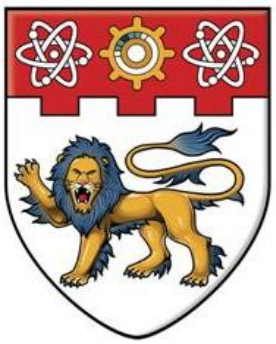
GENOMICS STUDY OF BIOSYNTHESIS OF MICROBIAL SECONDARY METABOLITES A CHAKRABORTTI 2016



NANYANG TECHNOLOGICAL UNIVERSITY

A GENOMICS-GUIDED STUDY OF THE BIOSYNTHESIS OF MICROBIAL SECONDARY METABOLITES

ALOLIKA CHAKRABORTTI

SCHOOL OF BIOLOGICAL SCIENCES

2016

A GENOMICS-GUIDED STUDY OF THE BIOSYNTHESIS OF MICROBIAL SECONDARY METABOLITES

ALOLIKA CHAKRABORTTI

A CHAKRABORTTI

School of Biological Sciences

A thesis submitted to the Nanyang Technological University in partial fulfillment of the requirement for the degree of Doctor of Philosophy
2016

Acknowledgement

While submitting this doctoral thesis, I would like to take the opportunity of thanking a few people who stood by me and extended their support and assistance to me at all times during the last few years.

First and foremost, I would like to thank my Ph.D-Supervisor Assoc. Prof. Liang Zhao-Xun for being such a source of motivation and encouragement. I am immensely thankful to him for giving me the opportunity and freedom to grow as an independent researcher and take up new challenges. At the same time, I am indebted to him for his invaluable insights and inputs on my research. It has been a complete privilege for me to work under his tutelage.

I am deeply thankful to my ex-colleagues – Dr. Ho Chun Loong and Dr. Yang Lifeng for their constant guidance and most importantly for standing by me through thick and thin in the last four years. I would also like to extend my gratitude to my ex-labmate and mentor Dr. Mary Chuah for her guidance. I would like to thank and acknowledge the support of Yeo Qin Yi, Denise Seow, Leow Min Li, Ding Yichen, Huang Peng and Cheang Qing Wei for working with me on this project at various stages. I want to express my gratitude towards all my other former and present lab-mates - Prabhadevi Venkataramani, Amanda Tan, Dr. Sun Hui Hua, Ishin Soehano, Swathi Pasunooti, Koh Siew Lee, Dr. Xu Ling Hui, Jamila Oppong, Yong Grace, Xin Lingyi, Pang Li Mei and Low Zhen Zie for their continuous and unfailing camaraderie. I would also like to thank Dr. Nishanth V. Menon, Dr. Manisha Mukherjee, Areetha D'Souza and Dr. Tan Jiazi for their help and support.

I am thankful to all my collaborators; Dr. Yoganathan Kanagasundaram (Bioinformatic Institute, A*STAR, Singapore), A/P Liu Xuewei (School of Physical and Mathematical Sciences, Nanyang Technological University), Asst/P Yang Liang (Singapore Centre on Environmental Life Science Engineering, Singapore), Jinming Li (Southern Medical University, China) and A/P Mark Turner (School of Agriculture and Food Sciences, University of Queensland). I would like to express my sincere gratitude to Dr. Yit Heng Chooi, Research fellow, (College of Medicine, Biology and Environment, Autralian National University) for his advice on the bioinformatic analysis during the early days of the project.

I am extremely thankful to my Thesis Advisory Committee (TAC) members Prof. James Tam, A/P Liu Chuan Fa, Dr. Tang Kai and Dr. Peter Cheung from the School of Biological Sciences, NTU for their valuable inputs and guidance.

I am indebted to Nanyang Technological University and Ministry of Education, Singapore for providing me with the NTU research scholarship and grants necessary for my research.

Last but not the least a special thanks to my family and friends. I am eternally thankful to my parents for being my biggest source of strength and inspiration at all times. Words are enough to express how grateful I am to all my friends for their being a part of my support-system in such a huge way. This acknowledgement would be incomplete without the mention of Dr. Arnab Kapat, my mentor at Reliance Life Sciences, India, for I would have probably not been here without his constant encouragement and nudging. I am thankful to him for always having some kind words to inspire and motivate me.

