccacccgaggcttagtcttagaacactagactgat
 Q  -  -  -  V  G  S  W  P  T  R  G  L  V  L  E  H  -  T  D
tgataccgctagaaactagcggccttcttcgccttgctatcaggctcatgatggcatgtg
 -  Y  R  -  K  L  A  A  F  F  A  L  L  S  G  S  -  W  H  V
cgacgaccaataaggacaggtatgtgcctgtgttcagctccgcttgcgcctaaatcaact
 R  R  P  I  R  T  G  M  C  L  C  S  A  P  L  A  P  K  S  T
gtcctaattcttagtgattatgtccctcttgtgaatctcattggagtatatttccagatt
 V  L  I  L  S  D  Y  V  P  L  V  N  L  I  G  V  Y  F  Q  I
cgagacgacctgatgaa
 R  D  D  L  M
```

P322:

```
CACCGTGGTCAAGGCCTCGAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAAGAATACATTTCGA
TGGTCAACAATAGTGAGTTAACCTGGGGCCAGTCTAAGCCTACGGTTAGACTCATTTCGTATCAGAAACCAGT
GGCCTCCTACGTATCGGTATTAAACTCATGATGGCATGTGCCACCACAAACGCGGATATGTATGTGTTTTTGG
CAGTGCATTTTGGGGTTTGAACTGAATGGCCATTTTCTAGCGATTATATTCCTTTGACAAACCTCATTGGAGT
CTATTTCCAAATCCGGGATGACCTTATGAA
```

5'3' Frame 1

```
caccgtggtcaaggcctcgaaatcttctggagagattccctcatctgtccctccgaggaa
 H  R  G  Q  G  L  E  I  F  W  R  D  S  L  I  C  P  S  E  E
gaatacatttcgatggtcaacaatagtgagttaacctggggccagtctaagcctacggtt
 E  Y  I  S  M  V  N  N  S  E  L  T  W  G  Q  S  K  P  T  V
agactcatttcgtatcagaaaccagtggcctcctacgtatcggtattaaactcatgatgg
 R  L  I  S  Y  Q  K  P  V  A  S  Y  V  S  V  L  N  S  -  W
catgtgccaccacaaacgcggatatgtatgtgtttttggcagtgcattttggggtttgaa
 H  V  P  P  Q  T  R  I  C  M  C  F  W  Q  C  I  L  G  F  E
ctgaatggccattttctagcgattatattcctttgacaaacctcattggagtctatttcc
 L  N  G  H  F  L  A  I  I  F  L  -  Q  T  S  L  E  S  I  S
aaatccgggatgaccttatgaa
 K  S  G  M  T  L  -
```

5'3' Frame 3

```
caccgtggtcaaggcctcgaaatcttctggagagattccctcatctgtccctccgaggaaga
   P  W  S  R  P  R  N  L  L  E  R  F  P  H  L  S  L  R  G  R
atacatttcgatggtcaacaatagtgagttaacctggggccagtctaagcctacggttag
 I  H  F  D  G  Q  Q  -  -  V  N  L  G  P  V  -  A  Y  G  -
actcatttcgtatcagaaaccagtggcctcctacgtatcggtattaaactcatgatggca
 T  H  F  V  S  E  T  S  G  L  L  R  I  G  I  K  L  M  M  A
tgtgccaccacaaacgcggatatgtatgtgtttttggcagtgcattttggggtttgaact
 C  A  T  T  N  A  D  M  Y  V  F  L  A  V  H  F  G  V  -  T
gaatggccattttctagcgattatattcctttgacaaacctcattggagtctatttccaa
 E  W  P  F  S  S  D  Y  I  P  L  T  N  L  I  G  V  Y  F  Q
atccgggatgaccttatgaa
 I  R  D  D  L  M
```

P345:

```
CATCGCGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAGGAGTATATCTCCA
TGGTTAACAATAGTTAGTGGCTTTTTCATTACTTGAGATGAGGACTCGATACTCATGATCATTTCGTTGAAAT
TTTCTAGAGACGGGCGGTCTTCTACGCATCGGCATCAAACTCATGATGGCATGCGCGACCACCAACACTGGAG
```

```
TGTACGTATTCCATCCTTCTCATTGTCCACTCTGGTCTTTCTGACATTTTCTATTTCTTCCAGTGACTTTATA
CCCTTAGTCAACCTCATCGGCATCTACTTCCAAATCAGAGACGATCTGATGAA
```

5'3' Frame 1

```
catcgcgggcagggtctggaattactttggcgtgattcgttgacgtgcccttcggaggag
 H  R  G  Q  G  L  E  L  L  W  R  D  S  L  T  C  P  S  E  E
gagtatatctccatggttaacaatagttagtggctttttcattacttgagatgaggactc
 E  Y  I  S  M  V  N  N  S  -  W  L  F  H  Y  L  R  -  G  L
gatactcatgatcatttcgttgaaattttctagagacgggcggtcttctacgcatcggca
 D  T  H  D  H  F  V  E  I  F  -  R  R  A  V  F  Y  A  S  A
tcaaactcatgatggcatgcgcgaccaccaacactggagtgtacgtattccatccttctc
 S  N  S  -  W  H  A  R  P  P  T  L  E  C  T  Y  S  I  L  L
attgtccactctggtctttctgacattttctatttcttccagtgactttataccttagt
 I  V  H  S  G  L  S  D  I  F  Y  F  F  Q  -  L  Y  T  L  S
caacctcatcggcatctacttccaaatcagagacgatctgatgaa
 Q  P  H  R  H  L  L  P  N  Q  R  R  S  D  E
```

5'3' Frame 2

```
catcgcgggcagggtctggaattactttggcgtgattcgttgacgtgcccttcggaggagg
  I  A  G  R  V  W  N  Y  F  G  V  I  R  -  R  A  L  R  R  R
agtatatctccatggttaacaatagttagtggctttttcattacttgagatgaggactcg
 S  I  S  P  W  L  T  I  V  S  G  F  F  I  T  -  D  E  D  S
atactcatgatcatttcgttgaaattttctagagacgggcggtcttctacgcatcggcat
 I  L  M  I  I  S  L  K  F  S  R  D  G  R  S  S  T  H  R  H
caaactcatgatggcatgcgcgaccaccaacactggagtgtacgtattccatccttctca
 Q  T  H  D  G  M  R  D  H  Q  H  W  S  V  R  I  P  S  F  S
ttgtccactctggtctttctgacattttctatttcttccagtgactttataccttagtc
 L  S  T  L  V  F  L  T  F  S  I  S  S  S  D  F  I  P  L  V
aacctcatcggcatctacttccaaatcagagacgatctgatgaa
 N  L  I  G  I  Y  F  Q  I  R  D  D  L  M
```

Because these fragments contain intron(s), cDNA sequences for the fragments were generated using proposed splice sites.[34] The typical 5' splice site starts from the sequence GTxxGT, while the typical 3' splice site ends in CAG or TAG. The cleavage occurs at the 5' end of GTxxGT and the 5' end of (C/T)AG. Figure 2.21 shows how the cDNA sequences were proposed according to the splice sites.

```
P322            CACCGTGGTCAAGGCCTCGAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAA 60
P345            CATCGCGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAG 60
P319            ---------------CTTGATTTGTTCTGGAGAGATACCTTTACCTGCCCGACGGAGGAA 45
                               ** **  *  * *** * *** *  * *  ** **  * *****

P322            GAATACATTTCGATGGTCAACAATAGTGAGTTAACCTGGGGCC-AGTCTAAG-------C 112
P345            GAGTATATCTCCATGGTTAACAATAGTTAGTGGCTTTTTCATT-ACTTGAGATGAGGACT 119
P319            GAGTATGTTTCTATGGTCAATAATAGTGAGTTGGTTCTTGGCCCACCCGAGGCTTAGTCT 105
                ** **  * ** ***** ** ****** ***            *    *

P322            CTACGGTTAGACTCATTTCGTA--------TCAGAAACCAGTGGCCTCCTACGTATCGGT 164
P345            CGATACTCATGATCATTTCGTTGAAATTTTCTAGAGACGGGCGGTCTTCTACGCATCGGC 179
P319            TAGAACACTAGACTGATTGATA-----CCGCTAGAAACTAGCGGCCTTCTTCGCCTTGCT 160
                           **  *          *** **  * ** ** ** **  * *

P322            ATTAAACTCATGATGGCATGTGCCACCACAAACGCGGATATGTATGTGTTT-TTGGCAGT 223
P345            ATCAAACTCATGATGGCATGCGCGACCACCAACACTGGAGTGTACGTATTC-CATCCTTC 238
P319            ATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGACAGGTATGTGCCTGTGTTCAGC 220
                ** *  ************** ** ** ** **    *    *** **         *

P322            GCATTTTGGGGTTTGAACTGAATGGCCATTT--------TCTAGCGATTATATTCCTTTG 275
P345            TCATTGTCCACTCTGGTCTTTCTGACATTTTCTATTTCTTCCAGTGACTTTATACCCTTA 298
P319            TCCGCTTGCGC-CTAAATCAACTGTCCTAAT-------TCTTAGTGATTATGTCCCTCTT 272
                 *    *      *        ** *    *           ** ** * * * **  *

P322            ACAAACCTCATTGGAGTCTATTTCCAAATCCGGGATGACCTTATGAA 322
P345            GTCAACCTCATCGGCATCTACTTCCAAATCAGAGACGATCTGATGAA 345
P319            GTGAATCTCATTGGAGTATATTTCCAGATTCGAGACGACCTGATGAA 319
                   ** ***** **  * ** ***** **  * ** ** ** *****
```

Figure 2.21 Proposed splice sites for p319, p322 and p345. Splice sites are indicated by the colored letters. The 1st intron is the sequence between the blue letters, while the 2nd intron is the sequence between the red letters.

According to the splice sites, the cDNA sequences were proposed, with p319 corresponding to a 200 bp cDNA, both p322 and p345 with a 215 bp cDNA. The sequences of these three cDNAs and the protein translations of these sequences are shown below.

Proposed cDNA from p319:

```
CTTGATTTGTTCTGGAGAGATACCTTTACCTGCCCGACGGAGGAAGAGTATGTTTCTATGGTCAATAATAAAA
CTAGCGGCCTTCTTCGCCTTGCTATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGACAGTGATTATGT
CCCTCTTGTGAATCTCATTGGAGTATATTTCCAGATTCGAGACGACCTGATGAA

cttgatttgttctggagagataccttt acctgcccgacggaggaagagtatgtttctatg
 L  D  L  F  W  R  D  T  F  T  C  P  T  E  E  E  Y  V  S  M
gtcaataataaaactagcggccttcttcgccttgctatcaggctcatgatggcatgtgcg
 V  N  N  K  T  S  G  L  L  R  L  A  I  R  L  M  M  A  C  A
acgaccaataaggacagtgattatgtccctcttgtgaatctcattggagtatatttccag
 T  T  N  K  D  S  D  Y  V  P  L  V  N  L  I  G  V  Y  F  Q
attcgagacgacctgatgaa
 I  R  D  D  L  M
```

Proposed cDNA from p322:

```
CACCGTGGTCAAGGCCTCGAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAAGAATACATTTCGA
TGGTCAACAATAAAACCAGTGGCCTCCTACGTATCGGTATTAAACTCATGATGGCATGTGCCACCACAAACGC
GGATATCGATTATATTCCTTTGACAAACCTCATTGGAGTCTATTTCCAAATCCGGGATGACCTTATGAA

caccgtggtcaaggcctcgaaatcttctggagagattccctcatctgtccctccgaggaa
 H  R  G  Q  G  L  E  I  F  W  R  D  S  L  I  C  P  S  E  E
gaatacatttcgatggtcaacaataaaaccagtggcctcctacgtatcggtattaaactc
 E  Y  I  S  M  V  N  N  K  T  S  G  L  L  R  I  G  I  K  L
atgatggcatgtgccaccacaaacgcggatatcgattatattcctttgacaaacctcatt
 M  M  A  C  A  T  T  N  A  D  I  D  Y  I  P  L  T  N  L  I
ggagtctatttccaaatccgggatgaccttatgaa
 G  V  Y  F  Q  I  R  D  D  L  M
```

Proposed cDNA from p345:

```
CATCGCGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAGGAGTATATCTCCA
TGGTTAACAATAAGACGGGCGGTCTTCTACGCATCGGCATCAAACTCATGATGGCATGCGCGACCACCAACAC
TGGAGTTGACTTTATACCCTTAGTCAACCTCATCGGCATCTACTTCCAAATCAGAGACGATCTGATGAA

catcgcgggcagggtctggaattactttggcgtgattcgttgacgtgcccttcggaggag
 H  R  G  Q  G  L  E  L  L  W  R  D  S  L  T  C  P  S  E  E
gagtatatctccatggttaacaataagacgggcggtcttctacgcatcggcatcaaactc
 E  Y  I  S  M  V  N  N  K  T  G  G  L  L  R  I  G  I  K  L
atgatggcatgcgcgaccaccaacactggagttgactttatacccttagtcaacctcatc
 M  M  A  C  A  T  T  N  T  G  V  D  F  I  P  L  V  N  L  I
ggcatctacttccaaatcagagacgatctgatgaa
 G  I  Y  F  Q  I  R  D  D  L  M
```

Translations of the three constructed cDNA sequences were found to have the conserved motifs ‘HRGQGMD’ and ‘QIRDDY’ or ‘MDLFWRD’ and ‘QIRDDY’ in the same frames. Translations of these sequences were also found to share high similarity with known fungal GGPP synthases such as *Neurospora crassa* (*N.c.*) (Figures 2.22). The presence of the KT motif in each is consistent with our hypothesis that these sequences represent real prenyltransferase genes. They also share high similarity with each other, with p319 and p322 being 78% identical, p319 and p345 being 75% identical and p322 and p345 being 85% identical (Figures 2.23).

```
P322    HRGQGLEIFWRDSLICPSEEEYISMVNNKTSGLLRIGIKLMMACATTNADIDYIPLTNLI 60
P345    HRGQGLELLWRDSLTCPSEEEYISMVNNKTGGLLRIGIKLMMACATTNTGVDFIPLVNLI 60
P319    -----LDLFWRDTFTCPTEEEYVSMVNNKTSGLLRLAIRLMMACATTNKDSDYVPLVNLI 55
N.c.    HRGQGMDLFWRDTLTCPTEDDYLEMVSNKTGGLFRLGIKLMQAESRS--PVDCVPLVNII 58
             ::::***:: **:*::*:.**.***.**:*:.*:** * : :    * :**.*:*

P322    GVYFQIRDDLM 71
P345    GIYFQIRDDLM 71
P319    GVYFQIRDDLM 66
N.c.    GLIFQIADDY- 68
        *: *** **
```

Figure 2.22. Alignment of the translated sequences from the proposed spliced genes with the GGPP synthase from the fungus *Neurospora crassa* (*N.c.*). KT motif is highlighted in red.

```
P322    HRGQGLEIFWRDSLICPSEEEYISMVNNKTSGLLRIGIKLMMACATTNADIDYIPLTNLI 60
P345    HRGQGLELLWRDSLTCPSEEEYISMVNNKTGGLLRIGIKLMMACATTNTGVDFIPLVNLI 60
P319    -----LDLFWRDTFTCPTEEEYVSMVNNKTSGLLRLAIRLMMACATTNKDSDYVPLVNLI 55
        *:::***:: **:****:*******.****:.*:********* . *::**.***

P322    GVYFQIRDDLM 71
P345    GIYFQIRDDLM 71
P319    GVYFQIRDDLM 66
        *:*********
```

Figure 2.23. Alignment of the translated sequences from the proposed spliced genes for p319, p322 and p345.

3'-Rapid amplification of the cDNA ends (3'-RACE) encoding the GGPP synthase

Although the BLAST searches strongly suggested that they are *ggs* gene fragments, more experiments were needed to confirm their identities. Extension of the sequences using 3'-RACE should help confirm their identities as well as clearly defining the splice sites. For the 3'-RACE method that was used, the RNA population was reverse transcribed to cDNA using the adaptor with oligo(dT) residue at the 3' end and the adaptor primer sequence at the 5' end. PCR was carried out using a gene-specific primer as the forward primer and the adaptor sequence as the reverse primer.

Because these three *ggs* genes share high similarity, the primers for the PCR amplifications must be designed with significant differences, so that the individual templates can be separately amplified. Also, nested PCR was performed to increase the specificity. The first 20 bp at the 5'-end of the sequences were avoided for designing the primers because they were generated from degenerate primers and could have some errors. However, because of the short lengths of these three sequences (about 300 bp) and other requirements for the primers, the 5'-end of the p319 sequence was still used to design the primer for the first round PCR, in the hope that the relatively low PCR annealing temperatures can help generate PCR products and that nested PCR can help generate a specific band. The regions used to design the primers are shown in Figure 2.24. Six primers were designed and their sequences are shown in Table 2.5. The three primers for the second round PCR contain some restriction sites for later cloning into available vectors. The adaptor used for the 3'-RACE contains *Xho*I, *Sal*I, and *Spe*I restriction sites and oligo-dT_{17}.

```
P322 CATCGGGGTCAAGGTATGGAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAA 60
P345 CATCGGGGTCAGGGTATGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAG 60
P319 ---------------ATGGATTTATTTTGGGGGGATACCTTTACCTGCCCGACGGAGGAA 45
                     *****  *  * *** * *** *  * *  ** **  * *****

P322 GAATACATTTCGATGGTCAACAATAAAACCAGTGGCCTCCTACGTATCGGTATTAAACTC 120
P345 GAGTATATCTCCATGGTTAACAATAAGACGGGCGGTCTTCTACGCATCGGCATCAAACTC 120
P319 GAGTATGTTTCTATGGTCAATAATAAAACTAGCGGCCTTCTTCGCCTTGCTATCAGGCTC 105
     ** **  * ** ***** ** ***** **  * ** ** ** **  * *  ** *  ***

P322 ATGATGGCATGTGCCACCACAAACGCGGATATCGATTATATTCCTTTGACAAACCTCATT 180
P345 ATGATGGCATGCGCGACCACCAACACTGGAGTTGACTTTATACCCTTAGTCAACCTCATC 180
P319 ATGATGGCATGTGCGACGACCAATAAGGACAGTGGTTATGTCCCTCTTGTGAATCTCATT 165
     ********** ** ** ** **    *     *  * * * **  *    ** *****

P322 GGAGTCTATTTCCAAATACGCGACGACTACATGAA 215
P345 GGCATCTACTTCCAAATACGCGACGACTACATGAA 215
P319 GGAGTATATTTCCAGATACGCGACGACTAATCGAA 200
     **  * ** ***** **************   ***
```

Figure 2.24. The primers for *P. passeckerianus* 3'-RACE (highlighted in red). The blue underlined one is for a second attempt at 3'-RACE PCRs.

Table 2.5. The primers for *P. passeckerianus* 3'-RACE.

Templates	1st round PCR	2nd round PCR
P319	TTTATTTTGGGGGGATACCT	CCTTGCTATCAGGCTCATGA
P322	AATCTTCTGGAGAGATTCCC	CGCGGATATCGATTATATTC
P345	ATTACTTTGGCGTGATTCGT	TGGCATGCGCGACCACCAAC

In order to get good results, the RNA was heated to 65 °C for 5 min to get rid of the secondary structure before performing the reverse transcription. For PCR, 50 °C and 55 °C annealing temperature were tried. It was found that there was no obvious difference for the PCR from these different conditions (Figure 2.25).

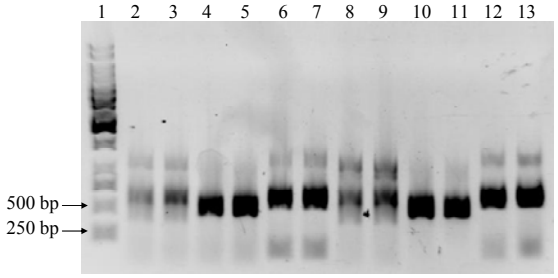


Figure 2.25. 3'-RACE 2nd round PCR result from nested PCR for *P. passeckerianus ggs* genes p319, p322 and p345. Lanes 2 to 7 were PCR results using the RNA that was not heated to 65 °C before performing reverse transcription. Lanes 8 to13 were PCR results using the RNA heated to 65 °C for 5 min to get rid of the secondary structure before performing reverse transcription. Lanes 2, 3, 8 and 9 were results for p319. Lanes 4, 5, 10 and 11 were results for p322, while lanes 6, 7, 12 and 13 were results for p345. Lanes 2, 4, 6, 8, 10 and 12 were results using 50°C annealing temperature for PCR. Lanes 3, 5, 7, 9, 11 and 13 were results using 55°C annealing temperature for PCR. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

For p322 and p345, the RACE reaction gave single strong broad bands, while for p319, two weaker bands were observed. The PCR products were digested with restriction enzymes and cloned into vectors. The p345 PCR product was

successfully cloned into a *Sph*I and *Sal*I digested pGEMT-EASY vector. A 582 bp cDNA (without counting polyA at the 3' end) was obtained. The sequencing results for eight different colonies showed that the numbers of A at the 3' end of PCR products varied from 15 to 37 (the number of As in the mRNA polyA tail could range from 20 to 250). The PCR products that appeared as a strong and broad band were actually the combination of multiple bands of similar sizes with varied numbers of A at the 3' end (T at the 5' end of complementary strands). The varied numbers of A/T depend on the position of A on the mRNA template that hybridized with the oligo-dT primer.

The p322 RACE PCR products were digested with *Cla*I and *Xho*I and cloned into *Cla*I and *Xho*I digested pBluescript KS vector. The digested PCR products were found to have similar sizes as the original ones, indicating that there were no internal *Cla*I and *Xho*I restriction sites. Unexpectedly, the clones did not give the right digestion patterns. Because p319 did not yield a single band and cloning the p322 was unsuccessful, new gene-specific primers for the second round PCR were designed for p322 and p319 which included an additional *Sph*I restriction site at the 5' end. The primer sequence is ACATGCATGCCGCGGATATCGATTATATTC for p322 and ACATGCATGCTAAGGACAGTGGTTATGTCC for p319. The first round PCR still used the same primers and the other PCR conditions were the same as used previously, but only the 55 °C annealing temperature was used. P322 gave a strong band again, while p319 still did not give single strong PCR product (Figure 2.26).

The second round PCR product of p322 was cloned into a *Sph*I and *Sal*I digested pGEMT-EASY vector. A 449 bp cDNA was obtained.

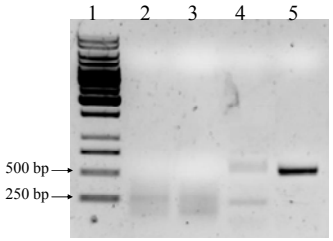


Figure 2.26. 3'-RACE PCR result from nested PCR for *P. passeckerianus ggs* genes p319, and p322. Lanes 2 and 3 were results from PCR 1st round. Lanes 4 and 5 were results from PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 and 3 respectively. Lanes 2 and 4 were results for *ggs* gene p319. Lanes 3 and 5 were results for *ggs* gene p322. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

Further along in the project, the p319 sequence used to design the first round PCR primer was found to have four bases in error. This could be the reason that there was no successful RACE PCR amplification for p319.

The obtained p322 and p345 cDNA sequences were found to share about 85% identity to a known fungal GGPP synthase cDNA from the fungus *Cryptococcus neoformans*, which further supports their identity as GGPP synthase genes.

Genome walking

Genome walking was used to extend the sequence of each of the three PCR products. Genome walking is a method that extends the sequence by walking upstream or downstream along the genomic DNA from a known sequence.[35] Genomic DNA was digested with four blunt-end restriction enzymes *Dra*I, *Eco*RV,

*Pvu*II and *Stu*I to give four different genomic DNA library pools, sdDL1 to sdDL4, respectively for *S. araneosa* and pdDL1 to pdDL4 respectively, for *P. passeckerianus*. Unique adaptors were ligated to the blunt ends to provide known sequence for the genome walking. Gene specific primers for nested PCR were designed on the basis of the obtained sequence. For the first time genome walking, the primers were designed from the information from p319, p322 and p345. The gene specific primers are 26 ~ 28 bp long, and the T_ms are 70 °C to ensure specific amplifications with the 67 °C PCR annealing temperature. Also, the GC content is 40-60%, and there are no more than three G's and C's in the last six positions at the 3' end of the primer. Tables 2.6 and 2.7 show the primers and the other information for PCR and the PCR results are shown in Figures 2.27 and 2.28. Not every library pool gave PCR products and the sizes of the fragments varied depending on the distribution of the restriction sites on the sequence. The largest strong single PCR fragment out of all four library pools was cloned and sequenced, and the sequence information was used to design primers for successive genome walking. For the first and second genome walking experiments, all of the strong single fragments regardless of their sizes were cloned and sequenced in order to test the method and reliability of the results. All of the strong single fragments turned out to be correct and real amplifications; the common sequences they shared were identical.

Table 2.6. PCR primers for genome walking.

DNA library	Sample names		1st round PCR primers	Sample names		2nd round PCR primers
SdDL1 (*Dra*I digested genomic DNA)	Sord224: upstream downstream	Sd1-1-1A Sd1-1-2A	Sord224-32-58 Sord224-154-180	Sord224 upstream downstream	Sd1-1-1B Sd1-1-2B	Sord224-66-91 Sord224-107-133
	Sord247: upstream downstream	Sd1-2-1A Sd1-2-2A	Sord247-37-63 Sord247-185-211	Sord247 upstream downstream	Sd1-2-1B Sd1-2-2B	Sord247-67-93 Sord247-131-157
SdDL2 (*Eco*RV digested genomic DNA)	Sord224: upstream downstream	Sd2-1-1A Sd2-1-2A	Sord224-32-58 Sord224-154-180	Sord224 upstream downstream	Sd2-1-1B Sd2-1-2B	Sord224-66-91 Sord224-107-133
	Sord247: upstream downstream	Sd2-2-1A Sd2-2-2A	Sord247-37-63 Sord247-185-211	Sord247 upstream downstream	Sd2-2-1B Sd2-2-2B	Sord247-67-93 Sord247-131-157
SdDL3 (*Puv*II digested genomic DNA)	Sord224: upstream downstream	Sd3-1-1A Sd3-1-2A	Sord224-32-58 Sord224-154-180	Sord224 upstream downstream	Sd3-1-1B Sd3-1-2B	Sord224-66-91 Sord224-107-133
	Sord247: upstream downstream	Sd3-2-1A Sd3-2-2A	Sord247-37-63 Sord247-185-211	Sord247 upstream downstream	Sd3-2-1B Sd3-2-2B	Sord247-67-93 Sord247-131-157
SdDL4 (*Stu*I digested genomic DNA)	Sord224: upstream downstream	Sd4-1-1A Sd4-1-2A	Sord224-32-58 Sord224-154-180	Sord224 upstream downstream	Sd4-1-1B Sd4-1-2B	Sord224-66-91 Sord224-107-133
	Sord247: upstream downstream	Sd4-2-1A Sd4-2-2A	Sord247-37-63 Sord247-185-211	Sord247 upstream downstream	Sd4-2-1B Sd4-2-2B	Sord247-67-93 Sord247-131-157
PdDL1 (*Dra*I digested genomic DNA)	Pleur319: upstream downstream	Pd1-1-1A Pd1-1-2A	Pleur319-93-118 Pleur319-269-295	Pleur319 upstream downstream	Pd1-1-1B Pd1-1-2B	Pleur319-149-175 Pleur319-227-253
	Pleur345: upstream downstream	Pd1-2-1A Pd1-2-2A	Pleur345-91-117 Pleur345-157-183	Pleur345 upstream downstream	Pd1-2-1B Pd1-2-2B	Pleur345-153-179 Pleur345-117-91
	Pleur322: upstream downstream	Pd1-3-1A Pd1-3-2A	Pleur322-90-116 Pleur322-251-277	Pleur322 upstream downstream	Pd1-3-1B Pd1-3-2B	Pleur322-161-187 Pleur322-187-161
PdDL2 (*Eco*RV digested genomic DNA)	Pleur319: upstream downstream	Pd2-1-1A Pd2-1-2A	Pleur319-93-118 Pleur319-269-295	Pleur319 upstream downstream	Pd2-1-1B Pd2-1-2B	Pleur319-149-175 Pleur319-227-253
	Pleur345: upstream downstream	Pd2-2-1A Pd2-2-2A	Pleur345-91-117 Pleur345-157-183	Pleur345 upstream downstream	Pd2-2-1B Pd2-2-2B	Pleur345-153-179 Pleur345-117-91
	Pleur322: upstream downstream	Pd2-3-1A Pd2-3-2A	Pleur322-90-116 Pleur322-251-277	Pleur322 upstream downstream	Pd2-3-1B Pd2-3-2B	Pleur322-161-187 Pleur322-187-161
PdDL3 (*Puv*II digested genomic DNA)	Pleur319: upstream downstream	Pd3-1-1A Pd3-1-2A	Pleur319-93-118 Pleur319-269-295	Pleur319 upstream downstream	Pd3-1-1B Pd3-1-2B	Pleur319-149-175 Pleur319-227-253
	Pleur345: upstream downstream	Pd3-2-1A Pd3-2-2A	Pleur345-91-117 Pleur345-157-183	Pleur345 upstream downstream	Pd3-2-1B Pd3-2-2B	Pleur345-153-179 Pleur345-117-91
	Pleur322: upstream downstream	Pd3-3-1A Pd3-3-2A	Pleur322-90-116 Pleur322-251-277	Pleur322 upstream downstream	Pd3-3-1B Pd3-3-2B	Pleur322-161-187 Pleur322-187-161
PdDL4 (*Stu*I digested genomic DNA)	Pleur319: upstream downstream	Pd4-1-1A Pd4-1-2A	Pleur319-93-118 Pleur319-269-295	Pleur319 upstream downstream	Pd4-1-1B Pd4-1-2B	Pleur319-149-175 Pleur319-227-253
	Pleur345: upstream downstream	Pd4-2-1A Pd4-2-2A	Pleur345-91-117 Pleur345-157-183	Pleur345 upstream downstream	Pd4-2-1B Pd4-2-2B	Pleur345-153-179 Pleur345-117-91
	Pleur322: upstream downstream	Pd4-3-1A Pd4-3-2A	Pleur322-90-116 Pleur322-251-277	Pleur322 upstream downstream	Pd4-3-1B Pd4-3-2B	Pleur322-161-187 Pleur322-187-161

Table 2.7. PCR primers for genome walking.

Primer name	Primer sequences
pleur319-93-118	AAGCGGAGCTGAACACAGGCACATAC
pleur319-149-175	ATGAGCCTGATAGCAAGGCGAAGAAGG
pleur319-227-253	AATAGTGAGTTGGTTCTTGGCCCACCC
pleur319-269-295	TTTACCTGCCCGACGGAGGAAGAGTAT
pleur345-91-117	ACCAGAGTGGACAATGAGAAGGATGGA
pleur345-153-179	ATCATGAGTTTGATGCCGATGCGTAGA
pleur345-117-91	TCCATCCTTCTCATTGTCCACTCTGGT
pleur345-157-183	GTCTTCTACGCATCGGCATCAAACTCA
pleur322-90-116	CCCAAAATGCACTGCCAAAAACACATA
pleur322-161-187	CGATACGTAGGAGGCCACTGGTTTCTG
pleur322-187-161	CAGAAACCAGTGGCCTCCTACGTATCG
pleur322-251-277	TGTCCCTCCGAGGAAGAATACATTTCG

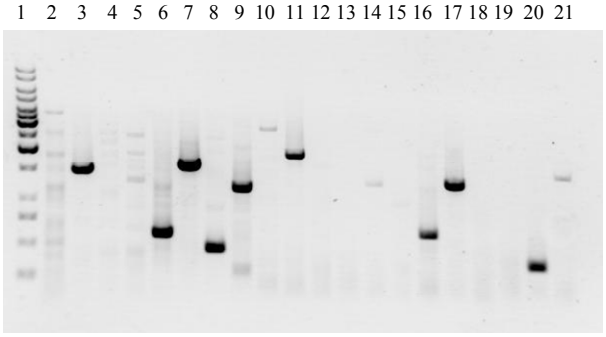


Figure 2.27. Genome walking nested PCR 2nd round products amplified from *P. passeckerianus* and *S. araneosa* genomic DNA. Lanes 2 to 9 were from *S. araneosa*, while lanes 10 to 21 were from *P. passeckerianus*. Lane 2 was Sd1-1-1B. Lane 3 was Sd1-1-2B. Lane 4 was Sd1-2-1B. Lane 5 was Sd1-2-2B. Lane 6 was Sd2-1-1B. Lane 7 was Sd2-1-2B. Lane 8 was Sd2-2-1B. Lane 9 was Sd2-2-2B. Lane 10 was Pd1-1-1B. Lane 11 was Pd1-1-2B. Lane 12 was Pd1-2-1B. Lane 13 was Pd1-2-2B. Lane 14 was Pd1-3-1B. Lane 15 was Pd1-3-2B. Lane 16 was Pd2-1-1B. Lane 17 was Pd2-1-2B. Lane 18 was Pd2-2-1B. Lane 19 was Pd2-2-2B. Lane 20 was Pd2-3-1B. Lane 21 was Pd2-3-2B. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

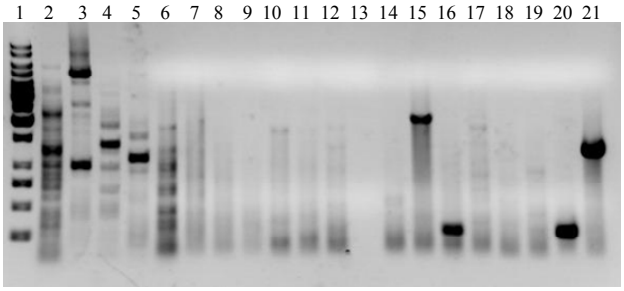


Figure 2.28. Genome walking nested PCR 2nd round products amplified from *P. passeckerianus* and *S. araneosa* genomic DNA. Lanes 2 to 9 were from *S. araneosa*, while lanes 10 to 21 were from *P. passeckerianus*. Lane 2 was Sd3-1-1B. Lane 3 was Sd3-1-2B. Lane 4 was Sd3-2-1B. Lane 5 was Sd3-2-2B. Lane 6 was Sd4-1-1B. Lane 7 was Sd4-1-2B. Lane 8 was Sd4-2-1B. Lane 9 was Sd4-2-2B. Lane 10 was Pd3-1-1B. Lane 11 was Pd3-1-2B. Lane 12 was Pd3-2-1B. Lane 13 was Pd3-2-2B. Lane 14 was Pd3-3-1B. Lane 15 was Pd3-3-2B. Lane 16 was Pd4-1-1B. Lane 17 was Pd4-1-2B. Lane 18 was Pd4-2-1B. Lane 19 was Pd4-2-2B. Lane 20 was Pd4-3-1B. Lane 21 was Pd4-3-2B. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

There were no strong and specific PCR amplifications for p345 in the first round of genome walking (Figures 2.27 and 2.28). Therefore, new GenomeWalker libraries were constructed using different blunt-end restriction enzymes. The enzymes *Hpa*I and *Sma*I were used even though they are six-basepair enzymes. These two new libraries did give some strong, single bands, about 200 bp (Figure 2.29 lane 22) for upstream and 800 bp for downstream (Figure 2.29 lane 25).

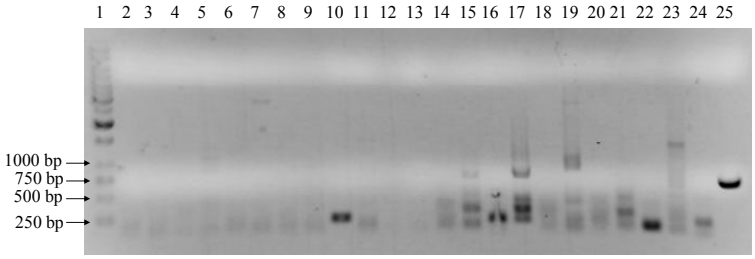


Figure 2.29. Genome walking nested PCR results for p345 using both old and new

(Figure 2.29 continued) libraries. Lanes 2 to 13 were products from PCR 1st round. Lane 2 was p345 upstream amplified from *Dra*I digested library pool. Lane 3 was p345 downstream amplified from *Dra*I digested library pool. Lane 4 was p345 upstream amplified from *Eco*RV digested library pool. Lane 5 was p345 downstream amplified from *Eco*RV digested library pool. Lane 6 was p345 upstream amplified from *Pvu*II digested library pool. Lane 7 was p345 downstream amplified from *Pvu*II digested library pool. Lane 8 was p345 upstream amplified from *Stu*I digested library pool. Lane 9 was p345 downstream amplified from *Stu*I digested library pool. Lane 10 was p345 upstream amplified from *Hpa*I digested library pool. Lane 11 was p345 downstream amplified from *Hpa*I digested library pool. Lane 12 was p345 upstream amplified from *Sma*I digested library pool. Lane 13 was p345 downstream amplified from *Sma*I digested library pool. Lanes 14 to 25 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 13, respectively. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

Successive genome walking was conducted using the gene-specific primers designed from the newly obtained sequences and walking continued both upstream and downstream of the putative *ggs* genes. Because the primers were designed from the far ends of the sequences of the library pools gave the longest fragments, theoretically, only the rest of the library pools were needed for the successive genome walking. However, sometimes all library pools were still used in order to make sure no data was lost, especially when the PCR product was not strong or more than one band was generated in the former reaction. Tables 2.8 and 2.9, and Figures 2.30 and 2.31 show the PCR primers and 2nd round PCR results for the successive PCRs for s224, p319 and p322.

Table 2.8. Genome walking PCR primers for upstream sequences of s224, p319 and p322.

Template	PCR	Primer name	Primer sequences
S224	1st round	S224u-1	CCGATTATTGCATCTGCTACCCTCACA
	2nd round	S224u-2	GCGTCTCATCACTGCGTTTTGTCATTT
P319	1st round	p319u-1	ACTCGCCCAGCCTCAGAAAATTCTACC
	2nd round	p319u-2	ACCGGACTTGAATCTGAATGGATGGAG
P322	1st round	p322u-1	GGAGTTTCCACGTCCTTCCGAGTAGTG
	2nd round	p322u-2	GCCCTAATTTCAAGGAGGGGGAAGGTA

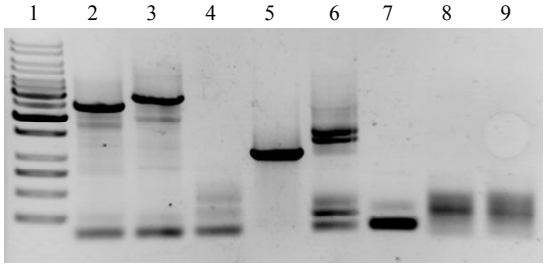


Figure 2.30. Genome walking nested PCR 2nd round products for upstream sequences of s224 using *S. araneosa* library pools, and for p319 and p322 using *P. passeckerianus* library pools. Lanes 2 to 4 were amplification for s224. Lanes 5 and 6 were amplification for p319. Lanes 7 to 9 were amplification for p322. Lane 2 was amplified from *Dra*I digested library pool sdDL1. Lane 3 was amplified from *Eco*RV digested library pool sdDL3. Lane 4 was amplified from *Stu*I digested library pool sdDL4. Lane 5 was amplified from *Eco*RV digested library pool pdDL2. Lane 6 was amplified from *Stu*I digested library pool pdDL4. Lane 7 was amplified from *Eco*RV digested library pool pdDL2. Lane 8 was amplified from *Pvu*II digested library pool pdDL3. Lane 9 was amplified from *Stu*I digested library pool pdDL4. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

Table 2.9. PCR primers for genome walking.

Amplifications	PCR	Primer name	Primer sequences
P319 upstream	1st round	P319u3rd-1	CACTCAAGGTCAATGCTGGGACAAGATA
	2nd round	P319u3rd-2	GGAAGAGAAGCTCACGACAGCCCAATA
P319 downstream	1st round	P319d2nd-1	CGACCAACGAGCTGAAATCGAACTCTA
	2nd round	P319d2nd-2	TGTGTGGTCCGCCATGAATCTAGACATA
P322 upstream	1st round	P322u3rd-1	TAAAGAGGTGCACTTTGCCGGTGTTC
	2nd round	P322u3rd-2	TTCACGAAGTCTGGAGTGGCAAGGAT
P322 downstream	1st round	P322d2nd-1	GCCTCCTAGCTTGTGCTTCGCAATATC
	2nd round	P322d2nd-2	CAGATTTGCCGGTTAGTTGTTCGGAGTA

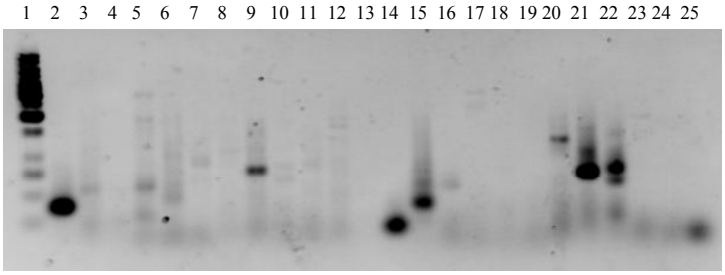


Figure 2.31. Genome walking nested PCR products for p319 and p322 using *P. passeckerianus* library pools. Lanes 2 to 13 were PCR 1st round results. Lane 2 was amplified for p319 upstream from *Dra*I digested library pool pdDL1. Lane 3 was amplified for p319 upstream from *Pvu*II digested library pool pdDL3. Lane 4 was amplified for p319 upstream from *Stu*I digested library pool, pdDL4. Lane 5 was amplified for p322 upstream from *Dra*I digested library pool pdDL1. Lane 6 was amplified for p322 upstream from *Pvu*II digested library pool pdDL3. Lane 7 was amplified for p322 upstream, *Stu*I digested library pool pdDL4. Lane 8 was amplified for p319 downstream from *Eco*RV digested library pool pdDL2. Lane 9 was amplified for p319 downstream from *Pvu*II digested library pool, pdDL3. Lane 10 was amplified for p319 downstream from *Stu*I digested library pool pdDL4. Lane 11 was amplified for p322 downstream from *Dra*I digested library pool pdDL1. Lane 12 was amplified for p322 downstream from *Eco*RV digested library pool pdDL2. Lane 13 was amplified for p322 downstream from *Stu*I digested library pool pdDL4. Lanes 14 to 25 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 13, respectively. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

The genome walking method was successful for obtaining the full sequences for the p319, p322 and p345 projected *ggs* genes and some initial sequences flanking these genes. From the DNA sequences obtained from genome walking and by comparing these sequences with those of other known fungal *ggs* genes, all of the fungal *ggs* gene conserved motifs were located in the obtained sequences for p319, p322 and p345. The stop codons for the p322 and p345 projected *ggs* genes were previously located from the 3'-RACE. The start codons for the p319, p322 and p345 projected *ggs* genes and the stop codon for the p319 projected *ggs* gene were

proposed by referring to the sequences of known fungal *ggs* genes. The proposed genomic DNA sequence of 1646 bp for the full putative *ggs* gene derived from p319, which was designated as Pleur319, that of 1583 bp gene derived from p322, which was designated as Pleur322 and that of 1367 bp gene derived from p345, which was designated as Pleur345, were obtained. The sequences of these putative *ggs* genes are shown in the appendix. Full cDNA sequences corresponding to the three putative *ggs* fragments were obtained from the DNA sequences obtained from genome walking and by using proposed splice sites.[34] The translations of the three sequences were expected to have high identity with known fungal GGPP synthases as illustrated in Figure 2.32. The BLAST search results strongly support their identities as GGPP synthases, but biochemical tests would be necessary to confirm this. The top five homologs out of many for the three potential *ggs* genes, designated as Pleur319, Pleur322 and Pleur345 respectively, are shown in Table 2.10. The translations of Pleur319, Pleur322 and Pleur345 show high identity with each other, with Pleur319 and Pleur322 70% identical, Pleur319 and Pleur345 66% identical, and Pleur322 and Pleur345 75% identical.