Table of contents

List of figures

List of Tables

Abbreviations

A domain	Adenylation domain
ACP	Acyl carrier protein
AT	Acyl transferase domain
ATP	Adenosine tri-phosphate
Aw	*Aspergillus westerdijkiae*
C Domain	Condensation domain
C-MeT	C-methyltranferase
Core	Core domain of PR-PKS
DH	Dehydratase domain
DNA	Deoxyribo-nucleic acid
ER	Enolyreductase
EDTA	Ethylenediaminetetraacetic acid
FAS	Fatty acid synthase
HPLC	High performance liquid chromatography
iPKS	Iterative polyketide synthases
ISP2 medium	International Streptomyces medium
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KS	Beta-ketoacyl-synthase domain
KR	Keto-reductase domain
LC-MS	Liquid chromatography Mass Spectrometry
LC-HRMS	Liquid chromatography High Resolution Mass Spectrometry
LB broth	Luria Bertanii broth
LMM broth	Lactose minimal media broth
NRPS	Non-ribosomal peptide synthatases
Nj	*Nocardia jinanensis*
NMR	Nuclear magnetic resonance

ORF	Open reading frame
OTA	Ochratoxin A
PCP	Peptidyl carrier protein
PKS	Polyketide synthases
PT	Product template
PCR	Polymerase chain reaction
RNA	Ribo-nucleic acid
SAT	Starter-unit acyl transferase
SAM broth	Semi-synthetic medium broth
SM	Secondary metabolism/metabolite
TLC	Thin layer chromatography
TE/CLC	Thioesterase/Claisen cyclase
UV-Vis	Ultraviolet-visible spectroscopy

Abstract

Microorganisms such as filamentous fungi and actinomycetes are prolific producers of structurally diverse secondary metabolites. Many microbial natural products have been developed into commercially valuable compounds as pharmaceutical agents, pesticides, food preservative and cosmetic ingredients. After several decades of screening for bioactive compounds, the current field of microbial natural product discovery is plagued with the problem of re-discovering already known compounds. To circumvent the re-discovery challenge, we took a genomics-guided approach in this study to uncover the secondary metabolite synthesizing potential of the filamentous fungus *Aspergillus westerdijkiae* and the rare actinomycetes *Nocardia jinanensis*. Sequencing and bioinformatic analysis of the genomes of *A. westerdijkiae* and *N. jinanensis* revealed the remarkable ability of the two microorganisms in producing secondary metabolites, as evidenced by the large number of secondary metabolism biosynthetic gene clusters found in the genomes. Comparative genome analysis with other related environmental and pathogenic strains yielded insight into the factors determining the life-style and pathogenicity of the two microorganisms. A number of structurally known and unknown compounds were identified from the culture broth of the two microorganisms by metabolite profiling and compound isolation. By genome mining and gene knockout, we identified the biosynthetic gene clusters for a putative siderophore and notoamide, and validated the biosynthetic gene clusters for the mycotoxin ochratoxin A (OTA) and the peptidyl alkaloid circumdatins in *A. westerdijkiae*. Biosynthetic mechanisms were proposed for OTA and circumdatins with functions assigned to individual biosynthetic enzymes. From the culture broth of *N. jinanensis*, we isolated several cyclic dipeptides, the alkaloid flazine and a lipopeptide. The putative biosynthetic mechanism for flazine and the lipopeptide were proposed. Considering the large number of biosynthetic gene clusters observed in the two genomes, many of the biosynthetic pathways seem to be cryptic or silent. The genome sequencing and characterization of the secondary metabolites described in this dissertation sets the stage for exploring the full biosynthetic potential of the two highly prolific producers of secondary metabolite for the discovery of novel natural products and biosynthetic enzymes.

Chapter 1 - Introduction

Microbial secondary metabolites

Microbial secondary metabolites are a class of low-molecular-weight natural products produced by the secondary metabolic pathways of a variety of microorganisms including bacteria and fungi. [1, 2]. Unlike the products of primary metabolism, these compounds are not absolutely essential for the growth and survival of the organism under nutrient-rich conditions, but are considered to be crucial under nutrient-deficiency or stress. A number of reports suggest that the production of secondary metabolites is correlated with growth and development of microorganisms [3]. For example, the stationary phase accompanied by sporulation in some microbes has been associated with the higher yield of secondary metabolites [4]. The production of these small molecules is often thought to give the organisms a competitive edge over the others in the natural environment [5]. These low-molecular-weight compounds are endowed with a variety of bioactivities and have been exploited as antibiotics, antitumor agents, cholesterol-lowering drugs, immunosuppressants, antihelmintic agents, herbicides, insecticides, toxins, virulence factors and so on [6]. According to a review, approximately 31,600 microbial secondary metabolites have been reported from 1900 onwards, and about 20,000 of them possess some kind of biological activity [7]. Although thousands of structurally diverse secondary metabolites are manufactured by microorganisms using combinatorial chemistry, they are essentially synthesized from only a few key precursor molecules such as amino acids and acyl CoAs. The remarkable functional diversity of secondary metabolites has been attributed to the enormous structural variation and complexity [8]. Over half of the drugs in the market today are secondary metabolite or their derivatives [6]. Secondary metabolites have evolved to bind various biological targets in the natural environment and as a result many are promising as drug leads [6]. Given their great pharmaceutical potential, discovery of new bioactive secondary metabolites remain an exciting area of research.

Secondary metabolite based therapeutics have been known to mankind through traditional medicines over centuries. In modern times, the field of secondary metabolism research (especially antimicrobial drugs) gained tremendous impetus after the discovery of penicillin from the fungus Penicillium in 1928 by Alexander Fleming [9]. The decades

that followed witnessed a surge in the discovery of numerous secondary metabolites from various microorganisms. This trend continued for about three to four decades after which there has been a sharp decline in the rate of finding novel secondary metabolites [10]. Infact, the field of microbial natural product discovery is currently plagued with the problem of re-discovery of already known compounds. On the other hand, the rampant use of antibiotics and the property of natural selection in living organisms have concomitantly led to the emergence of various drug resistant microbes [11]. While the problem of compound re-discovery has led to a discovery "void" in the field of secondary metabolism research, the emergence of multi-drug resistant microbes and the rise in new clinical manifestations related to cancer and various other other diseases, necessitates the rapid discovery of secondary metabolites with novel biaoctivities. In this context, while discovery of novel molecules remains a very important priority, a better understanding of the biosynthetic mechanism of these compounds is also very essential for the designing and synthesis of more effective small molecules using metabolic engineering and synthetic biology [12]. Metabolic engineering has been instrumental in the synthesis of more-effective analogs of a number of useful microbial secondary metabolites including penicillin and cephalosporin [12]. The knowledge of detailed biosynthetic mechanism is also essential because many of the secondary metabolites are toxins and virulence factors and a detailed understanding of their biosynthesis could be helpful to tackle the problems they entail. Secondary metabolites with the ability to act as biopesticides and bio-herbicides are very important both industrially and environmentally (since they are not as harmful to the environment as chemical pesticides), and hence the discovery of more such small molecules is of relevance to biologists.

Several factors apart from re-discovery of known natural products triggered the decline in novel secondary metabolite discovery, including the technical challenges associated with isolation, purification and structure elucidation of secondary metabolites from microbial fermentations [13]. At a time when secondary metabolite discovery programs were falling quite rapidly in terms of numbers during the late 1990s, the whole genome sequencing of microbes began to gain attention [13]. One of the very first microbial genome to be sequenced was that of *Haemophilus influenza* in 1995 [14]. Since then there has been no looking back and a huge number of microbial genomes have been sequenced and reported. At present there are more than 500 complete microbial genome sequences present in various databases [13]. This information would help to increase our

understanding of the genetics and enzymology underlying microbial secondary metabolite biosynthesis. The genomics-guided approach has facilitated the identification and analysis of gene clusters, likely to encode for secondary metabolite biosynthetic pathways in sequenced microbial genomes [15]. Genome mining of several microorganisms including cyanobacteria, actinomycetes, fungi and others have revealed numerous secondary metabolite biosynthetic gene clusters (some which might be cryptic), which could be endowed with the potential to synthesize a large group of diverse small molecules which have been never known or observed before (the core unit responsible for secondary metabolite production is a biosynthetic gene cluster, which codes for biosynthetic enzymes, resistance determinants, transporters and regulatory elements) [16].