```
Pleur322 RPYTYTVSAPGKDIRRKLIDSFNLWLNVPDDKLDIIVNIVNMLHNASLLLDDIEDDSQLR 109
Pleur345 EPFTYITSNPGKDIRSKLIDAFNVWLNVPADKLIVITKIVNMLHAASLLVDDIEDDSQLR 87
Pleur319 EPYAYTVTAPGKGIRQILIEAFNVWLNVPEDKLLLIKNVVDMLHNASLLIDDIEDDSHLR 91
Phoma    -PYDYIAAKPGKEVRTLLLACFDEWLQVPPESLEVIGQVVRMLHTASLLIDDIQDNSELR 117
          *: * .: *** :*  *: .*: **:** :.* :* ::* *** ****:***:*:*.**

Pleur322 RGQPVAHKIYGVPQTINAANYVWLKSYLEVARLHQLDPSNATDFDAIVTEELIWLHRGQG 169
Pleur345 RGQPVAHKIYGIPQTINTANYVFFLAFQELFKLLRTP-----DLDAIVNEELLSLHRGQG 142
Pleur319 RGQPAAHKVYGVPRTMNAAHYVWLNVYKELSKFDEPN---------SSNHEMIYLHRGQG 142
Phoma    RGKPVAQNIFGTALTINSANYVYFLALEKLNSLKNPN------ITDIFTEELLRLHRGQA 171
         **:*.*::::* . *:*:*:**::    ::  : .             ..*:: *****.

Pleur322 LEIFWRDSLICPSEEEYISMVNNKTSGLLRIGIKLMMACATTNADIDYIPLTNLIGVYFQ 229
Pleur345 LELLWRDSLTCPSEEEYISMVNNKTGGLLRIGIKLMMACATTNTGVDFIPLVNLIGIYFQ 202
Pleur319 LDLFWRDTFTCPTEEEYVSMVNNKTSGLLRLAIRLMMACATTNKDSDYVPLVNLIGVYFQ 202
Phoma    MDLYWRDTLTCPTEEEYFEMVANKTGGLFRLMYRMMKAESSMP--IDLLPVVELLGVIFQ 229
         ::: ***:: **:****..** ***.**:*:  ::* * ::     * :*:.:*:*: **

Pleur322 IRDDLMNLDSTEYAQNKGFAEDLTEGKFSFPIVHGIHADTSNRHLYNIIQKRPENSMLKV 289
Pleur345 IRDDLMNLQSTEYEKNKGFAEDLTEGKFSFPIVHGVQADTSNRHILNVLQKRPSTPTLKI 262
Pleur319 IRDDLMNLDSTEYAQKKGFAEDLTEGKFSFPIIHGIQADTSSHFLLDVLRKRSGEDAFKI 262
Phoma    VVDDYKNLCSREYGKLKGFGEDLTEGKFSFPVIHSIRSNPEDLQLLHVLQQKSSNEHVKL 289
         : **  ** * ** : ***.***********::*.::::...  : .::::::.    .*:
```

Figure 2.32. Alignment of *P. passeckerianus* GGPP synthases with a representative fungal GGPPS from *Phoma betae*. Key conserved GGPPS regions are underlined. N-terminal and C-terminal extensions with low identity have been omitted.

Table 2.10. The top five homologs of the translated proteins of potential *ggs* genes from *P. passeckerianus*.

ggs genes	Homologues	Organisms	Identities	Similarities
Pleur322	GGPP synthase	*Mucor circinelloides*; Fungi; Zygomycota.	61%	74%
	GGPP synthase	*Cryptococcus neoformans*; Fungi; Basidiomycota.	59%	73%
	GGPP synthase	*Neotyphodium lolii*; Fungi; Ascomycota.	53%	71%
	GGPP synthase	*Epichloe festucae*; Fungi; Ascomycota.	53%	71%
	GGPP synthase	*Gibberella fujikuroi*; Fungi; Ascomycota.	52%	72%
Pleur319	GGPP synthase	*Mucor circinelloides*; Fungi; Zygomycota.	57%	75%
	GGPP synthase	*Cryptococcus neoformans*; Fungi; Basidiomycota.	57%	72%
	GGPP synthase	*Aspergillus clavatus*; Fungi; Ascomycota.	54%	71%
	GGPP synthase	*Neosartorya fischeri*; Fungi; Ascomycota.	53%	71%
	GGPP synthase	*Gibberella fujikuroi*; Fungi; Ascomycota.	52%	69%
Pleur345	GGPP synthase	*Cryptococcus neoformans*; Fungi; Basidiomycota.	68%	79%
	GGPP synthase	*Mucor circinelloides*; Fungi; Zygomycota.	63%	80%
	GGPP synthase	*Epichloe typhina*; Fungi; Ascomycota.	58%	77%
	GGPP synthase	*Gibberella fujikuroi*; Fungi; Ascomycota.	57%	77%
	GGPP synthase	*Neotyphodium lolii*; Fungi; Ascomycota.	57%	76%

Although the genome walking method provided some initial sequences flanking the Pleur319, Pleur322 and Pleur345 *ggs* genes, this method did not work in an efficient manner. Several rounds of genome walking were necessary and the length of the extended fragments was often quite small, which was also observed. So the method of construction and screening a *P. passeckerianus* genomic DNA library was used later on to further extend the sequences of these three fragments.

Construction and screening of a genomic DNA library

Construction of a genomic DNA library

The entire gene cluster for pleuromutilin is predicted to include a diterpene synthase gene, a dedicated *ggs* gene, two P-450 monooxygenases and an acyltransferase, along with regulatory and resistance genes. The aphidicolin gene cluster from *Phoma betae*, which is 15.6 Kb, includes a diterpene synthase gene, a dedicated *ggs* gene, two P-450s and a transporter gene and a transcription factor.[30] The partial gibberellin gene cluster in *Gibberella fujikuroi* is 15 Kb which includes a *ggs* gene, a diterpene synthase gene, and three cytochrome P-450 genes. Therefore, the gene cluster for pleuromutilin should be approximately 15-20 Kb.[36]

A genomic DNA library was constructed for *P. passeckerianus*. The number of colonies for getting at least double coverage was calculated using the formula N = Ln (1-P) / Ln (1- f), where P is the desired probability (expressed as a fraction); f is the proportion of the genome contained in a single clone; and N is the required number of fosmid clones.[37] The sizes of fungal genomes characterized to date are between 12.5 Mb and 40 Mb. For a 20 Mb genome, the number of clones required to ensure a 99% probability of a given DNA sequence being contained within a fosmid library composed of 40 Kb inserts are: N = Ln (1-0.99) / Ln (1- [4 x 10^4 / 20 x 10^6]) = 2300 colonies. For a 40 Mb genome, N = Ln (1-0.99) / Ln (1- [4 x 10^4 / 40 x 10^6]) = 4600 colonies. Therefore, in order to get at least double coverage, about 4000-8000 clones should be screened. A genomic DNA library of 6,500 colonies was constructed using a standard library construction procedure.[37]

Screening the genomic DNA library using homologous probes

The sequence information from the genomic walking experiments was used to design probes to screen the library by Southern analysis. The DNA fragments were obtained from available plasmids prepared for the genomic walking. A DNA fragment of 854 bp from Pleur345 was used as a probe and five positive fosmids were obtained which were later on found to be overlapping by Southern blotting. An 1133 bp DNA fragment obtained from Pleur322 was used as a probe and one positive fosmid was obtained. A 1075 bp DNA fragment obtained from Pleur319 was used as a probe and one positive was identified (Figure 2.33). The sequences of the three probes are shown in the Appendix. Confirmation of these positive fosmid DNAs was shown by Southern blotting using the same probes that were used to generate the positive fosmids, and by sequencing using the primers used in the genome walking experiments. The sequences of several Kb flanking both ends of Pleur319 and Pleur345, including several open reading frames, were obtained. However, no genes were found that seem to be related to the biosynthesis of pleuromutilin. The obtained sequence from the Pleur322 probe is truncated downstream of the Pleur322 *ggs* gene. Therefore, genome walking was used later on to extend the sequence (see *Genome walking for truncated Pleur322 downstream*). In addition, in case any interesting genes are located at the edges of the isolated fosmids, end sequencing of the positive fosmids was also performed from the PCC1FOS vector using the PCC1FOS vector forward and reverse primers. No interesting genes were found from these sequencing runs.

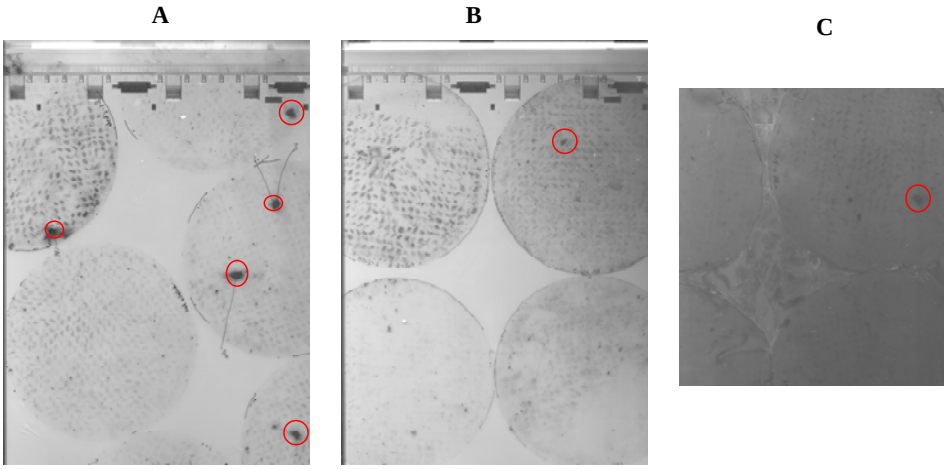


Figure 2.33. Genomic DNA library results. A: genomic DNA library screening results using a probe generated from Pleur345; B: genomic DNA library screening results using a probe generated from Pleur319; C: genomic DNA library screening results using a probe generated from Pleur322. The red circled ones are the positive colonies.

Screening the genomic DNA library using heterologous probes

In case the diterpene synthase genes are located somewhere outside of the sequenced regions, heterologous probes prepared from the putative diterpene synthase gene of the fungus *Botrytis cinerea* were used to screen the *P. passeckerianus* library. These heterologous probes were also used in Southern blotting to probe the positive Pleur319, Pleur322 and Pleur345 fosmids. *Botrytis cinerea* belongs to Ascomycetes fungi and is a pathogen that affects many plant species, such as grapes.[38] The putative diterpene synthase gene from *B. cinerea* was used because it was convenient to get a culture of the fungus from Dr. Henrik Stotz, OSU Department of Horticulture.

For known diterpene synthase genes, about 1.6 kb at the 3' end typically does not contain introns. The 3' end of the diterpene synthase gene from the fungus *B.*

cinerea shares about 50% identity with those of *Gibberella, Phoma*, and *Aspergillus*. Therefore, heterologous probes were prepared from the 3' end of the putative *B. cinerea* diterpene synthase gene. Genomic DNA was isolated and PCR was conducted using the primers designed from the cDNA sequence. The primer sequences and the sizes of the fragments are shown in Table 2.11. The PCR did give strong single bands of correct sizes (Figure 2.34). These three fragments of 911 bp, 1133 bp and 530 bp were used as probes to screen the *P. passeckerianus* genomic DNA library and also the positive Pleur319, Pleur322 and Pleur345 fosmids. Unfortunately, no positive hybridizations were found.

Table 2.11. PCR primers and the sizes of the fragments.

Forward primers	Reverse primers	Sizes
Botrytis-1858: GGCAAAGACAAGTACGTCTT	Botrytis-R8: GAACTCCGGAAAGCCGACAGAGTT	911 bp
Botrytis-1858: GGCAAAGACAAGTACGTCTT	Botrytis-R9: TATATCTGACCATACAGGTC	1133 bp
Botrytis-2461: TTCAAGGAGAATCTGCCTCT	Botrytis-R9: TATATCTGACCATACAGGTC	530 bp

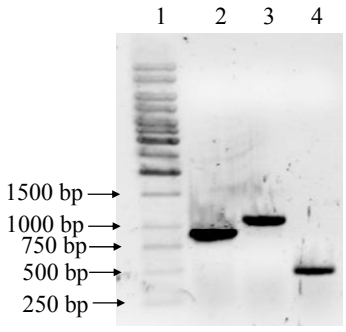


Figure 2.34. Amplified PCR fragments from *Botrytis cinerea* which would be used as heterologous probes. Lane 2 was PCR product (911 bp) with primers Botrytis-1858 and Botrytis-R8. Lane 3 was PCR product (1133 bp) with primers Botrytis-1858 and Botrytis-R9. Lane 4 was PCR product (530 bp) with primers Botrytis-2461 and Botrytis-R9. Lane 1 contains a 1 kb DNA ladder (Fermentas).

Genome walking to extend the Pleur322 downstream sequence

The positive fosmid DNA from screening the genomic DNA library using a Pleur322 probe did not give enough information because the fosmid insert terminated directly downstream of the Pleur322 *ggs* gene. Therefore, another round of genome walking was used to extend the sequence. New library pools were constructed with freshly harvested *P. passeckerianus* mycelia. These new library pools were tested using the gene-specific primers used with the old library pools, which are pleur322-251-277 for the 1st round PCR and pleur322-187-161 for the 2nd round PCR. The sizes of the PCR bands are the same as those obtained previously (Figures 2.27 and 2.28). Sequencing further confirmed the PCR results. The PCR result is shown in Figure 2.35.

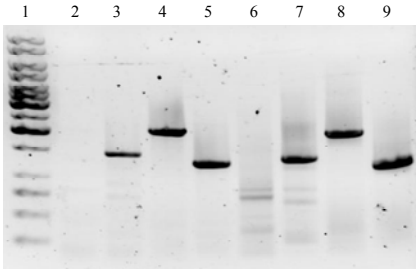


Figure 2.35. Genome walking nested PCR results for Pleur322 downstream. Lanes 2 to 5 were results from PCR 1st round. Lane 2 was amplified from *Dra*I library pool. Lane 3 was amplified from *Eco*RV library pool. Lane 4 was amplified from *Pvu*II library pool. Lane 5 was amplified from *Stu*I library pool. Lanes 6 to 9 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 5, respectively. Lane 1 contains a 1 kb DNA ladder (Fermentas).

Because no fragment was successfully amplified by walking successively downstream of the obtained fragment in the previous experiment (Figure 2.31), the PCR conditions were modified. The previous experiments followed the PCR

condition recommended by the GenomeWalker Kit, which recommends that when using the thermal cycler with a hot lid, the following two-step cycle parameters should be used: 94 °C, 2 sec and 72 °C, 3 min for 7 cycles, and then 94 °C, 2 sec and 67 °C, 3 min for 32 cycles, followed by the final extension of 67 °C, 7 min. It was found that when using this program, the thermal cycler could never get to the 94 °C denaturing temperature before the temperature started to go down. And whether the 94 °C, 2 sec cycle could completely denature the template is in question if its size is large. Therefore, a 95 °C, 25 sec denaturing condition was tried which was recommended when using a thermal cycler without a hot lid. The PCR result is shown in Figure 2.36. The amplified fragments were much stronger than those using the 94 °C, 2 sec denaturing condition (compared to Figure 2.35 lanes 2 to 5) and no non-specific fragments were amplified. Therefore, it was speculated that the modified PCR conditions may lead to some new results for those genome walking experiments that did not give PCR amplification. Successive genome walking was conducted using the same gene-specific primers, P322d2nd-1 and P322d2nd-2 (Table 2.9) that were used previously but did not give any amplification (Figure 2.31). This time, the amplifications were successful and gave strong single bands. The PCR results are shown in Figure 2.37.

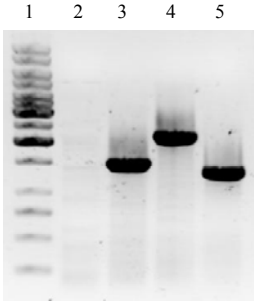


Figure 2.36. Genome walking PCR results for Pleur322 downstream. Lanes 2 to 5 used 95 °C 25 sec denaturing PCR condition; the other conditions were the same as those of lanes 2 to 5 respectively in Figure 2.35. Lane 1 contains a 1 kb DNA ladder (Fermentas).

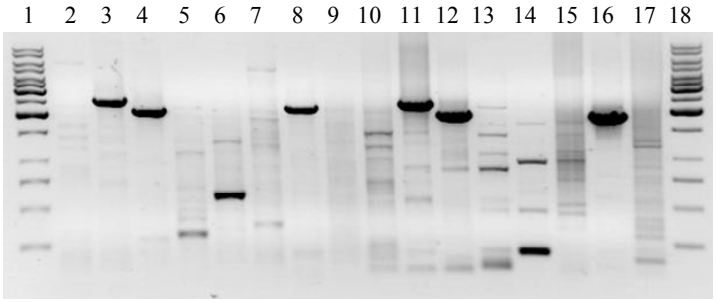


Figure 2.37. Genome walking nested PCR results for Pleur322 downstream (lanes 2 to 4 for the 1st round). The rest were results for *S. araneosa*. Lanes 2 to 4 were results from PCR 1st round. Lane 2 was from *Dra*I library pool. Lane 3 was from *Eco*RV library pool. Lane 4 was from *Stu*I library pool. Lanes 10 to 12 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 4, respectively. Lane 1 contains a 1 kb DNA ladder (Fermentas).

Gene organization near P. passeckerianus ggs genes

The combined sequencing results from genome walking and sequencing of fosmids are illustrated in Figure 2.38. About 12.6 kb of sequence was obtained surrounding Pleur319, about 11.3 kb surrounding Pleur322 and about 7.8 kb surrounding Pleur345. About 6 Kb sequence upstream and 5 Kb sequence downstream of Pleur319 putative *ggs* gene were obtained. The Pleur319 region has a *cis*-prenyltransferase gene and two hypothetical fungal protein genes (one is a partial gene) downstream and an ATP-dependent transporter and an ubiA (4-hydroxybenzoate polyprenyltransferase) homolog upstream (Figure 2.39). This suggests that the Pleur319 *ggs* gene is associated with a different isoprenoid pathway, possibly one involved with ubiquinone biosynthesis. For the Pleur345 *ggs* gene, more than 3 Kb sequences upstream and downstream of Pleur345 putative *ggs* gene were obtained. The upstream partial hypothetical protein hits

with a large hypothetical fungal protein whose gene size is 7.5 Kb, which indicates that it is not likely that there is a diterpene synthase gene upstream of this hypothetical protein. The downstream gene encoding the hypothetical protein does not seem to be related to isoprenoid biosynthesis by searching the homologous genes and checking the genes flanking those homologous genes from known sequences. Besides, in other fungal diterpene clusters, the diterpene synthase gene was located upstream of the *ggs* gene, but not downstream. For the Pleur322 *ggs* gene, about 6 Kb sequence upstream and 3.7 Kb sequence downstream of Pleur322 putative *ggs* gene were obtained. One open reading frame, a hypothetical protein, was found downstream of the *ggs* gene, and about 1 Kb of sequence was obtained downstream of this hypothetical protein. Two open reading frames, an ATP-dependent transporter and a (partial) hypothetical protein, were found upstream of *ggs* gene. Homolog searches do not reveal any connections between these genes and isoprenoid biosynthesis. Although the possibility can not be completely ruled out that the genes related to the pleuromutilin biosynthesis are located further away from the *ggs* gene, the chance is small according to the available information. Therefore, no further genomic DNA library sequencing and genome walking were performed.

Figure 2.39 shows that the gene organizations in several known and putative fungal diterpene biosynthetic gene clusters. Several of these clusters have been described during the course of this study. The gibberelin gene clusters in *Gibberella fujikuroi*[36] and *Phaeosphaeria* sp.[39] have been identified. The cluster in

Gibberella fujikuroi contains a dedicated *ggs* gene and related genes (P-450), while the one in *Phaeosphaeria* sp. does not contain a dedicated *ggs* gene, but the diterpene synthase is clustered with P-450 genes. Two putative diterpene synthase genes, #1 and #2, were found from *Aspergilus nidulans* through the genome mining projects of the Broad Institute. This indicates that a single fungus may have more than one diterpene gene cluster. For the #2 diterpene synthase gene, there is no *ggs* gene in the cluster. But there is a *ggs* gene upstream of the #1 diterpene synthase gene and these two genes are separated by a HMG-CoA reductase. There is also a *ggs* gene upstream of the putative diterpene synthase gene in *Aspergilus oryzae*. The *ggs* gene and the diterpene synthase gene are separated by two genes in this example, a P-450 and a small hypothetical protein. This clusters show the growing variation in the nature of fungal diterpene clusters. As can be seen, some clusters contain a dedicated *ggs* while others do not. In the ones that do have a dedicated *ggs*, the diterpene synthase gene always appears upstream of the *ggs*.

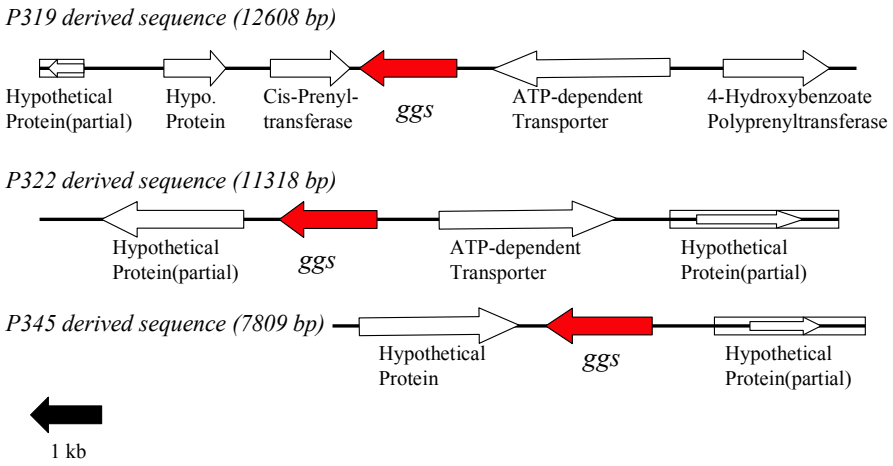


Figure 2.38. Gene organization near *P. passeckerianus ggs* genes.

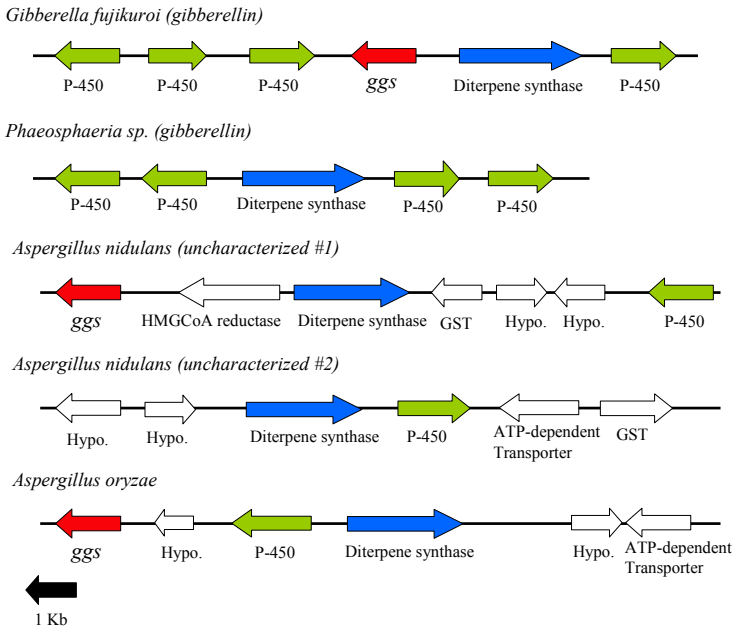


Figure 2.39. Known and putative fungal diterpene biosynthetic gene clusters. Hypo.: hypothetical protein.

Suppression Subtractive Hybridization (SSH)

Because the *ggs* gene does not appear to be clustered with the diterpene synthase gene, another method, Suppression Subtractive Hybridization (SSH) was tried. Suppression Subtractive Hybridization is a technique to find differentially expressed genes.[40] In this method, two populations of mRNA are compared and clones of genes that are expressed in one population, but not in the other, are revealed. The PCR-Select® cDNA Subtraction Kit (Clontech, CA, USA) was used and a subtracted library was formed from cDNA prepared from a non-producing culture and from cDNA from a production culture. The cDNAs were prepared from both a non-producing *P. passeckerianus* culture and a production culture,

which in this research are the cultures from the 2nd day that did not produce pleuromutilin and from the 7th day that produced pleuromutilin, which were confirmed by both mass spectrometry and NMR. The cDNAs were digested with *Rsa*I, which is a four basepair enzyme that gives blunt-ended fragments. This digestion results in shorter cDNA fragments which would lead to final PCR products of between 300 and 800 bp. The digested cDNA from the non-producing culture is called the tester DNA, while that from production culture is called the driver DNA. The tester DNA was divided into two parts, tester 1 and tester 2, and two adaptors, adaptor 1 and adaptor 2R were ligated to tester 1 and tester 2 respectively. The driver was not ligated to any adaptor.

Two hybridizations were conducted. In the first hybridization, excess amount of driver was added to each adaptor-ligated tester and the samples were heat denatured and allowed to anneal. Non-target cDNAs which were present in both the tester and driver cDNAs formed hybrids. Therefore, the resulting single-stranded cDNAs are dramatically enriched for differentially expressed sequences. In the second hybridization, the two samples from the first hybridization and the freshly-denatured driver which further enriched the differentially expressed sequences, were mixed together. Hybrid molecules which consist of differentially expressed cDNAs were formed. In addition, the concentration of high and low abundance sequences was equalized after two rounds of hybridizations because reannealing is faster for the more abundant molecules due to the second-order kinetics of hybridization.[41] The protruded ends were filled in and nested PCR was

performed. Only hybrid molecules with adaptor 1 on one end, and adaptor 2R on other end would be able to be PCR-amplified. The PCR mixture was enriched for differentially expressed cDNAs and differentially expressed transcripts that varied in abundance in the original mRNA sample should be present in roughly equal proportions.

The SSH PCR results are shown in Figure 2.40. For the *P. passeckerianus* subtracted sample (Figure 2.40 lanes 2 and 7), there is a light smear in the first round PCR and a much brighter smear in the second. The smear ranges from about 300 bp to 800 bp in size. There is a strong band of about 500 bp which may be a single band or combination of the bands of similar sizes. For the unsubtracted *P. passeckerianus* sample (Figure 2.40 lanes 3 and 8), the smear in the second round PCR was much lighter than that in the first round PCR because the non-specific amplification was reduced. The bright band around 500 bp that appeared in the subtracted sample, which should represent differentially expressed gene(s), did not show up in the unsubtracted sample. For the SSH Kit mRNA Control, 1 Kb DNA ladder was added to become the tester cDNA (Figure 2.40 lanes 4 and 9), and some similar size bands as the 1 Kb DNA ladder (Figure 2.40 lanes 1 and 12) appeared in the 2^{nd} round PCR. The reason that the sizes of the bands differ somewhat from the DNA ladder could not be explained by Clontech Technical Services. The unsubtracted control cDNA (Figure 2.40 lanes 5 and 10) did not show any similar size band and the smear is lighter. For the SSH Kit PCR Control subtracted cDNA (Figure 2.40 lanes 6 and 11), the specific bands are much

brighter in the 2nd round PCR than in the 1st round. The PCR results suggest that differentially expressed transcripts were indeed amplified. Therefore, the overall gel results suggest that the SSH method has worked although sequencing of the subtracted library will be necessary to find out the gene or genes that are associated with the pleuromutilin diterpene synthesis.

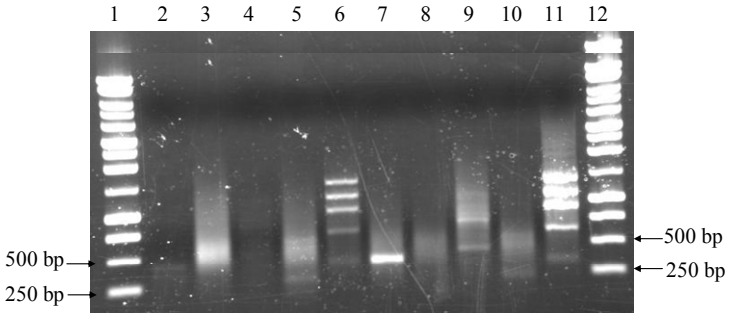


Figure 2.40. SSH nested PCR results for *P. passeckerianus* and controls. Lanes 2 to 6 were results from PCR 1st round. Lane 2 was result from subtracted *P. passeckerianus* cDNA. Lane 3 was result from unsubtracted *P. passeckerianus* cDNA. Lane 4 was result from unsubtracted SSH Kit Control mRNA derived cDNA added 1 kb ladder (Promega) to become subtracted cDNA. Lane 5 was result from unsubtracted SSH Kit Control mRNA derived cDNA. Lane 6 was result from SSH Kit PCR Control subtracted cDNA. Lanes 7 to 11 were from nested PCR 2nd round amplifications of 1/10 dilutions of the products shown in lanes 2 to 6, respectively. Lanes 1 and 12 contain 1 kb DNA ladder (Promega).

Conclusion

Three methods have been attempted to identify the diterpene synthase gene for the formation of pleuromutilin. The initial method to directly identify the diterpene synthase gene using degenerate primers was unsuccessful. The second method which utilizes an indirect approach via the fungal *ggs* genes, did generate some results. Three putative *ggs* genes have been found and about 10 Kb sequences

derived from these three *ggs* genes were obtained. Unfortunately, no diterpene biosynthetic gene has yet been found. Therefore, the SSH method was attempted to search for the differentially expressed genes in the culture that did not produce pleuromutilin and that did produce pleuromutilin. Sequencing of the subtracted library is in process. If fungal diterpene gene(s) or any of the related genes, such as P-450, acyltransferase or resistance genes are found, methods such as screening a genomic DNA library or genome walking will be used to obtain further sequence. If the SSH method is not successful, it may be necessary for the individual who takes over this project to use a reverse genetics approach to find the pleuromutilin diterpene synthase.

Materials and methods

General chemistry

Chemicals were purchased from Sigma-Aldrich, Fisher Scientific and VWR International.

General molecular biology

Routine molecular biology procedures including DNA manipulation, transformation, growth and maintenance of *Escherichia coli* (*E. coli*), competent cell preparation and electroporations were conducted according to standard techniques.[42] DNA precipitation and concentration used the following protocol:

200 µL of DNA solution, 20 µL (1/10 of DNA volume) of 3M sodium acetate, 3 µL of 10 mg / mL glycogen and 500 µL of EtOH, incubate on ice for 20 min. Plasmids were propagated in *Escherichia coli* DH10B. Restriction enzymes, T4 DNA ligase and *Taq* polymerase were purchased from various suppliers (Promega; Fermentas) and used according to the manufacturer's protocols. All PCRs were conducted in an Eppendorf Mastercycler gradient PCR thermal cycler. QIAprep® spin miniprep and QIAquick® gel extraction kits (Qiagen, Valencia, CA) were used for DNA purification. Sequencing of DNA inserts was conducted by the Center for Genome Research and Biocomputing (CGRB) at Oregon State University.

Fungal strains and growth conditions

P. passeckerianus Pilat (ATCC # 34646) was obtained from ATCC (American Type Culture Collection). Stock cultures were maintained on YM agar plates [Yeast Mold agar, Difco 0711. The recipe contains (/L): yeast extract 3.0 g, malt extract 3.0 g, peptone 5.0 g, dextrose 10.0 g, agar 20.0 g]. *P. passeckerianus* Pilat can be grown under shaking and static conditions. *P. passeckerianus* Pilat is grown on YM broth (Yeast Mold broth, Difco 0712) and YM agar at 25 °C. Under the shaking condition, the 1 liter culture was grown in a rotary shaker with 200 rpm constant shaking. A 100 mL 3-day culture in the same medium was used as inoculum. Under static condition, the cultures were grown on the bench at room temperature (around 25 °C) without shaking.

Isolation and purification of pleuromutilin

Extraction of culture and isolation of pleuromutilins

After 7 days the mycelium was removed by filtration. The culture filtrate was extracted with an equal volume of dichloromethane (CH_2Cl_2) three times. The combined organic phases were dried with anhydrous $MgSO_4$ and then evaporated to give the crude extract. The crude extract was then subjected to flash chromatography on a silica gel (Merck 60, 0.063 to 0.2 mm particle size) column (30 × 2 cm) and eluted with hexane / EtOAc 2:1 1000 mL. The fractions showing the same spots under UV and cerium sulfate stain, on silica TLC (thin layer chromatography) with hexane / EtOAc 1:1 solvent system, were combined. The R_f of pleuromutilin was 0.31 with hexane / EtOAc 1:1 solvent system on silica TLC. The combined eluents were dried with anhydrous $MgSO_4$, evaporated to give the partial purified compound. The partial purified compound was dissolved in EtOAc, treated with 2% (w/v) of active charcoal and concentrated *in vacuo*.

Mass Spectrometry (MS) of pleuromutilin

MS was performed on a ThermoFinnigan LCQ Advantage LC-MS system, consisting of a solvent pump, an autosampler, a PDA detector and an ion trap mass spectrometer detector. The system was controlled by a PC running Xcalibur 1.3 software. The MS was performed using (+) and (-) APCI with direct injection.

NMR analyses

The 1D and 2D NMR, ^{1}H, ^{13}C, COSY, DEPT, HSQC and HMBC were performed on Bruker AC 300 and AM 400 spectrometers. Chemical shifts are shown as δ. Reference is based on the NMR solvent as an internal standard ($CHCl_3$ at 7.26 ppm for ^{1}H NMR and $CDCl_3$ at 77.0 ppm for the center line of the triplet in ^{13}C NMR).

Time course of pleuromutilin production

Small (5 ml) cultures from the 2nd to 10th days were each extracted with dichloromethane (3 × 5) ml. The combined CH_2Cl_2 extracts were dried with anhydrous $MgSO_4$, evaporated in vacuum and then redissolved in 100 μL of isopropyl alcohol. MS was performed at (+) and (-) APCI.

Nucleic acid analysis

The extraction of genomic DNAs from fungi followed a literature procedure.[43] The mycelium was frozen in liquid nitrogen and ground into fine powder with a mortar and pestle. The mycelium powder was added TES buffer (100 mM Tris, pH 8.0, 10 mM EDTA, 2% SDS) and 50-100 μg Proteinase K; incubate at 37 °C for 1 hour and then 60 °C for 30 min. The lysate was adjusted to 1.4 M salt concentration with 5 M NaCl and added 1/10 volume of 10% CTAB and then incubated for 10 min at 60 °C. The solution was added 1 volume of phenol / chloroform and the

DNA was extracted into the aqueous layer. The DNA was then precipitated with isopropyl alcohol.

General preparation for handling RNA

All solutions were treated with 0.1% DEPC (diethylpyrocarbonate). DEPC 0.1 mL was added to 100 mL of the solution to be treated and stired overnight. The solution was autoclaved for 25 minutes to remove any trace of DEPC. Glassware was cleaned with a detergent, thoroughly rinsed, and oven baked at 240 °C overnight. Plasticware and the electrophoresis tank were thoroughly rinsed with RNase Zap (Biohit INC), followed by RNase-free water and then rinsed with ethanol and allowed to dry.

Formaldehyde agarose (FA) gel electrophoresis

Composition of FA gel buffers: 10× FA gel buffer: 200 mM 3-[N-morpholino]propanesulfonic acid (MOPS), 50 mM sodium acetate, 10 mM EDTA, adjust pH to 7.0 with NaOH. 1x FA gel running buffer: 100 mL 10× FA gel buffer, 20 ml 37% (12.3 M) formaldehyde and 880 ml RNase-free water. 5× RNA loading buffer: 16 µL saturated aqueous bromophenol blue solution, 80 µL 500 mM EDTA (pH 8.0), 720 µL 37% (12.3 M) formaldehyde, 2 mL 100% glycerol, 3.084 mL formamide, 4 mL 10× FA gel buffer and add RNase-free water to 10 mL. To prepare FA gel (1.2% agarose) of size 10× 14 x 0.7 cm, mix: 1.2 g agarose, 10 mL 10× FA gel buffer and add RNase-free water to 100 mL.

Total RNA isolation

Total RNAs were extracted from fungi by using the RNeasy Plant Mini Kit (Qiagen). The mycelium was harvested form the 2^{nd} day or the 7^{th} day culture by filtration and washed with ddH_2O twice. The mycelium was frozen in liquid nitrogen and ground into fine powder with a mortar and pestle. The mycelium powder was added lysis buffer; the resulting lysate was transfered to the spin column. After RNA binding and cleaning up, the total RNA was eluted from the column. Messenger RNA was purified from the resulting total RNA by using Oligotex mRNA Mini Kit (Qiagen). The kit contains Oligotex Suspension to which the $dC_{10}T_{30}$ oligonucleotides are covalently linked and thus mRNA can bind to the Oligotex Suspension and purified from total RNA. The cDNA was generated from mRNA by using the Omniscript RT Kit (Qiagen) (see ***Direct identification of the DS gene:*** *Reverse transcription-polymerase chain reaction (RT-PCR) for the diterpene synthase gene search*.)

Direct identification of the DS gene

Reverse transcription-polymerase chain reaction (RT-PCR) for the diterpene synthase gene search

OneStep RT-PCR Kit (Qiagen) and Omniscript Reverse Transcriptase (Qiagen) were used. For the OneStep RT-PCR Kit method, OneStep RT-PCR Enzyme Mix (containing Omniscript Reverse Transcriptase, Sensiscript Reverse Transcriptase and HotStar *Taq* DNA Polymerase) was used. The reverse transcription and PCR were performed in one tube. The degenerate primers used were designed from

known or putative fungal diterpene synthase genes. The primers for the PCR were added to the reaction mixture. RT-PCR was carried out in a final volume of 50 μL containing 1× RT-PCR buffer, 0.4 mM of each deoxynucleoside triphosphate (dNTP), 0.6 μM of each primer, 2 μL of Enzyme Mix, and 50 ng of mRNA. The reverse transcription was performed at 50 °C (the unique buffer allows this high reverse transcription temperature and improves the efficiency by disrupting the secondary structure) for 30 min and then, the temperature was raised to 95 °C for 15 min to inactivate the reverse transcriptase and activate the *Taq* Polymerase. The PCR program was carried out. The cycling conditions were as follows: 1 cycle of 1 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s annealing (varied temperatures between 45 and 55 °C), and 1 min at 72 °C; a final extension for 5 min at 72 °C.

RT-PCR was also performed by a two-step RT-PCR protocol. Reverse transcriptions were performed by following the Omniscript Reverse Transcription Kit (Qiagen). Omniscript Reverse Transcriptase was used which has RNA-directed DNA-polymerase activity, DNA-directed DNA-polymerase activity and RNase H activity. The reverse transcription reaction was performed in a final volume of 20 μL containing 1× RT buffer, 0.5 mM of each deoxynucleoside triphosphate (dNTP), 1 μM oligo-dT primer, 1 μL of Omniscript Reverse Transcriptase, and 2 μg of mRNA. The reaction mixture was incubated at 37 °C for 1 hr. For PCR, different annealing temperatures were tried: each degree from 42 °C to 52 °C and 54 °C, 56 °C, 58 °C, 60 °C, 64 °C. Both normal PCR and nested PCR were tried. PCRs were carried out in a final volume of 25 μL containing 1.5 mM $MgCl_2$, 0.4

mM of each deoxynucleoside triphosphate (dNTP), 0.4 μM of each primer, 2.5 U of Expand HF (Boehringer, Indianapolis IN, USA) or 0.5 U of *Taq* DNA polymerase (Fermentas), and 50 ng of cDNA. Templates used in the second round PCR are 1:50 dilution of the reaction mixture from the first round of PCR. The cycling conditions were as follows: 1 cycle of 1 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s annealing (varied temperatures, see above), and 1 min at 72 °C; and a final extension for 5 min at 72 °C.

Identification of the diterpene gene cluster via the ggs gene

Polymerase chain reaction (PCR) amplifications

For generating GGPP synthase genes by using degenerate primers, nested PCRs were carried out in a final volume of 25 μL containing 1.5 mM $MgCl_2$, 0.4 mM of each deoxynucleoside triphosphate (dNTP), 0.6 mM of each primer, 0.5 U of *Taq* DNA polymerase (Fermentas), and 50 ng of genomic DNA. The primers used in first round and second round PCR are as described in the text (Table 2.4). Templates used in second round PCR are 1:50 dilution of the reaction mixture from the first round of PCR. The cycling conditions were as follows: 1 cycle of 2 or 5 min (2 min for sheared DNA and 5 min for unsheared DNA) at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 45 ~ 48 °C, and 1 min at 72 °C; and a final extension for 5 min at 72 °C.

3'-Rapid amplification of the cDNA ends (3'-RACE) encoding the GGPP synthase genes

3'-RACE was used to extend the three *ggs* PCR fragments from *P. passeckerianus*. mRNA was isolated from the fungus *P. passeckerianus*. Reverse transcriptions were performed using the Omniscript Reverse Transcriptase (Qiagen). The experiments were performed according to the procedure as described.[44] The adaptor 5'-GACTCGAGTCGACATCGA-$(dT)_{17}$-3' was used for the reverse transcription using the oligo-d(T) part. This adaptor sequence contains *Xho*I, *Sal*I, and *Spe*I recognition sites. PCR was performed by using the primer 5'-GACTCGAGTCGACATCGA-3' from the adaptor and gene-specific primers designed from the obtained *P. passeckerianus ggs* genes p319, p322 and p345. Nested PCRs were carried out in a final volume of 50 µl containing 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate (dNTP), 0.2 µM of each primer, 1 U of *Taq* DNA polymerase (Fermentas), and 50 ng of cDNA. Templates used in second round PCR are 1:50 dilution of the reaction mixture from the first round of PCR. The cycling conditions were as follows: 1 cycle of 2 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 50 and 55 °C, and 1 min at 72 °C; and a final extension for 10 min at 72 °C. The 3'-RACE PCR product of p322 was digested with *Cla*I (Roche) and *Xho*I (Roche); while the p345 product was digested with *Sph*I (NEB) and *Sal*I (Roche). Double digestion were performed at 37 °C for 3 hr using 2 µL of PCR products, 1 µL of each enzyme and 1 µL of Roche Buffer H. The digested p322 was cloned into *Cla*I (Roche) and *Xho*I (Roche) digested

pBluescript KS. The digested p345 was cloned into *Sph*I (NEB) and *Sal*I (Roche) digested pGEMT-EASY vector (Promega).

Two new gene-specific primers for the second round PCR were designed for p322 and p319 in which a *Sph*I restriction site was added at the 5' end. Nested PCR conditions were the same as before, but only the 55 °C annealing temperature was used. The second round PCR product of p322 was cloned into *Sph*I (NEB) and *Sal*I (Roche) digested pGEMT-EASY vector (Promega).

Genome walking

A GenomeWalker library (Clontech, CA, USA) was constructed with genomic DNA according to the manufactuer's protocol. Genomic DNA was isolated and separate aliquots of genomic DNA were digested with four different four-basepair blunt-end restriction enzymes, *Dra*I, *Eco*RV, *Pvu*II and *Stu*I, which give four different genomic DNA pools. In each pool, GenomeWalker adaptors were ligated to each end of the DNA. Nested PCR was conducted by using adaptor primers and gene specific primers designed from the identified sequences.

First round PCR was performed in a 50 µL reaction volume by using 1 µL of 50x Advantage 2 polymerase mix (Clontech, CA, USA) with 5 µL of 10x Advantage 2 PCR buffer, 1 µL of 10mM dNTP, 1 µL of 10 µM of each primer, and 1 µL of each DNA library using a two-step program: denatuation at 95 °C for 25 sec, anealing and extension at 72 °C for 3 min for 7 cycles, followed by denatuation at 94 °C or 95 °C for 2 sec or 25 sec, anealing and extension at 67 °C for 3 min for

32 cycles, and then final extension at 67 °C for 7 min. The second round PCR was performed in a 50 μL reaction volume under the same conditions, except using 1 μL of the diluted primary PCR products (dilute 1 μL of each primary PCR product into 49 μL of deionized H_2O) as a template.

The series of PCRs was started with the adapter primers, AP1 and AP2, and with two nested gene specific primers designed on the basis of the GGPPS sequence. Genome walking was continued with the primers designed from the newly-obtained sequences.

Cloning and sequence analysis of the PCR products

The PCR products were purified by agarose gel electrophoresis and ligated into a pGEMT-EASY vector which has a T overhang (Promega, Madison, WI, USA). The ligation products were introduced into *E. coli* DH10B competent cells by electroporation and plated on LB / ampicillin (100 μg / mL) plates with 40 μL of X-gal (20 mg / mL) / IPTG (10 mg / mL) solution for the selections of the colonies with inserts. White colonies were selected. Homology searches were performed by using BLAST (http//www.ncbi.nlm.nih.gov/BLAST/) and the amino acid sequences were aligned with Clustal W.

Construction and screening of a genomic DNA library

A genomic DNA library was constructed by following the instructions for the CopyControl Fosmid library Production Kit (Epicentre). The genomic DNA was

isolated from 7th day *P. passeckerianus* mycelia and randomly sheared. The sheared DNA was end-repaired to 5'- phosphorylated blunt-end DNA using End-It™ DNA End-Repair Kit (Epicentre) containing End-Repair Enzyme Mix (T4 DNA polymerase and T4 polynucleotide kinase), 10× buffer, ATP and dNTPs. The end-repaired DNA was purified, concentrated and resuspended in water. The DNA was run on a 0.8% Low Melting Point agarose gel. The correct size of about 40 Kb genomic DNA was isolated using a DNA Extraction Kit (Fermentas). The DNA was then ligated into the Cloning-Ready CopyControl pCC1FOS vector using Fast-Link™ DNA Ligation Kit (Epicentre). The ligated DNA was packaged using MaxPlax™ Lambda Packaging Extracts, introduced into EPI300-T1 *E. coli* and plated on LB / chloramphenicol (12.5 µg / mL) plates.

The colonies were picked into 96-well plates and then transferred to 150 x 15 mm petri dishes. The colonies were transferred to Hybond™-N+ nylon membranes (Amersham Biosciences) and soaked with denaturation solution (0.5 M NaOH and 1.5 M NaCl), neutralization solution (1.5 M NaCl, 1.0 M Tris, adjusted to pH 7.5), 2× SSC (0.3 M NaCl, 0.03 M sodium citrate, adjusted to pH 7.0), and then UV crosslinked. The membranes were treated with proteinase K and washed with water to remove the cell debris. DNA fragments of about 800 bp and 1 kb prepared from p319, p322 and p345 sequences obtained from genome walking were used as probes and were labeled, hybridized and detected by using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche). The probe DNA 1.5 µg was diluted with water to a volume of 16 µl, denatured in boiling water for 10 min and chilled

in ice-water. DIG-High Prime 4 μl was added to the DNA and incubated overnight. The DIG-labeled DNA probe was denatured, added to the membranes and hybridization was allowed to go on overnight. The next day, after a series washes and blocking, the immunological detection was carried out using Anti-DIG-AP (anti-digoxigenin-alkaline phosphatase conjugate) and CSPD (chemiluminescent substrate for alkaline phosphatase).

Southern blotting

Southern blotting was carried out following the standard protocol.[42, 45] The DNA samples were separated on a 0.8% agarose gel and processed following the procedure of depurination, denaturation, neutralization and capillary transfer to the Hybond™-N+ nylon membranes (Amersham Biosciences), and then UV crosslinked. The labeling of the DNA probes, hybridization and immunological detection are the same as for screening the library.

Suppression Subtractive Hybridization (SSH) for pleuromutilin biosynthesis genes

Suppression Subtractive Hybridization (SSH) experiments follow the protocol of PCR-Select™ cDNA Subtraction Kit (Clontech, CA, USA). A subtracted library was formed from cDNA prepared from a 2^{nd} day culture that did not produce pleuromutilin and from cDNA from a 7^{th} day culture, which was confirmed by NMR to produce pleuromutilin.

Total RNA was isolated from both cultures using RNeasy Midi Kit (Qiagen). Messenger RNA was purified from total RNA by using Oligotex mRNA Mini Kit (Qiagen). Both mRNA populations are converted into single-stranded cDNA and then double-stranded cDNA, followed by blunt-ended restriction enzyme Rsa I digestion. The cDNA that contains differentially expressed transcripts is called tester (7 days old), while the other is called driver (2 days old). The digested and purified tester cDNA is then divided into two portions, tester 1 and tester 2, which were then ligated to adaptor 1 and adaptor 2R respectively. The driver cDNA is not ligated to the adaptors. Two hybridizations are then performed. In the first hybridization, excess amount of driver was added to each adaptor-ligated tester and the samples were heat denatured and allowed to anneal. In the second hybridization, the two samples from the first hybridization and the freshly-denatured driver were mixed together. The protruded ends were filled in and nested PCR was performed using different adaptor primers. The generated subtraction library was ligated into a pGEMT-EASY vector (Promega, Madison, WI, USA). The ligation products were introduced into ElectroMax DH10B competent cells (Invitrogen) by electroporation and plated on LB / ampicillin (100 μg/mL) plates with 40 μL of X-gal (20 mg/mL) / IPTG (10 mg/mL) solution for the selections of the colonies with inserts. White colonies were selected and transferred to 96 well plates with 200 μL super broth / ampicillin (100 mg/L). The plates were incubated at 37 °C until a clear growth is observed. The liquid cultures were then replicated on LB / ampicillin (100 μg/mL) plates with 40 μL of X-gal

(20 mg/mL) / IPTG (10 mg/mL) solution for the selections of the colonies with inserts for the second time. The next day, the white colonies were selected and transferred to 96 well plates with super broth liquid media, plus 10% glycerol. The plates were incubated at 37 °C until a clear growth was observed. The glycerol stock plates were sent for sequencing at the University of Washington High-Throughput Genomics Unit.

References

1. Mukhopadhyay, A.; Peterson, R. T. "Fishing for new antimicrobials," *Curr. Opin. Chem. Biol.* **2006,** *10*, 327-33.
2. Newman, D. J.; Cragg, G. M.; Snader, K. M. "Natural products as sources of new drugs over the period 1981-2002," *J. Nat. Prod.* **2003,** *66*, 1022-37.
3. Berkow, R. *The Merck Manual of Medical Information - Home Edition* **1999**.
4. Monaghan, R. L.; Barrett, J. F. "Antibacterial drug discovery--then, now and the genomics future," *Biochem. Pharmacol.* **2006,** *71*, 901-9.
5. Soulsby, E. J. "Resistance to antimicrobials in humans and animals," *Brit. J. Med.* **2005,** *331*, 1219-20.
6. French, G. L. "Clinical impact and relevance of antibiotic resistance," *Adv. Drug Deliver. Rev.* **2005** *57*, 1514-27.
7. Kavanagh, F.; Harvey, H.; Robbins, W. J. "Antibiotic substances from Basidiomycetes. VIII. *Pleurotus mutilis* Sacc and *Pleurotus passeckerianus* Pilat," *Proc. Natl. Acad. Sci. USA.* **1951,** *37*, 570-4.
8. Berner, H.; Vyplel, H.; Schulz, G.; Stuchlik, P. "Chemie der Pleuromutiline, IX. Konfigurationsumkehr der Methylgruppe am Kohlenstoff 6 im tricyclischen Gerüst des Diterpens Pleuromutilin," *Monatsh. Chem.* **1983,** *114*, 1125-36.
9. Lobova, D.; Smola, J.; Cizek, A. "Decreased susceptibilty to tiamulin and valnemulin among Czech isolates of *Brachyspira hyodysenteriae*," *J. Med. Microbiol.* **2004,** *53*, 287-91.
10. Egger, H.; Reinshagen, H. "New pleuromutilin derivatives with enhanced antimicrobial activity. II. Structure-activity correlations," *J. Antibiot.* **1976,** *29*, 923-7.
11. Heilmann, C.; Jensen, L.; Jensen, J. S.; Lundstrom, K.; Windsor, D.; Windsor, H.; Webster, D. "Treatment of resistant mycoplasma infection in immunocompromised patients with a new pleuromutiin antibiotic," *J. Infect. Dis.* **2001,** *43*, 234-8.
12. Butler, M. S.; Buss, A. D. "Natural Products - The Future Scaffolds for Novel Antibiotics," *Biochem. Pharmacol.* **2006,** *71*, 919-29.
13. Rittenhouse, S.; Biswas, S.; Broskey, J.; McCloskey, L.; Moore, T.; Vasey, S.; West, J.; Zalacain, M.; Zonis, R.; D., P. "Selection of retapamulin, a novel pleuromutilin for topical use," *Antimicrob. Agents Chemother.* **2006,** *50*, 3882-5.
14. Kosowska-Shick, K.; Clark, C.; Credito, K.; McGhee, p.; Dewasse, B.; Bogdanovich, T.; Appelbaum, P. C. "Single- and multistep resistance selection studies on the activity of retapamulin compared to other agents against *Staphylococcus aureus* and *Streptococcus pyogenes*," *Antimicrob. Agents Chemother.* **2006,** *50*, 765-9.

15. Schlunzen, F.; Pyetan, E.; Fucini, P.; Yonath, A.; Harms, J. M. "Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin," *Mol. Microbiol.* **2004,** *54*, 1287-94.
16. Poulsen, S. M.; Karlsson, M.; Johansson, L. B.; Vester, B. "The pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl transferase centre on the ribosome," *Mol. Microb.* **2001,** *41*, 1091-9.
17. Long, K. S.; Hansen, L. H.; Jakobsen, L.; Vester, B. "Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center," *Antimicrob. Agents Chemother.* **2006,** *50*, 1458-62.
18. Pringle, M.; Poehlsgaard, J.; Vester, B.; Long, K. S. "Mutations in ribosomal protein L3 and 23S ribosomal RNA at the peptidyl transferase centre are associated with reduced susceptibility to tiamulin in *Brachyspira* spp. isolates," *Mol. Microbiol.* **2004,** *54*, 1295-306.
19. Kawaide, H.; Imai, R.; Sassa, T.; Kamiya, Y. "*ent*-kaurene synthase from the fungus *Phaeosphaeria* sp, L487 - cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase in fungal gibberelln biosynthesis," *J. Biol. Chem.* **1997,** *272*, 21706-12.
20. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Sassa, T. "Cloning of a full-length cDNA encoding *ent*-kaurene synthase from *Gibberella fujikuroi*: functional analysis of a bifunctional diterpene cyclase," *Biosci. Biotechnol. Biochem.* **2000,** *64*, 660-4.
21. Oikawa, H.; Toyomasu, T.; Toshima, H.; Ohashi, S.; Kawaide, H.; Kamiya, y.; Ohtsuka, M.; Shinoda, S.; Mitsuhashi, W.; Sassa, T. "Cloning and functional expression of cDNA encoding aphidicolan-16β-ol synthase: a key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin," *J. Am. Chem. Soc.* **2001,** *123*, 5154-5.
22. Naegell, p. "Zur kenntnis des pleuromutilins," *Diss. Nr. ETH Zurich* **1961**, 3206.
23. Arigoni, D. "Structure of a new type of terpene," *Gazz. Chim. Ital.* **1962,** *92*, 884-901.
24. Birch, A. J.; Cameron, D. W.; Holzapfel, C. W.; Richards, R. W. "Diterpene nature of pleuromutilin," *Chem. Ind.* **1963**, 374-5.
25. Birch, A. J.; Holzapfel, C. W.; Richards, R. W. "The structure and some aspects of the biosynthesis of pleuromutilin," *Tetrahedron. Suppl. (part II)* **1966,** *8*, 358-87.
26. Buzzolini, M. "Zur biogenesis des pleuromutilin und der lagopodine," *Diss. Nr. ETH Zurich* **1966**, 3797.
27. Bonavia, G. "Pleuromutilin, sterochemie und detaillierte biosynthesis," *Diss. Nr. ETH Zurich* **1968**, 4189.
28. Knauseder, F.; Brandl, E. "Pleuromutilins. Fermentation, structure and biosynthesis," *J. Antibiot.* **1975,** *29*, 125-31.

29. Paquette, L. A.; Wiedeman, P. E.; Bulman-Page, P. C. "(+)-Pleuromutilin synthetic studies. Degradative and de novo acquisition of a levorotatory tricyclic lactone subunit," *J. Org. Chem.* **1988,** *53*, 1441-50.
30. Toyomasu, T.; Nakaminami, K. "Cloning of a gene cluster responsible for the biosynthesis of diterpene aphidicolin, a specifc inbitor of DNA polymerase," *Biosci. Biotechnol. Biochem.* **2004,** *68*, 146-52.
31. Young, C. A.; McMillan, L. "Molecular cloning and genetic analysis of an indolediterpene gene cluster from *Penicillium paxilli*," *Mol. Microbiol.* **2001,** *39*, 754-64.
32. Zhang, S.; Monahan, B. J.; Tkacz, J. S.; B., S. "Indole-diterpene gene cluster from *Aspergillus flavus*," *Appl. Environ. Microbiol.* **2004,** *70*, 6875-83.
33. Young, C. A.; Bryant, M. K.; Christensen, M. J.; Tapper, B. A.; Bryan, G. T.; Scott, B. "Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass," *Mol. Gen. Genomics* **2005,** *274*, 13–29.
34. Kupfer, D. M.; Drabenstot, S. D.; Buchanan, K. L.; Lai, H.; Zhu, H.; Dyer, D. W.; Roe, B. A.; Murphy, J. W. "Introns and splicing elements of five diverse fungi," *Eukaryot. Cell* **2004,** *3*, 1088-100.
35. Siebert, P. D.; Chenchik, A.; Kellogg, D. E.; Lukyanov, K. A.; Lukyanov, S. A. "An improved method for walking in uncloned genomic DNA," *Nucleic Acids Res.* **1995,** *23*, 1087-8.
36. Tudzynski, B.; Holter, K. "Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster," *Fung. Genet. Biol.* **1998,** *25*, 157–70.
37. EPICENTRE. "CopyControl Fosmid Library Production Kit protocol,".
38. Agrios, G. N. *Plant Pathology* **1998,** *3rd edition, Academic Press, San Diego*, 403-7.
39. Kawaide, H. "Biochemical and molecular analyses of gibberellin biosynthesis in fungi," *Biosci. Biotechnol. Biochem.* **2006,** *70*, 583-90.
40. Diatchenko, L.; Lau, Y. F.; Campbell, A. P.; Chenchik, A.; Moqadam, F.; Huang, B.; Lukyanov, S.; Lukyanov, K.; Gurskaya, N.; Sverdlov, E. D.; Siebert, P. D. "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries," *Proc. Natl. Acad. Sci. USA.* **1996,** *93*, 6025-30.
41. Anderson, M. L. M. "Nucleic acid hybridization," **1999**, Bios Scientific Publishers; New York: Springer.
42. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning: a laboratory manual* **1989**, Second edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.
43. Moller, E. M.; Bahnweg, G.; Sandermann, H.; Geiger, H. H. "A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies and infected plant tissues," *Nucleic Acids Res.* **1992,** *20*, 6115-6.

44. Frohman, M. A.; Dush, M. K.; Martin, G. R. "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer," *Proc. Natl. Acad. Sci. USA* **1998,** *85*, 8998-9002.
45. Amersham Biosciences. "Hybond-N+ protocol,".

CHAPTER THREE

Sordarin Biosynthesis in *Sordaria araneosa* Cain

Antifungal compounds

Fungal infections range from superficial conditions of the skin to life-threatening diseases. Until the 1970s, most fungal infections were considered to be limited to skin infections and thus were considered non-life-threatening. Therefore, the demand for new antifungal drugs was limited and hence there was little progress in the development of the antifungal drugs. However, the prevalence of systemic fungal infections has increased significantly during the past decade especially in immunocompromised patients, such as AIDS, cancer and transplant patients whose immune systems are damaged and whose natural barriers are broken[1, 2] The unsatisfactory activities and limited commercial diversity of antifungal drugs in the market pose a demand for new antifungal drugs.

Serious invasive fungal infections are normally caused by several different types of fungi, including *Candida* spp*., Cryptococcus neoformans, Aspergillus* spp*., Pneumocystis carinii* and *Histoplasma capsulatum*. The available antifungal agents in the market for the treatment of these life-threatening fungal infections include the imidazole and triazole groups of antifungal drugs, such as ketoconazole and itraconazole,[3] the polyenes such as amphotericin B,[4] the allylamines such as

terbinafine,[5] and the echinocandins[6] such as caspofungin. The structures of some antifungal compounds are shown in Figure 3.1. In addition to the limited diversities, some commercially available antifungal agents show some limitations, such as the significant nephrotoxicity (kidney damage) of amphotericin B[7] and emerging resistance to the imidazoles and triazoles.[8, 9] The development of new antifungal agents, preferably naturally occurring compounds with novel mechanisms of action, is an urgent medical need.

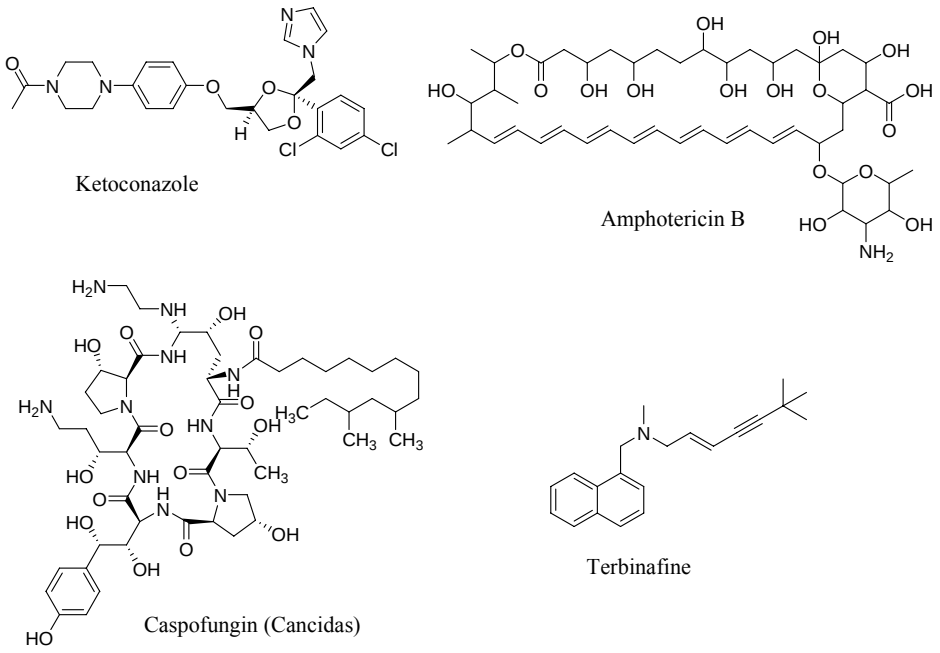


Figure 3.1. Examples of antifungal compounds.

Most of the antifungal drugs in the market target fungal ergosterol. Ergosterol is a component of fungal cell membranes, serving the same function that cholesterol serves in animal cells**.** Animal cells contain cholesterol instead of ergosterol and so

are much less susceptible to the antifungal drugs targeting ergosterol. The polyenes bind to the ergosterol in the fungal cell membrane. The cell membrane changes from a fluid state to a more crystalline state. As a result, the cell's contents leak out and the cell dies.[4] The imidazoles and triazoles inhibit the enzyme cytochrome P450 14α-demethylase which converts lanosterol to ergosterol, and is required for fungal cell membrane synthesis.[10] The allylamines inhibit the enzyme squalene epoxidase, another enzyme required for ergosterol synthesis.[11]

The echinocandins function by a separate mechanism. They are the inhibitors of glucan synthesis in the cell wall.[6]

Inhibition of fungal protein synthesis is an alternative approach to antifungal agents.[10] Protein synthesis has always been considered one of the most attractive targets in the development of antimicrobial agents.[12] However, at the same time, protein synthesis is not an easy target for antifungal agents because of the eukaryotic nature of fungi and therefore the high similarity between fungal and mammalian protein synthesis machineries. Two fungal factors, EF3 and EF2, that belong to the fungal elongation factor family, have been found to be good targets for antifungal drug discovery.[13] These two elongation factors are essential for protein synthesis. The elongation factor EF3 is only required by fungal ribosomes, and thus does not interfere with the normal functions of other eukaryotes such as humans and other mammals.[14, 15] Although the protein sequences of EF2 are highly conserved in all eukaryotes, the sordarins have been found to be selective inhibitors of EF2 function in fungal protein synthesis. Sordarins exclusively impair

fungal protein synthesis machinery, but not those of bacteria and other eukaryotes, such as mammal and plants.[16]

Transcription and translation in eukaryotes

The eukaryotic genome includes exons and introns. After transcription, the introns in the initial product (pre-mRNA) will be removed before translation occurs. Translation, which results in protein synthesis, takes places on the ribosome. Translation involves three steps: initiation, elongation and termination. In the initiation step, an amino acid is attached to transfer RNA (tRNA) to form the aminoacylated tRNA. The mRNA is bound to the ribosome and tRNA which has the anticodon that matches the mRNA codon, brings the amino acids to the ribosome one at a time.

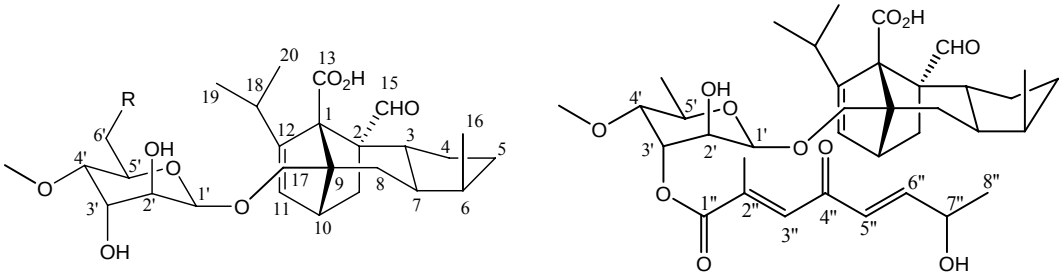
The functional eukaryote 80s ribosomes are composed of a 60s subunit (large subunit) and a 40s subunit (small subunit). The 40s subunit contains a 18s rRNA (17s in yeast) plus proteins. The 60s subunit contains 28s (26s in yeast), 5s and 5.8s rRNAs, plus proteins. Among the 60s subunit proteins, the rpP0 and rpL12, plus the rpP1a, rpP1b, rpP2a, and rpP2b tetramer, compose the ribosomal stalk. The stalk plays a prominent role in the elongation step of protein synthesis.[17] Protein rpP0 plays a central role in stalk architecture, interacting with rpL12 and 26 S rRNA (in the yeast) at its amino terminal and with the other four P-proteins at its carboxyl-terminal terminal.[18, 19] The molecular detail of the interaction between

stalk components and elongation factors is still unknown. Recent studies indicate that elongation factors make multiple contacts with stalk components; elongation factor 2 (EF2) is located at the base of the stalk.[20, 21]

In eukaryotes, the elongation step needs three elongation factors, EF1a, EF1bc and EF2. The nascent polypeptide chain is attached to the tRNA in the P (peptidyl) site. The next amino acid which is in a complex with EF1a, comes to the A (aminoacyl) site, where the peptide bond is formed between the polypeptide chain attached to the tRNA in the P site and the amino acid attached to the tRNA in the A site. Then translocation takes places. The tRNA originally in the P site is moved to the E (exit) site and the tRNA that is originally in the A site and now has the nascent polypeptide chain attached to it, is moved to the P site. In the elongation step, the ribosome changes from the pre-translocational (PRE) to the post-translocational (POST) state as the P site and A site bound tRNAs move to the E site and the P site respectively. The EF2 catalyzes the translocation. The mechanism of translocation and the catalytic mechanism of EF2 have not been yet fully revealed. It is suggested that binding of EF2 induces large-scale conformational changes in the ribosome that might be related to tRNA movement.[22] Through the X-ray structure of eukaryote EF2, in complex with the antibody sordarin[23] and the molecular model of the yeast 80S ribosome,[24, 25] large conformational changes within EF2 and the 80S ribosome are revealed. Interference with the interaction between EF2 and the ribosome is a potential mechanism for new antifungal agents.

Introduction to the sordarins

Sordarin (Figure 3.2) was originally isolated from the fungus *Sordaria araneosa* Cain in 1971.[26] This strain also produces neosordarin and hydroxysordarin (Figure 3.2).[27] The structures of sordarin and its derivatives are composed of the diterpene sordaricin**,** the aglycone core of sordarin, and a sugar moiety, sordarose, with varied sidechains on the 3' hydroxyl group.



R= H Sordarin
R= OH Hydroxysordarin

Neosordarin

Figure 3.2. Sordarin and its derivatives.

Antifungal function and mechanism

Sordarin and its derivatives show antifungal activity both *in vivo* and *in vitro*. They have been shown to inhibit most fungal pathogens known to affect immunocompromised patients, including *Candida albicans*,[28, 29] *Cryptococcus neoformans*[30] and *Pneumocystis carinii*,[31, 32] the causal agent of pneumonia. It has also been demonstrated that sordarin derivatives have potent antifungal activity against important endemic fungal pathogens such as *Histoplasma capsulatum*,[33]

Blastomyces dermatidis,[34] *Paracoccidioides brasiliensis*[34] and *Coccidioides immitis*.[35, 36]

Antifungal mechanism

The sordarins target and bind to Elongation Factor 2 (EF2), blocking the interaction of EF2 with the ribosomal P-protein stalk.[36, 37] It was shown that sordarin induces a large conformational change in EF2 and this conformational change traps the EF2 on the 80S ribosome.[23, 38]

The high affinity binding of sordarin to EF2 requires the presence of ribosomes. Mutations in either EF2 or the large ribosomal stalk protein rpP0 lead to resistance to sordarin, although EF2 is the principal determinant of sordarin specificity.[16, 39-41] Because of the differences between fungal and mammalian cells, this multiple interaction may explain the high degree of selectivity of this class of compounds between fungal and mammalian cells.[34]

Sordarin and most of its natural derivatives only differ in the modification of the sordarose sugar (Figure 3.2).[27, 42-44] The sugar moiety plays a role in enhancing the binding of sordarin in the target binding site and its derivatives with an acyl sidechain enhance the lipophilicity and increase activity, which could be related to improved cell uptake.[45, 46] Both aldehyde and carboxylic acid groups on the diterpene moiety have been shown to be essential for the retention of the activities of sordarin compounds, presumably due to the interactions of these groups with particular amino acid residues at the binding pocket of the targets.[36]

General biosynthetic pathway

The proposed biosynthetic scheme for formation of sordaricin **6** from GGPP **1** is shown in Figure 3.3. It is an ionization-initiation reaction. Sordaricin has been shown to be biologically derived from cycloaraneosene **4**.[47] The bond between C8 and C9 is broken to give an intermediate, such as **5**, and two new bonds are formed, between C7 and C10, and between C12 and C18, to give **6**. It is speculated that the biogenetic route from **5** to **6** might proceed by means of an enzyme-mediated intramolecular Diel-Alder reaction.[47, 48] The steps from **1** to **4** are likely to be mediated by the diterpene synthase.

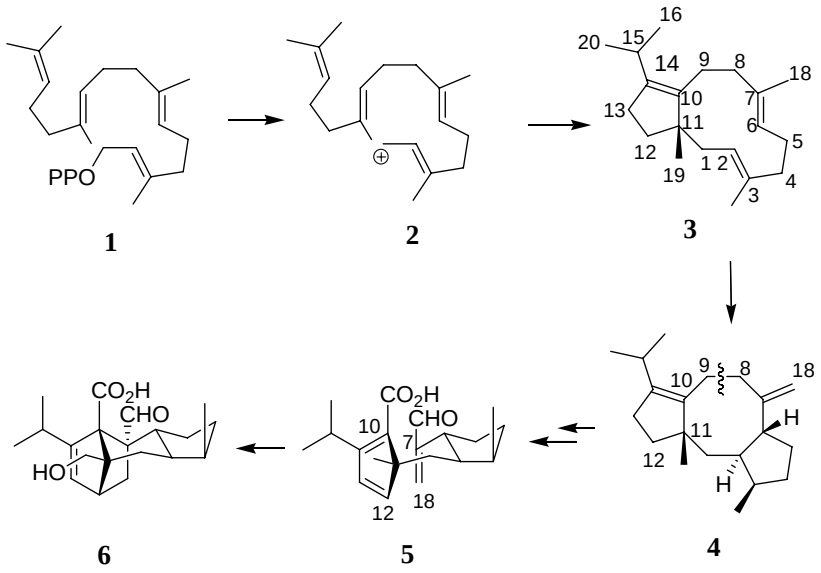


Figure 3.3. Proposed biosynthetic scheme for formation of sordaricin from GGPP.

Results and discussion

The growth of *Sordaria araneosa* and production of sordarins

Sordaria araneosa Cain can be grown under both shaking and static conditions. Under the shaking condition, the fungus can grow into a lot of white, round aggregations which can be up to 1 cm diameter in seven days. Under the static condition, the white mycelium covered the surface of the liquid in 500 mL Erlenmeyer flasks about three weeks after inoculation (Figure 3.4). Compared to the fungus *Pleurotus passeckerianus* Pilat, the growth rate of *S. araneosa* was faster; more mycelia or thicker mycelia were produced. Sordarin and its derivatives were produced under both shaking and static conditions. The growth rate of the mycelia and the yield of the compounds were not compared for these two conditions. The cultures used for this project were grown under shaking condition.[27]

Figure 3.4. *Sordaria araneosa* Cain grown under static condition for 20 days.

Determination of the presence of sordarins

In order to determine that the fungus obtained from ATCC was the correct organism and that it does produce the sordarin secondary metabolites under our

laboratory conditions, these compounds were isolated and their structures were confirmed by NMR and MS.[27] The presence of sordarin and several of its derivatives was demonstrated.

The growth medium that was used was reported to produce sordarin, neosordarin and hypoxysordarin.[27] Both the culture medium and the mycelia were extracted. The amount of mycelial extract was minimal, so extraction of the mycelia was omitted in subsequent studies. For LC-MS, reverse-phase HPLC was conducted. Isocratic elution was carried out using 70:30 acetonitrile:H_2O for 45 min and then 100% of acetonitrile for 25 min at a flow rate of 0.5 ml/min. The MS was performed using (+) and (-) APCI and ESI. The presence of sordarins was proved by mass spectrometry (Figure 3.5, A, B, C and D). The peaks eluted from the column at around 12 min, 21 min and 34 min, and the masses match hydroxysordarin, sordarin and neosordarin, respectively.[27] Table 3.1 summarizes the MS (-) APCI results of sordarin and its derivatives. For sordarin, the base peak is *m/z* 491 $(M-H)^-$. For hydroxysordarin, the base peak is *m/z* 507 $(M-H)^-$. For neosordarin, the major peaks are *m/z* 657 (base peak, M-H) and *m/z* 491 (loss of sidechain).[27] There is another major peak of *m/z* 676. Because there is another peak which overlaps the neosordarin peak in the HPLC trace (Figure 3.5A), this *m/z* 676 peak may be something unrelated to sordarin. Similar HPLC conditions were used to isolate sufficient quantities of sordarin for analysis by NMR. The typical peak of the C15 aldehyde group at 9.7 ppm in ^1H NMR was used to confirm the presence of sordarin and its derivatives in the crude extracts (Figure

3.6). Another important peak in ^{1}H NMR is the C11 double bond proton peak at 6.08 ~ 6.12 ppm (doublet, J = 3). The ^{13}C NMR spectra were also acquired to provide further confirmation of the structures. Neosordarin can be easily distinguished from sordarin and hydroxysordarin in both ^{1}H NMR and ^{13}C NMR because of the double bonds in its distinct sidechain. The sordarin and hydroxysordarin can be distinguished in the ^{13}C NMR by the C6' shift, with the sordarin C6' at 18.1 ppm and hydroxysordarin at 62.4 ppm. For the numbering of sordarin and its derivatives, see Figure 3.2.

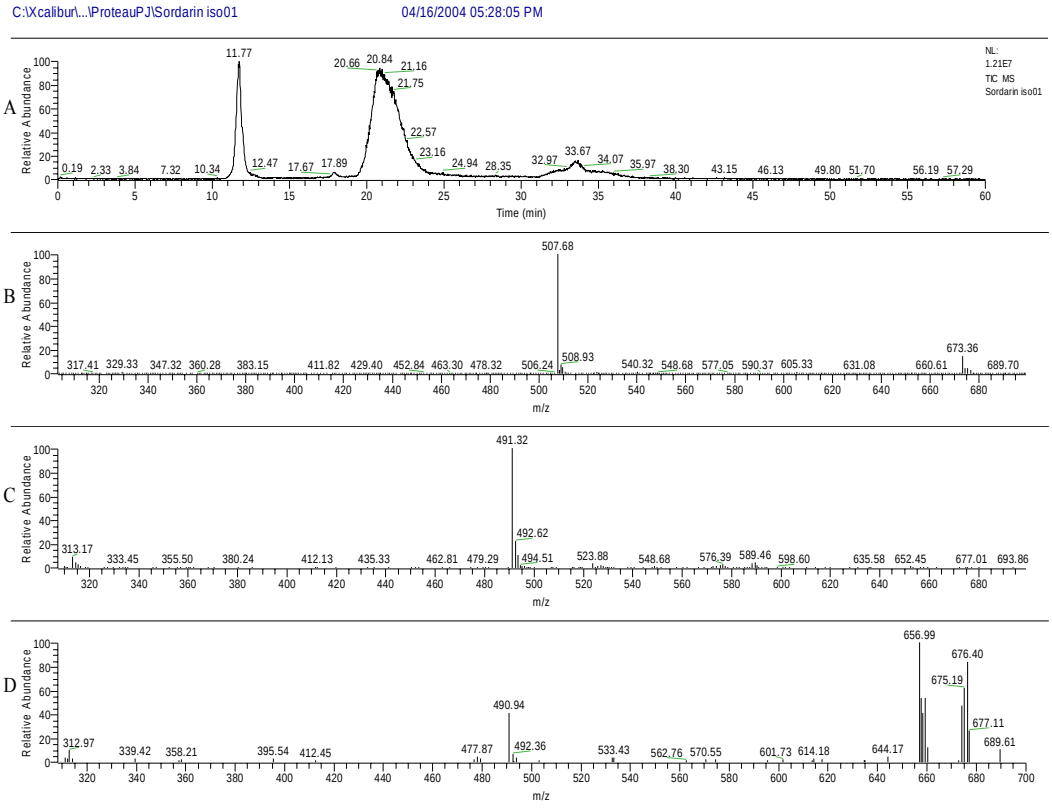


Figure 3.5. LC-MS spectra of sordarins. A. Ion chromatogram spectrum of sordarins. The first peak is hydroxysordarin, the second peak is sordarin and the third one is neosordarin. B. MS spectrum of hydroxysordarin. C. MS spectrum of sordarin. D. MS spectrum of neosordarin.

Table 3.1. MS (-)APCI results of sordarin and its derivatives.

	Molecular formula	Molecular weight	Major peak(s)
sordarin	$C_{27}H_{40}O_8$	492	491
hydroxysordarin	$C_{27}H_{40}O_9$	508	507
neosordarin	$C_{36}H_{50}O_{11}$	658	657, 491

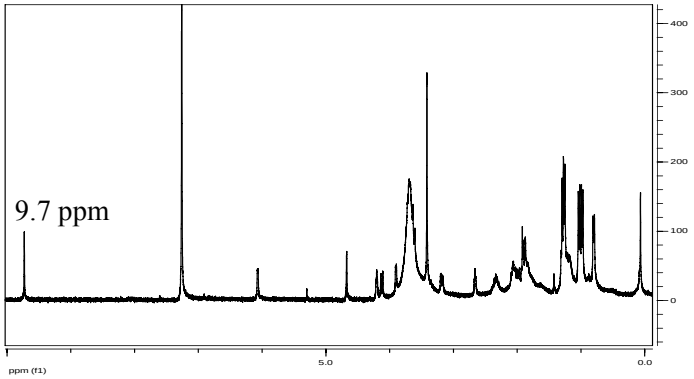


Figure 3.6. ^{1}H NMR spectrum of the crude mixture of sordarin and its derivatives (key aldehyde proton at 9.7 ppm).

Time course of sordarin production

The daily crude extracts from 1st day to 6th day cultures, and the crude extracts of 8th, 10th and 12th day cultures, were used to detect the presence of sordarin and its derivatives by mass spectrometry without further purification and separation. The presence of either sordarin or its derivatives indicates that the genes for the biosynthesis of sordarin have been turned on. Both (+) APCI and (-) APCI were conducted. The peaks belonging to sordarin and/or its derivatives were present from the 4th day and became much stronger on the 6th day. The existence of the sordarins continued through at least the 12th day. For the (+) APCI results, the m/z 493 peak is the sordarin $[M+H]^+$; the *m/z* 333 peak represents sordaricin; additional mass peak that are seen are *m/z* 271 (base peak) and 315. The 271, 315

and 333 peaks could represent the sordarin and any of its derivatives as has been observed in previous studies.[27] The 279 peak belongs to unknown substance and does not seem to be related to the sordarins. The MS (+) APCI results from the 1st day and 8th day cultures are shown in Figure 3.7 A and B respectively. Table 3.2 summarizes the MS (+) APCI and (-) APCI results of sordarin and its derivatives for the time course of production. The relative intensities of the different sordarin peaks may represent different derivative production varied with time. During the time course of sordarin production and under the (+) APCI, the *m/z* 271 peak was the base peak for the whole time course; the peak of *m/z* 333 was the 2nd largest peak for most of time, but the *m/z* 493 peak which should represent sordarin increased largely and exceeded the peak of *m/z* 333 from the 10th day. The (-) APCI pattern was relatively simple compared to the (+) APCI, with the peak of 491 existing from the 4th day, which should represent the sordarin $(M-H)^-$, or the neosordarin for the loss of sidechain on the sordarose. The peak of *m/z* 657, which represent neosordarin, showed up as small peaks on the 4th day, and the peak of 507, which represent hydroxysordarin, was present from the 8th day. According to the time course of production, the 7th day culture was chosen for the isolation of mRNA.

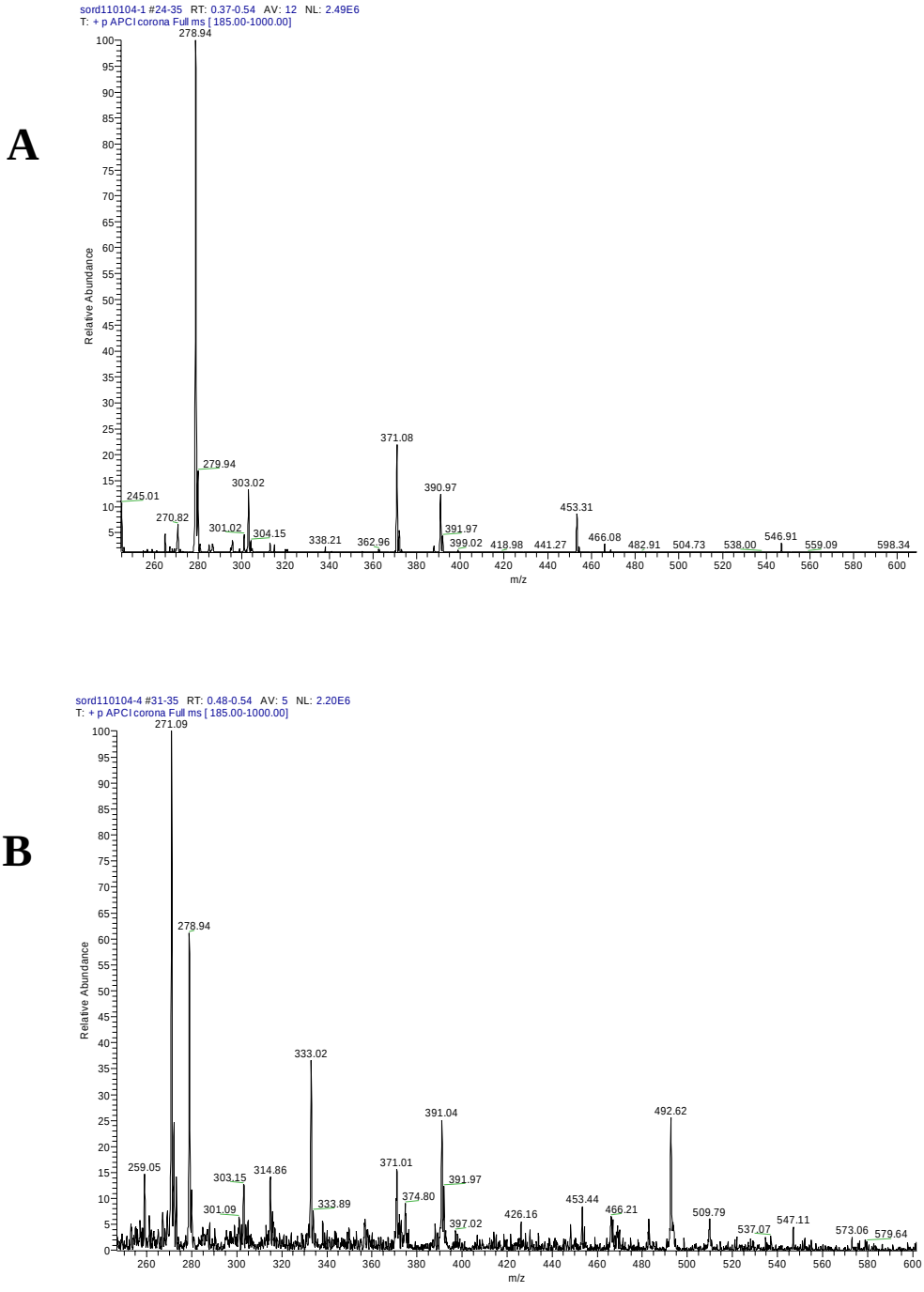


Figure 3.7. MS data of *S. araneosa* cultures. A. MS data of 1st day *S. araneosa* culture. B. MS data of 8th day *S. araneosa* culture.

Table 3.2. Sordarin production.

Day	Peaks of sordarins (+) APCI	Peaks of sordarins (-) APCI	
1st			No sordarin peaks
2nd			No sordarin peaks
3rd			No sordarin peaks
4th	271, 315, 333, 493	491, 657	Sordarin peaks are obvious
5th	271, 315, 333, 493	491	Sordarin peaks become stronger
6th	271, 315, 333, 493	491	Sordarin peaks become stronger
8th	271, 315, 333, 493	491, 507	Sordarin peaks are strong
10th	271, 315, 333, 493	491, 507	Sordarin peaks are strong
12th	271, 315, 333, 493	491, 507	Sordarin peaks are present

Attempted cloning of the sordarin biosynthetic gene cluster

Direct identification of the diterpene synthase gene

The same procedure tried in Chapter 2 to identify the pleuromutilin diterpene synthase gene was also utilized in the search for the sordarin diterpene synthase gene. Unfortunately, no specific PCR products were formed from *S. araneosa* DNA either (see Figure 2.17).

Identification of the diterpene gene cluster via the ggs gene

Most of the procedures were the same as in Chapter 2. Most of these experiments were conducted along with the *P. passeckerianus* samples and thus most of the sordarin results are depicted in figures in Chapter 2.

The same fungal *ggs* degenerate primers F1, F2, R1, R2, GGPPS27, GGPPS28 and GGPPS29 were successfully used to obtain two fragments which represent potential *ggs* genes from *S. araneosa.* These two fragments from *S. araneosa* have sizes of 224 bp and 272 bp (designated as s224 and s272 respectively) (Figure 3.8, the band in the red box). The s224 product was generated from the primers F1 and

GGPPS29 for the first round PCR, and GGPPS27 and GGPPS28 for the second round PCR. The fragment s272 was generated from the primers F1 and R2 for the first round PCR, and GGPPS27 and GGPPS29 for the second round PCR (Figure 3.9, the band in red box). The nucleotide sequences and the translated protein sequences for the *S. araneosa* PCR fragments s224 and s272 are shown below and the protein sequences contain regions that matched the "HRGQGMD" and "QIRDDY" conserved regions (highlighted in red). Unlike the *P. passeckerianus*, no introns are present. BLAST searches showed that these two fragments from *S. araneosa* are potential *ggs* genes, both containing the conserved KT motif. The sequences of these fragments are shown below and also in the appendix.

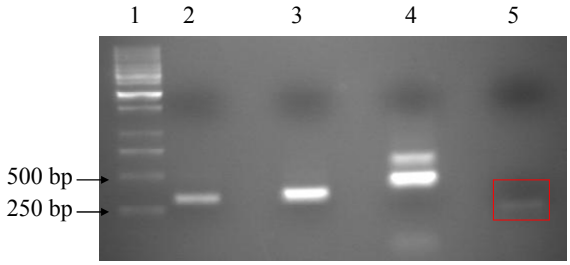


Figure 3.8. Nested PCR 2nd round products amplified from genomic DNAs of *P. passeckerianus* and *S. araneosa*. Lanes 2 to 4 were from *P. passeckerianus* (see Chapter 2), while lane 5 was from *S. araneosa*. Lane 5 was a PCR (46 °C annealing temperature) product with primers (F1, GGPPS29) for the 1st round and (GGPPS27, GGPPS28) for the 2nd round. The PCR product was a faint band slightly smaller than 250 bp. Lane 1 contains a 1 Kb DNA ladder (Fermentas).

S224:

CATCATGGTCAGGGCCTCGATATACGCTGGCGGGATCATCTCAAGACAACCAAGCTCCCGCTGATAGAAGAATA
CATGGAAATGATCATGAACAAAACCGGAGGATTATTTCGGCTTGCCGTTAAACTCTTGGGAGCTTTCAGCACCC
CACAAGACAAGCCAACTGTGTTGGCCATAGCTAATCTCTTCGGGATTGTATTCCAGATCCGTGATGACCTTATG
AA

catcatggtcagggcctcgatatacgctggcgggatcatctcaagacaaccaagctcccg
 H H G Q G L D I R W R D H L K T T K L P
ctgatagaagaatacatggaaatgatcatgaacaaaaccggaggattatttcggcttgcc
 L I E E Y M E M I M N K T G G L F R L A
gttaaactcttgggagctttcagcaccccacaagacaagccaactgtgttggccatagct
 V K L L G A F S T P Q D K P T V L A I A
aatctcttcgggattgtattccagatccgtgatgaccttatgaa
 N L F G I V F Q I R D D L M

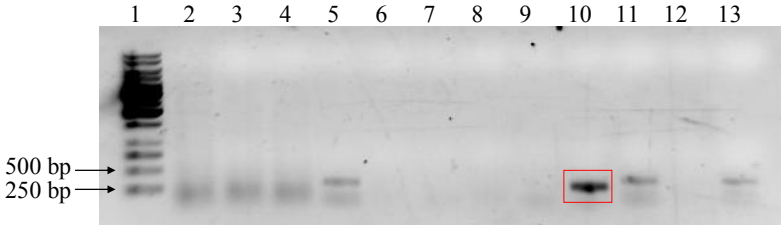


Figure 3.9. Nested PCR results for *S. araneosa*. Lanes 2 to 4 were products from PCR 1st round. Lane 2 was PCR product amplified with primers GGPPS27 and GGPPS29. Lane 3 was PCR product with primers F1 and R2. Lane 4 was PCR products with primers F1 and GGPPS29. Lane 5 was from nested PCR 2nd round amplification of 1/50 dilution of the products shown in lanes 2 and used primers GGPPS27 and GGPPS28 for the 2nd round. Lanes 6 to 11 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 3. Lane 6 used primers F2 and R1 for the 2nd round. Lane 7 used primers F2 and GGPPS29 for the 2nd round. Lane 8 used primers F2 and GGPPS28 for the 2nd round. Lane 9 used primers GGPPS27 and R1 for the 2nd round. Lane 10 used primers GGPPS27 and GGPPS29 for the 2nd round. Lane 11 used primers GGPPS27 and (Figure 3.9 continued) GGPPS28 for the 2nd round. Lanes 12 and 13 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 4. Lane 12 used primers F2 and GGPPS28 for the 2nd round. Lane 13 used primers GGPPS27 and GGPPS28 for the 2nd round. Lane 1 contains a 1 kb DNA ladder (Fermentas). The band highlighted in the red box is the s272 band.

S272:

CACCGGGGCCAAGGGATGGACCTCTTCTGGAGAGACACTCTGACCTGCCCGACCGAAGACGAGTACCTCGAGA
TGGTATCCAACAAGACTGGGGGGCTGTTCAGGCTTGGTATTAAGCTGATGCAGGCTGAGTCCCGGAGTCTTGT
GGACTGCGTACCGCTTGTCAGCATCATGGGCTTAATCTTCCAGATTGCGGATGACTACCAGAACCTCTGGAGC
CGGCGGTATACTGCCAACAAAGGCATGTGTGAGGACCTCACCGAGGGGAAGTT