Recent headways in "-omics" technology that have been introduced over the last two decades have allowed the establishment of various research areas pertaining to secondary metabolism ("-omics" encompasses the various genes and their products such as transcripts, proteins, or metabolites) [17]. Genomics is concerned with genes, their variation, and function; and transcriptomics deals with information at the mRNA (transcript) level present at a particular point of time depending on environmental signals and other bio-physiological parameters [18]. Proteomics deals with the expression, function, and regulation of an entire set of proteins [18]. Moreover, as the proteins within cells are the functional units, their expression is strongly influenced by environmental signals and physiological conditions, and thus, proteomics is a complementary technology to genomic and transcriptomic research. Metabolomics considers the detailed metabolic analysis of the organism in question. Based on the huge progress in "-omics" technologies, new approaches are aimed at ensuring optimal engineering of the cell factory to achieve optimized metabolite production [19]. Genomics, transcriptomics, proteomics, and metabolomics are playing a pivotal role as research tools in systems biology of secondary metabolism for enhancing the production of secondary metabolites in engineered hosts [17]. The unraveling of numerous uncharacterized microbial secondary metabolic gene clusters has attracted synthetic biologists who would want to exploit recent advances in sequencing technologies and synthesis procedures to achieve unprecedented control over the some of the important metabolic pathways [17]. In particular, rare actinomycetes or microbes from unusual/unexplored habitat are promising candidates for the search of novel secondary metabolites. A deeper analysis of the genetics and biochemistry involved in the various microbial metabolic pathways

identified in these microorganisms can be crucial for small molecule discovery and development [16].

Since the announcement of the first microbial genome sequence, genome mining approach has been observed to be largely responsible for the revitalization of secondary metabolite discovery research. A careful selection of microbial subjects and the combination of genome mining with advanced analytical methodologies hold the future of secondary metabolite discovery and elucidation of biosynthetic mechanisms entailed in each of the major metabolic pathways. The availability of untapped biological resources (genomes), smart screening methods, high-resolution and sensitive separation of compounds along with structural analysis, metabolic engineering and synthetic biology offer the perfect combination of all the right ingredients necessary for the in-depth study of microbial secondary metabolism [16, 20].

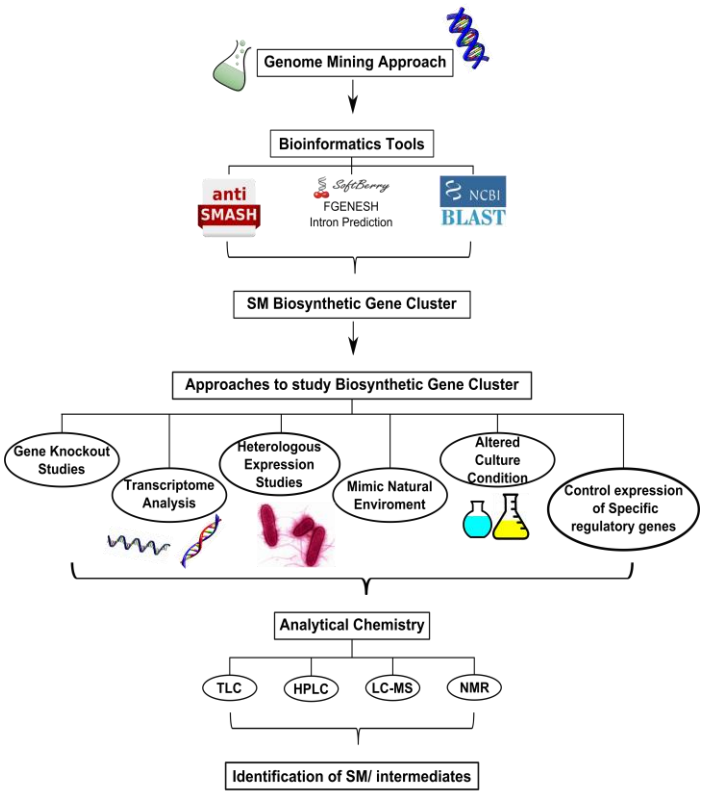


Figure 1.1: The schematic representation of the genomics-guided approach to study secondary metabolism.

Key concepts

Before going into the details of the different aspects of secondary metabolism research and what entails genomics-guided approach to study microbial secondary metabolism in the forth-coming sections and chapters it might be useful to go through the following

concepts briefly - (1) the different classes of secondary metabolites depending on their biosynthetic origin, (2) the enzymes that are responsible for secondary metabolite synthesis; and (3) regulation of secondary metabolism.

Classes of secondary metabolites

Secondary metabolites can be grouped according to their biosynthetic origins.