```
caccggggccaagggatggacctcttctggagagacactctgacctgcccgaccgaagac
 H  R  G  Q  G  M  D  L  F  W  R  D  T  L  T  C  P  T  E  D
gagtacctcgagatggtatccaacaagactgggggctgttcaggcttggtattaagctg
 E  Y  L  E  M  V  S  N  K  T  G  G  L  F  R  L  G  I  K  L
atgcaggctgagtcccggagtcttgtggactgcgtaccgcttgtcagcatcatgggctta
 M  Q  A  E  S  R  S  L  V  D  C  V  P  L  V  S  I  M  G  L
atcttccagattgcggatgactaccagaacctctggagccggcggtatactgccaacaaa
 I  F  Q  I  A  D  D  Y  Q  N  L  W  S  R  R  Y  T  A  N  K
ggcatgtgtgaggacctcaccgaggggaagtt
 G  M  C  E  D  L  T  E  G  K
```

Another PCR product of 247 bp, which was designated as s247, was also obtained in these experiments. However, later experiments (genome walking, Figure 2.27 and 2.28) indicated that this fragment does not belong to a GGPP synthase gene.

Genome Walking

Genome walking was used to extend the sequences of the s224 and s272 PCR products. Most of s224 genome walking experiments were conducted with those from *P. passeckerianus* and share the same figures in Chapter 2. For the first time genome walking, the primers were designed from the sequences of s224 and s272. A single round of genome walking experiments was also performed with s247.

S224

The designed gene-specific primers are shown in Table 3.3. The other information for the 1st round genome walking refers to Table 2.6. The PCR result is shown in Figures 2.27 and 2.28.

Table 3.3. Genome walking PCR primers for s224 and s247.

Primer name	Primer sequences
sord224-32-58	TCCCGAAGAGATTAGCTATGGCCAACA
sord224-66-91	CTTGTCTTGTGGGGTGCTGAAAGCTC
sord224-107-133	ACAAAACCGGAGGATTATTTCGGCTTG
sord224-154-180	GACAACCAAGCTCCCGCTGATAGAAGA
sord247-37-63	TAGGCAATTCATGCACGCCATTTTGTA
sord247-67-93	CAGACAGTCCCATGTACGGAAGGTTGA
sord247-131-157	TGTATAGCGGCTGCAAAATCCTGCATA
sord247-185-211	TACGACTTGGTGTTGAACCGGATTGTG

Successive genome walking was conducted using the gene-specific primers designed from the newly obtained sequences and walking continued both upstream and downstream. The 2nd time genome walking PCR primers refer to Table 2.8 and the 2nd round PCR results for s224 refer to Figure 2.30. The 3rd time genome

walking PCR primers are shown in Table 3.4 and 2nd round PCR results for s224 are shown in Figure 3.10.

Table 3.4. The 3rd time genome walking PCR primers for s224.

Primer name	Primer sequences
sord224-32-58	TCCCGAAGAGATTAGCTATGGCCAACA
sord224-66-91	CTTGTCTTGTGGGGTGCTGAAAGCTC
sord224-107-133	ACAAAACCGGAGGATTATTTCGGCTTG
sord224-154-180	GACAACCAAGCTCCCGCTGATAGAAGA
sord247-37-63	TAGGCAATTCATGCACGCCATTTTGTA
sord247-67-93	CAGACAGTCCCATGTACGGAAGGTTGA
sord247-131-157	TGTATAGCGGCTGCAAAATCCTGCATA
sord247-185-211	TACGACTTGGTGTTGAACCGGATTGTG

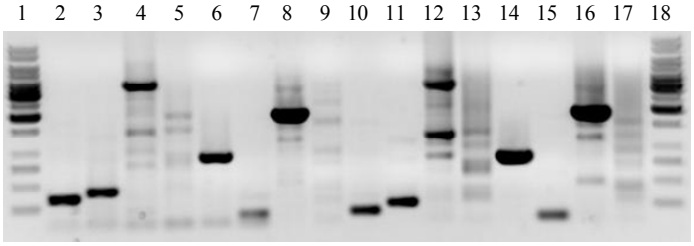


Figure 3.10. Genome walking nested PCR results for s224. Lanes 2 to 9 were results from PCR 1st round. Lanes 2 to 5 were results for downstream of s224. Lanes 6 to 9 were results for upstream of s224. Lanes 2 and 6 were from *Dra*I library pool sdDL1. Lanes 3 and 7 were from *Eco*RV library pool sdDL2. Lanes 4 and 8 were from *Pvu*II library pool sdDL3. Lanes 5 and 9 were from *Stu*I library pool sdDL4. Lanes 10 to 17 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 9, respectively. Lanes 1 and 18 contain 1 kb DNA ladder (Fermentas).

S247

For the fragment of s247, sequence surrounding the original PCR product was obtained, with 1.1 Kb of sequence downstream and 1.2 Kb of sequence upstream. Sequence analysis revealed that although the original PCR product had some similarities to a GGPP synthase gene, the true sequence through the degenerate

primer regions was not consistent with a GGPP synthase gene. This PCR product was not pursued further.

S272

The s272 *ggs* gene fragment was generated several months after the s224 *ggs* gene was found and when the s247 product was shown to be a false positive. In order to get reliable results, new GenomeWalker library pools were constructed. The *Eco*RV digested library pool (sdDL2) was lost during the preparation step, so only three library pools were used for PCR. The primers for PCR are shown in Table 3.5 and the PCR result is shown in Figure 3.11.

Table 3.5. Genome walking PCR primers for s272.

PCR	Primer names	Primer sequences
Upstream, 1st round	S272-85-112u1	TGGAAGATTAAGCCCATGATGCTGACA
Upstream, 2nd round	S272-138-165u2	GGACTCAGCCTGCATCAGCTTAATACCA
Downstream, 1st round	S272-138-164d1	GAGTCTTGTGGACTGCGTACCGCTTGT
Downstream, 2nd round	S272-186-212d2	CCAGATTGCGGATGACTACCAGAACCT

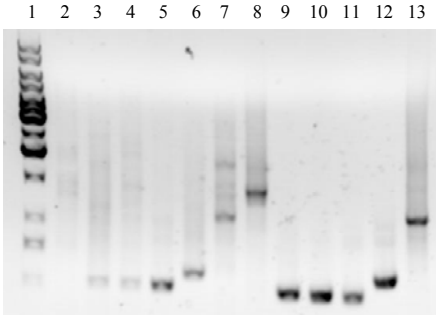


Figure 3.11. Genome walking nested PCR results for s272. Lanes 2 to 7 were results from PCR 1st round. Lanes 2 to 4 were results for upstream of s272. Lanes 5 to 7 were results for downstream of s272. Lanes 2 and 5 were from *Dra*I library pool sdDL1. Lanes 3 and 6 were from *Pvu*II library pool sdDL3. Lanes 4 and 7

(Figure 3.11 continued) were from *Stu*I library pool sdDL4. Lanes 8 to 13 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 7, respectively. Lane 1 contains 1 kb DNA ladder (Fermentas).

The longest bands for s272 upstream and downstream were generated from sdDL1 (No. 8) and sdDL4 (No. 13), so the PCR for extension the next time should use the remaining library pools. But because the band from the upstream (sdDL1 pool) was not very strong, this library pool was also used in the PCR generation in the second round (Figure 3.12).

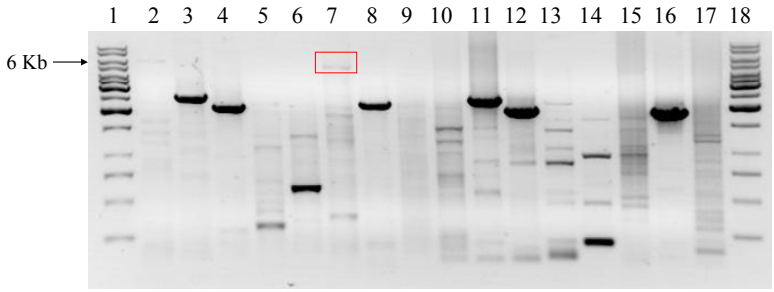


Figure 3.12. Genome walking nested PCR results for s272 (and p322 downstream). Lanes 2 to 9 were results from PCR 1st round. Lanes 2 to 4 were results for p322 (refer to Figure 2.37). Lanes 5 to 7 were results for upstream of s272. Lanes 8 and 9 were results for downstream of s272. Lane 5 was from *Dra*I library pool sdDL1. Lane 6 was from *Pvu*II library pool sdDL3. Lane 7 was from *Stu*I library pool sdDL4. Lane 8 was from *Dra*I library pool sdDL1. Lane 9 was from *Stu*I library pool sdDL4. Lanes 10 to 17 were from nested PCR 2nd round amplifications of 1/50 dilutions of the products shown in lanes 2 to 9, respectively. Lanes 1 and 18 contain 1 kb DNA ladder (Fermentas).

For the upstream sequence of s272, there is a strong band of about 0.3 Kb from the 2nd round PCR (Figure 3.12 lane 14), a strong band of about 0.65 Kb (Figure 3.12 lane 6) and a weak band of about 5.7 Kb (Figure 3.12 lane 7) from the 1st round

PCR. According to the GenomeWalker Kit, the upper limit of PCR product is about 6 Kb.[49] The ~ 5.7 Kb band triggered our interest. The reason that it was weak may be due to its large size. If this 5.7 Kb band is the right band, it will contain much more information. This ~ 5.7 Kb fragment was sequenced along with the 0.3 Kb fragment from the 2nd round PCR and the 0.65 Kb fragment from the 1st round PCR. It was found that the overlapped sequences of these fragments are identical and also identical with the sequence obtained from the previous round of genome walking. This indicated that the ~ 5.7 Kb fragment is the right fragment. Further sequencing gave continued sequence of the gene that was partially sequenced in the prior round of genome walking originated from, which further confirmed the authenticity of the fragment. Therefore, this fragment was fully sequenced and gave about ~ 5.7 Kb of sequence information upstream of s272.

From the DNA sequence obtained from genome walking and by comparing this sequence with those of known fungal *ggs* genes, all the fungal *ggs* gene conserved motifs were located in the obtained sequences for s224 and s272. The ATG start codons were proposed by referring to the sequences of known fungal *ggs* genes. The proposed genomic DNA sequence of 990 bp for the full proposed *ggs* gene derived from s224, which was designated as Sord224, and that of the 1305 bp gene derived from s272, which was designated as Sord272, were obtained. The sequences of these putative *ggs* genes are shown in the appendix. Translations of these sequences were found to share high similarity with known fungal GGPP synthases such as *Epichloe typhina* (*E. t.*) (Figures 3.13). They also share 43%

identity with each other. The BLAST search results strongly support their identities as GGPP synthases, although biochemical tests would be necessary to confirm this. Five homologs out of many for the potential *ggs* genes Sord224 and Sord272, are shown in Table 3.6.

```
Sord272   PYDYLNEHPGK--DFRSQLVKAFD-AWLEVP-PESLEVITKVVGMLHTASLLVDNVEDFS 56
E.t.      PFDYVAGHPGK--DFRSQLISSFN-AFLQVP-PDRLEIITRAVGMLHESSLLVDDVQDSS 56
Sord224   PYDYLAAIPGTNGDIRTQFIKSFNDLYFTIDNHDTIATISEVITVFHNSSLLIDDIEDGS 60
          *:**:   **.  *:*:*::.:*:  :: :   : :  *:..: ::* :***:*:::* *

Sord272   LLRRGFPVAHSIFGIPQTINSSNYVYFCALQELQKLK---NPK---------AVSVFTEE 104
E.t.      ELRRGFPVAHNIFGVPQTINSANYIYFVALQELQKLG---NPP---------IIAIFADE 104
Sord224   EFRRGLPAAHTKYGVALTLNCGNLMYFVALQTAQQKLPQYHPSDNKNDIALSILNILVEE 120
          :***:*.**. :*:. *:*..* :** ***  *:     :*           : ::.:*

Sord272   LINLHRGQGMDLFWRDTLT---CPTEDEYLEMVSNKTGGLFRLGIKLMQAESRS--LVDC 159
E.t.      LVNLHRGQGMDLFWRDTLT---CPTEDDYLEMVGNKTGGLFRLGIKLMQAESAA--AVDC 159
Sord224   LLNLHHGQGLDIRWRDHLKTTKLPLIEEYMEMIMNKTGGLFRLAVKLLGAFSTPQDKPTV 180
          *:***:***:*: *** *.    *  ::*:**: *********.:**: * * .