Polyketides are synthesized by polyketide synthases (PKS) and sometimes by the PKS-NRPS (non-ribosomal peptide synthetases) hybrid systems using acyl CoA and malonyl CoA building blocks. Polyketide-based drugs and products account for more than $20 billion in sales annually [21]. Polyketides are made naturally by bacteria, fungi and plants. Some of the pharmaceutically important polyketides include the antibiotics tetracycline and erythromycin [22], the anti-cancer drugs doxorubicin [23], the anti-oxidants EGCG and resveratrol [24] and the cholesterol-lowering lovastatin [25].

Non-ribosomal peptides (NRP) form a large group of secondary metabolites and are synthesized by the NRPSs (non-ribosomal peptide synthatases) using amino acid building blocks which could be both proteinogenic and non-proteinogenic. Non-proteinogenic amino acids are those which are not found naturally in the genetic codes; for example: beta alanine and gamma aminobutyric acid (GABA). NRPs have a wide range of biological properties and can be toxins, antibiotics or siderophores. The antibiotics surfactin [26] and gramicidin [27], the novel siderophore vibriobactin [28] produced by *Vibrio cholera* and the immunosuppressive agent cyclosporine A are some of the examples of NRPs [29].

Terpenoids are a class of volatile secondary metabolites synthesized using isoprene subunits, dimethylallyldiphosphate and isopentenyl diphosphate which are then condensed by terpene synthases/terpene cyclases. The best-known terpenes come from plant; but fungi also is known to synthesize several important terpenes, including aristolochenes, carotenoids, gibberellins, indole-diterpenes and trichothecenes [30]. All terpenes, which are composed of several isoprene units, can be linear or cyclic, saturated or unsaturated, and are modified in various ways [30].

Saccharides are secondary metabolites that contain carbohydrate as building blocks and their synthesis involves enzymes necessary to make the saccharide backbone. One of the best examples in this regard is the biosynthesis of the bactericidal molecule kanamycin

[31]. A number of other aminoglycosidic antibiotics which contain amino sugars as their primary building blocks are classified under this class of compounds [31].

Hybrid metabolites further add to the variety of secondary metabolite structures. The combination of different biosynthetic pathways results in the synthesis of these hybrid molecules. One example is the synthesis of aflatoxin where a polyketide synthase incorporates a fatty acid as a starter unit [32, 33]. From a structural point of view, the fusion of polyketide and nonribosomal peptide biosynthetic pathways is particularly interesting. Some of the examples of bacterial PKS-NRPS hybrid molecules include bleomycin [34], epothilone [35], yersiniabactin [36], and rapamycin [37]. A large number of PKS-NRPS hybrids have been identified in fungi some of which are endowed with important bioactivity [34]. Fusarin C [37], cyclopiazonic acid [38], and flavipucine[39] are some of the fungal PKS-NRPS products.

Biosynthetic mechanisms and enzymes

Usually secondary metabolism genes are found clustered together. These genes clusters encode for proteins that synthesize the backbone or skeleton of the secondary metabolite product i.e. PKS and NRPS along with other ancillary enzymes that are responsible for the step by step modification of the backbone scaffold to generate the final product. In addition to the genes essential for secondary metabolite biosynthesis, secondary metabolite gene clusters also usually contain a number of regulatory genes. These genes encode for transcription factors for the regulation of the biosynthesis of the secondary metabolites. There are several major families of biosynthetic enzymes and have been discussed below.

Polyketide synthases (PKS) are enzymes which are responsible for the synthesis of structurally diverse polyketides using acyl-coenzyme (CoA) substrates. This is accomplished with the condensation of acetyl CoA or propionyl CoA (starter unit) with a malonyl CoA (extender unit) [40]. This reaction is followed by decarboxylation of the extender unit. Repetitive condensation and decarboxylation along with other modifications like reduction, hydration, methylation and cyclisation completes the polyketide synthesis [40]. The type of modification depends on the domain organization of the PKS protein.

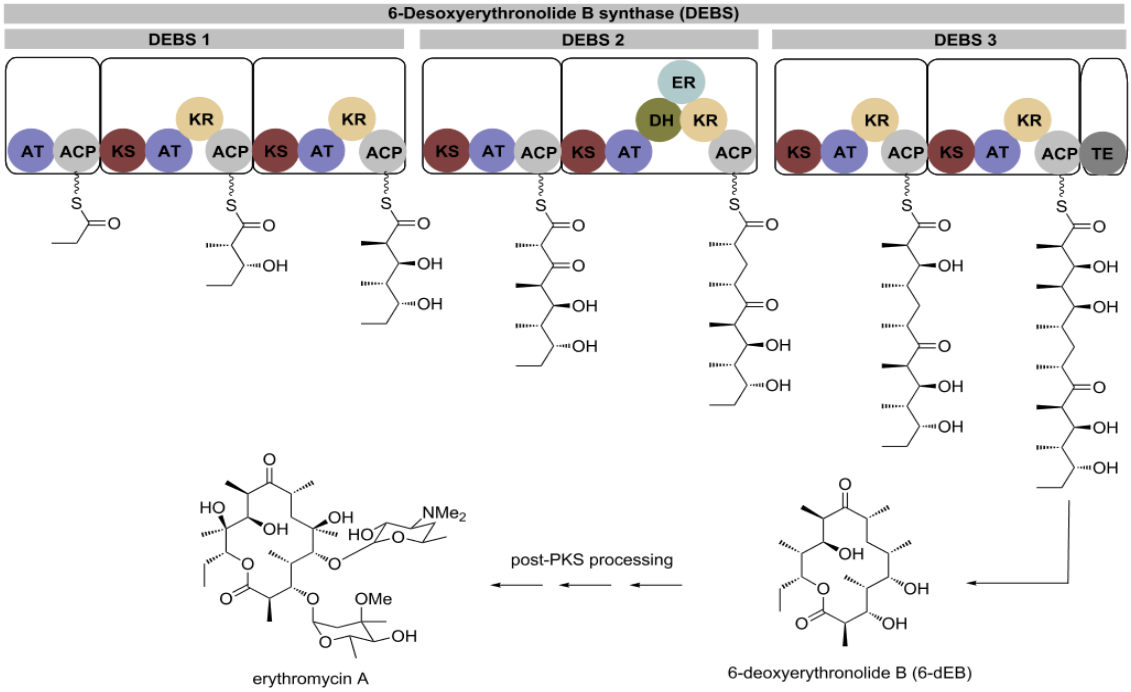


Figure 1.1: Biosynthesis of the antibiotic erythomycin (uses modular PKS) (KS: β-ketoacyl synthase, AT: Acyltransferase, ACP: Acyl carrier protein, TE: Thioesterase, KR: Ketoreductase, DH: Dehydratase ER: Enoylreductase [22].

PKSs are generally categorized into three classes that include modular type I PKS, iterative type I PKS, type II PKS and type III PKS. Modular type I PKSs are processive in nature and consist of multiple modules wherein each catalytic domain is used only once for the synthesis of the product. The product synthesis in this case is in an assembly-line manner [41]. Erythromycin is an example of drug molecule produced by this category of PKS enzyme [22].

Unlike modular PKS, iterative type I PKS use a single set of domain iteratively for product synthesis [42]. Iterative PKSs can be further sub-divided into non-reducing PKS (NR-PKS), partially reducing PKS (PR-PKS) and highly reducing PKS (HR-PKS). Iterative type I PKS (iPKS) are analogous to fatty acid synthases (FAS) in vertebrates. One of the intriguing aspects about iPKS is that some catalytic domains are "programmed" to function only at certain stage during the chain extension process [43]. In conjunction with the use of different domains, this selective nature largely accounts for the observed structural diversity of iPKS-derived polyketides [43]. The best studied iPKSs include the NR-iPKS from the biosynthetic pathway of the mycotoxins aflatoxin [44] and the HR-iPKS biosynthetic pathway of the cholesterol-lowering drug lovastatin [25, 43].