Sord272   VPLVSIMGLIFQIADDYQNLWSRRYTANKGM-CEDLTEGKFSFPVIH 205
E.t.      VPLVNLIGLIFQIRDDYMNLSSREYSDNKGL-CEDLTEGKFSFPIIH 205
Sord224   LAIANLFGIVFQIRDDLMNLVDDKYSHMKGMKGEDLVEGKLSLPILH 227
          :.:..::*::*** **  ** . .*:  **:  ***.***:*:*::*
```

Figure 3.13. Alignment of *S. araneosa* GGPP synthases with a representative fungal GGPPS from *Epichloe typhina* (*E. t.*). Key conserved GGPPS regions are underlined. N-terminal and C-terminal extensions with low identity have been omitted.

Table 3.6. The five homologs of the *ggs* genes Sord224 and Sord272.

ggs genes	Homologues	Organisms	Identities	Similarities
Sord224	GGPP synthase	*Candida albicans*; Fungi; Ascomycota.	52%	70%
	GGPP synthase	*Penicillium paxilli*; Fungi; Ascomycota.	42%	62%
	GGPP synthase	*Aspergillus nidulans*; Fungi; Ascomycota.	41%	61%
	GGPP synthase	*Epichloe typhina*; Fungi; Ascomycota.	41%	57%
	GGPP synthase	*Gibberella fujikuroi*; Fungi; Ascomycota.	40%	57%
Sord272	GGPP synthase	*Neurospora crassa*; Fungi; Ascomycota.	71%	82%
	GGPP synthase	*Gibberella zeae*; Fungi; Ascomycota.	67%	84%
	GGPP synthase	*Epichloe typhina*; Fungi; Ascomycota.	67%	84%
	GGPP synthase	*Penicillium paxilli*; Fungi; Ascomycota.	67%	82%
	GGPP synthase	*Magnaporthe grisea*; Fungi; Ascomycota.	65%	78%

Gene organization near S. araneosa ggs genes

The combined sequencing results from genome walking are illustrated in Figure 3.14. About 9.5 Kb of sequence was obtained for the Sord224 region, and about 9 Kb of sequence was obtained for the Sord272 region.

About 5.3 Kb sequence upstream and 3.2 Kb sequence downstream of Sord224 putative *ggs* gene were obtained. For both of the Sord224 and Sord272 genes, there are no diterpene synthase genes in close proximity to the *ggs* genes. Three hypothetical proteins (one is partial) were found upstream of the Sord224 *ggs* gene. These hypothetical proteins (or partial hypothetical proteins) do not seem to be

related to isoprenoid biosynthesis based on BLAST searches and by checking the genes flanking those homolog genes from known sequences. Three open reading frames, a diphthamide methyltransferase, a hypothetical protein and a partial alpha-aminoadipate reductase gene were found downstream of the *ggs* gene. The diphthamide methyltransferase is related to general function of the Elongation Factor 2, but not related to the self-resistance of the fungus. The obtained sequence of the partial alpha-aminoadipate reductase was about 350 bp, while the whole size of alpha-aminoadipate reductase is more than 4 Kb by the homology search. Therefore, it is not likely that there is a diterpene synthase gene downstream of this alpha-aminoadipate reductase.

About 5.3 Kb of sequence upstream of Sord272 putative *ggs* gene was obtained and two hypothetical proteins were found. These hypothetical proteins do not seem to be related to isoprenoid biosynthesis based on the homology searches. About 2.4 Kb of sequence downstream of the Sord272 putative *ggs* gene were obtained and a hypothetical protein with strong hits to the membrane peroxin, which is not related to the isopenoid biosynthesis, was found. Because in all the known diterpene clusters, the diterpene synthase gene was located upstream of the *ggs* gene, it is not likely that there is a diterpene synthase gene downstream of this hypothetical protein. Although the possibility can not be completely ruled out that the genes related to the sordarin biosynthesis are located further away for the *ggs* gene, the chance is small according to the information from known fungal diterpene clusters. Therefore, no further genome walking was performed. Despite

the lack of success with the direct approach to fungal diterpene synthase genes and the apparent lack of clustering of a *ggs* gene with the sordarin diterpene biosynthetic gene cluster, future success may be obtained through SSH studies or a reverse genetics approach.

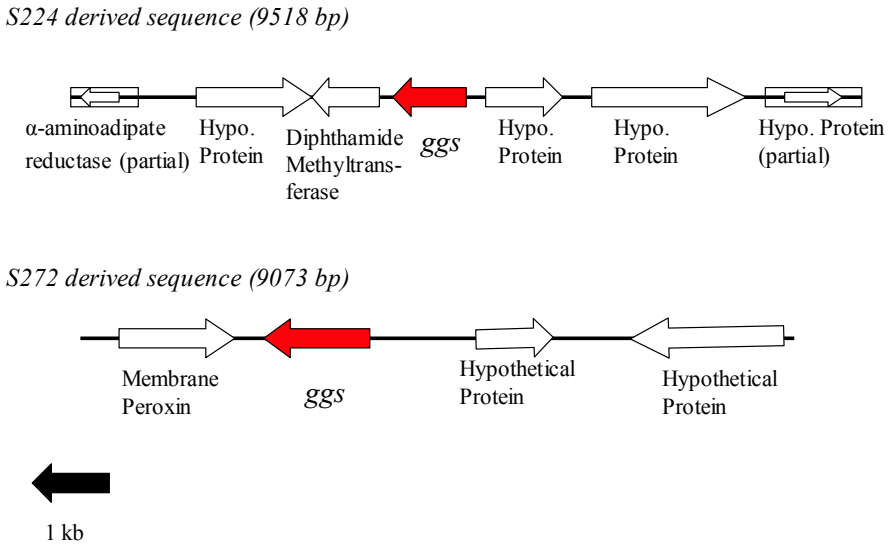


Figure 3.14. Gene organization near *S. araneosa ggs* genes.

Materials and methods

General chemistry

Chemicals were purchased from Sigma-Aldrich, Fisher Scientific and VWR International.

NMR analyses

The 1D and 2D NMR, ^{1}H, ^{13}C, COSY, DEPT, HSQC and HMBC were performed on Bruker AC 300 and AM 400 spectrometers.

General molecular biology

Routine molecular biology procedures including DNA manipulation, transformation, growth and maintenance of *Escherichia coli* (*E. coli*), competent cell preparation and electroporations were conducted according to standard techniques.[50] Plasmids were propagated in *Escherichia coli* DH10B. Restriction enzymes, T4 DNA ligase and *Taq* polymerase were purchased from various suppliers and used according to the manufacturer's protocol. All PCRs were conducted in an Eppendorf Mastercycler gradient PCR thermal cycler. QIAprep® spin miniprep and QIAquick® gel extraction kits (Qiagen, Valencia, CA) were used for DNA purification. Sequencing of DNA inserts was conducted by the Center for Genome Research and Biocomputing (CGRB) of Oregon State University. Chemicals were purchased from Sigma-Aldrich, Fisher Scientific and VWR International.

Fungal strains and growth conditions

Sordaria araneosa Cain (ATCC # 36386) was obtained from ATCC (American Type Culture Collection). *Sordaria araneosa* Cain is maintained on YpSs agar [Yeast protein Soluble starch agar, Difco 0739. The recipe contains (/L): yeast

extract 4.0 g, soluble starch 15.0 g, K_2HPO_4 1.0 g, $MgSO_4 \cdot 7H_2O$ 0.5 g, agar 20.0 g] and grown at 26 °C in liquid medium that contains following elements (per liter): glucose 20 g, malt extract 2 g, peptone 2 g, malt extract 2 g, KH_2PO_4 2 g and $MgSO_4 \cdot 7H_2O$ 2 g. Both shaking and static conditions were used to grow the fungus. Under the shaking condition, the 1 liter culture was grown in a rotary shaker with 200 rpm constant shaking. A 100 mL 3-day culture in the same medium was used as inoculum. Under the static condition, the cultures were grown on the bench at room temperature without shaking.

Isolation and purification of sordarin and its derivatives

Extraction of culture

The procedure followed the literature procedure with some modifications.[27] *Sordaria araneosa* Cain (ATCC 36386) was fermented at 26 °C, 200 rpm in a 1 liter medium composed of (g/litre): Glucose 20, malt extract 2, peptone 2, yeast extract 2, KH_2PO_4 2, $MgSO_4 \cdot 7H_2O$. A 3-day culture of *S. araneosa* in the same medium was used as inoculum. After five days the mycelium was removed by filtration. The Diaion HP20 (Mitsubishi) adsorbing resin (40 g resin / 1 L medium) was added into the filtrate and the mixture was shaken overnight on a rotary shaker at room temperature. The HP20 resin was packed into a column (20 × 3 cm), which was eluted with acetone/water 1:1 300 ml and then acetone 200 mL. The color of the resin turned from yellow back to white. The original eluent was dark brown, while the final eluent was colorless. The pooled eluents were concentrated

in vacuo to evaporate the acetone. The resulting slurry was diluted with water to about 200 mL and extracted with EtOAc (3 × 200 mL). The combined organic phases were dried with anhydrous $MgSO_4$ and evaporated to give the crude extract. The crude extract was subjected to flash chromatography on a silica gel (Merck 60, 0.063 to 0.2 mm particle size) column (30 × 2 cm) and eluted with hexane / EtOAc 30:70 200 mL and then 10:90 250 mL. The fractions showing the same spots under UV and cerium sulfate stain, on silica TLC (thin layer chromatography) with hexane / EtOAc 30:70 solvent system, are combined. The R_f of sordarin was 0.30 with hexane / EtOAc 3:7 solvent system on silica TLC. The combined organic phases were dried with anhydrous $MgSO_4$, evaporated to give the partially purified componds and then subjected to HPLC (condition see below, *LC-MS isolation and purification of sordarin and its derivatives*), LC-MS and NMR.

Extraction of mycelia

The mycelia were extracted with methanol: H_2O (9:1) (4 × 200ml). The methanol was removed in vacuo and the aqueous phase (pH= 5.5) was adjusted to pH 2-3 using 6 M HCl and extracted with methylene chloride (3 × 200ml). The combined organic layers were dried with $MgSO_4$ and concentrated in vacuo to give yellow oil. The extract was combined with the filtrate extract and then subjected to flash chromatography using the procedure described above.

LC-MS isolation and purification of sordarin and its derivatives

Reversed phase HPLC analyses were performed using a Hamilton PRP-1 reverse phase column (10 μm; 10.0 × 250 mm) with guard column and a Beckman Ultrasphere C18 column (5 μm; 4.6 × 250 mm) with guard column. LC-MS was performed on a ThermoFinnigan LCQ Advantage LC-MS system, consisting of a solvent pump, an autosampler, a PDA detector and an ion trap mass spectrometer detector. The system was controlled by a PC running Xcalibur 1.3 software. Isocratic elution was carried out using 70% of acetonitrile for 45 min and then 100% of acetonitrile for 25 min at a flow rate of 0.5 mL / min. The effluent was monitored at 254 nm and meanwhile scanned from 200 nm to 350 nm with the PDA detector. The MS was performed using (+) and (-) APCI and ESI. The hydroxysordarin, sordarin and neosordarin eluted from the Hamilton PRP-1 reverse phase column at around 12 min, 21 min and 34 min, respectively.

Time course of sordarin production

The 5 mL cultures from the 1st day to the 12th day were passed through a 6 cc (50 mg) C_{18} Sep-Pak (J. T. Baker) respectively and the eluents were passed through the Sep-Pak column two more times. The column was washed with 30 ml ddH_2O, and eluted with 20 mL CH_3CN. The combined CH_3CN eluent was dried in vacuo and redissolved in 500 μL of CH_3CN. MS was performed in both (+) and (-) APCI mode with direct injection.

Nucleic acid analysis, RNA handling and isolation and all the materials and methods in direct identification of the diterpene synthase gene and identification of the diterpene gene cluster via the *ggs* gene were the same as in Chapter 2.

References

1. Beck-Sague, C. M.; Jarvis, W. R. "Secular trends in the epidemiology of nosocomial fungal infections in the United States. 1980-90. National nosocomial infections surveillance system," *J. Infect. Dis.* **1993,** *167*, 1247-51.
2. Diamond, R. D. "The growing problem of mycoses in patients infected with the human immunodeficiency virus," *Rev. Infect. Dis.* **1991,** *13*, 480-6.
3. De Beule, K.; Van Gestel, J. "Pharmacology of itraconazole," *Drugs* **2001,** *61*, 27-37.
4. Brajtburg, J.; Powderly, W. G.; Kobayashi, G. S.; Medoff, G. "Amphotericin B: current understanding of mechanisms of action," *Antimicrob. Agents Chemother.* **1990,** *34*, 183-8.
5. Perez, A. "Terbinafine: broad new spectrum of indications in several subcutaneous and systemic and parasitic diseases," *Mycoses* **1999,** *42*, 111-4.
6. Nyfeler, R.; Keller, S. W. "Metabolites of microorganisms.143. Echinocandin B, a novel polypeptideantibiotic from *Aspergillus nidulans* var. echinulatus: isolation and structural components," *Helv. Chim. Acta.* **1974,** *57*, 2459-77.
7. Georgopapadakou, N. H.; Walsh, T. J. "Human mycoses: drugs and targets for emerging pathogens," *Science* **1994,** *264*, 371-3.
8. Carledge, J. D.; Midgley, J.; Gazzard, B. G. "Clinically significant azole cross-resistance in Candida isolates from HIV-positive patients with oral candidiasis," *AIDS* **1997,** *11*, 1839-44.
9. Muller, F. C.; Weig, M.; Peter, J.; Walsh, T. J. "Azole cross - resistance to ketoconazole, fluconazole, itraconazole, and voriconazole in clinical *Candida albicans* isolates from HIV-infected children with *oropharyngeal candidosis*," *J. Antimicrob. Chemother.* **2000,** *46*, 338-41.
10. Vicente, M. F.; Basilio, A.; Cabello, A.; Pelaez, F. "Microbial natural products as a source of antifungals," *Clin. Microbiol. Infec.* **2003,** *9*, 15-32.
11. Petranyi, G.; Ryder, N. S.; Stutz, A. "Allylamine derivatives: new class of synthetic antifungal agents inhibiting fungal squalene epoxidase," *Science* **1984,** *224*, 1239-41.
12. Hall, C. C.; Bertasso, A. B.; Watkins, J. D.; Georgopapadakou, N. H. "Screening assays for protein synthesis inhibitors," *J. Antibiotics* **1992,** *45*, 1697-9.
13. Belfield, G. P.; Ross-Smith, N. J.; Tuite, M. F. "Translation elongation factor-3 (EF-3): an evolving eukaryotic ribosomal protein?," *J. Mol. Evol.* **1995,** *41*, 367-87.

14. Kamath, A.; Chakraburtty, K. "Role of yeast Elongation Factor 3 in the Elongation Cycle," *J. Biol. Chem.* **1989,** *264*, 15423-8.
15. Skogerson, L.; Engelhardt, D. "Dissimilarity in chain elongation factor requirements between yeasts and rat liver ribosomes," *J. Biol. Chem.* **1977,** *252*, 1471-5.
16. Justice, M. C.; Hsu, M.; Tse, B.; Ku, T. B., J.; Schmatz, D.; Nielsen, J. "Elongation factor 2 as a novel target for selective inhibition of fungal protein synthesis," *J. Biol. Chem.* **1998,** *273*, 3148-51.
17. Gudkov, A. T. "The L7/L12 ribosomal domain of the ribosome: structural and functional studies," *FEBS Lett.* **1997,** *407*, 253-6.
18. Ballesta, J. P. G.; Remacha, M. "The large ribosomal subunit stalk as a regulatory element of the eukaryotic translational machinery," *Prog. Nucleic Acid Re.* **1996,** *55*, 157-93.
19. Santos, C.; Ballesta, J. P. G. "The highly conserved protein P0 carboxyl end is essential for ribosome activity only in the absence of proteins P1 and P2," *J. Biol. Chem.* **1995,** *270*, 20608-14.
20. Stark, H.; Rodnina, M. V.; Rinkeappel, J.; Brimacombe, R.; Wintermeyer, W.; Vanheel, M. "Visualization of elongation factor Tu on the *Escherichia coli* ribosome," *Nature* **1997,** *389*, 403-6.
21. Wilson, K. S.; Noller, H. F. "Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing," *Cell* **1998,** *92*, 131-9.
22. Frank, J.; Agrawal, R. K. "A ratchet-like inter-subunit reorganization of the ribosome during translocation," *Nature* **2000,** *406*, 318-22.
23. Jorgensen, R.; Ortiz, P. A.; Carr-Schmid, A.; Nissen, P.; Kinzy, T. G.; Andersen, G. R. "Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase," *Nat. Struct. Biol.* **2003,** *10*, 379-85.
24. Beckmann, R.; Spahn, C. M. T.; Eswar, N.; Helmers, J.; Penczek, P. A.; Sali, A.; Frank, J.; Blobel, G. "Architecture of the protein-conducting channel associated with the translating 80S ribosome," *Cell* **2001,** *107*, 361-72.
25. Spahn, C. M. T.; Beckmann, R.; Eswar, N.; Penczek, P. A.; Sali, A.; Blobel, G.; Frank, J. "Structure of the 80S ribosome from *Saccharomyces cerevisiae*-tRNA-ribosome and subunit-subunit interactions," *Cell* **2001,** *107*, 373-86.
26. Hauser, D.; Sigg, H. P. "Isolierung und abbau von sordarin," *Helv. Chim. Acta* **1971,** *54*, 1178-90.
27. Davoli, P.; Engel, G.; Werle, A.; Sterner, O.; Anke, T. "Neosodarin and hydroxysodarin, two new antifungal agents from *Sordaria araneosa*," *J. Antibiot.* **2002,** *55*, 377-82.
28. Martinez, A.; Regadera, J.; Jimenez, E.; Santos, I.; Gargallo-Viola, D. "Antifungal efficacy of GM237354, a sordarin derivative, in experimental

oral candidiasis in immunosuppressed rats," *Antimicrob. Agents Chemother.* **2001,** *45*, 1008-13.

29. Martinez, A.; Ferrer, S.; Santos, I.; Jimenez, E.; Sparrowe, J.; Regadera, J.; Gomez de las Heras, F.; Gargallo-Viola, D. "Antifungal activities of two new azasordarins GW471552 and GW471558, in experimental models of oral and vulvovaginal candidiasis in immunosuppressed rats," *Antimicrob. Agents Chemother.* **2001,** *45*, 3304-9.
30. Torres-Rodriguez, J. M.; Morera, Y.; Baro, T.; Lopez, O.; Alia, C.; T., J. "In vitro susceptibility of Cryptococcus neoformans serotypes to GM 237354 derivative of the sordarin class," *Mycoses* **2002,** *45*, 313-6.
31. Herreros, E.; Martinez, C. M.; Almela, M. J.; Marriott, M. S.; Gomez de las Heras, F.; Gargallo-Viola, D. "Sordarins: in vitro activities of new antifungal derivatives against pathogenic yeasts, *Pneumocystis carinii*, and filamentous fungi," *Antimicrob. Agents Chemother.* **1998,** *42*, 2863-9.
32. Aviles, P.; Alioquat, E.; Martinez, A.; Dei-cas, E.; Herreros, E.; Dujardin, L.; Gargallo-Violai, D. "In vitro pharmacodynamic parameters of sordarin derivatives in comparison with those of marketed compounds against *Pneumocystis carinii* isolated from rats," *Antimicrob. Agents Chemother.* **2000,** *44*, 1874-7.
33. Graybill, J. R.; Najvar, L.; Fothergill, A.; Bocanegra, R.; Gomez de las Heras, F. "Activities of sordarins in murine histoplasmosis," *Antimicrob. Agents Chemother.* **1999,** *43*, 1716-8.
34. Gargallo-Viola, D. "Sordarins as antifungal compounds," *Curr. Opin. Anti-infect. Invest. Drugs* **1999,** *1*, 297-305.
35. Clemons, K. V.; Stevens, D. A. "Efficacies of sordarin derivatives GM193663, GM211676, and GM237354 in a murine model of systemic coccidioidomycosis," *Antimicrob. Agents Chemother.* **2000,** *44*, 1874-7.
36. Dominguez, J. M.; Kelly, V. A.; Kinsman, O. S.; Marriot, M. S.; Gomez de las Heras, F.; Martin, J. J. "Sordarins: a new class of antifungal with selective inhibition of the protein synthesis elongation cycle in yeasts," *Antimicrob. Agents Chemother.* **1998,** *42*, 2274-8.
37. Dominguez, J. M.; Martin, J. J. "Identification of elongation factors as the essential protein targeted by sordarins in *Candida albicans*," *Antimicrob. Agents Chemother.* **1998,** *42*, 2279-83.
38. Spahn, C. M.; Gomez-Lorenzo, M. G.; Grassucci, R. A.; Jorgensen, R.; Andersen, G. R.; Beckmann, R.; Penczek, P. A.; Ballesta, J. P.; Frank, J. "Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation FREE," *Embo J.* **2004,** *23*, 1008-19.
39. Capa, L.; Mendoza, A.; Lavandera, J. L.; Gomez de las Heras, F.; Garcia-Bustos, J. F. "Translation Elongation Factor 2 is part of the target for a new family of antifungals," *Antimicrob. Agents Chemother.* **1998,** *42*, 2694-9.
40. Gomez-Lorenzo, M. G.; Garcia-Bustos, J. F. "Ribosomal P-protein stalk function is targeted by sordarin antifungals," *J. Biol. Chem.* **1998,** *273*, 25041-4.

41. Justice, M. C.; Ku, T.; Hsu, M. J.; Carniol, K.; Schmatz, D.; Nielsen, J. "Mutations in ribosomal protein L10e confer resistance to the fungal-specific eukaryotic elongation factor 2 inhibitor sordarin," *J. Biol. Chem.* **1999,** *274*, 4869-75.
42. Kinsman, O. S.; Chalk, P. A.; Jackson, H. C.; Middleton, A.; Shuttleworth, B. A.; Rudd, A. M.; Jones, H. M.; Noble, H. G.; Wildman, M. J.; Dawson, C.; Stylli, P. J.; Sidebottom, B.; Lamont, S. L.; Hayes, M. V. "Isolation and characterization of an antifungal antibiotic (GR135402) with protein synthesis inhibition," *J. Antibiot.* **1998,** *51*, 41-9.
43. Kennedy, T. C.; Webb, G.; Cannell, R. J. P.; Kinsman, O. S.; Middleton, R. F.; Sidebottom, P. J.; Taylor, N. L.; Dawson, M. J.; Buss, A. D. "Novel inhibitors of fungal protein synthesis produced by a strain of Graphium putredinis. Isolation characterization and biological properties," *J. Antibiot.* **1998,** *51*, 1012-8.
44. Daferner, M.; Mensch, M.; Anke, T.; Sterner, O. "Hydroxysodarin, a new sodarin derivative from *Hypoxylon croceum*," *Z. Naturforsch* **1999,** *54c*, 474-80.
45. Tse, B.; Balkovec, J. M.; Blazey, C. M.; Hsu, M.-J.; Nielsen, J.; Schmatz, D. "Alkyl side-chain derivatives of sordaricin as potent antifungal agents against yeast," *Bioorg. Med. Chem. Lett.* **1998,** *8*, 2269-72.
46. Cuevas, J. C.; Lavandera, J. L.; Martos, J. L. "Design and synthesis of simplified sordaricin derivatives as inhibitors of fungal protein synthesis," *Bioorg. Med. Chem. Lett.* **1999,** *9*, 103-8.
47. Jenny, L.; Borschberg, H. "Synthesis of the dolabellane diterpene hydrocarbon (±)-*δ*-araneosene," *Helv. Chim. Acta* **1995,** *78*, 715-31.
48. Mander, L. N.; Thomson, R. J. "Total synthesis of sordaricin," *Org. Lett.* **2003,** *5*, 1321-4.
49. Clontech. "BD GenomeWalker Universal Kit user manual," **2004**.
50. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning: a laboratory manual* **1989**, Second edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

CHAPTER FOUR

General Conclusion

This dissertation describes studies aimed at discovering the diterpene synthase enzymes which provide the backbone structures for pleuromutilin and sordarin, two fungal diterpene antibiotics. Pleuromutilin is produced by the fungus *Pleurotus passeckerianus* Pilat and this family of compounds inhibits mainly Gram-positive bacteria and mycoplasms. The pleuromutilin class of antibiotics has been used in veterinary medicine for a long time and now is being reinvestigated for use in humans. Sordarin is produced by the fungus *Sordaria araneosa* Cain. Sordarin and its derivatives are antifungal compounds.

Pleuromutilin and sordarin possess interesting structural scaffolds. They are derived from the common precursor GGPP, but undergo different biosynthetic pathways to give the different core structures. A greater understanding of the biosynthetic enzymes for these two compounds may allow us to eventually manipulate the corresponding genes, creating new scaffolds for bioactive molecules.

In this dissertation, attempts to identify the biosynthetic gene clusters for the formation of pleuromutilin and sordarin are described. Three methods were used to locate the diterpene synthase genes.

The first method utilized PCR (Polymerase Chain Reaction) and degenerate primers designed from identified and putative fungal diterpene synthase genes. A variety of conditions were tried and degenerate primers were redesigned, but no desired PCR products were found. The reason that this method did not work might be the limited number of diverse fungal diterpene synthase genes available; the degenerate primers might not be suitable for the specific diterpene synthase genes from *P. passeckerianus* and *S. araneosa*.

The second method was based on locating diterpene synthase genes through the related *ggs* (GGPP synthase) genes. This approach has worked for several fungal clusters. This method is based on the hypothesis that the terpene biosynthesis genes in sordarin and pleuromutilin-producing fungi are clustered with a dedicated *ggs* gene. Many fungal *ggs* gene sequences are available and they contain regions that are highly conserved. Therefore, degenerate primers can be designed that are more likely to give specific PCR products from genomic DNA. Three putative *ggs* fragments were identified from *P. passeckerianus* and two fragments were identified from *S. araneosa*. The BLAST searches strongly supported their *ggs* gene identities. These putative *ggs* genes from *P. passeckerianus* and *S. araneosa* were used as starting points to locate the pleuromutilin and sordarin biosynthetic genes by probing and screening a genomic DNA library and by genome walking methods. About 10 kb of sequence extended from each fragment was obtained. No genes associated with the pleuromutilin and sordarin biosyntheses were yet found. These results and a recent review[1] suggest that fungal diterpene gene clusters do

not necessarily need a dedicated *ggs*. The biosynthesis genes for the formation of pleuromutilin and sordarin seem to fall into the category of those that do not require a dedicated *ggs* gene.

The third method is the Suppression Subtractive Hybridization (SSH) method which is a technique to find differentially expressed genes.[2] This method was tried with *P. passeckerianus*, the pleuromutilin producer. A subtracted library was formed from cDNA prepared from a 2^{nd} day culture that did not produce pleuromutilin and from cDNA from a 7^{th} day culture that produced pleuromutilin. The pleuromutilin biosynthetic genes should be among the genes in the subtracted library. The PCR amplification results suggest that the SSH method has worked, but sequencing of clones from the subtracted library will be necessary to identify genes that are associated with the pleuromutilin diterpene synthesis.

In conclusion, three methods have been attempted to identify the diterpene synthase genes for the formation of pleuromutilin and sordarin. Although no diterpene biosynthetic genes have been found to date, a solid foundation has been established for further studies in this area.

References

1. Kawaide, H. "Biochemical and molecular analyses of gibberellin biosynthesis in fungi," *Biosci. Biotechnol. Biochem.* **2006,** *70*, 583-90.
2. Diatchenko, L.; Lau, Y. F.; Campbell, A. P.; Chenchik, A.; Moqadam, F.; Huang, B.; Lukyanov, S.; Lukyanov, K.; Gurskaya, N.; Sverdlov, E. D.; Siebert, P. D. "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries," *Proc. Natl. Acad. Sci. USA.* **1996,** *93*, 6025-30.

Bibliography

1. Abe, I.; Rohmer, M.; Prestwich, G. D. "Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes," *Chem. Rev.* **1993,** *93*, 2189-206.
2. Adam, P.; Hecht, S.; Eisenreich, W.; Kaiser, J.; Grâwert, T.; Arigoni, D. "Biosynthesis of terpenes. Studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase," *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 12108-13.
3. Agrios, G. N. *Plant Pathology* **1998,** *3rd edition, Academic Press, San Diego*, 403-7.
4. Ait-Ali, T.; Swain, S. M.; Reid, J. B.; Sun, T. P.; Kamiya, Y. "The LS locus of pea encodes the gibberellin biosynthesis enzyme *ent*-kaurene synthase A " *Plant J.* **1997,** *11*, 443-54.
5. Alexander, R. W. "Teasing apart the taxol pathway," *Trends Biochem. Sci.* **2001,** *26*, 152.
6. Altincicek, B.; Duin, E. C.; Reichenberg, A.; Hedderich, R.; Kollas, A. K.; Hintz, M. "LytB protein catalyses the terminal step of the 2-C-methyl-D-ehrtol-4-phosphate pathway of isoprenoid biosynthesis," *FEBS Lett.* **2002,** *532*, 437-40.
7. Amersham Biosciences. "Hybond-N+ protocol,".
8. Andersen, G. R.; Nissen, P.; Nyborg, J. "Elongation factors in protein biosynthesis," *Trends Biochem. Sci.* **2003,** *28*, 434-41.
9. Anderson, M. L. M. "Nucleic acid hybridization," **1999**, Bios Scientific Publishers; New York: Springer.
10. Arigoni, D. "Structure of a new type of terpene," *Gazz. Chim. Ital.* **1962,** *92*, 884-901.
11. Aviles, P.; Alioquat, E.; Martinez, A.; Dei-cas, E.; Herreros, E.; Dujardin, L.; Gargallo-Violai, D. "In vitro pharmacodynamic parameters of sordarin derivatives in comparison with those of marketed compounds against *Pneumocystis carinii* isolated from rats," *Antimicrob. Agents Chemother.* **2000,** *44*, 1874-7.
12. Bach, T. J. "Some new aspects of isoprenoid biosynthesis in plants - a review," *Lipids* **1995,** *30*, 191-202.
13. Ballesta, J. P. G.; Remacha, M. "The large ribosomal subunit stalk as a regulatory element of the eukaryotic translational machinery," *Prog. Nucleic Acid Re.* **1996,** *55*, 157-93.
14. Beckmann, R.; Spahn, C. M. T.; Eswar, N.; Helmers, J.; Penczek, P. A.; Sali, A.; Frank, J.; Blobel, G. "Architecture of the protein-conducting channel associated with the translating 80S ribosome," *Cell* **2001,** *107*, 361-72.
15. Beck-Sague, C. M.; Jarvis, W. R. "Secular trends in the epidemiology of nosocomial fungal infections in the United States. 1980-90. National

nosocomial infections surveillance system," *J. Infect. Dis.* **1993,** *167*, 1247-51.

16. Belfield, G. P.; Ross-Smith, N. J.; Tuite, M. F. "Translation elongation factor-3 (EF-3): an evolving eukaryotic ribosomal protein?," *J. Mol. Evol.* **1995,** *41*, 367-87.
17. Bensen, R. J.; Johal, G. S.; Crane, V. C.; Tossberg, J. T.; Schnable, P. S.; Meeley, R. B.; Briggs, S. P. "Cloning and characterization of the maize An1 gene," *Plant Cell* **1995,** *7*, 75-84.
18. Berkow, R. *The Merck Manual of Medical Information - Home Edition* **1999**.
19. Berner, H.; Vyplel, H.; Schulz, G.; Stuchlik, P. "Chemie der Pleuromutiline, IX. Konfigurationsumkehr der Methylgruppe am Kohlenstoff 6 im tricyclischen Gerüst des Diterpens Pleuromutilin," *Monatsh. Chem.* **1983,** *114*, 1125-36.
20. Birch, A. J.; Cameron, D. W.; Holzapfel, C. W.; Richards, R. W. "Diterpene nature of pleuromutilin," *Chem. Ind.* **1963**, 374-5.
21. Birch, A. J.; Holzapfel, C. W.; Richards, R. W. "The structure and some aspects of the biosynthesis of pleuromutilin," *Tetrahedron. Suppl. (part II)* **1966,** *8*, 358-87.
22. Bloch, K. "Sterol molecule: structure, biosynthesis, and function," *Steroids* **1992,** *57*, 378-82.
23. Bloch, K.; Chaykin, S.; Phillps, J. W.; deWaard, A. "Mevalonic acid pyrophosphate and isopentenylpyrophosphate," *J. Biol. Chem.* **1959,** *234*, 2595-604.
24. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. "Plant terpenoid synthases: molecular biology and phylogenetic analysis," *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 4126-33.
25. Bonavia, G. "Pleuromutilin, sterochemie und detaillierte biosynthesis," *Diss. Nr. ETH Zurich* **1968**, 4189.
26. Brajtburg, J.; Powderly, W. G.; Kobayashi, G. S.; Medoff, G. "Amphotericin B: current understanding of mechanisms of action," *Antimicrob. Agents Chemother.* **1990,** *34*, 183-8.
27. Butler, M. S.; Buss, A. D. "Natural Products - The Future Scaffolds for Novel Antibiotics," *Biochem. Pharmacol.* **2006,** *71*, 919-29.
28. Buzzolini, M. "Zur biogenesis des pleuromutilin und der lagopodine," *Diss. Nr. ETH Zurich* **1966**, 3797.
29. Cane, D. E. "Isoprenoid biosynthesis. Stereochemistry of the cyclization of allylic pyrophosphates," *Acc. Chem. Res.* **1985,** *18*, 220-6.
30. Cane, D. E. "Enzymic formation of sesquiterpenes," *Chem. Rev.* **1990,** *90*, 1089-103.
31. Cane, D. E.; Kang, I. "Aristolochene synthase: purification, molecular cloning, high-level expression in Escherichia coli, and characterization of the *Aspergillus terreus* cyclase," *Arch. Biochem. Biophys.* **2000,** *376*, 354-64.

32. Capa, L.; Mendoza, A.; Lavandera, J. L.; Gomez de las Heras, F.; Garcia-Bustos, J. F. "Translation Elongation Factor 2 is part of the target for a new family of antifungals," *Antimicrob. Agents Chemother.* **1998,** *42*, 2694-9.
33. Carledge, J. D.; Midgley, J.; Gazzard, B. G. "Clinically significant azole cross-resistance in Candida isolates from HIV-positive patients with oral candidiasis," *AIDS* **1997,** *11*, 1839-44.