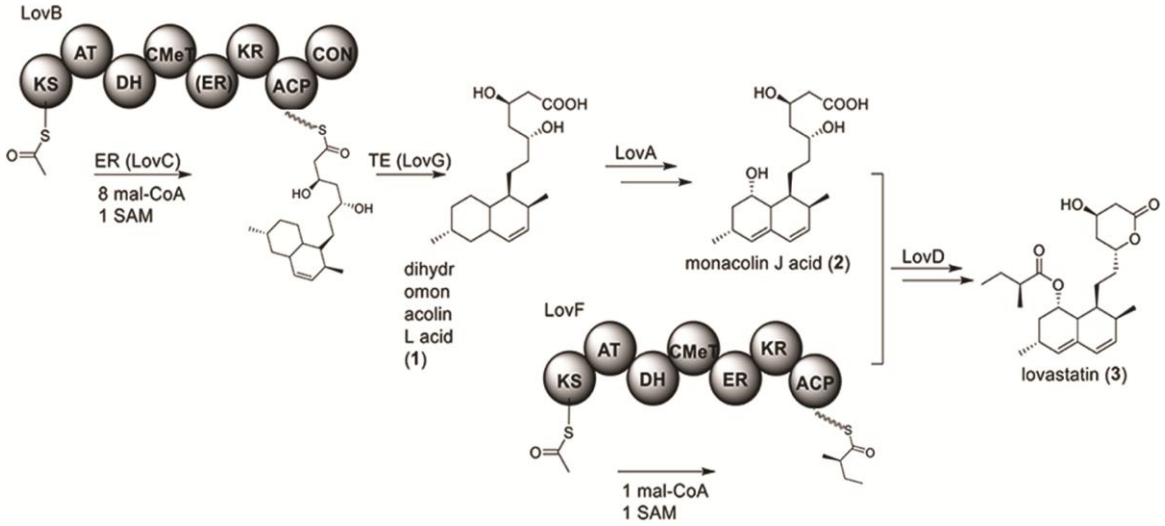


Figure 1.2: Lovastain biosynthesis by iterative type I PKS in *Aspergillus terreus* [25].

Actinobacteria are known for producing some of the important polyphenol antibiotics that are synthesized by type II PKS. They consist of multi-enzyme complexes wherein each enzyme has a specific function. Type II polyketide synthases contain at least one minimal PKS unit consisting of two ketosynthase domains (KS alpha and KS beta) and an acyl-carrier protein (ACP). The presence of additional enzyme domains coding for cyclization, reduction etc. modify the product of the minimal PKS unit to give the final product [45]. Anthracyclin with anti-cancer properties, the antibiotic tetracycline and the antitumor landomycin A are all synthesized by type II PKS [46].

Type III PKS are also known as chalcone synthases. They have been found in both plants and microorganisms. They are multimodular proteins and iterative in nature. They use their single active site iteratively for priming, condensation and cyclization [47]. The type III PKS in plants usually uses 4-coumaroyl CoA as the starter unit and malonyl CoA as the extender unit for product synthesis whereas the fungal type III PKS 2′-oxoalkylresorcylic acid synthase (ORAS) uses long-chain fatty acyl-CoA as starter units [47, 48]. Curcuminoid and chalcone derived from plants are examples of type III PKS synthesized metabolites.

Non-ribosomal peptide synthetases (NRPS) are responsible for the synthesis of an important class of secondary metabolites called the NRP (non-ribosomal peptides). These

enzymes are multi-modular proteins in which each module is involved in the incorporation of a specific amino acid into the growing chain. Typically, a NRPS consists of a loading module comprising of an adenylation domain (A) which recruits a specific amino acid and activates it using an ATP. The A domain is followed by a thiolation or peptidyl carrier protein (PCP/PP) which covalently binds to the activated amino acid by a phosphopantetheinyl arm [42]. The subsequent modules generally carry an A domain, a PCP domain and a condensation domain (C). The C domain is responsible for forming a peptide bond between two amino acids linked onto adjacent modules [42]. The final module carries an additional TE or thioesterase domain which ensures the cleavage of the resultant peptide from the PCP domain of the termination module.