34. Chen, A. P.; Kroon, P. A.; Poulter, C. D. "Isoprenyl diphosphate synthases: protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure," *Protein Sci.* **1994,** *3*, 600-7.
35. Clemons, K. V.; Stevens, D. A. "Efficacies of sordarin derivatives GM193663, GM211676, and GM237354 in a murine model of systemic coccidioidomycosis," *Antimicrob. Agents Chemother.* **2000,** *44*, 1874-7.
36. Clontech. "BD GenomeWalker Universal Kit user manual," **2004**.
37. Croteau, R.; Ketchum, R. E. B.; Long, R. M.; Kaspera, R.; Wildung, M. R. "Taxol biosynthesis and molecular genetics," *Phytochem. Rev.* **2006,** *5*, 75-97.
38. Cuevas, J. C.; Lavandera, J. L.; Martos, J. L. "Design and synthesis of simplified sordaricin derivatives as inhibitors of fungal protein synthesis," *Bioorg. Med. Chem. Lett.* **1999,** *9*, 103-8.
39. Daferner, M.; Mensch, M.; Anke, T.; Sterner, O. "Hydroxysodarin, a new sodarin derivative from *Hypoxylon croceum*," *Z. Naturforsch* **1999,** *54c*, 474-80.
40. Davis, E. M.; Croteau, R. "Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes," *Top. Curr. Chem.* **2000,** *209*, 53-95.
41. Davisson, V. J.; Neal, T. R.; Poulter, C. D. "Farnesyl-diphosphate synthase. Catalysis of an intramolecular prenyl transfer with bisubstrate analogs," *J. Am. Chem. Soc.* **1993,** *115*, 1235-45.
42. Davisson, V. J.; Poulter, C. D. "Farnesyl-diphosphate synthase. Interplay between substrate topology, stereochemistry, and regiochemistry in electrophilic alkylations," *J. Am. Chem. Soc.* **1993,** *115*, 1245-60.
43. Davoli, P.; Engel, G.; Werle, A.; Sterner, O.; Anke, T. "Neosodarin and hydroxysodarin, two new antifungal agents from *Sordaria araneosa*," *J. Antibiot.* **2002,** *55*, 377-82.
44. De Beule, K.; Van Gestel, J. "Pharmacology of itraconazole," *Drugs* **2001,** *61*, 27-37.
45. Deacon, J. W. *Fungal Biology (4th ed)* **2005**, Malden, MA: Blackwell Publishers.
46. Dewick, P. M. "Medicinal natural products: a biosynthetic approach," **2001**, John Wiley & Sons Ltd.
47. Diamond, R. D. "The growing problem of mycoses in patients infected with the human immunodeficiency virus," *Rev. Infect. Dis.* **1991,** *13*, 480-6.
48. Diatchenko, L.; Lau, Y. F.; Campbell, A. P.; Chenchik, A.; Moqadam, F.; Huang, B.; Lukyanov, S.; Lukyanov, K.; Gurskaya, N.; Sverdlov, E. D.;

Siebert, P. D. "Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries," *Proc. Natl. Acad. Sci. USA.* **1996,** *93*, 6025-30.

49. Dogbo, A.; Camara, B. "Purification of isopentenyl pyrophosphate isomerase and geranyl geranyl pyrophosphate synthase from Capsicum chromoplasts by affinity chromatography," *Biochim. Biophys. Acta* **1967,** *920*, 140-8.
50. Dominguez, J. M.; Kelly, V. A.; Kinsman, O. S.; Marriot, M. S.; Gomez de las Heras, F.; Martin, J. J. "Sordarins: a new class of antifungal with selective inhibition of the protein synthesis elongation cycle in yeasts," *Antimicrob. Agents Chemother.* **1998,** *42*, 2274-8.
51. Dominguez, J. M.; Martin, J. J. "Identification of elongation factors as the essential protein targeted by sordarins in *Candida albicans*," *Antimicrob. Agents Chemother.* **1998,** *42*, 2279-83.
52. Dominguez, J. M.; Kelly, V. A.; Kinsman, O. S. "Sordarins: a new class of antifungals with selective inhibition of the protein synthesis elongation cycle in yeasts," *Antimicrob. Agents Chemother.* **1998,** *42*, 2274-8.
53. Egger, H.; Reinshagen, H. "New pleuromutilin derivatives with enhanced antimicrobial activity. II. Structure-activity correlations," *J. Antibiot.* **1976,** *29*, 923-7.
54. Eisenreich, W.; Bacher, A.; Arigoni, D.; Rohdich, F. "Biosynthesis of isoprenoids via the non-mevalonate pathway," *Cell Mol. Life Sci.* **2004,** *61*, 1401-26.
55. Eisenreich, W.; Rohdich, F.; Bacher, A. "Deoxyxyluose phosphate pathway to terpenoids," *Trends Plant Sci.* **2001,** *6*, 78-84.
56. Eisenreich, W.; Schwarz, M.; Cartayade, A.; Arigoni, D.; Zenk, M. H.; Bacher, A. "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms," *Chem. Biol.* **1998,** *5*, R221-R33.
57. EPICENTRE. "CopyControl Fosmid Library Production Kit protocol,".
58. Fehr, T.; Acklin, W. "Isolation of 2 new indole derivatives from the mycelia of *Claviceps paspali*," *Helv. Chim. Acta* **1966,** *49*, 1907-10.
59. Flesch, G.; Rohmer, M. "Prokaryotic hopanoids: the biosynthesis of the bacteriohopane skeleton. Formation of isoprenic units from two distinct acetate pools and a novel type of carbon/carbon linkage between a triterpene and D-ribose," *Eur. J. Biochem.* **1988,** *175*, 405-11.
60. Frank, J.; Agrawal, R. K. "A ratchet-like inter-subunit reorganization of the ribosome during translocation," *Nature* **2000,** *406*, 318-22.
61. French, G. L. "Clinical impact and relevance of antibiotic resistance," *Adv. Drug Deliver. Rev.* **2005** *57*, 1514-27.
62. Frohman, M. A.; Dush, M. K.; Martin, G. R. "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer," *Proc. Natl. Acad. Sci. USA* **1998,** *85*, 8998-9002.

63. Gargallo-Viola, D. "Sordarins as antifungal compounds," *Curr. Opin. Anti-infect. Invest. Drugs* **1999,** *1*, 297-305.
64. Georgopapadakou, N. H.; Walsh, T. J. "Human mycoses: drugs and targets for emerging pathogens," *Science* **1994,** *264*, 371-3.
65. Gomez-Lorenzo, M. G.; Garcia-Bustos, J. F. "Ribosomal P-protein stalk function is targeted by sordarin antifungals," *J. Biol. Chem.* **1998,** *273*, 25041-4.
66. Graebe, J. E.; Hedden, P.; Gaskin, P.; MacMillan, J. "The biosynthesis of a C19-gibberellin from mevalonic acid in a cell-free system from a higher plant," *Planta* **1974,** *120*, 307-9.
67. Graybill, J. R.; Najvar, L.; Fothergill, A.; Bocanegra, R.; Gomez de las Heras, F. "Activities of sordarins in murine histoplasmosis," *Antimicrob. Agents Chemother.* **1999,** *43*, 1716-8.
68. Grove, J. F. "Microcyclic trichothecenes," *Nat. Prod. Rep.* **1993,** *10*, 429-48.
69. Grove, J. F. "Non-macrocyclic trichothecenes, part 2," *Prog. Chem. Org. Nat. Prod.* **1996,** *69*, 1-70.
70. Gudkov, A. T. "The L7/L12 ribosomal domain of the ribosome: structural and functional studies," *FEBS Lett.* **1997,** *407*, 253-6.
71. Hall, C. C.; Bertasso, A. B.; Watkins, J. D.; Georgopapadakou, N. H. "Screening assays for protein synthesis inhibitors," *J. Antibiotics* **1992,** *45*, 1697-9.
72. Hamano, D. T.; Kuzuyama, Y.; Itoh, T.; Funhata, N. K.; Seto, H. "Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic, terpentecin," *J. Bacteriol.* **2001,** *183*, 6085-4.
73. Hauser, D.; Sigg, H. P. "Isolierung und abbau von sordarin," *Helv. Chim. Acta* **1971,** *54*, 1178-90.
74. Hedden, P.; Kamiya, Y. "Gibberellin biosynthesis: enzymes, genes and their regulation," *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1997,** *48*, 431-60.
75. Heilmann, C.; Jensen, L.; Jensen, J. S.; Lundstrom, K.; Windsor, D.; Windsor, H.; Webster, D. "Treatment of resistant mycoplasma infection in immunocompromised patients with a new pleuromutiin antibiotic," *J. Infect. Dis.* **2001,** *43*, 234-8.
76. Herreros, E.; Martinez, C. M.; Almela, M. J.; Marriott, M. S.; Gomez de las Heras, F.; Gargallo-Viola, D. "Sordarins: in vitro activities of new antifungal derivatives against pathogenic yeasts, *Pneumocystis carinii*, and filamentous fungi," *Antimicrob. Agents Chemother.* **1998,** *42*, 2863-9.
77. Houghton, P. "Herbal products: Ginkgo," *Pharm. J.* **1994,** *253*, 122-3.
78. Inouye, S.; Abe, S.; Yamaguchi, H. "Fungal terpenoid antibiotics and enzyme inhibitors," *Mycology Series* **2004,** *Handbook of Fungal Biotechnology (2nd Edition). 20*, 379-99.
79. Jenny, L.; Borschberg, H. "Synthesis of the dolabellane diterpene hydrocarbon (±)-*δ*-araneosene," *Helv. Chim. Acta* **1995,** *78*, 715-31.

80. Jiang, Y.; Proteau, P.; Poulter, D.; Ferro-Novick, S. "BTSl encodes a geranyl geranyl diphosphate synthase in *Saccharomyces cerevisiae*," *J. Biol. Chem.* **1995,** *270*, 21793-9.
81. Jorgensen, R.; Ortiz, P. A.; Carr-Schmid, A.; Nissen, P.; Kinzy, T. G.; Andersen, G. R. "Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase," *Nat. Struct. Biol.* **2003,** *10*, 379-85.
82. Justice, M. C.; Hsu, M.; Tse, B.; Ku, T. B., J.; Schmatz, D.; Nielsen, J. "Elongation factor 2 as a novel target for selective inhibition of fungal protein synthesis," *J. Biol. Chem.* **1998,** *273*, 3148-51.
83. Justice, M. C.; Ku, T.; Hsu, M. J.; Carniol, K.; Schmatz, D.; Nielsen, J. "Mutations in ribosomal protein L10e confer resistance to the fungal-specific eukaryotic elongation factor 2 inhibitor sordarin," *J. Biol. Chem.* **1999,** *274*, 4869-75.
84. Kamath, A.; Chakraburtty, K. "Role of yeast Elongation Factor 3 in the Elongation Cycle," *J. Biol. Chem.* **1989,** *264*, 15423-8.
85. Kavanagh, F.; Harvey, H.; Robbins, W. J. "Antibiotic substances from Basidiomycetes. VIII. *Pleurotus mutilis* Sacc and *Pleurotus passeckerianus* Pilat," *Proc. Natl. Acad. Sci. USA.* **1951,** *37*, 570-4.
86. Kawaide, H. "Biochemical and molecular analyses of gibberellin biosynthesis in fungi," *Biosci. Biotechnol. Biochem.* **2006,** *70*, 583-90.
87. Kawaide, H.; Imai, R.; Sassa, T.; Kamiya, Y. "*ent*-kaurene synthase from the fungus *Phaeosphaeria* sp, L487 - cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase in fungal gibberelln biosynthesis," *J. Biol. Chem.* **1997,** *272*, 21706-12.
88. Keller, N. P.; Hohn, T. M. "Metabolic pathway gene clusters in filamentous fungi," *Fung. Genet. Biol.* **1997,** *21*, 17-29.
89. Kellogg, B. A.; Poulter, C. D. "Chain elongation in the isoprenoid biosynthetic pathway," *Curr. Opin. Chem. Biol.* **1997,** *1*, 570-8.
90. Kennedy, T. C.; Webb, G.; Cannell, R. J. P.; Kinsman, O. S.; Middleton, R. F.; Sidebottom, P. J.; Taylor, N. L.; Dawson, M. J.; Buss, A. D. "Novel inhibitors of fungal protein synthesis produced by a strain of Graphium putredinis. Isolation characterization and biological properties," *J. Antibiot.* **1998,** *51*, 1012-8.
91. Kinsman, O. S.; Chalk, P. A.; Jackson, H. C.; Middleton, A.; Shuttleworth, B. A.; Rudd, A. M.; Jones, H. M.; Noble, H. G.; Wildman, M. J.; Dawson, C.; Stylli, P. J.; Sidebottom, B.; Lamont, S. L.; Hayes, M. V. "Isolation and characterization of an antifungal antibiotic (GR135402) with protein synthesis inhibition," *J. Antibiot.* **1998,** *51*, 41-9.
92. Kleinig, H. "The role of plastids in isoprenoid biosynthesis. Annu. Rev. Plant Physiol.," *Plant Mol. Biol.* **1989,** *40*, 39-59.
93. Knauseder, F.; Brandl, E. "Pleuromutilins. Fermentation, structure and biosynthesis," *J. Antibiot.* **1975,** *29*, 125-31.

94. Kosowska-Shick, K.; Clark, C.; Credito, K.; McGhee, p.; Dewasse, B.; Bogdanovich, T.; Appelbaum, P. C. "Single- and multistep resistance selection studies on the activity of retapamulin compared to other agents against *Staphylococcus aureus* and *Streptococcus pyogenes*," *Antimicrob. Agents Chemother.* **2006,** *50*, 765-9.
95. Kupfer, D. M.; Drabenstot, S. D.; Buchanan, K. L.; Lai, H.; Zhu, H.; Dyer, D. W.; Roe, B. A.; Murphy, J. W. "Introns and splicing elements of five diverse fungi," *Eukaryot. Cell* **2004,** *3*, 1088-100.
96. Kuzuguchi, T.; Morita, Y.; Sagam, I.; Sagam, H.; Ogua, K. "Human geraylgerayl diphosphate synthase. cDNA cloning and expression," *J. Biol. Chem.* **1999,** *274*, 5888-94.
97. Kuzuyama, T.; Seto, H. "Diversity of the biosynthesis of the isoprene units," *Nat. Prod. Rep.* **2003,** *20*, 171-83.
98. Kuzuyama, T.; Seto, H. "Diversity of the biosynthesis of the isoprene units," *Nat. Prod. Rep.* **2003,** *20*, 171-83.
99. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R. "Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes," *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 13172-7.
100. Lesburg, C. A.; Caruthers, J. M.; Paschall, C. M.; Christianson, D. W. "Managing and manipulating carbocations in biology: terpenoid cyclase structure and mechanism," *Curr. Opin. Struct. Biol.* **1998,** *8*, 695-703.
101. Lichtenthaler, H. K. "The plants' 1-deoxyD-xylulose-5-phosphate pathway for biosynthesis of isoprenoids," *Fett-Lipid* **1998,** *100*, 128-38.
102. Lobova, D.; Smola, J.; Cizek, A. "Decreased susceptibilty to tiamulin and valnemulin among Czech isolates of *Brachyspira hyodysenteriae*," *J. Med. Microbiol.* **2004,** *53*, 287-91.
103. Long, K. S.; Hansen, L. H.; Jakobsen, L.; Vester, B. "Interaction of pleuromutilin derivatives with the ribosomal peptidyl transferase center," *Antimicrob. Agents Chemother.* **2006,** *50*, 1458-62.
104. Luthra, R.; Luthra, P. M.; Kumar, S. "Redefined role of mevalonate-isoprenoid pathway in terpenoid biosynthesis in higher plants," *Curr. Sci.* **1999,** *76*, 133-5.
105. Mahmoud, S. S.; Croteau, R. "Strategies for transgenic manipulation of monoterpene biosynthesis in plants," *Trends Plant Sci.* **2002,** *7*.
106. Mander, L. N.; Thomson, R. J. "Total synthesis of sordaricin," *Org. Lett.* **2003,** *5*, 1321-4.
107. Mann, J. "Secondary metabolism," **1987**, Clarendon Press. New York: Oxford University Press.
108. Mann, J. "Chemical aspects of biosynthesis," **1994**, Oxford University Press.
109. Margulis, L.; Schwartz, K. V. "Five kingdoms: an illustrated guide to the phyla of life on earth," **1998**, W.H. Freeman, New York.
110. Martinez, A.; Ferrer, S.; Santos, I.; Jimenez, E.; Sparrowe, J.; Regadera, J.; Gomez de las Heras, F.; Gargallo-Viola, D. "Antifungal activities of two

new azasordarins GW471552 and GW471558, in experimental models of oral and vulvovaginal candidiasis in immunosuppressed rats," *Antimicrob. Agents Chemother.* **2001,** *45*, 3304-9.

111. Martinez, A.; Regadera, J.; Jimenez, E.; Santos, I.; Gargallo-Viola, D. "Antifungal efficacy of GM237354, a sordarin derivative, in experimental oral candidiasis in immunosuppressed rats," *Antimicrob. Agents Chemother.* **2001,** *45*, 1008-13.
112. Mathews, C. K.; Holde, K. E. V.; Ahern, K. G. *Biochemistry* **1999,** *3rd edition. Addison Wesley Longman Inc.*
113. McGarvey, D. J.; Croteau, R. "Terpenoid metabolism," *Plant Cell* **1995,** *7*, 1015-26.
114. Mende, K.; Homann, V.; Tudzynski, B. "The geranylgeranyl diphosphate synthase gene of *Gibberella fujikuroi*: Isolation and expression," *Mol. Gen. Genet.* **1997,** *255*, 96-105.
115. Misawa, N.; Satomi, Y.; Kondo, K.; Yokoyama, A.; Kajiwar, S.; Saito, T.; Ohta, T.; Mild, W. "Strcture and functional analysis of a marne bactenal carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level " *J. Bacteriol.* **1995,** *177*, 6575-84.
116. Moller, E. M.; Bahnweg, G.; Sandermann, H.; Geiger, H. H. "A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies and infected plant tissues," *Nucleic Acids Res.* **1992,** *20*, 6115-6.
117. Monaghan, R. L.; Barrett, J. F. "Antibacterial drug discovery--then, now and the genomics future," *Biochem. Pharmacol.* **2006,** *71*, 901-9.
118. Mukhopadhyay, A.; Peterson, R. T. "Fishing for new antimicrobials," *Curr. Opin. Chem. Biol.* **2006,** *10*, 327-33.
119. Muller, F. C.; Weig, M.; Peter, J.; Walsh, T. J. "Azole cross - resistance to ketoconazole, fluconazole, itraconazole, and voriconazole in clinical *Candida albicans* isolates from HIV-infected children with *oropharyngeal candidosis*," *J. Antimicrob. Chemother.* **2000,** *46*, 338-41.
120. Naegell, p. "Zur kenntnis des pleuromutilins," *Diss. Nr. ETH Zurich* **1961**, 3206.
121. Newman, D. J.; Cragg, G. M.; Snader, K. M. "Natural products as sources of new drugs over the period 1981-2002," *J. Nat. Prod.* **2003,** *66*, 1022-37.
122. Nyfeler, R.; Keller, S. W. "Metabolites of microorganisms.143. Echinocandin B, a novel polypeptideantibiotic from *Aspergillus nidulans* var. echinulatus: isolation and structural components," *Helv. Chim. Acta.* **1974,** *57*, 2459-77.
123. Oikawa, H.; Toyomasu, T.; Toshima, H.; Ohashi, S.; Kawaide, H.; Kamiya, y.; Ohtsuka, M.; Shinoda, S.; Mitsuhashi, W.; Sassa, T. "Cloning and functional expression of cDNA encoding aphidicolan-16*β*-ol synthase: a key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin," *J. Am. Chem. Soc.* **2001,** *123*, 5154-5.

124. Pandit, J.; Danley, D. E.; Schulte, G. K.; Mazzalupo, S.; Pauly, T. A.; Hayward, C. M.; Hamanaka, E. S.; Thompson, J. F.; Harwood, H. J., Jr. "Crystal structure of human squalene synthase. a key enzyme in cholesterol biosynthesis " *J. Biol. Chem.* **2000,** *275*, 30610-7.
125. Paquette, L. A.; Wiedeman, P. E.; Bulman-Page, P. C. "(+)-Pleuromutilin synthetic studies. Degradative and de novo acquisition of a levorotatory tricyclic lactone subunit," *J. Org. Chem.* **1988,** *53*, 1441-50.
126. Parker, E. J.; Scott, D. B. "Indole-diterpene biosynthesis in Ascomycetous fungi," *Handbook of industrial mycology* **2004**, 405-26.
127. Perez, A. "Terbinafine: broad new spectrum of indications in several subcutaneous and systemic and parasitic diseases," *Mycoses* **1999,** *42*, 111-4.
128. Petranyi, G.; Ryder, N. S.; Stutz, A. "Allylamine derivatives: new class of synthetic antifungal agents inhibiting fungal squalene epoxidase," *Science* **1984,** *224*, 1239-41.
129. Poulsen, S. M.; Karlsson, M.; Johansson, L. B.; Vester, B. "The pleuromutilin drugs tiamulin and valnemulin bind to the RNA at the peptidyl transferase centre on the ribosome," *Mol. Microb.* **2001,** *41*, 1091-9.
130. Poulter, C. D.; Argyle, J. C.; Mash, E. A. "Farnesyl pyrophosphate synthetase. Mechanistic studies of the 1'-4 coupling reaction with 2-fluorogeranyl pyrophosphate," *J. Biol. Chem.* **1978,** *253*, 7227-33.
131. Poulter, C. D.; Satterwhite, D. M. "Mechanism of the prenyl transfer reaction. Studies with (E)- and (Z)-3-trifluoromethyl-2-buten-1-yl pyrophosphate," *Biochemistry* **1977,** *16*, 5470-8.
132. Poulter, C. D.; Wiggins, P. L.; Le, A. T. "Farnesylpyrophosphate synthetase. A stepwise mechanism for the 1'-4 condensation reaction," *J. Am. Chem. Soc.* **1981,** *103*, 3926-7.
133. Pringle, M.; Poehlsgaard, J.; Vester, B.; Long, K. S. "Mutations in ribosomal protein L3 and 23S ribosomal RNA at the peptidyl transferase centre are associated with reduced susceptibility to tiamulin in *Brachyspira* spp. isolates," *Mol. Microbiol.* **2004,** *54*, 1295-306.
134. Reardon, D.; Farber, G. K. "The structure and evolution of alpha/beta barrel proteins," *FASEB J.* **1995,** *9*, 497-503.
135. Rittenhouse, S.; Biswas, S.; Broskey, J.; McCloskey, L.; Moore, T.; Vasey, S.; West, J.; Zalacain, M.; Zonis, R.; D., P. "Selection of retapamulin, a novel pleuromutilin for topical use," *Antimicrob. Agents Chemother.* **2006,** *50*, 3882-5.
136. Rittenhouse, S.; Singley, C.; Hoover, J.; Page, R.; Payne, D. "Use of the surgical wound infection model to determine the efficacious dosing regimen of retapamulin, a novel topical antibiotic," *Antimicrob. Agents Chemother.* **2006,** *50*, 3886-8.
137. Rohmer, M. "Isoprenoid biosynthesis via the mevalonate-dependent route," *Prog. Drug Res.* **1998,** *50*, 136-54.

138. Rohmer, M. "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants," *Nat. Prod. Rep.* **1999,** *16*, 565-74.
139. Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. "Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate," *Biochem. J.* **1993,** *295*, 517-24.
140. Rohmer, M.; Sutter, B.; Sahm, H. "Biosynthesis of the side-chain of bacteriohopanetetrol and of a carbocyclic pseudopentose from 13C-labelled glucose in *Zymomonas mobilis*," *J. Chem. Soc., Chem. Commun.* **1989,** *19*, 1471-2.
141. Ruzicka, L. "The isoprene rule and the biogenesis of terpenic compounds," *Experientia* **1953,** *9*, 357-67.
142. Sacchettini, J. C.; Poulter, C. D. "Creating isoprenoid diversity," *Science* **1997,** *277*, 1788-9.
143. Sagami, H.; Morita, Y.; Ogura, K. "Purification and properties of geranylgeranyl-diphosphate synthase from bovine brain," *J. Biol. Chem.* **1994,** *269*, 20561-6.
144. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning: a laboratory manual* **1989**, Second edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.
145. Santos, C.; Ballesta, J. P. G. "The highly conserved protein P0 carboxyl end is essential for ribosome activity only in the absence of proteins P1 and P2," *J. Biol. Chem.* **1995,** *270*, 20608-14.
146. Schlunzen, F.; Pyetan, E.; Fucini, P.; Yonath, A.; Harms, J. M. "Inhibition of peptide bond formation by pleuromutilins: the structure of the 50S ribosomal subunit from *Deinococcus radiodurans* in complex with tiamulin," *Mol. Microbiol.* **2004,** *54*, 1287-94.
147. Siebert, P. D.; Chenchik, A.; Kellogg, D. E.; Lukyanov, K. A.; Lukyanov, S. A. "An improved method for walking in uncloned genomic DNA," *Nucleic Acids Res.* **1995,** *23*, 1087-8.
148. Singh, N.; Luthra, R.; Sangwan, R. S.; Thakur, R. S. "Metabolism of monoterpenoids in aromatic plants," *Curr. Res. Med. Arom. Plants* **1989,** *11*, 174-97.
149. Skogerson, L.; Engelhardt, D. "Dissimilarity in chain elongation factor requirements between yeasts and rat liver ribosomes," *J. Biol. Chem.* **1977,** *252*, 1471-5.
150. Soulsby, E. J. "Resistance to antimicrobials in humans and animals," *Brit. J. Med.* **2005,** *331*, 1219-20.
151. Spahn, C. M.; Gomez-Lorenzo, M. G.; Grassucci, R. A.; Jorgensen, R.; Andersen, G. R.; Beckmann, R.; Penczek, P. A.; Ballesta, J. P.; Frank, J. "Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation FREE," *Embo J.* **2004,** *23*, 1008-19.
152. Spahn, C. M. T.; Beckmann, R.; Eswar, N.; Penczek, P. A.; Sali, A.; Blobel, G.; Frank, J. "Structure of the 80S ribosome from *Saccharomyces*

cerevisiae-tRNA-ribosome and subunit-subunit interactions," *Cell* **2001,** *107*, 373-86.

153. Stark, H.; Rodnina, M. V.; Rinkeappel, J.; Brimacombe, R.; Wintermeyer, W.; Vanheel, M. "Visualization of elongation factor Tu on the *Escherichia coli* ribosome," *Nature* **1997,** *389*, 403-6.
154. Sun, T. P.; Kamiya, Y. "The Arabidopsis GA1 locus encodes the cyclase *ent*-kaurene synthetase A of gibberellin biosynthesis," *Plant Cell* **1994,** *6*, 1509-18.
155. Tarshis, L. C.; Proteau, P. J.; Kellogg, B. A.; Sacchettini, J. C.; Poulter, C. D. "Regulation of product chain length by isoprenyl diphosphate synthases," *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 15018-23.
156. Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. "Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase," **2004,** *432*, 118-22.
157. Torres-Rodriguez, J. M.; Morera, Y.; Baro, T.; Lopez, O.; Alia, C.; T., J. "In vitro susceptibility of Cryptococcus neoformans serotypes to GM 237354 derivative of the sordarin class," *Mycoses* **2002,** *45*, 313-6.
158. Toyomasu, T.; Kawaide, H.; Ishizaki, A.; Shinoda, S.; Sassa, T. "Cloning of a full-length cDNA encoding *ent*-kaurene synthase from *Gibberella fujikuroi*: functional analysis of a bifunctional diterpene cyclase," *Biosci. Biotechnol. Biochem.* **2000,** *64*, 660-4.
159. Toyomasu, T.; Nakaminami, K. "Cloning of a gene cluster responsible for the biosynthesis of diterpene aphidicolin, a specifc inbitor of DNA polymerase," *Biosci. Biotechnol. Biochem.* **2004,** *68*, 146-52.
160. Tse, B.; Balkovec, J. M.; Blazey, C. M.; Hsu, M.-J.; Nielsen, J.; Schmatz, D. "Alkyl side-chain derivatives of sordaricin as potent antifungal agents against yeast," *Bioorg. Med. Chem. Lett.* **1998,** *8*, 2269-72.
161. Tudzynski, B.; Holter, K. "Gibberellin biosynthetic pathway in *Gibberella fujikuroi*: evidence for a gene cluster," *Fung. Genet. Biol.* **1998,** *25*, 157–70.
162. Tudzynski, B.; Homann, V.; Feng, B.; Marzluf, G. A. "Isolation, characterization and disruption of the areA nitrogen regulatory gene of *Gibberella fujikuroi*," *Mol. Gen. Genet.* **1999,** *261*, 106-14.
163. Tudzynski, B.; Kawaide, H.; Kamiya, Y. "The gibberellin biosynthesis in *Gibberella fujikuroi*: cloning and characterization of the copalyl diphosphate synthase gene," *Curr. Genet.* **1998,** *34*, 234-40.
164. Verbist, L. "The antimicrobial activity of fusidic acid," *J. Antimicrob. Chemother.* **1990,** *25 (Suppl B)*, 1-5.
165. Vicente, M. F.; Basilio, A.; Cabello, A.; Pelaez, F. "Microbial natural products as a source of antifungals," *Clin. Microbiol. Infec.* **2003,** *9*, 15-32.
166. Vogel, B. S.; Wildung, M. R.; Vogel, G.; R., C. "Abietadiene Synthase from Grand Fir (*Abies grandis*). cDNA isolation, characterization, and bacterial expression of a bifunctional diterpene cyclase involved in resin acid biosynthesis," *J. Biol. Chem.* **1996,** *271*, 23262-8.

167. Wang, D. Y.; Wu, Y.; Wu, Y. L.; Li, Y.; Shan, F. "Synthesis, iron(II)-induced cleavage and in vivo antimalarial efficacy of 10-(2-hydroxy-1-naphthyl)-deoxoqinghaosu," *J. Chem. Soc. Perkin Trans.* **1999,** *1*, 1827-31.
168. Wendt, K. U.; Schulz, G. E. "Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes," *Structure* **1998,** *6*, 127-33.
169. Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. "Enzyme mechanisms for polycyclic triterpene formation," *Angew. Chem., Int. Ed.* **2000,** *39*, 2812-33.
170. Williams, D. C.; McGarvey, D. J.; Katahira, E. J.; Croteau, R. "Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair," *Biochemistry* **1998,** *37*, 12213-20.
171. Wilson, K. S.; Noller, H. F. "Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing," *Cell* **1998,** *92*, 131-9.
172. Witting, L. A.; Porter, J. W. "Intermediates in the conversion of melaonic acid to squalene by a rat liver enzyme system," *J. Biol. Chem.* **1959,** *234*, 2841-6.
173. Yamaguchi, S.; Saito, T.; Abe, H.; Yamane, H.; Murofushi, N.; Kamiya, Y. "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*, 203-13.
174. Young, C. A.; Bryant, M. K.; Christensen, M. J.; Tapper, B. A.; Bryan, G. T.; Scott, B. "Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass," *Mol. Gen. Genomics* **2005,** *274*, 13–29.
175. Young, C. A.; McMillan, L. "Molecular cloning and genetic analysis of an indolediterpene gene cluster from *Penicillium paxilli*," *Mol. Microbiol.* **2001,** *39*, 754-64.
176. Zhang, S.; Monahan, B. J.; Tkacz, J. S.; B., S. "Indole-diterpene gene cluster from *Aspergillus flavus*," *Appl. Environ. Microbiol.* **2004,** *70*, 6875-83.

APPENDIX

Sequences of p319, p322, p345, s224 and s272

Three different sequences, p319, p322 and p345, representing fragments of *ggs* genes from *P. passeckerianus* and two sequences, s224 and s272, from *S. araneosa* were obtained from the nested PCR using the degenerate primers designed from fungal *ggs* genes and their sequences are shown below.

p319: 319 bp

CTTGATTTGTTCTGGAGAGATACCTTTACCTGCCCGACGGAGGAAGAGTATGTTTCTATGGTCAAT
AATAGTGAGTTGGTTCTTGGCCCACCCGAGGCTTAGTCTTAGAACACTAGACTGATTGATACCGCT
AGAAACTAGCGGCCTTCTTCGCCTTGCTATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGA
CAGGTATGTGCCTGTGTTCAGCTCCGCTTGCGCCTAAATCAACTGTCCTAATTCTTAGTGATTATG
TCCCTCTTGTGAATCTCATTGGAGTATATTTCCAGATTCGAGACGACCTGATGAA

P322: 322 bp

CACCGTGGTCAAGGCCTCGAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAAGAATAC
ATTTCGATGGTCAACAATAGTGAGTTAACCTGGGGCCAGTCTAAGCCTACGGTTAGACTCATTTCG
TATCAGAAACCAGTGGCCTCCTACGTATCGGTATTAAACTCATGATGGCATGTGCCACCACAAACG
CGGATATGTATGTGTTTTTGGCAGTGCATTTTGGGGTTTGAACTGAATGGCCATTTTCTAGCGATT
ATATTCCTTTGACAAACCTCATTGGAGTCTATTTCCAAATCCGGGATGACCTTATGAA

p345: 345 bp

CATCGCGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAGGAGTAT
ATCTCCATGGTTAACAATAGTTAGTGGCTTTTTCATTACTTGAGATGAGGACTCGATACTCATGAT
CATTTCGTTGAAATTTTCTAGAGACGGGCGGTCTTCTACGCATCGGCATCAAACTCATGATGGCAT
GCGCGACCACCAACACTGGAGTGTACGTATTCCATCCTTCTCATTGTCCACTCTGGTCTTTCTGAC
ATTTTCTATTTCTTCCAGTGACTTTATACCCTTAGTCAACCTCATCGGCATCTACTTCCAAATCAG
AGACGATCTGATGAA

s224: 224 bp

CATCATGGTCAGGGCCTCGATATACGCTGGCGGGATCATCTCAAGACAACCAAGCTCCCGCTGATA
GAAGAATACATGGAAATGATCATGAACAAAACCGGAGGATTATTTCGGCTTGCCGTTAAACTCTTG
GGAGCTTTCAGCACCCCACAAGACAAGCCAACTGTGTTGGCCATAGCTAATCTCTTCGGGATTGTA
TTCCAGATCCGTGATGACCTTATGAA

s272: 272 bp

CACCGGGGCCAAGGGATGGACCTCTTCTGGAGAGACACTCTGACCTGCCCGACCGAAGACGAGTAC
CTCGAGATGGTATCCAACAAGACTGGGGGGCTGTTCAGGCTTGGTATTAAGCTGATGCAGGCTGAG
TCCCGGAGTCTTGTGGACTGCGTACCGCTTGTCAGCATCATGGGCTTAATCTTCCAGATTGCGGAT

GACTACCAGAACCTCTGGAGCCGGCGGTATACTGCCAACAAAGGCATGTGTGAGGACCTCACCGAG
GGGAAGTT

Sequences of Pleur319, Pleur322, Pleur345, Sord224 and Sord272

The proposed genomic DNA sequence of the full *ggs* gene Pleur319 of 1646 bp which was derived from p319, that of the Pleur322 of 1583 bp derived from p322, that of the Pleur345 of 1367 bp derived from p345, that of the Sord224 of 990 bp derived from s224, and that of the Sord272 of 1292 bp derived from s272, are shown below.

Pleur319: 1646 bp

ATGTCAAAACTGACCGGGGCTGAGTTTGACGTCCTGCATAGTTCAGAATATGTTTCCTCCGAGATC
CTAACCCATGAATATTCTAACCCGCGAGAGGTTTAATGAAGTCAATGTCGGGAGCCCTGGAAGACC
CATAAAACCCACTTAGAGTCACCACCCAGATCCAAGTGTCTCCCACTTGCAAACTTTTAATGTCGC
TCTACAACAACATTCTCCGTAGTAAACTTGGTCACGAACAACCATGGCAGCCTCAGAGCGAAAAGG
CGAGTGAACTTCGGAGTGTTCCTGACTTCTGAATTTGACGCTCCTCGAAGGCCATTACGGAGCCAT
ATGCCTATACTGTAACTGCACCAGGAAAAGGCATACGTCAAATCCTCATTGAGGCCTTCAATGTTT
GGCTAAACGTGCCAGAGGATAAGTTGCTTCTCATAAAGAACGTCGTGGATATGCTCCATAATGCCA
GCTTGCTGTGAGTGGTCTCCTGGTATCTCGACGGTAGTTGACCGATTGGTGGTTAATCATAGTATC
GACGATATCGAAGATGATTCTCATCTAAGGAGGGGCCAACCGGGTAAGGCTCTAAGAAATATGTCA
TTAATAGTGATTGACCAATATCACTAGCTGCGCACAAAGTTTATGGCGTTCCGCGAACCATGAACG
CTGCGCATTACGTCTGGCTGAATGTGTATAAGGAGCTTTCGAAGTTCGACGAGCCAAACTCTTCGA
ATCGTGATGCTCCCCACTTGACAGATATTGTTACCGGTATGCACAGTACTATCCAACTACATTTGA
TTGCTGATGAGAATACTAGACGAAATGATATATTTGCATCGCGGCCAAGGCCTTGATTTGTTCTGG
AGAGATACCTTTACCTGCCCGACGGAGGAAGAGTATGTTTCTATGGTCAATAATAGTGAGTTGGTT
CTTGGCCCACCCGAGGCTTAGTCTTAGAACACTAGACTGATTGATACCGCTAGAAACTAGCGGCCT
TCTTCGCCTTGCTATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGACAGGTATGTGCCTGT
GTTCAGCTCCGCTTGCGCCTAAATCAACTGTCCTAATTCTTAGTGATTATGTCCCTCTTGTGAATC
TCATTGGAGTATATTTCCAGATTCGAGACGACCTGATGAATCTTGACAGCACGGAGGTGAGCGAAC
GTCTACTAGGTTAGCGGCCTACTTAGTTTGTTATAATTTCAGTATGCCCAAAAGAAAGGCTTCGCA
GAGGACCTCACCGAAGGGAAGTTTTCGTTCCCTATCATACACGGCATCCAAGCAGACACTTCAAGT
CATTTTCTACTTGGTAAGTGCCAGTTTTTCGCAGGCTTTCCTCCTACGTAACCCTCGAAACTAACG
CTCCGTCGATAGATGTACTCCGAAAGAGGTCAGGGGAAGACGCGTTCAAGATTAGTGCCATTGAAT
ACCTGCGTCATCGAACAAAGTCCTTCGAGTATACCCAAGAGGTTCTGGATTCCTTGGAGATGCAGG
CTTTCGACGAGATCAGGCAATTCGGGGGAAACTGGCAGCTTGATCAGATGTTGCGCCTTCTTCGGG
TCCCTACTCCAAAACCAGAGTGTCCTCTTATCGAAGATTTTGATGACTTGGACCTGGACTAG

Pleur322: 1583 bp

ATGGAGTTTAGTCTGACATCTTGCATGCTCAAGATGTGACGTCTGATCTTCTTTTGACAATTATGA
ACATTTCGGCGGTCATTCCCTAAGCCATTGTCAACGCAAAGACACATACTAGTATAAGAAGCTCCT
CTCTTTATCCAGACGGAATCAACCCCTATCCGTGAAACCACTCATGTCAATCTACGACGACATCCT
CAGTAAACTTGGCGGCGATCAGCCCTGGTCCTTTCAGAACGAAAAGGTAAGGCCCTCCAGAGTTGG
TTCCCAGTCCTAAATCCAACGTCATTACAAGGCTATTTTGGAGCCATATACATACACAGTCTCTGC
CCCAGGGAAAGACATACGTCGGAAGCTCATCGACTCCTTTAATCTCTGGCTCAACGTTCCAGATGA
CAAGTTGGATATCATAGTCAACATCGTGAACATGTTACACAACGCCAGCTTGCTGTAAGCTTCGCC
GGGTGGTTCAGTGATCACTTACTGACTACTTCCGCTCACAGTCTTGACGACATCGAAGACGACTCT
CAGCTGCGGAGGGGCCAGCCAGGTAATCACTTCACTATTTGACGAACGTTTTCCTCGCTGATCAAA
TTTCTTTCAGTTGCACACAAAATCTACGGTGTTCCCCAGACCATTAACGCGGCCAACTACGTTTGG
CTGAAATCGTATCTAGAGGTCGCGAGATTACACCAACTCGACCCCTCTAATGCCACAGATTTTGAC
GCCATTGTTACCGGTGTGTACCGAACTCGCAAACGAGATTCGATATCCAATATTGATGGCGAAATC
AGAGGAGCTCATCTGGCTCCACCGTGGTCAAGGCCTCGAAATCTTCTGGAGAGATTCCCTCATCTG
TCCCTCCGAGGAAGAATACATTTCGATGGTCAACAATAGTGAGTTAACCTGGGGCCAGTCTAAGCC
TACGGTTAGACTCATTTCGTATCAGAAACCAGTGGCCTCCTACGTATCGGTATTAAACTCATGATG
GCATGTGCCACCACAAACGCGGATATGTATGTGTTTTTGGCAGTGCATTTTGGGGTTTGAACTGAA
TGGCCATTTTCTAGCGATTATATTCCTTTGACAAACCTCATTGGAGTCTATTTCCAAATCCGGGAT
GACCTTATGAACCTCGACAGCACTGAGGTTCGGGAACCTTTTCCGCTGTCGAAGCCCCAACTAACT
TACTCCATTTCTAGTACGCACAAAACAAGGGCTTTGCTGAGGACCTCACTGAAGGAAAGTTCTCAT
TCCCAATCGTACACGGCATCCACGCCGACACTTCGAACCGCCACTTATACAGTTAGTACCTTCATT
AATCAGCTCTACACCATCCTGACCGCCCAACCTCGCCCTTGAAAGATATAATCCAGAAGAGGCCCG
AGAACTCAATGCTCAAGGTCCACGCCATCGAATACTTGCGCAACCGAACAAAGTCCTTTGACTATA
CCCAAACGGTTTTGGATTCTTTGGAGGTACAGACTTTGGACGAGATTAGACGACTTGGGGGTAATC
GGGGGCTCGATCAGATCCTCCGTTATCTTCATGTCGAAACACAGAAACCGGAGTCTCCAATCTAA

Pleur345: 1367 bp

ATGTCTCGCTACGACAATATCCTTGAGAAACTCCACAAGGAGGAGTATACTTGGTCGAGGCGGAAT
GAAGCTTCCATCCTTGAGCCGTTCACCTACATCACCTCCAACCCCGGCAAAGACATACGCAGCAAG
CTCATTGACGCGTTCAATGTCTGGTTGAATGTCCCTGCTGATAAGCTGATCGTCATCACAAAGATC
GTCAACATGCTTCACGCTGCCAGCTTGCTGTACGTTCACAATTCCTCTCAATGCACCTTCTCGTAC
TCACCCCAACCCTCTTCACAGCGTCGACGACATAGAAGACGACTCCCAACTCCGCCGCGGCCAGCC
CGTCGCGCACAAGATCTACGGTATCCCGCAAACCATCAACACCGCCAACTACGTCTTCTTCCTCGC
CTTCCAAGAGCTCTTCAAACTCCGTGACGACGAGCACGCCCTCAACCGCGAATACCGTCCTCCCTC
CCCGGGATCAGTCCGCACACCCGATCTCGACGCGATCGTCAATGAGGAGCTTTTGTCGCTGCATCG
CGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAGGAGTATATCTC
CATGGTTAACAATAGTTAGTGGCTTTTTCATTACTTGAGATGAGGACTCGATACTCATGATCATTT
CGTTGAAATTTTCTAGAGACGGGCGGTCTTCTACGCATCGGCATCAAACTCATGATGGCATGCGCG
ACCACCAACACTGGAGTGTACGTATTCCATCCTTCTCATTGTCCACTCTGGTCTTTCTGACATTTT
CTATTTCTTCCAGTGACTTTATACCCTTAGTCAACCTCATCGGCATCTACTTCCAAATCAGAGACG
ATCTGATGAACTTGCAGAGTACAGAAGTCGGTGTACCCTTATCTTCTAATACAAACACCTACTGAC
CCCGACTATCCATTCAGTATGAGAAGAACAAAGGCTTCGCAGAAGATCTAACCGAAGGAAAATTCT
CGTTCCCAATCGTTCACGGCGTCCAAGCTGATACCTCGAACCGGCACATCCTTAGTAAGTTCTTCT
CTCCCCTTCCGAGTCCGCCAGAGGCCGCTGTCCTCTCCACCTTTTCACCCTTTCTGGTACTTTACC
CCGAGATGAACGTATCTAACCCTTCTCTCCCAGACGTCCTCCAGAAACGCCCAAGCACGCCAACGC
TCAAGATCCACACTATCGAGTACCTCCGCAACCGAACCAAGTCCTTCGAGTACACACAAGGAGTTA
TGGACTCGCTCGAATCACAGACGTTGAAAGAGATTGAGCGGCTTGGGGGTAACCCTGGGTTGGATA
AGATTATGACTCTGCTGCACGTGGATCCGCCGAAGATGGAGGAGTAA

Sord224: 990 bp

ATGGAACTCATCGATACACTTATCTACTCGCAACCGCAAATGATTCCTGACGCCATTCGTCTTCCG
TACGATTATCTTGCTGCGATTCCAGGCACTAATGGCGATATCAGAACCCAGTTTATCAAAAGCTTT
AACGATCTCTACTTTACCATCGACAACCATGATACTATCGCTACCATTAGCGAAGTTATAACTGTG
TTCCATAACTCGTCACTCCTAATTGACGACATTGAAGATGGCTCTGAGTTTCGTCGGGGTCTTCCT
GCTGCCCATACAAAGTACGGGGTGGCTCTGACTCTCAATTGTGGTAATTTGATGTACTTTGTAGCT
CTTCAAACAGCTCAACAGAAGCTCCCTCAATACCACCCTAGCGACAACAAAAATGATATTGCATTG
TCCATACTGAATATTCTCGTTGAAGAGTTGCTCAATTTGCATCATGGTCAGGGCCTCGATATACGC
TGGCGGGATCATCTCAAGACAACCAAGCTCCCGCTGATAGAAGAATACATGGAAATGATCATGAAC
AAAACCGGAGGATTATTTCGGCTTGCCGTTAAACTCTTGGGAGCTTTCAGCACCCCACAAGACAAG
CCAACTGTGTTGGCCATAGCTAATCTCTTCGGGATTGTATTCCAGATCCGTGATGACCTTATGAAC
TTAGTTGACGACAAATACTCACATATGAAGGGCATGAAAGGTGAAGATTTGGTGGAAGGCAAGTTG
AGCTTGCCAATTCTCCATTGTTTACACACCCTGACAGAGGACTCTCCGGTTAGAACACTCCTCTAT
CAGATGGATATGAAACAACGCCAGCAAAATCTTCAGCTTGTGGAAGAAGCAGTGGAGTACATGAAG
GTGGAAACCAAATCACTAGAGTATACTCAGAATTTATTGCGGCTCTACGGAGCCAAAATCATAGAG
TTGGTGGGTCCAGACACAAACTCGTTTTTGGTACAAATAGTAGACAAGTTGCGAAATTCCATATAG

Sord272: 1292 bp

ATGGCCCGCCTCTTCCTCTTCGGGCTCCGCTCAGCCAACCCGGTCAAAACTCCGCCCTCTTCCAAT
CTCCCAACGCAATACCTCCAAGAACTTCTTCTGTTGTGACATCTGCTACCTCCTCCTCGACCCCTC
CAGGCCCCACCAAGGCGGCTCTTCCATCCCTCAGCGAAAGCGACTGGCTTGCAAATAACCTCAAGG
GCCAACAGCACTCCTCTTCTACCGCTGGTCATACTCATCAAAGAAGCCGAAAAGTCTCGATAACGT
CGTCTGTCGGCACCCGCACCATTACACAAGCATCCTCTATCCCTCCCGCCAACATGGCCTCAGCCC
CTCCTACCCCGCCAGACCCGAAGCGCTTCGCGACCGAAGATTTCTCTTCTCGCCGGGCATGGTCCC
CAGAAAAGGACAAAGTCCTTACCGGGCCATATGACTACCTCAACGAGCACCCGGGCAAAGACTTCC
GCTCCCAGCTGGTCAAGGCCTTCGATGCCTGGCTCGAGGTACCCCCCGAGTCCCTAGAGGTGATCA
CCAAGGTAGTCGGGATGCTCCACACCGCGTCACTCCTAGTGGACAACGTCGAAGATTTCTCCCTGC
TGCGCCGTGGCTTCCCCGTAGCCCACTCGATATTCGGCATTCCTCAGACCATCAACTCATCCAACT
ATGTCTACTTCTGTGCCCTCCAAGAGTTGCAGAAGCTCAAGAACCCCAAGGCAGTCTCTGTCTTTA
CCGAGGAGCTCATCAACCTCCACCGGGGCCAAGGGATGGACCTCTTCTGGAGAGACACTCTGACCT
GCCCGACCGAAGACGAGTACCTCGAGATGGTATCCAACAAGACTGGGGGGCTGTTCAGGCTTGGTA
TTAAGCTGATGCAGGCTGAGTCCCGGAGTCTTGTGGACTGCGTACCGCTTGTCAGCATCATGGGCT
TAATCTTCCAGATTGCGGATGACTACCAGAACCTCTGGAGCCGGCGGTATACTGCCAACAAAGGCA
TGTGTGAGGACCTCACCGAGGGGAAGTTTTCGTTTCCCGTCATCCACTCGATACGGTCTGATCCAA
ACAATCTCCAGCTTCTGAATATTCTCAAGCAAAAGACTACGGATGAGGAAGTCAAGCGATACGCGG
TAGCCATCATGGAGCGGACTGGGAGTTTTCATTATACACGGGAGGTGCTGGAGGTGCTGATCGAAA
GGGCACGGAAGCTCACTGATGAGATTGATGAAGGTCGGGGTCTGGCCAAGGGTGTACATGCCATCC
TGGACCGCATGGTCATCCTGCAGGAGGATGGGGTGTAG

Sequences of the Probes for Screening the *Pleurotus passeckerianus* Pilat Genomic DNA Library

Three probes were used to screen the *Pleurotus passeckerianus* Pilat genomic DNA library. These probes were obtained from available plasmids from genomic walking. These probes are a 854 bp fragment from Pleur345, a 1133 bp one from

Pleur322 and a 1075 bp one from Pleur319. The sequences of these three probes are shown below.

Probe 1: *DNA fragment obtained from Pleur345, 854 bp*

GGTCTTTCTGACATTTTCTATTTCTTCCAGTGACTTTATACCCTTAGTCAACCTCATCGGCATCTA
CTTCCAAATCAGAGACGATCTGATGAACTTGCAGAGTACAGAAGTCGGTGTACCCTTATCTTCTAA
TACAAACACCTACTGACCCCGACTATCCATTCAGTATGAGAAGAACAAAGGCTTCGCAGAAGATCT
AACCGAAGGAAAATTCTCGTTCCCAATCGTTCACGGCGTCCAAGCTGATACCTCGAACCGGCACAT
CCTTAGTAAGTTCTTCTCTCCCCTTCCGAGTCCGCCAGAGGCCGCTGTCCTCTCCACCTTTTCACC
CTTTCTGGTACTTTACCCCGAGATGAACGTATCTAACCCTTCTCTCCCAGACGTCCTCCAGAAACG
CCCAAGCACGCCAACGCTCAAGATCCACACTATCGAGTACCTCCGCAACCGAACCAAGTCCTTCGA
GTACACACAAGGAGTTATGGACTCGCTCGAATCACAGACGTTGAAAGAGATTGAGCGGCTTGGGGG
TAACCCTGGGTTGGATAAGATTATGACTCTGCTGCACGTGGATCCGCCGAAGATGGAGGAGTAAGG
GTAGAGCATTTGGCGATGAGACCGTGATGAGAAAGATTAGTCTAGGATATGCTTATCTCGTTTTTT
TCTGGAACAGTGAATTATTCTTGGACAATACAATCAACGTATCTATCTATCTGTAACTCTATGATG
TCGCACAATCTTTGAAGTCTCTTGCGGTTCTCTCTACCCACAATATACAGTCCGATACTACCATGC
CCCGTCTTTGCCTTGTCCGAGTCGACCCAACCTCCTTAGATCCCCGCATCACCACCCTCTGA

Probe 2: *DNA fragment obtained from Pleur322: 1133 bp*

AAACGTTAGAAGGCGATCTGGCACCGGGCTGAGCTTTATCTACCTTCCCCCTCCTTGAAATTAGGG
CTTTGACTGGCGCATCGTGTCTGAGTTCACTACTCGGAAGGACGTGGAAACTCCTCAAGACACAAG
TCTCTGGAAATTTGAAATCGCTAGAACCATGGAGTTTAGTCTGACATCTTGCATGCTCAAGATGTG
ACGTCTGATCTTCTTTTGACAATTATGAACATTTCGGCGGTCATTCCCTAAGCCATTGTCAACGCA
AAGACACATACTAGTATAAGAAGCTCCTCTCTTTATCCAGACGGAATCAACCCCTATCCGTGAAAC
CACTCATGTCAATCTACGACGACATCCTCAGTAAACTTGGCGGCGATCAGCCCTGGTCCTTTCAGA
ACGAAAAGGTAAGGCCCTCCAGAGTTGGTTCCCAGTCCTAAATCCAACGTCATTACAAGGCTATTT
TGGAGCCATATACATACACAGTCTCTGCCCCAGGGAAAGACATACGTCGGAAGCTCATCGACTCCT
TTAATCTCTGGCTCAACGTTCCAGATGACAAGTTGGATATCATAGTCAACATCGTGAACATGTTAC
ACAACGCCAGCTTGCTGTAAGCTTCGCCGGGTGGTTCAGTGATCACTTACTGACTACTTCCGCTCA
CAGTCTTGACGACATCGAAGACGACTCTCAGCTGCGGAGGGGCCAGCCAGGTAATCACTTCACTAT
TTGACGAACGTTTTCCTCGCTGATCAAATTTCTTTCAGTTGCACACAAAATCTACGGTGTTCCCCA
GACCATTAACGCGGCCAACTACGTTTGGCTGAAATCGTATCTAGAGGTCGCGAGATTACACCAACT
CGACCCCTCTAATGCCACAGATTTTGACGCCATTGTTACCGGTGTGTACCGAACTCGCAAACGAGA
TTCGATATCCAATATTGATGGCGAAATCAGAGGAGCTCATCTGGCTCCACCGTGGTCAAGGCCTCG
AAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAAGAATACATTTCGATGGTCAACAATA
GTGAGTTAACCTGGGGCCAGTCTAAGCCTACGGTTAGACTCATTTCGTATCAGAAACCAGTGGCCT
CCTACGTATCG

Probe 3: *DNA fragment obtained from Pleur319: 1075 bp*

ATCTTATTGGGCTGTCGTGAGCTTCTCTTCCTTCTTGCATTTCGATTTCACTTCTGAAAATGACGT
CCAGGGACTTCATCCAGGAGCACCAGCTTTTTTCCGGTACCTTGGCTTATTGTTCCTTGTGATTTC
TGTCTCCGAAAGTCAGATACTCTTCATCGCGTCTGTAGTCCCTTATCTTGTCCCAGCATTGACCTT

GAGTGCATTGTAAGTCCTTTACGTAACTATTGTCATATGGTTCTAATTCGGATTCTAGCGTGAATG
GCTTCTGGCTCAGTTTGAGCGGTTTCTTGTCAGTTTGTCCCCACTGATCAGTAGCCGCCGAAAACC
TAATGCTTTGACCACAGTATAAGTGCACGCTCGCTTCCAGGGTTCTGGCACAACTCGTTCCATTAT
GCTGACTACATGAAGTACGCCTTTGAACTGTTAAGCAGTGTCAGTCAACACGCCCCCCATATTCTA
TTCCGTGTGCCATTTCTGATGGGCCCATTATTTCCAGTCTGATCTACGCGGTGTCACCTTCGAGTG
CCAACAGAGTCCAGATGGATCTTGCGTCTGCGCTTACCCTTCATCAACGCCCGAAACCTGTACAGT
ATCAGGTGATGATGTTCTCAACAACCTAGGCATTGGGTCCATCTCCTACAGTAGGTGGGCAGGCTT
TCTTGTCGTGATATGGGCGGTCTGTCGTATACTGTGAGTATCCTTCCTGATTGTTTGGGGACCTAG
CCTGACTGAAATTTCAGGATGTACTTCGCCCTGAAGAGCGGTTCCAAGTGACGACACCTATTATGG
CTTTACCTAGATATTTTCATCGAATCCAATACTTATCTTATAAGTTAATCACGGGAAATGGCTACC
AGTATTAGCTTCCCGATACAATGGTTTAGATTCTTAGAGCTGAGGAAAAACGAACCGTTATTCTTA
CAGAACCGTGCAGCTGAGGAGGTTCTGACATGCAGACTGTTTTCATAGAATCTGAAGTTAGCTCAA
AAGTGAAAAGCCTTTGAAAAGCCGAACCTGCCGGGAAAGACTTAGGGAAGGTTAAACCCTCCATCC
ATTCAGATTCAAGTCCGGT

Sequences of Pleur319 region, Pleur322 region, Pleur345 region, Sord224 region and Sord272 region

Pleur319 region: 12440 bp

CCAACTGAAAGACTCCACCAATATAGATGACCAGGAGTGCAAGTGGCACAAAGAGAAAGGCTATCA
TATACCTACTCGCCCTATCACCGAACACTATTGGCAGTGTTTTTCTGCCGACTTTCCGGTCTCCCT
CCACATCTCGCAGGTCTTGAATGTGCATGATCATGGAAGTCCATATAATAACAGCGTTCCTCCAGC
GGTTGATCTGCGGGTTGGTTGGAACACCAGTAGCCTTTAGAGCACCATCTAGCTCCGCCCAAATAC
ATATGGGTATCGCGATTGTGTTCTTCCCAAACCAATGTTGACCCTGAGGGGTAAGTGCAAGGAAGC
CCGTCACAAGGATCCACGTTATCGTGGCGGGAGCGAAGGTAGGTTTCACTGCCGCTAACGTCACAA
ATGTAGTTGTAGCAATGGCCCATCGCACTTTTGCACCGAACATGGTCACCATGCCTGAGGGCAGTG
GGCGATCAGGTTTGTTTGCCTTGTCCTCCTCTGGATCTGACAGGATCTGAGCCGAAATATTAGCAA
AGAGTATGAAACTTGCAAGCCAAAGGAAGAGGTAGATGACAGCGCGACATAGTTCATCGGCTGGTG
TTGGCTGCTGGTCGTTATATCCTTTGAGGGCGACGGCGCCGTATAGGAATCCAGGGAGAAAAGCGT
TTGGTATATCGCGGTACGTGAAGGCATATAGAACTCTGGCTTCGTTATGCAGGCGGGGAGAGTATC
CGTCTCACGGACGACATGGCAATATAGGTGGCGGGAGAACGCTTGATAGTAGCACTTATATACCCA
ATAAACCGACATAACTTTCGCTACTTGTCGAATGAAGACTAACTTAACTCCGAAGTGCGGAAATGA
TCATCTACAGTTGATCTCATCAACGAAAGTAAATAAGAGTGACATAAGGAGATGCGAGGGTCGATT
GTTATCGACTCATCTATGTTGCTCCTGGCTTCTAGTGCCTAAAATTACATGATAACCTTGCATGAC
AAGAGTAATTTTTCGCTAACCCCGGTGGACTTTTAGAGAATGAAAGCTTCTGAGTTCAATTTGCAA
CTTCGTGTCCAGGCCAATTCAACTCCCGCTCTTGAGGCTGCCTGCCGGTTGCTTTTAAATTAGTCG
TTGCTTGGCTCCAACTCGTTGTTCAATATTCTCGTAGCTTGCTGTGTTTGTATAGTTCTCGTCTAA
GTTCGTAGAGAATTCACCGCCCATTTCATGCCCCAGTTTAAGTTGTAGTAAGGCCCAGTCTGATTC
ATTTCTCTGTTCGGGTCTGAGTAGTCGTGTCAGAACTGTGCCGGTGACCATTGGCGCCAGTGTGCA
TGCGTTGTCCCTTGCATTGGCCGGGAAAGCGAACAAACGTTCATCACAAACATCGCGTCCTGTATT
CATTAAGGAAGATGTAGAACACATCCAAGTCACTATCCTTGCCCCTGGAGGGCGGCGACAAATAGG
CAAGTCAGGCGAAATTGGATGGGGGTCTAAGCTTAAGTCCGAAAGATTCGGACGAAAGTCGGTTGG
GTGTGACACAGGCATTCTCCATTTGTCATGGTTGATTAACGCGTTGATTGTGCAGGTCATTGATAC
CGGAACGGGTGAACCCGAAGCGAATAGACGTTCACCCTTGCAATTGCGACATACAGAGATTACCCG
GTCTTCTGTACACCCAAAATTGATTGGCCATTATCACTTCTAGCTTCCTTATGACGCCCAAGTTTC
TTGGTCTCAGTAGCTTACATATTGTCTTCACAGCCTTGTTTATGAGTTTGAGGGATCTACGTCTTG
AGGAACAGGAGATACAATCGTTTTTTTAAAGTCGCGCCGAAATCAGGGCTCCCACCAATCTGTACA
GTCTCGGCTAGGCCCTATTTAGGCGTGGGAAAACGGAGTCAGCCGGCCGTCCACCGTGCTCATTTT
CTGGTGTTCGCCCATGTTCGGGACATTAAAGCTTCAACCTTCCTTGTTGGCCCAAATTATTCTGAC
TATTTCGCTCACAATGATCTCCGCTACTATTCTTATGATACCAGAATTGCTATTTCGTGCAAATAC

GGTGTCTGGACCGTGTTTTTACCAAAATCAGGATTTGTGATGCACACCAAATTGGGTATATTTTGC
CATTCATTCTTCACCAACTTATTTTCTTCGCATAAGAAAATTATAACAGTCTCTTATGGTGTCAAT
ACATAGACTACAGTCAATAATTTATGTGTGGAAAGCCTCTGTGCTTAGATGTGGTTGAGTGTACAG
ATGAGGGGCGGTGTACAGATGACCGGGTAACTCTGTAAGGAGATCTTGGCAAGATCTGGATTCTAG
CGTTCGAAGCTGCATGCAACCCTTACTTGCTTGACATGAAAAATGATCACTAAACCTGGCAAAACT
TTGGTCATGACTTCCGGGTCTGACATCCAACTCGCTGTCGCTCCTCATACTATTTAAGCAGGATCC
GACAGGTTCGAGCTAAGTGTCGCAAGCTGCCAGTGAACGACACATCAAATACCAATGACCACGCCT
TCGACGCTTTCCTACGAGAGACTGAGTTACGAGGTAAAGCTCAAGAATGGTGAAATGCGACGCCTG
GTGGATAGCGTTTCCGTTCAAGTAAAGGCTGGGGAAGTCTTGGCGATTCTGGTGCGTGGGTTGGTC
ATTTTCTTTGCTTTTTGTACTGAAGGCTCGTAGGGACCGTCCGGAGCTGGTAACCGCTGGTTCTTT
TGTGCAGAACAGCAACCCTGATGACTTTCCGGAAGGGAAAAGCACGTTACTTGAACTCATGTCCTT
TCGGAGGTCCAGTCTTCCGGGGGCTTCAGTGAGTCTTGCCTATTCATCTCTTGATATAGAGACCTT
AATCTAATATATAGGTCAAATTCAATGGAAAAGTTCTTCCCAAGGAAACTATGCACAAGATCTCTG
CATTCGTGGAACAGGATGATGCTCTTTTCGGGGTCCTGACGGTTCGGGAATCTATCGAGTACTCTC
TCCGTCTTCAGTAAGTCCTAGGTCCTTGGGCTTTACAGTACGAGACTAACATCAATCGACAGCTCG
CCCTTCCTCAAACGCAAGGATGTAGCTTCCCGAGTGAATAATGTAATCGCCGCATTAGGGCTATCA
TCCTGTAAAGACCAAAGGATAGGAACAGTACTATCCCGCGGCATCTCTAACGGCAAGTCATCAGGC
CCATTGACAGGTGTTCAACTACTCAAACTTCGGTAGGTCAGAAACGTAGGACAGCCGCTGCTTGCG
CAATGGTCACCTACCCGTAAGCCCCCGCAAGTCTGGCCCTAATACTCTTACTCACTCAATCTGATA
GACGCATCCTCTTTCTCGACGAGGTTACGTCTGGTGTAAGTGTTTGCTACACAATATAGCTATTTG
TGCCTCATGCTGATACATTGCTAGCTCGATAGCACGGCTGCACGGGAGGTGATTGCTTCAATCAAG
AACCTCGCGACCGCGGANGGGTATGATGTATAGCGGCCATACACCAACCCTCGATGGACACTTTGT
CGCATTTCNCAATCTAATATTGCTTTCCGGGCGGCAAAATGTGCTATGGAGGACCTGTCAATGAAT
TGGAGAAATTCTTCGACACCTGGGGGCATCCAGTTCCCCGCCTTGTCAGTATCCCTCTGATAGTTC
GAACATCATACGAACTGACTGCCTTCGTTTCAGCTCACTCCTCCCGAACACGCGCTGAACTTCTTG
AACAGCGATTTCGAGTATGCCGAGGATTCATACGACTCAAAGGCCACTGATGCTCTGGCAATGCAG
GATTTCTACCTTAATTCTTGGCGTGCTCATTTAAGTGATGGTGTTGAGCGACCTCATGACGATATC
CAGAAGACCATTGCACGCGACAGGGACACCCGACCCACAATGGCGATGATGTGGAATACCCTAGTT
CTTAGCGAGCGAACAGCGAAAAACTACGTTCGCAATATCCTCTCGTATCCAATGCGAGCTAGTATA
TTCTGCGGTGAGCACGTCCTTTCTCTGTCAACTCCTCCGGCTTCTGACAGCAACTACACAGGCTCC
GGATTGTTGGTTGCGTGAGTATCTCGTCTAGAGGGAGGTAGGGGCATGAGAGGTTAATTTCCGTTC
TCTATTGCCTAGGACAATATGGATTCGGCTGGACCTCCTAGACACAACTGTCAAAGATAGGCTCGC
TCTGCATTTCTACACGATCGCGTATCTGGCAATGGTAAGATGTCTTTTCTAGTTCTGCTTGAGATT
GTGTTTGACTTGCAGTTGCAGAGTTCGGTTGCGGTTGTGCCGTCCTGTACGTTCTCGCATTGTGTT
ACCTTTACCTCCACCTCTAATACTTTGCCCTTTAAAGATCTAGAGGAGCGTGCTGTCTACTATCGT
GAAACTAAGAACGGATTATATTCCGTAATGCCGTTCATGGTATCGAACATGGTGGTTAACGCCCCA
TTTTTATTCGGATGTACAGTCATCTATAACTGNTATCTTATTGGGCTGTCGTGAGCTTCTCTTCCT
TCTTGCATTTCGATTTCACTTCTGAAAATGACGTCCAGGGACTTCATCCAGGAGCACCAGCTTTTT
TCCGGTACCTTGGCTTATTGTTCCTTGTGATTTCTGTCTCCGAAAGTCAGATACTCTTCATCGCGT
CTGTAGTCCCTTATCTTGTCCCAGCATTGACCTTGAGTGCATTGTAAGTCCTTTACGTAACTATTG
TCATATGGTTCTAATTCGGATTCTAGCGTGAATGGCTTCTGGCTCAGTTTGAGCGGTTTCTTGTCA
GTTTGTCCCCACTGATCAGTAGCTGCCGAAAACCTAATGCTTTGACCACAGTATAAGTGCACGCTC
GCTTCCAGGGTTCTGGCACAACTCGTTCCATTATGCTGACTACATGAAGTACGCCTTTGAACTGTT
AAGCAGTGTCAGTCAACACGCCCCCCATATTCTATTCCGTGTGCCATTTCTGATGGGCCCATTATT
TCCAGTCTGATCTACGCGGTGTCACCTTCGAGTGCCAACAGAGTCCAGATGGATCTTGCGTCTGCG
CTTACCCTTCATCAACGCCCGAAACCTGTACAGTATCAGGTGATGATGTTCTCAACAACCTAGGCA
TTGGGTCCATCTCCTACAGTAGGTGGGCAGGCTTTCTTGTCGTGATATGGGCGGTCTGTCGTATAC
TGTGAGTATCCTTCCTGATTGTTTGGGGACCTAGCCTGACTGAAATTTCAGGATGTACTTCGCCCT
GAAGAGCGGTTCCAAGTGACGACACCTATTATGGCTTTACCTAGATATTTTCATCGAATCCAATAC
TTATCTTATAAGTTAATCACGGAAATGGCTACCAGTATTAGCTTCCCGATACAATGGTTTAGATTC
TTAGAGCTGAGGAAAAACGAACCGTTATTCTTACAGAACCGTGCAGCTGAGGAGGTTCTGACATGC
AGACTGTTTTCATAGAATCTGAAGTTAGCTCAAAAGTGAAAAGCCTTTGAAAAGCCGAACCTGCCG
GGAAAGACTTAGGGAAGGTTAAACCCTCCATCCATTCAGATTCAAGTCCGGTCTTTCGTATCCCCC
AACCAGTCAGTAATCCGGCTCCCGATCAGTTGCTTAATCAACCGTGCCACGAAGACCTTCTTCAGG

GAGGATCGTGGACCCTCTAGTCTAGCCAGAGAATAATGCAGGGCCAGACTGAAGAATCATTGCACT
TAAACCCGACAGACTCATCGAAGCAACATATCCTGATCCAAATGGGTCTATTCAGTGCTGCATCTG
CGCTTGGCGAGGAATGTAGTCTGCTCAACTTAAACGAATGAAGCCGTGGTGTAGGCGAACTTTGGT
AGAATTTTCTGAGGCTGGGCGAGTCTGGCCTACCCACCATTTGGTGAGTAATACGTGAGCCCAAAC
CGAAAGTTATAATTTCAACTGCTGGATTCAGATCCTATCCGGTTCAATGTCATGTATTGTGGGTTC
CTGGAAATTTGTTATGTCAAAACTGACCGGGGCTGAGTTTGACGTCCTGCATAGTTCAGAATATGT
TTCCTCCGAGATCCTAACCCATGAATNTTCTAACCCGCGAGAGGTTTAATGAAGTCAATGTCGGGA
GCCCTGGAAGACCCATAAAACCCACTTAGAGTCACCACCCAGATCCAAGTGTCTCCCACTTGCAAA
CTTTTAATGTCGCTCTACAACAACATTCTCCGTAGTAAACTTGGTCACGAACAACCATGGCAGCCT
CAGAGCGAAAAGGCGAGTGAACTTCGGAGTGTTCCTGACTTCTGAATTTGACGCTCCTCGAAGGCC
ATTACGGAGCCATATGCCTATACTGTAACTGCACCAGGAAAAGGCATACGTCAAATCCTCATTGAG
GCCTTCAATGTTTGGCTAAACGTGCCAGAGGATAAGTTGCTTCTCATAAAGAACGTCGTGGATATG
CTCCATAATGCCAGCTTGCTGTGAGTGGTCTCCTGGTATCTCGACGGTAGTTGACCGATTGGTGGT
TAATCATAGTATCGACGATATCGAAGATGATTCTCATCTAAGGAGGGGCCAACCGGGTAAGGCTCT
AAGAAATATGTCATTAATAGTGATTGACCAATATCACTAGCTGCGCACAAAGTTTATGGCGTTCCG
CGAACCATGAACGCTGCGCATTACGTCTGGCTGAATGTGTATAAGGAGCTTTCGAAGTTCGACGAG
CCAAACTCTTCGAATCGTGATGCTCCCCACTTGACAGATATTGTTACCGGTATGCACAGTACTATC
CAACTACATTTGATTGCTGATGAGAATACTAGACGAAATGATATATTTGCATCGCGGCCAAGGCCT
TGATTTGTTCTGGAGAGATACCTTTACCTGCCCGACGGAGGAAGAGTATGTTTCTATGGTCAATAA
TAGTGAGTTGGTTCTTGGCCCACCCGAGGCTTAGTCTTAGAACACTAGACTGATTGATACCGCTAG
AAACTAGCGGCCTTCTTCGCCTTGCTATCAGGCTCATGATGGCATGTGCGACGACCAATAAGGACA
GGTATGTGCCTGTGTTCAGCTCCGCTTGCGCCTAAATCAACTGTCCTAATTCTTAGTGATTATGTC
CCTCTTGTGAATCTCATTGGAGTATATTTCCAGATTCGAGACGACCTGATGAATCTTGACAGCACG
GAGGTGAGCGAACGTCTACTAGGTTAGCGGCCTACTTAGTTTGTTATAATTTCAGTATGCCCAAAA
GAAAGGCTTCGCAGAGGACCTCACCGAAGGGAAGTTTTCGTTCCCTATCATACACGGCATCCAAGC
AGACACTTCAAGTCATTTTCTACTTGGTAAGTGCCAGTTTTTCGCAGGCTTTCCTCCTACGTAACC
CTCGAAACTAACGCTCCGTCGATAGATGTACTCCGAAAGAGGTCAGGGGAAGACGCGTTCAAGATT
AGTGCCATTGAATACCTGCGTCATCGAACAAAGTCCTTCGAGTATACCCAAGAGGTTCTGGATTCC
TTGGAGATGCAGGCTTTCGACGAGATCAGGCAATTCGGGGGAAACTGGCAGCTTGATCAGATGTTG
CGCCTTCTTCGGGTCCCTACTCCAAAACCAGAGTGTCCTCTTATCGAAGATTTTGATGACTTGGAC
CTGGACTAGTTTATTGCTTATAGTATTTTATGTACTAAAGTGGTCCCGAAGCTCGCGAACGTACAC
GCATCCGAGGAAAACGCCAACTTGAATTTATCTTGTGCAATCTGCTGAGACTTGCTCATTAGGACT
GTGTGCATATTATTGTTTCGAACGGCCATTACGCAGGGTATTGTAGCTAAATACTGCAAATATCAC
ACATTGGGCTCGTCAACTTTGGAGCCCTGAGGGAGCCCTCCCCCAAATCTTTCGCTGATAATCAAG
CAAAATGCAGCCAAACGCAAAAGCTCCGAAGTCAGGCCAATTGGATGGATAAAAGTGAATTTGTGT
ATTTTCACAGCACTAATTCGAAACAGAGGTCAGCCATAGCCCTTGACCCCAAAGGATAATACCCTT
ACTTGCCAGAGGAGGAAATCACTCAGTCGCTTTGCACTGGAGGTGCGAATGAGGATATCCAGAGGG
GGACTCCTTGTTCCGCCCTCGGAGGCCATACGATTGGTCATGAGCTCCTGTTCAATGTCGTCTTCA
GTGATGCGACGCCTATTACGGAAAAGCCATCGAGTATTCCATAGTATGAGTCAATCCGATTTTAGC
CTAATTTGACCTAGTAGACGCACCCTTTCTTGTTTGGACTCCCGAGATCCTCCCCTAAGTACTTCG
TAACGGTGGTCTGAATAGCGTTGGTCATCTCATCTTGAGATGTGTAAGCTAGGCAAAGATTTAGGA
TCCTAATACCCGTCAATAAACTACCCATAGGAGACTCAATGAGAACCCACTTGGAGTAATTCTGGG
TCTTTTCCCCCGCTTCCTCAGCTATTCGCCTCACTGTTTCCGGTAAGAGTTCTTTCCGACCAACGA
GCTGAAATCGAACTCTATGCTTGTGAAGCACACCGCTGAAAAAGTTAATTAGTTGTGATGCTTGTG
TGGTCCGCCATGAATCTAGACATACCCAGGAGCAATCATCGTCTCTACTGCCTCCACGAATAGATC
CATTAGCGGGGCAACCTCCTCGGGGGAACGTTTAAAGTTCTCAATAGAGAAAGCATACGCCGAGAT
GCATTGGATACCAAGCGTGTACCCTACCTCTATCATCTGCTGATAACTGAGCCATCATCACTTGAA
GCTGGTAATCGTGAAGGTTACTTGCCGCCTTCAATGCATCCGTCCCTTCCCTATGTCCCTGCTGAA
CTGGTAGGCCATGGCCTCGAGCCCACCGCCTGTTGCCATCCATAATAAAAGCGACATGTTGAGGAA
CGGGCCCAGCAGAGAAGACACAAATGCCAAACCCTATGAAACGATCGTAAAGAGGCTTGTAGACGA
GTGTATCAAGGGCTTGGCTGAATCCGGTCAAGGTACTCAAAACGGTTACGAGAAGACTGTTCCAGT
TTCTGCCAAGGCCAATGCCTGCAATCAACAGCGAAGGAAGCCAGAGAATGATGAAATTAGAGCTGA
TTTCCATGATGAGAGGGCGATGGTCTCAGAGGGTATGGAAAGCTGCCAGCTAAGGCCCAGCTCCAT
TTTGTACTCGAGTGTGTTCTCTAAACTTTCGGAATTGTCAGTCAGAAAGGAGCAGTAGCGCTTGGC

TTGAACAGGGTCAAGCTAGAAGTATGTCAGTTGTTCGGTTGAACGAATTCTGACACACAAATTGTT
CTTTGAATTCAATGGATTCTGGAGACTATTCCAAAAGGTTCAAGACAAACATATCACGCGCCGGCT
GATTCGGAAAACAAAACATTCTGTCTGTTGTGACAGGGTCGTTCCGACGGCCCGAGTTTCCGACCA
ACGACTTGCAAATATTAACGCTCAGCTTTCAAAAATAGAGTGGATACCAAATTATCTATCATTGCG
GTTGGAATAAGTATGATTAAGTACAGGTAATTCGTCTTCAGACGAAGGAAATGAAGGAGATGGTAT
GAGTAGAATCGGCATAACTATAAGAATAAAGCAAAGCGAAGTCCTGATGAAGTCCGAATTCATGTT
TATGTTTTGAAAGAATGATACGCCTGAAGCCGAAGCCGGCCATTAGTTCCCGATTAGTTTATGATA
AACTTCGTCTTTCTCATAGACATCGTTCTCGCTCTCAATCACCGAGTGTTGGTGGCATGTGGCAGC
ATGCTTCCACCTTCGAAAAGTTTAGCAGAAAGAAGCCCCGAAAACCATGAGAGGCTTACCAGAATA
TCTCCTTAACCTTGCTCTCGATAGCCTTTATCCCTTCGATCGGGACCCATTTTGCGCATTTATGGC
ATTTACCCTCCTGAAGTTCGGCCCTTTCCTTCTTTCCAGGGTTAGACCTTTTCACTCGACGAAAGG
CGATGGGAGGAGAGAACGGCCGTCCGGTTGCCGCGGATATGCCTGTTGCAATAGTAAGCAACAGAC
ACAACGTACGATGGGATCTTACCATGGGAATACTGCATATGATAATTGAAGGCAGAGAACTTCATG
GCCAACCAGATTTTCTTGCCGCTACCACCCCGAGCCCGAGGTTCACAACAGACCGGACACAGGCCG
ACCTTATCCGTTCCCCTTCCCTTCACAAAACGTGGAGTGTACAGGTCCAGCGAGAATCGCGGAGGG
TTGTAGTTTCCACCTGGGTGTCGCCGGCTGAAAAGAAGTTAGTCGAATGAGAAAGAAGAAGGTATC
TCAATACTCACTCGCTGGAAGCCAAGTCGAGGTATGAACTGACGGGAAATCGTCGATAATACATAG
GGAATTCAGACGAGACTTCAATTTGTGAAGAAAAGACCCTGTTCGTGAATATTGGCTCAGAAGACA
CGTCATCTTCAGTCGGGGTCGAAGATTGCTCGTCATTATTCTCAAGGGAGGAAAGTGAGGAGAGAG
GGTGATCAGAAGGTTCCTCGGGTGACCTGAGAACAGGAGAATCGGGTTCGTAAACGCGTAGACTGT
CTCCGCCCAGAAGGAGCAATACGAAGACCCAGTTTTTGCTTCTTAGCTGTAGAACGTGCCTTTCGA
CGAAGTGGAGACGGAAACCCAGAGCTATCCGACAGAGGCGTTAGAGGCGTCAGAGGAGATGCCGAG
GTAATTTGAGTAGATCCGTAATTCGGTGTTTCTAAAGGTGTAGATGTGGTTTGCTTCGAGGCGGAT
AAAGCAGAGAAGGGGGAAGATAAAAGGGCACTATTCGCGTTCCTCTTGCTGGGAGTGTTGTAGAGT
CTCTTGCAAGGAGAACCGAGGGAAGGAAGCGGAGAGCCAAGGTTGGGTGGATAGTTCTCGGCAAGC
GGCCGCTCTGGGGAGGAGGCCGCCGAAGGGACCTCAGGTTCTGATACACGTATCTGTTTAATTCTT
TCGAGGAACGTTGGTTTAGTGTTTTGAGAATTGAAGGGAGGGTAGGCCGAGGCGACAGAGATCAGA
GCGCTTTCGGTTGAAGCATGGATGACGGGAGACACGTTGTTGACTATGGGAAACGGAGGCAGTTGT
GGAGAGTATGGCTGATAGTGTTCTGGAAAGAAGTTGGTAATATCCTTCAGAGCAGATCTTTGAGGT
TGATAGTCGAGAAGCACGGCTGGAGCGGGAGGCGATACAGAAAAGCGGTATAAAGGGCTAAGTAGC
CTATCGGAGGTGCCGAAATTTGGGCTGCTTGGGGCCACAAAAGGAGTAGGTGTCGAAGTAAAGAGG
TGCCGATTTTGTTGAAACTGAGCATGGTTTGATGTGTTTGGTAAATTGGTGGAAAAGGCTGGATAT
GATATGTCCAGCTCCAAAGGTGTGTAATAACGAAAGTTAAAGTTTGGGTCTACTGCAAATTCCTGA
GGTGCGGCCAAGTGGTGAGGATAGAAATCCGATTGAGAGGATCCAGACGAAGAAAGTGCTGGTGGC
GAATCGACGACCAGGGAAGATGCGTCCAGCCAATCGTTGAAGAGATTGTAATCAGATAAGATAGTG
GAAGTCATCAAGAGCAATAATGAGGTTGTGCCGCCGAGTATTCGAGGAAAAAAGCCCGGATTTTGA
GAGTTCGAGAGCGAAAAGGATCCCTCAAATCCAGTGCGTGAGCAATATAGATGGTGGTAAGAGAGA
AAAATTGTTTGTGACCTCTCGAAATTGTGCCAGTGAAGGAGGGGTGTTTGGACCTGTTTGGCTACG
AATCCCGACCCTTCACAGAATCCAACGTCACGCCAAACCGCACCTTGACAGCAGGTCTTCACGCAA
AATGGTGGTCGTCCTACACCTATTATCCACACTTTTCTTTCTTTTCTCGCGAATATATTACGTCAA
GGCAACTGCATTAACCACGGCGCTGGCAGCTAATGAACGACTTTGCTTTTATGCCGATGTCGACAA
GGCCGGGGAGAAGATTGGTGTAAGTGAAGCCATCAATTTCTACACGTCGACAGTTCAGGGGAATTA
ACGAATCCTACAGTTCTACTTTGCAGTATGTCTCGCTAGCTGGCTCTGTGATATACCAACCCAACA
CCTTAGAACAGGTACAATCAGGAGGCTCGTTTGATATCGACTTCGATGTGAAGGACCCGAACGACA
AGATAATACTGGACGGCGAACGGGAGCGGCAAGGGGACTATGTCCTGACGGCTAACACCATTGGCG
AATATGCCTTCTGCTTCGAGAACGACATCTCTACGCTCACGTCTAAATTGATCGACTTCGACATCA
TGGTTGAGAGCGAGCCGAGGAGAGAGGCACCAGCAAAACCAGGCCAAATCGCAGAACAGACGAATG
CATTGGAGGAAAGCATATTCAGGCTGAGCGGAATGCTGCAGAATATCAAGAGGTTGCAGAAACAGT
GGGTCTTTCTCCTCTCAAGAACCGCCAGCCGGTCTGACATGAATTGCAGCTTCCACACCAATGAGA
GGCGTGGCTTCTCGATAGTCGAATCAACACAG

Pleur322 region: 11318 bp

TCATGAATATCGTTGGGCGCGGGACTCACATTCCGTGCGACCTGCTTTCTACTGCCTCGATTATCT
CAGATAAAACTTCGGGTCTATCTCCAAGTAGCAGTCTGAGTTTAGGAAGCCGGGATATTTCGTGGT
GAAGAAAAAGGCGGATATCTTCGGGTTGGGCATGAACATGGAAGTCCGGAGCGGGAAGATTCGGCA
GATACAGCTCGCGGGACATGACGAATAAGGATCGTTGAAGCGAGACCAGGATATCCACCAGTTCGT
GGCAGACTTCTTCCTTTAGCTCGTCTAAGGCATCCACGACTATGAAAAGCTTGGGAAACAAACTGG
CCACTCCCCGGACTGTGGCCTGAAGTTCGGATGGGGTAAATGGTGTATTTTTCTGGATGGAGGGGG
CATAGAGCTTTTCGGCAACATAGCGAGCCGCTGCTTCATTGTCTTCGACAAGCTGCCGTAGTAATG
CAGCAAGGATTCGTGACACTGAGAGCTCGTCGTTAAACCGGCAGTAGGCATAAATAACCATAGAGG
GGGCCACATTCTCTGCTGGAGATGAACTGCCAAGGCCTTCGAGGTGGTGCACGACGATCGAACTAG
CGAACGTCAGTATGGGAACAGAACATGGGGCTCGATAACCTACGTTCTGATAGGTCTTCCGGCCCC
GGCTGCAGAGGATGATTGGGGTGACTGGAGTAGGGATCCTTGGTCGAGAACACTCACGTATACCCA
TTCCCCAAAAAACCTCGGATGCTTTGGCCTTCCAATTCTTGAAAACCGGGCTTTCGATGAGGTGTT
GGCCTGTACCAGGCGACACTTTCTCGAAGGTATCTTTCTGTATCGCACGATAATTCAGTGACGGTT
CGGATAACCAGTCAGCTATTTTGCGCTGGATACCTGTCAGGTTGAGGTCAGTCGAGAGGCAGACAG
GCGCCGGTTGAAATCGTCACCTGGCGTATTAGCCTCAACCCGAGGATTGCTAGAATTGACAGTGTT
GGTGATGGTTATATTGCCGCCAGCTGAGATATTGTTTGAGAGCATCATCATCGTCGCCAGATGGGG
CCCAAGTAGGTTCTCGGACGGGAGGGCATGCCTCAGGCGCTTTCCAGTGGTGGGTTGATCTTCATG
ATAACTATCTTGAGAGCCTTGTCGGCGTTTCATACGAAAGTAATCGAAGATTGGAATGAGGGGTTC
GGGATACAAGGCAGCTGGAGAGAGATGAAGGCAAGGGAAAGCAAAAAAAAACAGCGCCAAACGTCCA
GGTGCTGTGCAGCTATATGGCGGTGCAAGGGATGGTGAGCAACGGAGAAATCCCGGAGAGCTTCTA
ATCCATGTAAAACCTGAGCACGTAGATCTGCTCATTCGGAAAGGCTGTCCAAACAAGATGTGGTTT
AGACACCTCGACGAAGAATATGAACGACTGCGGAGATTGTATATCATTGATATCTAGGGATAGAGC
ACATTGACATAGACAGTAGATATTGAATGTCTTCGGGTGTAGTATCGAAAAGTACTGAGGAACTGG
TTGCTCAAGTCGCTGACTTGACAACTGTCTGCTTTGCTCAACGCCTGGGATGACTCATAGTCCGTG
GAACCTAGTAGATTTTGGTCGCTTGTCGCCTGAACGGAGATTAGCAAGGCCGTTCCCTTGAACGAT
TTGTTTTGCCTCATCAGCTACATAGATGAAATTTCGTATTGCTATGAAGAAATGTCAACGCCATGT
GAGATGCCAAGCAAGCAATACTCACGCTCATCGAGGGTACATCATCCAACGCCGCGAAAGAAAGAA
GGTTCTTCATACAGCAATATGACCAGTTTGCTTCTCGCTTATCATGGCCAGACGACTGAGATACAG
ATGACGGGATTTGAGCGGTTTATTACAGATAGAACAGCTCGATTTTCTTCTCGACGTCTGATTCGC
CTCATGAACAAAGTGTTGCAAATGGGCTGGTGGATGTCGATCGCTTGTCTTGGTGAGGCTACTTGA
TTATGGTAGGCGTCAGATTGGGGGCGGTCGCGTTCGCATTTTATCACAAGGATGAACCATCCTGTG
TATACTACAGAAAGACTGCCGCAGTAAACAATTTGTGGTAGGTTTTCTGTCTAACTTCGAAGTTGG
GTTGTCGCGATGATGCCGTTTCTCCGTCACAGGCCGTCACATGAAAGTGTCCACAGAAATCAAAGA
ATCGCAGTGCGGAAAACGAAACGATCACCAGTGCCAGTCTCCTCCTTAGACACCGCTCTGGACCTC
ATACTTTAATTCTGACTACTTATGCTTTTGGAACTAGAGCCTCTCTTAACTTCGTCGAATGCTTGT
TCAACTCATGGCCCGTCCACAAGTTTACGAACCCTTCAAATCTTTCAATACACTAATTTAAACGGC
TTGCCACTGGCAAATAGGTAACAAGTCGGAAGTCCAATGGTATTTGTGTACTATTGAGAACATAGG
AGAAATAAGCCACTACTTTCCATGGAGGATCAACTCGGGGCCATATTGGGCTCCTCTTTATGACCA
TGCACCAGGAGCCAACCTTCACTTGGAGCGGACCTTTAGAGTGATATAGGTCCTAACAAAACCGAT
TAGTATAGCTTATTACCCATACTTCAGTGCCACGTACAGGGTGCGGAGGGCTACCACCATAGACAC
CAGGACACAGGCCCATTTAGCATAGGAAATCCCTCCGATTTCGAGGTAATCAAGAACCTCTTGCCC
AGAGATGGTACATGCCCCTTCGACCGACGAAGGGTAAGCACAAGAGCAAGACCCGTTCGGCCGTTG
TTCGCACTGGAATGTAAGGTCTCTCATATCGGACTAAGTGAATTAATGGGGTCAGCCCGTCGTCTG
AATGACGATGAGGCAACGTAGAACTTACATTAGTCAGTAATTCAAACGCATATTTCGCATAATCGA
TGTGGTGAAACGTGTAGAACCAGAATCGAGGCAGAGTCCGAGCTTTGATGCTAAATACCGTTTTTC
TTAATAGGTTCTCATTGAGGTAAGCACGAAGACTTACAAGAAACCGCAGAGACACAGTAAAAGTCC
GTCGGCGCTGGTGTCTTGTTAATTACCGCCTTGGACATCCTACAGAAAAGACGCTCACAATGCACT
AATGGCTAGACTTGCCATAAAGCTTGGAATGAGGCTTGCGATAAAGACCGTCTGGCTCTCCCTACG
AGGACAGCGAGAAAGACGAAGACGAGTATCGAAAGAACGGTCCGGCACCAGAATGCAATCCCTATA
CCTCGTGAAAAAAGGCATTGATGGTTTCTAGTTGCACTTACCACAGACCAGTATACGATCAACGTG
AATGTGGTTGCGCATGCAAAAAGGAATGGTAGGTTGATCAAAGTGTTGGCAACCACAAAAGGGAGT
GTGGTGTAGAGACCATTCTTGCTCTCCCTATAGTATACGGCACGTTCTTCCAAGACTTTCCGTGAT
CAGTAGTTAGTATTCGGCTCCCCAACGTCTTATGGAGGGCTCACAGGCTGGAATACAGGCAACTGA

AATCTGCCATCAGGCGTCAGTATTCTGTGACATGTACACCTGAGTCGCTGAACTCACACTAGCCAA
AAAGCAGGGGATGAAGAAGTGCATAGCAAGCCGATCGTTGATGGTTGAATCCTTCAGTCCAAGGCG
GACCCAAATGGTGCTAGAAAAGCTGTATTCAAAAGAGTGAAGGCTTGAGTGAGTGGAAGGGTTATT
ACCTGACAAAAACCGCCATCGCTATAGGTCATGCAGGGGTTGAGTTAGTGTCATTAAGTGACGGAT
TTAGCCTAGGGCTCACCGATATACATGCCACCTCGAACGCCGTACGCCTGGACACTGCGCGAGTAG
TGGACAGTAGTCCTTTCTAGTAGAATCCCAACACGACGACCGACATCTGAGAGCCAGTGCTCACAC
TCCGATGTTAGCTCGACTTGGTACAAGTCCACCCTTGCTTCGGTCTGACCGACCTGATCTTTCGGG
GTCGTCGGAAGGGCTGAGATGAAGTAATCATGCATCTCGCCTGGGCAAGCATACTTTTCCTCCACC
TCCTCCGAAAGTGAAAAATCGTCGTTCAGAAAATTCATTGCATGCTCAAACGGAGTTGTCTGTAGA
AGAGTCAACATTGCGTCCCCCACAAGGAAGCTGGACGACCTACAAAGCGAGGCGTGGGCCTGCCCC
ATTTCTCGAGATATGCCTCAAGACTATTGACGCTACCAAAGATAGCATGGCTTGCCCTTCGCTAAC
ATTAGAAGGTTATCGAAGTTGGACATGATTTCCATAGACGGTTGGTGAATCGTGGCTACCACGATC
ATTCCCTCTGCCAAAGCTAGTTTTCGTATAGCAGCAATGACCTCCCTAGCGGCTGTACTATCGAGC
TACATAATTTTCTAAGTCTCGGCGAGTAAGGTGAGCTGGTCACTTCGCTTACCCCGGAGGTGACCT
CATCGAGGAATAGGATTTGGCTGGTGGGTGCTTCAATGTATGGGGCCAGAAGTCAACGAGGAAGAG
GGCTCACGGATATGTCACCATCGCACATGCTGCAGTGACGCGGCGCTTTTGACCTATACGTGCGTG
GATACAAAGTTGATGGAAGGACATCAAGGTATGAAGATACTATACATACCGTCTGATATTCCTCTG
CCAAACGCAGTACGGATTCGTTGGCCGGCGCAAGATTGCAAGCCCAGATCACTGACCACTCTGTCT
ACCCGTTCTATGACCTCTTTACGCCCCACGTTGGGAGAACTATGTCATCTGTCAATCAGTACCGTA
GCATCTGTAGCGAAAGACGTACTGCAGACGAAGACTGTACACGATCGATTCCCTGACAGTCAGAGT
TCCCAGAAGAGCGTCCTCCTGCTCCACAAATGCCGAAAGCTTGTGCATACTACGTGCGTTGAGGCG
CTGCCTATTGAGTNGGATTTAGTCCAGAGGATTATTAATGTCGCGGGCCCAGAAGGATGGTGTTGC
CCATACCGAGTGTCCGGGAGCTGAGGGCTTCTGAAGGTCATCAGATCGGAGAAGTGTGCTTTTTCC
TGGGTAAGTATCAGATAAGCCATTGNTTCGATTGACGAAATACAACGAGCATCCATACCCGCGCCA
GAGGGACCCTATAGACAAGAAAGAGGTTAGCGATGCATACACATAAACAAATGTAGCCTACCATAA
TCGCAAGAAGTTCACCAGCGCGGATGGTGACAGAGACATTGTCCACCAAAATGCGTGATTGATTTT
TCTTAGCAGGCACTTCGTAGGTCAATCCTTCATATGATAGCAAGGATCCGGTCATGGTGGCTCAGA
GCTAATCTCGTTGGGCCGACAACGACGTTCAACAGGGGTCAAAAGTTGGGTGGCCGCGTCCAGAGA
ACCCGAAAGTCAAAGATGGAAAGAAGGGGCCAGTGTTTGCTTTATCTGGGCCAGCTTTCCCTTTCT
GATTTGTTTCTGACAGCCATTAGAAGATAAGGGCGGAATTATCAACGAGCGTCTAGCCCTCTCGGG
AATGATGAAGACTTCATCTTAGTTGACTGATAGCTTGCATGGTGTTTCTGAAGGGTCCAATTGGTT
TTAGAAGGGAGCGCAGTAGATGGCGGTCGGTATGGAAGGGGCACCACAAAAGGCAAGACCCTTGCC
TGGAGCGGCCGATTGGTCCGATGGTTCCGATGCTGGCGTCAACCAATGTCAAGTGTTTGCCCCTTA
CCACCATCGTCGTCTGTGGCGGGACTATCGATCCGAATTCTAGGCTCTGACAGTCAGAAAATGGGG
CTTAAATGACTCCCAAGAAATTTCTATTCGCTTATACTGGTTAATAAACTCACCTCTGGAGATCTG
GATCGTAGAGATATCCTTGCCACTCCAGACTTCGTGAAAGCCTGAACACCGCAAAGTGCACCTCTT
TAAACGTTAGAAGGCGATCTGGCACCGGGCTGAGCTTTATCTACCTTCCCCCTCCTTGAAATTAGG
GCTTTGACTGGCGCATCGTGTCTGAGTTCACTACTCGGAAGGACGTGGAAACTCCTCAAGACACAA
GTCTCTGGAAATTTGAAATCGCTAGAACCATGGAGTTTAGTCTGACATCTTGCATGCTCAAGATGT
GACGTCTGATCTTCTTTTGACAATTATGAACATTTCGGCGGTCATTCCCTAAGCCATTGTCAACGC
AAAGACACATACTAGTATAAGAAGCTCCTCTCTTTATCCAGACGGAATCAACCCCTATCCGTGAAA
CCACTCATGTCAATCTACGACGACATCCTCAGTAAACTTGGCGGCGATCAGCCCTGGTCCTTTCAG
AACGAAAAGGTAAGGCCCTCCAGAGTTGGTTCCCAGTCCTAAATCCAACGTCATTACAAGGCTATT
TTGGAGCCATATACATACACAGTCTCTGCCCCAGGGAAAGACATACGTCGGAAGCTCATCGACTCC
TTTAATCTCTGGCTCAACGTTCCAGATGACAAGTTGGATATCATAGTCAACATCGTGAACATGTTA
CACAACGCCAGCTTGCTGTAAGCTTCGCCGGGTGGTTCAGTGATCACTTACTGACTACTTCCGCTC
ACAGTCTTGACGACATCGAAGACGACTCTCAGCTGCGGAGGGGCCAGCCAGGTAATCACTTCACTA
TTTGACGAACGTTTTCCTCGCTGATCAAATTTCTTTCAGTTGCACACAAAATCTACGGTGTTCCCC
AGACCATTAACGCGGCCAACTACGTTTGGCTGAAATCGTATCTAGAGGTCGCGAGATTACACCAAC
TCGACCCCTCTAATGCCACAGATTTTGACGCCATTGTTACCGGTGTGTACCGAACTCGCAAACGAG
ATTCGATATCCAATATTGATGGCGAAATCAGAGGAGCTCATCTGGCTCCACCGTGGTCAAGGCCTC
GAAATCTTCTGGAGAGATTCCCTCATCTGTCCCTCCGAGGAAGAATACATTTCGATGGTCAACAAT
AGTGAGTTAACCTGGGGCCAGTCTAAGCCTACGGTTAGACTCATTTCGTATCAGAAACCAGTGGCC
TCCTACGTATCGGTATTAAACTCATGATGGCATGTGCCACCACAAACGCGGATATGTATGTGTTTT

TGGCAGTGCATTTTGGGGTTTGAACTGAATGGCCATTTTCTAGCGATTATATTCCTTTGACAAACC
TCATTGGAGTCTATTTCCAAATCCGGGATGACCTTATGAACCTCGACAGCACTGAGGTTCGGGAAC
CTTTTCCGCTGTCGAAGCCCCAACTAACTTACTCCATTTCTAGTACGCACAAAACAAGGGCTTTGC
TGAGGACCTCACTGAAGGAAAGTTCTCATTCCCAATCGTACACGGCATCCACGCCGACACTTCGAA
CCGCCACTTATACAGTTAGTACCTTCATTAATCAGCTCTACACCATCCTGACCGCCCAACCTCGCC
CTTGAAAGATATAATCCAGAAGAGGCCCGAGAACTCAATGCTCAAGGTCCACGCCATCGAATACTT
GCGCAACCGAACAAAGTCCTTTGACTATACCCAAACGGTTTTGGATTCTTTGGAGGTACAGACTTT
GGACGAGATTAGACGACTTGGGGGTAATCGGGGGCTCGATCAGATCCTCCGTTATCTTCATGTCGA
AACACAGAAACCGGAGTCTCCAATCTAAATGTCACTGTATTTTTGTAATTTTACCAAGTATCTCAA
TCACGTTTCTCGCCTCGTACCACGAGCCACCTAAGGACTGTGTGGGTAAATTCTAAGATTAAACAT
AGAGCGGACCAAAAGTCATACCACAGAAAGAGAAAATGTGGACATTTAAGGGATACAATTTCGCAA
AGGGTAAATACAGCTACAGTCTAGCCTCTCTCGTCCTGCGAGGGTTTACTGCGTCGTTCTCTTTTG
CACAATACTTCGCGTACGGCTCGGCATCTTCTATTGGAGCTCGAGTCCCTCTGTGAGAAGGACTGA
ACTCTATAGCTTTGCCAATGCACATCGTCCCAGGAAGGAAATGTTCATTCGTTCTGGGGCTGGCGC
GAGCAAGTACAGTAGGGCCTTGTTATCAGCCGCCTGAAGCTGCAGCAAGTCAAGGATGATGGTCAA
TGCAGGCTGTCGACGATATTTCAAGTAGTATCTTGGAATCGGAAACTTGTGGAGCTTACATTATCT
CGGAAGGCCTTCAGTTTGGCAGATGGAAGCAGTTTAGCCCCTGGTCCATAAAATCCCAGCAACACG
TCATCGGATGCGCCTTTGGCAAGACATATCGGACGCTATGGGTGATCGTCTGCCATGGTACTGGTT
TTGGGTGCGTGAAGTCAAGAACGGGAGGAGTACTCCTCCTGTCAACGGCGATATCTCTGATGCATT
TCGCTGTGAGATCTGTCGGTAGGAAAGAGAAACGCTGGATAGAAAATGTCAACTCTTGAGTTGGAC
AAACAGGCTTTCAGCTAAGAAAAGCGTACTTGCTTCACTGAACATCGACTGTTCAAGCCCCAGACC
TGTGTGATCCAGTTTTCCATCAATGAAAGTAACCTTGGAGGATAGCAGGATGTCGAAGGGCAGGCC
CTTCTCTTGAGACTGTCGCTTGACACGTTCTAGGGATCTTCCTTCATTAGAACACGATTCAAGGCA
ATCTTGCTCCGGTCGACCGACAAGTATGGCAAGTCTGACCCAACGTTCCCTGGAGTTACCAGTCCA
CAGAAGCCACGTCGGGAAACTGGGAGAAAACTATATCTCTCAAGAAATGAGATCCCGCCTCCTAGC
TTGTGCTTCGCAATATCAGAAGGTCCTTCCCCCATCATTACAACTGCAAGATTTCGAAATGGGTGT
AATGCGATCCGATCATACGACCCAACCGGACCAACGCCTTTATGTTTTGGTCTTCAGGATTCTTTG
CAACCGAACTGCCAGGAAACGAGGCCACAGATTTGCCGGTTAGTTGTTCGGAGTATTCCCGTTTGT
TATCTGTGGTACCCAGCTGTATTAAGTGGTTCCCGGGACCCGATCTACCTCTGAGTCTGAGCCTCG
AGCCAGCCTGAGGAGGTGCGACTGATGGCGTTGAGAAAGAGTCCAGTGTTACACCTTGTGTCACGT
CCGCCTCACTGGGCTGAGCTCTCCGCACCAGCTTTCATGTCACCAGACCATTGTCGACCAGAACCT
TTTGTCGAACGCTCAGTTTGATCATGAGCTACTGAAGATGAAATTCTGTCACGACTCGACTTTTCG
GTTATCTCAGGAAGGGGGAACAGAGGCTCTTTGGGTCGCTCCGTGACACGGAAGCGAAGACGCAAC
CTCGTTACCCTTGAGCGCCATCCATATCTTCTCCGCGTGCCGTTCTAAGGCGGCTTTCTTGGGTCA
TCTCCGAACAGCGTGGCCCCTTCGCCAACATATTTAGCACTGAGACCAGCGAACAATAATCGAGCC
ACGGTCTCAGACTCATCCATGGATACACTGAAGCAGATTCACCCAGGCTTCTTCACCTTCTTTGGC
TGTTACTTCCTTTGCGCTGCTGGTATTGCCACCCGATCGCGGATGCTCCGTCCTGCATTCCTTATA
CCCATCGCAGGCGCGATATACGCCTCACAAACGCACAGCCCGAACCAGTTAGCTGCAGGATTCAAT
AATGTCCTTTTCGCAACTGTAGTGGACTATCTCCTGCTCTCCGATGTACAGCGTGATGTTTGGCGG
ATACATCTAGGAGCAAAGGGCGCGGACAAGCCCATCACGGACCAATCGTTCATGAAAAGGCTCGAA
TGGGCATTCCACCTCGGTCTCTCTAATCGCGGTGTCGGCTGGAAGCATGGCAATTGGACTGCTGCA
GACCCCTTGCTTTTCCGAGGGATGGAGTCACGCAGAGCTTTCCTCATCTCCAAAACGAAGTCGCTC
GTTCTCGGCCTCATCACCTTGTACACATCCTCCCTTCTAATCGGGGCAGACCCACGGTTCCAGAAG
GACGCGCCACCGTTCTCACAGCAGACAATGCTCCAGCGTTCCTGGTCGAATCTGGTGTTTGCTGCA
CAGTTTATGACTTTCTTGTCCATTTGCGGCAACGTCGGCCCTGGGTTTCTTGCAGTCGCAGTGGGA
TGGTCGGAACCTCGCAGCTGGCCGCCGATGTTTGGCAGCTTCAGTCATGCATATACTGTTCGTCGA
TTCTTGAGGCAAGTGTTCCGAAGTTTAACGAGTTCCTGCCCACTGACTTACTGCTCGTCTGGATCA
GATACACATGGGATCAGATGTTGCGTAGGGTAACCAGAACTCTTCACTTTCTATGTCTTCTGAAGT
TCGCAAGTTCTGACGAACTGTATCTATTGGCTAGATCATAACATCGTACAGCAAATTCTGCACCAA
GCGAATCTTCAATATTCAACCAGGCACCAAGTTTGCCCGGCGCATCAACGTCCTGCTTGGATTCCT
CTTCTCCGGCGCTCTCCACCTCTGCACGACAACTTCCATCCGACCTATGCCGTTTTTCCTCCTCCA
AACCACCGCGGTCCTTGCTGAAGACTACGTCTTAGAAAAGCTAAGAGATACCCCCAGTCGCTCATC
TGTTCAGTCTGACTGCTTGGCGGTGGATAGGTTATGTGTGGGTTTGGTGCTGGTTCGCATGGTCAT
TCTCATCGTCGCCGGACTGGTGGGCAAAATCGGGGTATGCAGGTGTGCCTATGTACCTGTTCAATC

AGCTATATACCGCATAGTTGTCTTTATTTGATTGAACACCCGTGCTGCCTTCATGAATCTACTTCG
GCCCGTAGAGAACGTCGTATCGTAGTCGATGCAGTTGAGAAATTAGATTCCTGGTTCATCACTGGC
ACATCAACGAGATTTGAATCATATATTCACAAGGTGTAATTACAGTTACAGTCTAGCCTCTCCCTC
CCTGCCATTGTAGGCCTCAGGTTTTGGGAAGAATCCAATCTTTCGCCAATACTTCGCATAGAGCTC
CGCATCTTGAGTCGAGATCGCTCGTGCTTCTCGGAGTGAAGGGCTGAACTCTACAGCATTGTCAAT
GCACATCGTACCGTGAGGGAAGTGTTCGTTCGTTTGGGGACCAGCACGGGCGAGCAAGGCCTTGTT
CTCAGTAGCGTGAAGCTGAAGCAAATCAACGACGGTCAATACAGGCTGTAAACGATATACTTAGCA
ACACCGTGGAACCGGAAAGTCCTGGAGCTCACATTATCTCGGAAGTCCTTCGGTGCAGCGGATGAA
AGCAGTTCGGCCACCTGGCGCATAAAGTCCCAGCCATTCATCGTTGGAAACGCCTTGTCCTTCGGC
ATAACATATCGGAGGCTACGAGTGAATACCTGCCATGGTACAGGTCGCGGGTGCGTGAAATTGAGA
ACTGGAGGCGTGCTCCTCCTGTCAATAGCGAT

Pleur345 region: 7809 bp

AGGGGACCATATTGGGCGAAGGTTGACCTTGAGTTGGGCTGTAATATGTTCTTGGTTAGCATGCTT
TCGTAGAATATCGTAGAAAATACCAACCAATGAGCCACCGTATGCTAACATCTGCACCAAATTCAT
CGTCGTCCTTCAGCATGGGACCCAATCTCCCGATCCTCAACACCCTTTCCCGGACTCCTTGGACGT
CCAGCGGCACTTCTTCTCCCTGCAGACATCGCTTAACGACATCCGATATTCCAGGTGGTGACTGCA
CTAGCGAAGAGCTGAGCAACCGTAGAGTGTTTAGCCGTAGCGCACGGGAGTGCGAGAGTAATGACG
TTTGTGAATGAGGATAAATCTCTTCTAGCGGGACTGGCGTTCGTTTCTTCGCAGGGCTGTTATGAA
CAATTCGTGAGTATTGTGATTCTTGAAGCAACAAAAATGGGTTTACCTTGCCTGAGCAAGCGCATA
TAGTCCACCTAATACCTCACTTTCTCCACCCCACCTAGCCATACACTTTCTCGTCCAACCCGTCAA
GTCAACTTTATCTTCCCATTGCTTGGCATCCATCTTCTCCAGACCTCGCATGCAGGCGCCGACGAC
CCACGCGGACTGTCCTTCGCTGAGATTTTCCATATCCTGCGGTGATTCAATAGTCCGGTTGATGAT
GTTGATGATAGGCTGAGCAACCCCAGACGAAAACATAGCCGAAATCGTAAGCAAGTCGTCAAGTTC
CGTTTCCTGTTGACCATCATCAGCACCTCTTCCTAACAGCAAAATTGAATCCCAAACCAACCTTCG
CAACATCCCCACCCTCCCATCCTTCCAACCTTGCCTTAACCCAAGCCCCGACCCTCTCCCTCCAAA
CTAAATCCGCCTCCTTCTCATTAACCTTCTTGCACCTCCTGACTTCAGATAAGAACCTCGCCATCC
TCGCGGGATCCATATCCAACATCTCTTGAGACGAGGTCACCTTCAGTAGCTGAGGTAGGGCAACGA
GTTTCCAGCCGCCCCAGTTGAGCTCAGCCAGAGCCGTGTGAATGCGAGTGCGAAGGAGNAACCATC
CTTGGCACCTGCCTCTAATGAGGACGTAGAAGGAGACGCGCATTGTGCCCATGTGCGTTGAAGGAA
CTTGTATCCTGGTCCAAGCCATAGCTGCATTTCTGAGGCGGTGAAGATGGAAGTTGTGAACCGGAG
GAGGGAGGCTTGCAGGGAAGACGCGATTGGAATTTTTGAACATTCCGTGAATATTCGAGCTAATTG
AGATTCTGTCGAAGCACGGTTAAGAATGTTCAAACACAGAACAGGAAGAGAAATTGACTTACGACT
CATTCGAGAACCCTGCCGCACAGCACAGGGGATAGCAAGGATTTCAAGCCATCGAAGAAGCTTGTC
TAACTCTGCTTCATTCATGCCTACTAATGATTTTGGGCATGCGATGAACTCCTTGATGAGTAAATC
CGCCACCGGGGCGAACTGCTCAGCCTTTTTGACATGATGTATTAGCGCCGTAGTCGTTCGTCGGAG
GAGTGTAAACGACGGTTTTGAGTTCTCAGCATTGATATGGTGCCGGAGTATAGCCAGAAGAATGGA
TCCGGTGGCGGTATGCAGAGTTTGAGAGACAGACTATTCATTGCAAAGAGTCAATATAAATGAATA
CCCGGAGCAACATGGAGAGGAATGATGGTCACCTTGCAAGCCGAAACAATAGCCCATGCGCTTGCA
TCTTCAATACCCTCCAGTTCTGAAGCAAGCAATGTTACGGCATTATCTCGCATAGAGGTCTTGAAT
CTCCGTAGAATGTTGCCCCAAACTTCTGCTGTGGCTCTCTGGATCTCTGGCAAACATTTTGGTAAA
GTTTCCCGTAAGGCAGTCCATGTTTGCTTCAGCAGCTCATCACTCGTGGAAGGGATGAGCAGATAC
TTGAATAGCGCTCCAAGAGTGTCAAGGAGGGCCTGGAGAGCCGGTGGGCTGATTGCACGTGGGAGG
AGTGCGAGAAGTCGGGTCACCAGGGAGGAATATATGGGCGAGAGTGTCGTTCGGAGGTCGTGTGCT
ATCTTTTGGAGTAGACCTATAGAATAAATTGTTAGCATTGTCCGATGATGCATTAGCAGGGAACAT
ACTCGAGTAAAGCTCTCAAGCATTCATCGTCTGCAACTTCGAATGCAGCTATCCATAGATCCGTGA
TTTCCTTCCAGTTGTGAAGCAACAAGGGCATGGACGCGGACAATGGGTCTGCTTTCTTCGCAAAAG
AGATGAAACTGGGAGCGAGATTGAGCTGCCTCCAGTGGTCCAGTGCTTGGTGAAAATTCGAATCAT
TTTCCTGCAGATAGTTCGCAGCCTTAGAAACATTAATGTGGCAGAGTTAACCGGCTAAGCTCACCT
CGAGTTCAACGTCGAGTCGCTGCTGGTTGTGTGCAGCAGTTGGATGCACATCTTTCAGGATTTCGT
TATAAGATTGGTGCTAAATGAACGCCCATCAGTTCCTTGCATGGCTCGCCTGGAATTCAGGTTACA
CACCTTGAACCGCTTGACCCTGGGCTGGTCTTCATGGTCAACGTCCATGTCGTCCATGGTGCCGTC
GTCGCTGCTTCCAACCTTCCTGAAGGCAATAGGAGCGTCGTTTTCTCGAGGAACCACGTTCAGCTA

CCAAATTTGGCGACCAAACCATGCCAAGTCCGAAATTTTGGCCCTTTGTTTGGATCTCATGGCGTT
GCATAACAAACCGCTTCCTTGTCCACGGGCCGAGATTACCGTAAGCGCTGGCACCGCCTTTGATAC
TGAATGACAGCAGTTGGACTGAGCCGCAGCTGTGGTCTGTCTTGACCTCTTTCCATCGTTTTCTTG
TCTACGCTACGGTCATCGCGGGACCTACAAAACGACGAATCTATAAGAAGCTCCCTCTGCCGCTTC
TCCCCGTAACCAAAGGCTTATCACGGCTGAACACTAGTGGCTCCAAGCGCCCAAAGTTACTCAGAT
CCTATACCATATTCTGCTATTGCCACCATTTCAATCATATCTCCACCATCTTTGGTCACTGAACTT
GTGGATCACCTAAGTACCTTGACGCCCACTTGAGGCGCTACCACATCGGCTAAGCGAGAGGGCATA
TTAAATTTGATATTTGTGGGATTTTACGGGATTAATCTGAGACCGGGCTCATAATCATGAACACCA
CACACGCCTATCCTTCTGGAGAAGAGGAAAGGTCGCTCAGTGGGTATATATGTCCCTGTCATTCTC
ATGTTCATGTTACTCATACCCTCTTCATCTTCATCTCTAGACTCCAGATCCATACGACCGTAAACC
AAGAATGTCTCGCTACGACAATATCCTTGAGAAACTCCACAAGGAGGAGTATACTTGGTCGAGGCG
GAATGAAGCTTCCATCCTTGAGCCGTTCACCTACATCACCTCCAACCCCGGCAAAGACATACGCAG
CAAGCTCATTGACGCGTTCAATGTCTGGTTGAATGTCCCTGCTGATAAGCTGATCGTCATCACAAA
GATCGTCAACATGCTTCACGCTGCCAGCTTGCTGTACGTTCACAATTCCTCTCAATGCACCTTCTC
GTACTCACCCCAACCCTCTTCACAGCGTCGACGACATAGAAGACGACTCCCAACTCCGCCGCGGCC
AGCCCGTCGCGCACAAGATCTACGGTATCCCGCAAACCATCAACACCGCCAACTACGTCTTCTTCC
TCGCCTTCCAAGAGCTCTTCAAACTCCGTGACGACGAGCACGCCCTCAACCGCGAATACCGTCCTC
CCTCCCCGGGATCAGTCCGCACACCCGATCTCGACGCGATCGTCAATGAGGAGCTTTTGTCGCTGC
ATCGCGGGCAGGGTCTGGAATTACTTTGGCGTGATTCGTTGACGTGCCCTTCGGAGGAGGAGTATA
TCTCCATGGTTAACAATAGTTAGTGGCTTTTTCATTACTTGAGATGAGGACTCGATACTCATGATC
ATTTCGTTGAAATTTTCTAGAGACGGGCGGTCTTCTACGCATCGGCATCAAACTCATGATGGCATG
CGCGACCACCAACACTGGAGTGTACGTATTCCATCCTTCTCATTGTCCACTCTGGTCTTTCTGACA
TTTTCTATTTCTTCCAGTGACTTTATACCCTTAGTCAACCTCATCGGCATCTACTTCCAAATCAGA
GACGATCTGATGAACTTGCAGAGTACAGAAGTCGGTGTACCCTTATCTTCTAATACAAACACCTAC
TGACCCCGACTATCCATTCAGTATGAGAAGAACAAAGGCTTCGCAGAAGATCTAACCGAAGGAAAA
TTCTCGTTCCCAATCGTTCACGGCGTCCAAGCTGATACCTCGAACCGGCACATCCTTAGTAAGTTC
TTCTCTCCCCTTCCGAGTCCGCCAGAGGCCGCTGTCCTCTCCACCTTTTCACCCTTTCTGGTACTT
TACCCCGAGATGAACGTATCTAACCCTTCTCTCCCAGACGTCCTCCAGAAACGCCCAAGCACGCCA
ACGCTCAAGATCCACACTATCGAGTACCTCCGCAACCGAACCAAGTCCTTCGAGTACACACAAGGA
GTTATGGACTCGCTCGAATCACAGACGTTGAAAGAGATTGAGCGGCTTGGGGGTAACCCTGGGTTG
GATAAGATTATGACTCTGCTGCACGTGGATCCGCCGAAGATGGAGGAGTAAGGGTAGAGCATTTGG
CGATGAGACCGTGATGAGAAAGATTAGTCTAGGATATGCTTATCTCGTTTTTTTCTGGAACAGTGA
ATTATTCTTGGACAATACAATCAACGTATCTATCTATCTGTAACTCTATGATGTCGCACAATCTTT
GAAGTCTCTTGCGGTTCTCTCTACCCACAATATACAGTCCGATACTACCATGCCCCGTCTTTGCCT
TGTCCGAGTCGACCCAACCTCCTTAGATCCCCGCATCACCACCCTCTGAACAAGCCGAAAACCGCA
CTTCGGCCTGCCTCGGTATCTTTGCACACTAGCAAACGTCTATCAGACTTTCCGGCGGCCTCATGA
AGCATTGTTCGGGAGCGGAGGAGGGTATATGGAAGCACCATGATGTCATATCTAATCTTGGCGAGC
GCTCCCCTTTCATGCCCGCGCCTACCATTGCGTCCGTGAAGATAGAGCAAAGATTACCGGAGGCGA
AGAGGGACCAATCATGATCGTACTGTCGATCCACGACCTGGGCTTGGACGGACGTTTCTTTTGATG
AAGTTCAGCCGGGTGGGGAAAAATGATACAGTACAATTGATTTACGCTGTTGGGGTATATGGCTAA
TATACCTGCAAACTAGAATTACAGAAACTATTGTTACTTTTATGTTTGGGGGGTTAATGCGAAGAA
TGAAGTAATGTACAGCAAGCGTCCATAGTCGCATCTTTCTCCTCAATCAACGACAACCAACCCCTC
AAAGCTCTTCCTCCCCACCGCCTTCTTCATAACATTAACTGGATTAAACTCGTCCTCTTTCTTCTT
CTCAACAACCACCGTCTCCCCATCCTCACCAATCCCCTCCCACTCCTCCTCCTCTACCACCTCCTC
ACCCATAAACACCTTCAACATCGCCGGACTCAACCCATTCATCATCGCCACCCCCTTCCTATCCGC
CTTCGCCTGCACAGTCCCATGATTCGCCAAATCCCAGAAAACCAACTCCGGAACATCATACCCCGC
CTCCTTATACACCTTCTCAATCGCATCATAATTCGTCTCCCACCCACTCTGCAACTGCGCCTGCAG
AACCTCAGCCTGGTCTGGCTGTCGTGCGTAGTACCCCTGAGCCACGGGCGAAGAGGCGTTCGCAGA
ATCGAACTGCATATCGGAGAAGACGAAGAGGCGCTTGATCATATCCTCGTTCTTGATGTTGTTCTT
CACCGCCAGTGGGAGCAGCAAATCCACAAACACCGAATGGAAATTCGTCGTCATACCCCAACTCGT
CTTCACCATCTCATTCGCCACAACATGCAGAGGTTTCTCCAGATCCAACTTGATATACTCCGGCTG
CGCCGAGAACGTGATGAATCCATTATTGAACGGCGGCTTCGCAAGGTTGGAGAGCACCAAAGACAA
AGCGACAGCGGGATAAATGGGCTCCACATCGCTCAGCCGCTTCGGCATCTTGGTCGAAGGTTCAAT
TCCACCCATCGAGCCAGAAACGTCGCATACAGCGAGGCAATTATCCAGTGCCCCGCTCTCCTTCAA

CCTCCCAATCAACGTCTTCCACTGCCCCTCAATAACCCTCAACGTCGGCTCCGCCCTCGCCTTCGC
CAACTTTTTCTTCACATCCCCAATCTTCAACTTCCCTCCCTCCTTCACCTCCGGCTCCTTGTAGTC
GACAAGTGCGCCAGCACTGACAGCCTCAGCAAGCAACTCGTGCGGCCAAAGCGTAGCCCCTGAAAT
CGTCTTCTTCCCCGACTCGACAGCGGTGAGGTACTTCTCGAACCCCTCCGGATCGTGCTCGAAGAA
CTTGTCCTTGTTGATCTGCATACATTTGGAGCTGACGCGGTTGTACTTGATCTCCTTCCACTTCTT
CCCAGACATCATCGGTTCTGGGACCAGGGCGGCTTTACGGAGAGGTGTCAGGACCCAGCGTTGGTA
GAAGGAGCGGAGGATAGAGACGGAAGCGTGGGGGTCGATATCGGGCTTGTTGAGTGAGGTGGGGTA
ATGAGAGGGGGTTTGGGCTTTCCGAAGAAGCAGAGCAATGGCGGTGGCGATATTCGTGACTTTATC
GTGAGCGCAGAGAGGGGTAGGAGCCCATTTCCCAACGAGGGTGAGTTGAGGAGATTTCTTGGTCTT
CCCGTCGACGTCATCGAAGTCTTCAGCTCATTGAACTCCTCAAGCGCGCATATATCCTTGGAAAGT
GCCTCGGAGAATAGTCGAGCAATGGCAATGTACAGGGCGCGGTAAGCGGGCTCTTTGAGCTTCGCC
TCAAGGTTGTCGTAGTTCGCCTTGTGGTTCTCCTGTCGAGCCTTTGCAGCTGCCTCAGCCCTTTCC
TTATCCTTCTCAAGAGAAATACGAATGTTCTCCTGCGGGTCCGTTCCCTTCTTCGCCGCTTTTCGC
TGCCGTCTGTTGGGGTAGGTGTATTTCTCTTTCTCGGAGTGGAGGAATTTGGGGGCGGCAACAGGA
TCGCCTAGTTGGCCGAGGGTAACGAGCGCTAGGATGTTCAGGAGGTCCTTCCAGTATCCATGAGAT
AGTCCTACAGCCTCCTCGCCTTCAACGGCCTTGCCTTTAGTGCAGACGGGCGAGATGAGTAGTGGG
AGGTTCGTGATGGCCGTTCGTGGATGCTTCTCGAGAAGCCAGCCGAAAGCGCTAGAAAAAGCACGC
ATATGATCAGTATCTGTGGCCGAGCCAGGGATTGGTTATGACTTACTTGTAGAAAAGCTCTTTATC
ACCCTTCCCATCATGAATACTCCGCGTATTCCATACGATCCTCAGCGTGAGCTCTGGATCCTCCTC
CCAAGCCCTATCCAAGCATCCATAGAAGTCATCCTTCGTGCGCTTGCTCTCCAGAAGCTGGAAAGC
ATCGAGTGTGGCAGATCCAGTGGTCGAATACGCCTGAGCGCCTTTGGCAGTATACATCTTATCTTC
CGTATCCGACTCCAGTTTATTAAGAGCATCGATCATAGGGTTAGATGGGGTATTCACGGCAGTGCT
GGTTGCCACTGTAGCATCGTCCGCCATGTCCACATCTTTCTCAGTCTTCGGTGGGAGCAGGACATC
GAGAAAGTTGAGATCGTAGAGCTCAGGATCTCGGGGAGCGTGATGATCTGTGAAGTGAAAGACATC
CTGGGTAATCTTATGCCGTCT

Sord224 region: 9518 bp

CTGCAAGTTTTCAAACCGACAGAGCAACCTTCCCCATTCTATATGCAAAAAGTTTCCCAAGTCCGA
GTTGTGGATATCGAGAGAGACAAGCTTCAGCAGCAATGTGGGACTCAAGCTGGCACGAGGATCTGC
CAAATGGATCAATAGCCTCAAGATCAGCACTCCCACTGCGGTTGGGAGTTCGATGGATGACTGGTC
GCGGCCACTACTTATCACGAGAGAGAAACCATTGATCATGTTGAGATATTGATGAGTGGTACTTTC
AGCTATGCCTTTCTGAGGAGCCACAAGACGAGTTAACATGTCACTAAAGGTGTTGAGCTCTTCCAT
AAATACCTTTGGATCCATAAGCTCTCTGGTGAAACAAAATGTCAATGAGTTGGGACAAGTAGAGAT
CAATGCTTTCTGCTTTAACGATACTGGGCTGCATGAAACCACCAGGGAAAGCAGGTAAATGTCGAT
ATAGTTTCGCTCCTGTCCAACATGTTTGGGCATATATGTCTCGGTAGAGTCCACCACTAGTGTCTT
CAGATAAAGCAGGCTGTACGCAAGGATAATAAGTGAATGGTAACGATCCCCATCATCTGTGTGCTC
AAGATCAGAAGCTAGTTTATTAATCTCAACAGTAGTGAGAAACGAAAATATGCGGAGAACAATTTC
TGGGGGAAGTCGAAGAAGCAAGTTGGCGAGATCCACCATAGTTGTAAACCAAGCCGAATTTTGGAT
CAGATAGCACCTGGTGCACAGATGTAGATACCAACGTACTTGGTGGTTATTCTTCCATTTCTTCTT
CGACACCGTCGATTTCTATCTCCATCATTTCCTCATCACTATCAAAGCCCTCGAAAGTGTCCCAGT
AGTTATAGCCTAGATCCTTATGCAACATTGCCTCTTCTTTATCTGTTCGTGCTTGTTGGCGTCTCT
TGAGCTGCAGTTGTTCGAGTCGGTTTTCTTGGTCCAAAATGACGCTAGATAGTTCAATGTTATTGA
ACGAGTCTGGCTCAATAGGTCCATGATTTGGTGGTAAAGTCGGAATGTTTTAAATTGTTTGCGGCC
CATGCGCACGGTTGTAGCGACGCCAAAATCGATTGAATAGTCGTAAAGCCCATCTCTGAAAGTTAT
ATACTACTAGGATGCGCCAGTCATAGTACTTAATAAGTGACATACGTGCTTTTTGTGCATCCATGT
ATTGCAATGCTAATTTTTTCACAATGTATGAATTGATATCATCTGTGGATGGAGCATGATCTAGAC
CTGGATTAACTTCTAGCTCTTCTTGATCGTCAAGAGACAACGAGCTAAACAAGCTGTCCAAGCGGA
TTCGTTTGGAACTGTCGTCATAGCTTGAAAGTCTTCGTTTCCTACTGGGCATGAACCATTAAAGAG
AAGCTTCTAGTATTATATAGTAGCGGATCTGCGACCCGGGGTTATATGCAGGTGCGGTCTGAAATG
CAGAGCTGCAAATTTATAATAGTAGTATCTATGCTAGCAAAATTCGTATATGCATGTCTATATAAT
TCGTCTAACTCGTATCCTATAACCCTTCCCGAAGACTGTGGCTGTCTCCGGTGGTTTTGCTTTCTG
ACTTGCTCACACCCAATTCTGTACATAGCTGATCATAGTATGCTTTGAGAGTTCGAAGATCTTGAA
TAGTGAGTGAATCAAAGTCTGTATCCTGGTACTTTGTGAGTTCTCTAATCGGAGCACTGAACACAG

AAGATTCTTCCTGAGATTCTACCGTCGAATCGTTGTGCTCCACCTCTGGAGCTGATTGGGATATCG
CCCCAAACCACCGATCCATTATCTGTATACTGGTGGTTCTGCTACGAGCTCCTTCGGCTCCTTTTA
GGTCGAGCGAAGTCTGGCTTGCATGGTTAGCGATGGTGGTAAAGCTTTGGTTTTGGCTGTGTGACT
GACCAATGCGGGTCATGAACTCGGTAACTCCTGTGGAAAACTTTTGTGTGATGGAGTTCCGCTTAG
GGTCAACATGATCATCCACCAGATTTGCTCGTTCAGAATCGAGTCTAACATCAATGCTATCTGTAT
CTCCCTCGCTAGTAGGGGTTCTGAACTGTGGAACGGACACCAGACTAAATCGTGTGATCAAACTCC
TCATAAGCAGCGAGCCGTCGTCTTCGGTAACCTCCATTGATGGAAAAGTTTCTCTTCTTGTCGGAC
CTGAATTGTTCCCCGAGTTGGACCGCAGACGTGTGGTGATTATAGACCCGGCATTCAAGCGGCTGC
TTATGGCATTTAAAAGTGAGGTAGAACGGCTGCGGTTATTGCGCAAACGGACAGGTGAAGAGCTGA
CGCTTGAACCACTTTCTTGCTTGATTTCGTCAGGGAATTTGAATTGGGTAAGCATATCTTTTTCCA
CGAGCTTGATAATAGAGGTGGACAAAATCTCAGCTTCATTTTGCAATTGCAACTCATCGCCAGTGA
CATTCAGGAGGAAGAACAAGGCTGCCTCGAAATTCGTCAAGTAATACATAAGCTCGCCATTGTTAA
GGTACTCTGCTTGAAGGTGTGCTGCAAGTTTCTTGTCGGAGTCAAAAAAATCAAAGTCTTTGACTG
GATGTTCTGTGTCATCGAGGTTGAGGTTCAAACACTCAAAGAGCAGCTTGCGAGCTCGCTTCTTGT
GGATCACCAGACGGGTCTCGGGTGCATAGTTGGAGAATGAAGAGTGAGTAGTGAAATGTTGAGGAT
CAATGCCAGTAGGGTCCAAAAGATTGAGAAAGTTTTTTATGAACATGAATGTGAGGTAGAGATCAC
AGCCCGATCCATGACATACAAGAAAATGGATGAGTATGGGAAGCAAATGATCAGCAGAGACATCTT
TACTTGTCTTCCCTGCTGAACATTCGGAAAGAAGTGTGATGAGACGCTGTTGCAAGTGCACCACCA
CATTCAACTTTTCACGTGGAGTCACCTTTTCTTTGACCAGTGCAAGTAGTCGCAGGAGCGTTTGTA
TTTCTTCCTGTTTATCCACGAGTGGATAAGGATCACCTTCTGAAAATGCTTCGAATGGAATATTCT
TCATGGCTTCTAACTTGTCGTCAAGAAATTCCTGGAACACTATATCGTGGTGATCAAACCTTTGAA
GCTTCTCATGTAACAGCAGGCCATCCGGCAGGGTGGATGCAGAAAATAGTTCGGTGTGAGCACATT
CGAGTGCCTTAGCCACAGTAAGGTATTTATAGTAATTGAACACTCGGAGTTGCATGAGAGATTCGT
TATCGGTTCGCTTTGAAAGCTTGATGGACGACAGTGAGTAGTTGGTGATGGCTAAAAGGTAATCGT
AGTGATGAGGATGGGCATTGAAATGATCCAGTTGGGCGATAAGACGCTTTTTGCTTGAATTTGACT
GGGTATACACATTGATTGCAAGTCCATTAAGGTCATCGTAAAAGCTATTGAAAAGAACTTTAAGCT
CTTCCAAGGAAAGTGGGGCCTGAAACCGTGGTTCTCGAAGATAACGGACAAACCCCTGAGTTAAGG
CAGAAATCAGGGGATGAAATGTGGTAGTAGTTTCAACTTCTTTCCGAATCGGGTGCAAAGGCGCAA
GTCTGGTGGCTGATTCCGAGTCCGGGACATTCGGATTGGAATGAGGAGTGTCCTCAGAAGTTGTCT
CTTCTGCCATCGAAGCCTCCGACTTCATTTGTGCCACTGTACTACTCAGGGCACCTGGCATGGAAA
CGTGTGAATGAGAGCGGATCAGAGGGATGTGCAGGGATGGCGCGGTGTCTGGGTGCAGAAACTGCT
GCGGTTAAGGAAAGTATGGTAGTATGCAAATATGATGTAGTCTATTGGCAAGATGATCGTAGTTGT
TTTATCGTATCTCGTTCTTTGAGATTTCATGCTCTTTCTCATTCTTGTACTTGTACTTGTACTTGA
TTTTCATTTCCATTAAGCACATGCCGGTCGTTCTCATTATGTTCGATCTGGTCGTTCTGCTCAAAC
TGACTGATCTCGTCAGTCTGGTCGTTCTGATCGTTTTGGTCGTTTTGGTTTTGTTCATTTTCGTCG
TTTTCGTCGCTTTCCGGCATCGCCTCTACCCGATTTAAGATACTCCAGTACCATGGCGGAAGCATG
AATAATCGAAGCAAATCGCTGGTGGTAAGCGGTCCGCGGCCCGCGTACTCCCGCTGGGTCTGTCTG
GTCTCAGACCATGCGATTTTTCGGCGTTCTTCGTTTTTGTCAAAGTAGACATTCGCAAGCAAAAAG
TCTCTGGAGAACCACCAAAAATGGCCCACGGTGTAGTTGACGAATTCGTTCCAGAGGAACGGCATG
TACACCGCCAGGCGGACAGCCGACGCCTCAAGCAGAAGTGAAGCGTAAACCTGATGAGCAAAGTCA
ATTTGGGTGATAGCGTGCATCACAGTTAAGCACCAAGGAACATAAGCTCCTGGAATGCTGAAAAGC
CCAAAAATCACCACCGGAGACTGAGGGTTCTGACGGCACCAGATATACCATATTACAGAATCGAGA
ACTGTGCTCAATTCGCTTACATTGATTCCAAGTTTGTAGTAACCGATGGTGACGACGACAATTATC
AGAGCAGCAATTGATGCCATGTAGTACAAGTAATCTGGGATTTTGTAGGCTTCCACCATTGCCCGT
AGCGTAGCATGTTTTGACTCACTTAATCGGGCCATTCGGCGTGGAGGAAACATGGATAACGGAGTG
GTAAATTTGCTCTCTAGATCCAGCGCCGATCCCATCATATAGTATACTCGAACCAAAAGACTGAAC
GAAAGACCTTCAAAATGACAAAACGCAGTGATGAGACGCCACGGCTGAGACCAGAAAGTTCTTTGA
GGCGTAAATAAAAGTTGCTTGGAACTACACCATCCTAAAGATGTGAGGGTAGCAGATGCAATAATC
AGTAACAACCAGTATCGGGTTACTGGAGGCACTCTGGTGACCAATTGGCCAACGCGGTCCATTTGT
GGAAAATGACGTAGAATTTACTGAGGAATGTTTCGCTAAATCGAAACAATGACGCTTCAAAGTTGC
GTTGCGCGGGCAAAATCCCAAAAGTCAAATAGGCTATTTGGGGCAAATTCGGCACTAGTTACCAAA
CTTGTTGACACCAAACCTTATTTAAGCTTGCTTAGAATTGACAAGAAATAGAGGCAAACAGGCAAA
CAGGGGCAATAATTGGACAAGAATTGGACAGGAATTGGACAAGCACTAGAGGCAACAATTAGAGCC
AACAATAGAAGTCCTATCTTTCAAGGCTAAACAACCACACTCACAGAACAAATTAATGACTATTTG

TCATCCTTGATCAACTTCGTGTTTCGCTTCTCATGGAACTCATCGATACACTTATCTACTCGCAAC
CGCAAATGATTCCTGACGCCATTCGTCTTCCGTACGATTATCTTGCTGCGATTCCAGGCACTAATG
GCGATATCAGAACCCAGTTTATCAAAAGCTTTAACGATCTCTACTTTACCATCGACAACCATGATA
CTATCGCTACCATTAGCGAAGTTATAACTGTGTTCCATAACTCGTCACTCCTAATTGACGACATTG
AAGATGGCTCTGAGTTTCGTCGGGGTCTTCCTGCTGCCCATACAAAGTACGGGGTGGCTCTGACTC
TCAATTGTGGTAATTTGATGTACTTTGTAGCTCTTCAAACAGCTCAACAGAAGCTCCCTCAATACC
ACCCTAGCGACAACAAAAATGATATTGCATTGTCCATACTGAATATTCTCGTTGAAGAGTTGCTCA
ATTTGCATCATGGTCAGGGCCTCGATATACGCTGGCGGGATCATCTCAAGACAACCAAGCTCCCGC
TGATAGAAGAATACATGGAAATGATCATGAACAAAACCGGAGGATTATTTCGGCTTGCCGTTAAAC
TCTTGGGAGCTTTCAGCACCCCACAAGACAAGCCAACTGTGTTGGCCATAGCTAATCTCTTCGGGA
TTGTATTCCAGATCCGTGATGACCTTATGAACTTAGTTGACGACAAATACTCACATATGAAGGGCA
TGAAAGGTGAAGATTTGGTGGAAGGCAAGTTGAGCTTGCCAATTCTCCATTGTTTACACACCCTGA
CAGAGGACTCTCCGGTTAGAACACTCCTCTATCAGATGGATATGAAACAACGCCAGCAAAATCTTC
AGCTTGTGGAAGAAGCAGTGGAGTACATGAAGGTGGAAACCAAATCACTAGAGTATACTCAGAATT
TATTGCGGCTCTACGGAGCCAAAATCATAGAGTTGGTGGGTCCAGACACAAACTCGTTTTTGGTAC
AAATAGTAGACAAGTTGCGAAATTCCATATAGCTTATACCAACACCTTTCCTATGAATCGACACCC
GCACACCAGCGCGTCCTCATCAACTCTTTCGCATCGCATTTCTTCAACATTTTACAATGCTCTACC
TAGTGGGTTTGGGACTCTCCTATGAGACAGATATAACAGTACGTGGCCTCGAGGTGGTCAGAAAAT
GTAAGCGTGTATATTTGGAGGCATACACTTCCATCCTCATGGCTGCAGATCAACAATCGTTGGAAA
AGTTCTATGGTAGAGAAGTGATTTTGGCCGACAGAGAACTTGTAGAAAGTGGTTCTGACGAAATCT
TGAAGGACGCCGATACCGAGGACATTGCATTTATGGTTGTGGGTGATCCTTTTGGAGCAACCACCC
ATTCAGATCTTATAATTCGTGCTAGAGAGCTTGGAATCAAGGTCGAAGCCATCCACAATGCTTCAG
TTATGAACGCAGTTGGTGCTTGTGGATTACAATTATATCAGTTTGGACAAACTGTTTCGCTTGTGT
TTTTCACTGAAACCTGGAAACCGGCCTCGTTTTATGATAAAGTGCTTGAAAACCGTAGAATTGGAC
TCCATACTCTCCTTTTGCTTGATATAAAAGTCAAAGAACAAAGCATAGAGAACATGGCTCGTGGCA
GATTGATTTATGAGCCACCTCGGTATATGGACATTGCCACTGCTGCCAAGCAAATGCTTGAGATTG
AAGAATTGAAGGGACAGAAGGCATATACTCCTGACTCTCCTTGTGTAGCCATCTACAGACTTGGCC
TGCCTACCCAGACGTTCAAAGCAGGAACGTTGGCGGAAATGGTGGAATATGATTCTGGAGAACCAT
TGCATTCATTGGTGTTGATGGGACGCCAAGTTCATGACTTGGAACTTGAATACTTGGACGAATTTG
TCAACGATCGTGAAGCATTCCGTAAGGCGGTGAAGGCGGACCAAGAGTACTTTAAACCCCCTCCAT
ACGTTCCAGAAGCTGAAGATTAATCGGCTTTTTGCTTAACATCTTGCCATACATTCTCGCGGATAT
CCCAAATCATCCAATCGTACCACCCGTAATAGGCAGGGAAAATCGTCACCAAATGCCACAATGAAT
GAGCATCCACCAAGTCATAAAAAATGGGTGCAAAATCAAAAATTTCCAATGCCATGCCACATATCA
CGATGACACATAAGAGTAGCGGATATAGCGAATAAAGCTTAGACGAGCGAAGGAAAAACGCACTGA
GAACAACCGTGTTCCACTTGGTGTAACGTGCAAAGTTCATGGACTTTTCAGGATGATTATGCTCCA
CATCGTAATACCGAATGTACAAATCGAAACACAACATAGCCCAGCATATGTTCTGTAAAATTCCAA
TAAATATATTGGCCCTCATATTGTAGGTGTACAGCCAGTCAGTCACCAAGTGGTACACATGACCAG
AGTACAAGGCCACACACAATGCGGTGAAAGCGGCCAGCCATAATTTCCGATCGGGAAGATATAATC
GATAAACTCGAGCAAACACAGCATGGAAACCAGTCAATACTGTGAGACCAGCCAAAAAGTAGTCTA
ATCTTTCTGTGAGTTCAAAGTCTCGAATGTGGAATATGGTGGAGCATATCCACGCTGCCAGGGTTA
CAACTGACACAATCAAGATATTATGAAAGTACAGTCCATAGTGAGGTGGTGAAGTATCAATTTGGT
GCTTTATTTTCCGCAATCCTTGCAAGTGAACAAGCAAATTACCAATGGAAAATATCATGGAGGCAA
ATTCCTGGATGCCGAAAACTCGTAAAAATGGCCATTTACCGTGAAATTGAAGCACTTCCTCCTTTT
CGTCCACGCGTTGCTGGGTGACCAAACGCTGACATTGATAGTCACAATTCTGTTCACACGTCCAAT
AAAGTAGTTTCAAGTGCTTTGGTAAGGGCAAGGGATCGAAATTCCAGTCGAATCTATTTTTTTCGT
TAAAATTGGCAACAGATCGATGATATCGATTTGGATCCGATTCGAAAAGTCTATAATTACACGTTA
TTGATTCACATCGGTGTCTGCACTCGTCGAATTCGTCCAAGTCGTCACCTGGCGAAGCAATAAATA
AGGGGAGATATACAAGAAATAGTAACCACCTGATCATTTTCAGAGGGTTGTTGAATATAAAAAAGA
GAGTCGCGAATTGAGGATTTTGGTGCACCAACGGGCAAGTAATGGTGCTTGTCCTAAAATGATTGA
GTCGAAAACTTGATGAGTATATGTATAAAAGTGCTCCGTCATGCTTTATTAGCCCTTATTGACCAA
TAGTAGTTGATAGGCTTTAGAGACTGGAGTTGATTTGTGCAATTGTATCATACAGTTTTAGCAAAC
TGACAAACAGCCAACATTCAGTACTAAATGGGTAAAATCACTCCGAATGTAACAGTTTGAATGTGT
ATTCTGAAAATGAAATTTCGTTGGAATTTAAACGTCTATTTCATCAGGCTTTATGCATTTCGCAGC
ACAAAAGTACTCAGACGGCTGCAATTTTAATCGTTGGTAAAAAAAAATAAGATTTCCTAAATCTTC

TTAGCGATTGCAATGATCATCATAGAATCCTGCTTTGCTTTCTGATTTCAGCCATATTAAAATATC
CAATTGGTTATGAATCACTTCCTCAAAATCAATCCCGTCTTAAATTTCTTTATATTTTCCAACCCT
AAATCTCTTCAAACCAACACAGACATTGCTACTACATCACACAAACATTTCTACCATCACACAAGA
CAACCTAGTTTCCTTTCTTAAAACATTTCCTCCTCTATGCTCATGACTCATAGACTGCATGCGACT
CTCGAGTCTCACATTTTTCACCACTACAGTACCAATCAAAATCATGTCTGACTTCTGGATCGACTA
CTTGGACAATCCTACGTTGTCAGTGCTCCCACATGATTTCCTCAAGCCCGTCAATAACTCGTCGGT
TGAAGCCAACTATCAAGTAAATCTTGCCTCCAATTTCTCCTTCACCGCTGGTCTTGCAACATTCGC
TGCACTTATCTACAGATTAACAGGAGATGAAGATGTGGTAATTGCGACAGATACGGAGGATGGCAA
GCCGCTTATTGTGAGATTGGCCATTGCTGAGTCCACTACTTTCAAAGAATTGCGTGCCAAAGTTGA
ACATGAGATCAAGACTGCAACCAGCCATATCAACTACCAAACGTTAGACGAAGTCGCCCGTCGTAT
CAAGGCATCCAAGG

Sord272 region: 9073 bp

GGTAGGCTTCGGCAAGCCACTAGTGCTGCTGCACTGGTCTTTTGGAATAATGTCATGCGCGACAAC
AGCCAATGCTGGTGTCCAAATCGTCCATGAGCGGTTTCAAGCGGTCGAAACGGGCGCCTTGCTGTC
GCCTCGATACTATCCCCCATCACCTGCCGAGCCGCTTCGAAAAAGGCAAAGTGTTTGTGCTGCGAA
TTACCATTCTTGTATGTTATCCCACCAACTGCCGTGAATCATGGATTTACGGATAGTGTAATCAAA
TACTCAAGATCAGATCTAACGATATGCTTCGGTTTCAGGTAACGAGCTTGGAGTTATTGGCGAAGG
CTTTTGCTGTAGTAATGATGAGTAGTACGTCAACGAAATCACCCAATCCGAGCATCCCTTCCTGGC
CTTTCCTGGTCTTATTGCTCGTAAATCCGATTTCAATGTCCATAAAGCCCGGGAGTACAGTTGGCT
TCTAAGATAGGCATACTAACTCCAAGACCAGCTGCATTGTTGACCCAGGAAACTCGAAGAAAGCAG
CATGCTGTAAGATAGGCTCGACATGTGGCTCGAAATGCTCAGAATGGCAGGTAGCATGTGCCTCAA
CAACAACCATCGCATCGGCTGGGACCACCTCGACTTCACTCGTCCCTGCATGTTGCGGGCGCCGAT
GTCCCCAGACTTCCATGTTTGGCTGTCCCGAATCATTCGGGAATGGATGCTGTTCCTACGGCTCAA
CATGTGCCCCAGACAAAAAGTGCATCTCCACGAGAAAGGAGTCTACCACAGAGTCTGCGATCGTCC
CTCTCATCACCCCCTGGCTGTACGACCAGCCAGATCAGCTGCGCCTCGAGTCTCGGCGGCGGTTGC
TGCGCTGTGACCCAGAGCTGCGTCAGACTCGAGAATGGCCCTGGATGTGCCGAGGTCACCATTACC
CCAACCGCGAGCGGTATCACTGCCGTCCCGTCTGACAATAACGAGCTGAGTGCGGGTGCCAAAGCC
GGTATCAGTGTCGGTGTTATTGTCGTAGCTGGCTTGGTGATTGGCGGTGTGACATACATGTGCCTA
CGAAGACGGAGGCGGAGACGGAGTGCGGCTGCTAGTAGCAGTGTCAATTCACGCCCTGGGAACGTT
GTTGGCGCATTGATAGGAGGGAGCGGTGGCGGGCGGGATATGACGGATCTGAATTCCGACTCGGTG
TCGCGCCCGGGCGGTGAATTACCTGGCGTTGCGCAGGATTATTTCGGCCCGGACCCACAGATCGGC
CCGTACTCTGACGCACACCACGATCAGCTATCTCCTGGGACGACTCCGGGTGCGGCGGAAAGAGGA
GGCGTCCCTATCATGCCTCATGGTCCGGGCGATATTGCCGCACCAGTCGAAATTGACTCCAGAGCG
GCCCAGACGAGTCCAGGGACTGGATCGATGGCGCATGCATCCCCCGAGAGTGGGGAAAGGGGTAAT
GGTAATACTTATTCCGATCACAGCGGGCCGAGGTTCCAGCCTCCGTCTAGCCATGTGGCTGTTGAG
GGCACATTTGAGCTCTACGGATCAGACCCCGACCCTCCAAGGGAGAGTGTCACACCATACGTGCCT
TCCCCGAGTGTCAACGATTCACAGGTGTCGCCCTCCGCCGAGACCGTACCGAGTCCAAGTTCCGGA
ATCGATGGTAGTACGGGGAGGAGATAGTGAGATGGATAATACCCGCCTGGCTAAGAGTCCAGTATC
TAATTAGGTGGAATGGTTTAATGTATGGACACGATATAATCTTTTGGGTTGAGTTCTTGAGATGTC
TCTTTTTTTTCATTCGTTGTTATATTCATGTATATACTTCTCTGGCATTTCGGCCTTCTCCAGCGT
AGTCCCATACTAGGGGTGTACCCTCTTGACCTATTATCTACATGTAGCCAGTTGCATCTAGCTGTT
TGCAGCTGTGGAGTCCACACAGCAGGTGTCGTGCTTACCGGGGAGCTTAACCGATGACCAAGACAG
CAATGTCGGAATGCTTACTTTGGTGACCGATGAGGATTATTAGGCTCAAGATAGGTCTGGTAATAT
AGTACATGGGTCTGACAAAGACTTGGTAGCCCGCAACCCATATACAGTCTGCAATATCGGCATATT
GAAAGAACCTCAAGAGAGAAGGCTGCGAATTTTTGAATAGTTATGGTATATCTATAACACAAAAGA
GAAATGAAAAGAGAGAATAACTAGACTGGTATATAAAGATATAATACAGAAAGACATGCTGACAAA
AGACATATTGTCCATTTCGGAATGCAACTCGCCCTGAAGCAATAAACACCAGGAATTCGGGTGGCA
ACCCGCAATATATGGCTCTGGAGTCATTGACAGCCCTCGCACATCATCGTCCATTAATTATCGTCC
TTCTGACTACCAGCCAGCAAGTTGTGATACCGGACTCCCCCAGGGGACGATAATCATCTCAACGCC
TCTCAGGAAAATCATCCTTTTGTTCTTGTTCTGTTGAGGTCCTGTCCTTCCCTCCATACTAGGTCG
ACCATGATGTCATCGTAATCTCACACTCTCCAATCATCTATGCCCGTATCCAGCTACATCATCCAC
TCCCAGCCATTACTCTCAAGATCGTTTCATCCATATATCAATCCAGCCATTAACCACATATATCCG

TATTTCGCTCCAATAAAACAAAACAACTCGCTACAGCCCCATACTCCCAACCAGAACCAAACCAAG
TGGGAAAAAAAATCACTTAGCACACGCCATCTCTACCAATCCCCAGCAGCAAATCATCAACTCAGT
ATACCCCCAAAAGGCTTGTACGTATCCAGGTCCTTAACCAAGAGATCCACAACGCCGGGCAGATAC
CGATCTACATATTGCTTCTTGGTCTCGACGCTCAACACGGCCCCGTTACTCTCACTGACCGACCTA
TCTCTCCCGCCGCCACCACCTCCAAAAGCCCGAACCAACTTACTCGGACTTCTTTGTAACAACCCA
TTATTATTCCCGTTCTGGTTTGACAAGGAAGCTTTCGGGATAGAAGGCGTTTCGGGGGTGAGATAA
AAGTAGAGAAACTCGGTGAGACGGAATTTGACCTCGCGGACGGTCTCGGGGTCTTTGAACAAGGAT
GTTATGGTTAGGAGGCCGTCTAAACGCTCAAAGACGCGCGTGTTTTGCGGGGAGTCGAGAAGGGCG
ACGACGAGGGTAATTATGGTTGCTGATTGGACTGCTGGACAGTTGACTGGTTCTATCAGGTCTAGG
AGGAGCTGTGAATAGGCAGATGGAAAGGTTAGCAAAGGAAGGGTAGGGGGATGGTTACGGCTGGCT
GGCTTACATTCATGTGCATCGTCCTAGAGAATAGCACTTTGCTGGGCGGATGTAGAAGGAGAACGC
CTTGTATCAAGTCGAGCGCATTGACTATAAGGAGGTCATATTGTCCGTCGCTGCCCCTGGCGACGA
GCCAGTCGAGGGTTGGGAGAAGGCGCTGCGCAATGTTCCACTCGAATCCATCCTGGAGTTTGAAGA
ATTCGCGAAAAGCTGGGTCGGAGGGGAGATCGGCTAGGACTTTGCGCGGCGGTCTTTGAGCTTCGG
ATGAGGATGACGGTATGCTGAGGGAATCGGCATGTGTTTGTCTGGAAGAGGGCCGAGGGCCGGAGA
GACAAATTTGCGCCAGAAGACCTTCTACTTGCTTGAGGCCCTTCTTGATCTTGACCCCGTCGTATG
ACGAGAGAATATCAAATGCTAACGAGAGTAGGGACTCCATTCTTTGTTGTGGCTTGGGTAAAGTGG
TATGACTGAGTCAGCTTCAGGCAAGAGATGCTTTGCGCCAACCGCGCGGGTCGCAGTTTGGATATT
TGAACCCCTGTGATGTTAAATCTGGTCGTTTTGAAGGATCTCCAATGTAGTTGAATAATGATTAAG
GTAATGCGACGTAAAAATAGCTGTAGTTGAAGTCTTGGTTGTGAAACCAAAGTCGTCAAGAGAGCT
TGCTTGTGGAGAGCTTCCACAGCCTTGACACAAACAATCGGCCTGGCCGCAAATACAGTGTCTGCG
CAAGGATGTACCTTCCTCTAATAGTGAGGCTTCAAGGGTGCCCCGGAGTGGCAAGGAAAATACTGG
TGGGTATCAAATGCAGGACATATGGATTGCAAGAGAAAAAAAAACCCGGCAAAGTCAATGAATCGG
GACCATCTATGGTTCTTTGGACAGTCGCCATGAAACGTTTCAAAGGCGCAATTGGTCCCTGGCTGC
CACAGCAACGGCATTCAACTGATAAATGGCCGATAAGACCGGCCCCTGCAGTTGGGTCACTTTGGT
GGTCCCGAGCCGGGGAAAGTGGAACTAGGCCCTGCAGAAACGCAGCTCGGCAGTCGCCCAGAAATC
AGGAACCAGAAGCTGGCTTGGACTCTTGGGTCAATCGAGATGCATCTTCACTGCTTGGTCTTCAGA
AGTTTCACCCATGACTTTATGAACTCGTCGTCTTTGAACACGAGGTAAACAGAGCCGCCTGAAAAC
ACAGCCGGAGACAACAAACATATAGGTGATCGGTCTTCTGGGCTCTCCATCCATTCGACCATGCGG
TCATGGAGATGGAACCACACAGACGGTATAAAAGGGTATAGCAATGCACCTCTGAAACCAGGCCCT
CCAAAGCCGTGATAATGTCAATGGATAAGCCACTGCTCAAGAACAAACAGTAGCGTATAAATACCT
TCTTCGTGCCTCTCCCTGCGCTCGGTCGTTCCCTTTGGAAAATGATTGAAAAGGGCCTAGGTAACC
TCGACCAAATATAACATGCCGGAGAAGTAAATTCTCTGCATTTTTTTTTTGCTTGTTTAGGTAAAA
ATCGATCATTCAAGGCCGTAAAAGTCGTCAACTCGTCGTTCCGAAGCTCCATGTCAGCAGCATCCA
AAACAAGGCTCACAGCTGTTCTCGCAGGCTCCACACCAGGCTCCCCACAATCGCCGCTAGTAACCG
ATCAGCAAGGACGGCCCGCCCAAATCAGGTAGCTGAGGGCCGAAGAACGGGTAACACAGGCAGCTA
CCGCCCGCCTGCGTCATCATCGCAACAGGTGTCTGTACCACACTGTTTTGCATGTAACTTTTCCGA
TAACGCACTTAGAGCAACCCCCCATACAAAATCGCCAACTAAGTTTAATTTCTATAAAGTACTAAA
CATTGTGATATGGTGAACAGAGATCGAATAGTAGATATCGTAGCGTATCATTGAACAAGGCTTGCG
CCACCAACGGACTCGTGCTCCTTGGGCCAACTGGTCCAAGCTGCTGTCCCAAAAACGATCGCCCCG
CCTCGGGAAACTCAAAAACCGGACGTCGACCACGCGACCCACCCACCAGCAATGTCTCTCTACAGA
CGGTTGTCCAAAAAACAGCCCGTCACCATGGCCCGCCTCTTCCTCTTCGGGCTCCGCTCAGCCAAC
CCGGTCAAAACTCCGCCCTCTTCCAATCTCCCAACGCAATACCTCCAAGAACTTCTTCTGTTGTGA
CATCTGCTACCTCCTCCTCGACCCCTCCAGGCCCCACCAAGGCGGCTCTTCCATCCCTCAGCGAAA
GCGACTGGCTTGCAAATAACCTCAAGGGCCAACAGCACTCCTCTTCTACCGCTGGTCATACTCATC
AAAGAAGCCGAAAAGTCTCGATAACGTCGTCTGTCGGCACCCGCACCATTACACAAGCATCCTCTA
TCCCTCCCGCCAACATGGCCTCAGCCCCTCCTACCCCGCCAGACCCGAAGCGCTTCGCGACCGAAG
ATTTCTCTTCTCGCCGGGCATGGTCCCCAGAAAAGGACAAAGTCCTTACCGGGCCATATGACTACC
TCAACGAGCACCCGGGCAAAGACTTCCGCTCCCAGCTGGTCAAGGCCTTCGATGCCTGGCTCGAGG
TACCCCCCGAGTCCCTAGAGGTGATCACCAAGGTAGTCGGGATGCTCCACACCGCGTCACTCCTAG
TGGACAACGTCGAAGATTTCTCCCTGCTGCGCCGTGGCTTCCCCGTAGCCCACTCGATATTCGGCA
TTCCTCAGACCATCAACTCATCCAACTATGTCTACTTCTGTGCCCTCCAAGAGTTGCAGAAGCTCA
AGAACCCCAAGGCAGTCTCTGTCTTTACCGAGGAGCTCATCAACCTCCACCGGGGCCAAGGGATGG
ACCTCTTCTGGAGAGACACTCTGACCTGCCCGACCGAAGACGAGTACCTCGAGATGGTATCCAACA

AGACTGGGGGGCTGTTCAGGCTTGGTATTAAGCTGATGCAGGCTGAGTCCCGGAGTCTTGTGGACT
GCGTACCGCTTGTCAGCATCATGGGCTTAATCTTCCAGATTGCGGATGACTACCAGAACCTCTGGA
GCCGGCGGTATACTGCCAACAAAGGCATGTGTGAGGACCTCACCGAGGGGAAGTTTTCGTTTCCCG
TCATCCACTCGATACGGTCTGATCCAAACAATCTCCAGCTTCTGAATATTCTCAAGCAAAAGACTA
CGGATGAGGAAGTCAAGCGATACGCGGTAGCCATCATGGAGCGGACTGGGAGTTTTCATTATACAC
GGGAGGTGCTGGAGGTGCTGATCGAAAGGGCACGGAAGCTCACTGATGAGATTGATGAAGGTCGGG
GTCTGGCCAAGGGTGTACATGCCATCCTGGACCGCATGGTCATCCTGCAGGAGGATGGGGTGTAGG
TATACCGTGTAACATGCATGCGGGATGTTTAAACGGTGGGCATATTTGCTTGTATTTTGCAGTTGG
ATGATTGTTCAGGTTCTGGTATAATACGGCGAAAAAGGAGTTTAGGACAGCTGGATATACGGGCAT
TTGGCAGGTTTATATAATACCCCTTTTATTCTTAGATCAATCATCTTAAATAGCAGACCTCGCCTG
AACAGGCTGAACATGTCTAGATCTTGTAATACCCAGAGTTGCCGTGCATTTTGAAAATAGGCGTGC
GAGTATAACCGATATATATGGTGAAAAGCTCAAGCGACTAACTCCTAGAGCGGAAGAAACTTGAAT
CGTAACAATCATTAATCATATGCAAACGCCTACATGGTGAATATACCAGTGATGTTTCCTAATCTA
GCCACTGCCTGTGATGTTCTTTCTTTTGACCATCCTGACCCATCTTCGCCTCCTCCATTCTCCCCT
CCGTCCAACACCCTCATGGTCACCCTCCTCCCAGCTAAGGTTACTCGTCCCAACTGTAACCCGATG
CTCCCTTCGTACCGGATCAGGCAAATCACCTCCGCCCGTGACAATCCCGCCGTTAGCACCACCATT
AGCACCTCCTTCATTTTCATTATAGATATCATAGACCCATCTCTCGCCCTCCGTTTCGACTTCCAC
GCCTGTGATGATCCTCTCCTCCACCCACTCTCGACTCCACAGGTCAAGAGCCCACTTCTTCTCACT
CCATTCCCAGCCTTCAGGGGGCAGGATATCTTCGAAAAACCGCGTCCCGCGCGGACGGTCTCCCGC
GATTCGGCCCGCGGATAAGGGATCATAAGGCGATGGACAAAACAACCACGGCTCCCACTCACCACT
CCCCTCGCCCAGCCTTTGTAGTTCGAAAATCTCAACCTCCCTGGTCTCGGGTGCTTCATCGAGAAT
AATGTCCCCCTTCACCCAATTATCAAAAGCCTCCTCAGCTAGCTGTACCGGATTAGGCATCTCAGT
CTTGTTTCCTACCCCGTCCTTGTCCGCTGTGCCTTTCTCGTCCTGGTGAGCACCAGTATACTGCTC
TTTCATCTGCGCCATCCAAACAGCAACAATAGGATGGCCGCTCCAGATGGCCGCCCAACCAGTTAC
CAGCGCCACGGCCTTCCATGGAACCAAGTGCGACGCGATGCTCATAAACACACATATTCCAAACAG
GAGTACAAACACGCCCGAACTCAACGCCTCGTCTGAGAAATTTGTCACGGGAACCACCAGCTCAAT
CACCTTATCATGGGCGACGCTAAAGTCCTCCATGACGTTTTGAAGATCGCCCATGTTACGGAAGAA
GTCTTTGCTCAGCTCCTTGACGGGTTTCACCGTCCTCGCCGGTGCTAGGGGAGGTCCTCGCGCGCT
GTACCCCATGTTGCGGACTCGGTCGGGGTGGTGTATGGTGACCACCGTCATCAGATAATATCGAAG
ACGGGGGAGGCGGATGGCGGGCCGATAAAGGATGGTACTAGAACGAAGAGCAGAATGCGGCGACTG
GAAGGATGGTGAAGAGGTATGGGTCTAGACAGACAAAGGTGTACACTGAGAGGAGGGATAGCGTGT
GGGAGGGTCTCCGCCAGGAGAGGACGCGTAATAACTTGTTCTGGAAGCGGAATACGACGCCTATCC
TGGCATTGAAGCGGCGGAAGTTGTTGGACATGGCTGTTACCGTGAAGCCTGGTCGGGAGGCATAGT
CTGGACTCTCGTCGCCGCCGGTTATCGTGGGAATGGTGTTCATGCCATTGTGGATCGAGTCGTCGG
CTGGAAGCATTTGGGAAAGCAGTCTATTGGGAAGAATAGAGCGATAAGTATCTACTGGTATGCTGT
GAAGGATTTTGAGAACAATAATGACGAAACTCACTTCTCAAGCAGCTTGTCCTGGAGACTGGCTGC
TTTGAAAATACCTCTCCGTATACTCCTCGCCGCCTTGGCCGAGTCGGCCGTCGTAGATTCAGACTG
CGCCGCCATCCCGGTATTCATTCTCGTTGTGGACGTGTTGTCGTCGGGGTCGTAGATTTCTGTGAG
AGCTTCTTCCTCCTCGTCTACGGTGCCGAGGAGCTCTGTCGCCAGCACGTCCATCTTAAATTTATT
GACTCTGGGCGGGTTTCGGGAAGAAAATGGTGGCTGGCCGGTGGCTTAGAAGGATGACGGATTCAG
AGTTTGGCTAGTTAAGGTTTGTTCGTTCGACCTCGTTTTTGGCCCGGTATGCTCTGGGGACCCGCG
AGTAGAAGGCGGGCCGGGGCGACGTCACACCTGGGTAACCACGTACAATGCGCTACACAGCGCGTG
TCTCTGGAGTCTGTTTTTTGAAGCTGCCCTGGTTTGTGCTGGTTATGAAACTTGGGGATCAGAACA
AATGAAGAGTTACAGAGTAACGGGGGCATTT