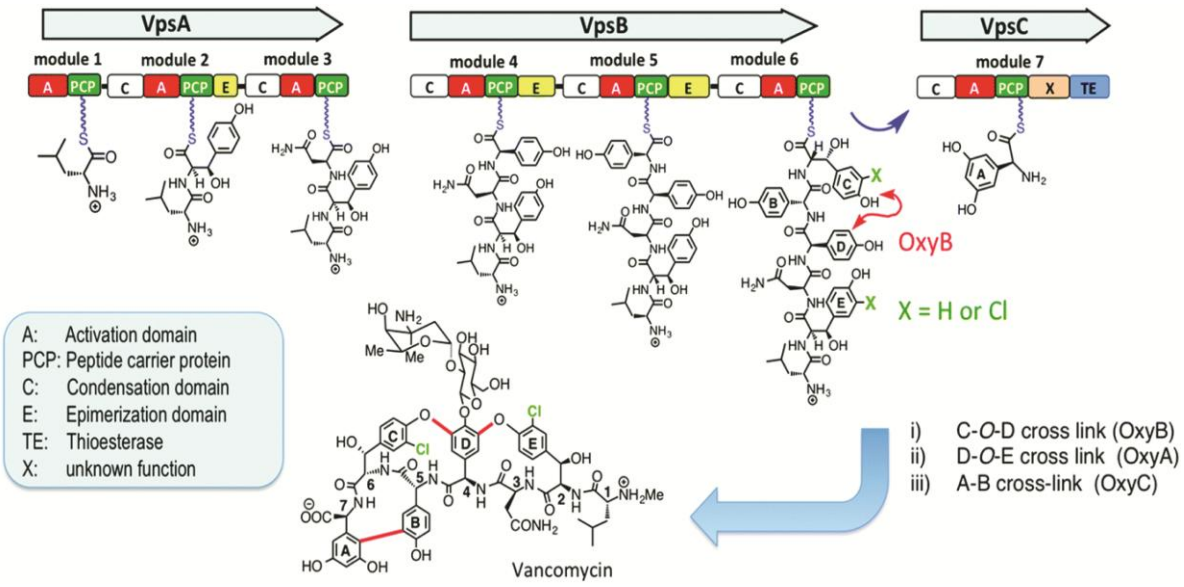


Figure 1.3: Biosynthesis of Vancomycin by NRPS [49].

A number of NRPS contain additional domains responsible for residue modifications leading to structural diversification of product molecules. These include domains for heterocyclisation (Cy), epimerization (E), reduction (R), formylation (F), methylation (M) and oxidation (O) [42]. Apart from PKS and NRPS genes in isolation, sometimes PKS-NRPS hybrids are also found in nature, which can confer greater diversity to the repertoire of secondary metabolite structures. As a matter of fact, a lot of commercially

vital secondary metabolites are synthesized by gene clusters containing both PKS and NRPS genes [50].

Terpene synthases/cyclases are responsible for the synthesis of terpene metabolites. The substrates for terpene synthases are IPP (isopentenyl diphosphate), DMAPP (dimethylallyl diphosphate), GPP (geranyl diphosphate), FPP (farnesyl diphosphate) or GGPP (geranylgeranyl diphosphate) [51]. Another group of enzymes, the prenyltransferases produce geranyl diphosphates GPP, FPP and GGPP which are used by terpene synthases to manufacture terpenes [51]. Isopentenyl diphosphate and dimethylallyl diphosphate (DMAPP) are products of the mevalonate pathway and act as building blocks for the linear polyprenyl diphosphates [30]. The linear polyprenyl diphosphates are precursors of various steroids, carotenoids and coenzyme Q in a variety of organisms. A class of isoprenyl diphosphate synthases is responsible for chain elongation. DMAPP and the isoprenoid intermediates are also the starting points for a wide range of secondary metabolites, including indole alkaloids, monoterpenes, sesquiterpenes and diterpenes. Terpenes are synthesized by cyclization of the isoprenoids [30]. The main enzyme in terpene synthesis is terpene cyclase, which is essential for the production of a variety of terpenes from different diphosphates. Several fungal terpene cyclases have been characterized, including a bifunctional terpene cyclase, a trichodiene synthase and an aristolochene cyclase [30, 52-54].

Ancillary enzymes include the tailoring enzymes like cytochrome P450, oxidoreductases, hydrolases, halogenases, epoxidases etc. that are responsible for the further modification of the secondary metabolite backbone molecule (synthesized by PKS or NRPS) to give the final product which is a biologically active compound. The secondary metabolite gene clusters also contain certain transporters which play a role in translocation of substrates and secretion of secondary metabolites across the membrane e.g: ATP-binding cassette (ABC) transporters [55].

Regulation of secondary metabolism

Regulation of secondary metabolism is often very complex and intricate. While some of the secondary metabolism gene clusters are under the influence of specific regulators, in some of them the regulation is via a common or global regulator. A large number of secondary metabolism gene clusters remain cryptic i.e. they do not express under usual laboratory conditions most likely due to the lack of highly specific triggers which are

usually present in the natural environment. Some studies have also found a correlation between secondary metabolite production and developmental stage of the producing organism [3]. In filamentous fungi *Claviceps purpurea,* it has been observed that the production of certain toxic secondary metabolites have a direct relation with sporulation [3]. In Streptomyces (belonging to the class actinomycetes) secondary metabolite synthesis has been found to coincide with aerial hyphae formation in surface cultures [56]. In *Streptomyces coelicolor,* signaling molecules like ppGpp (guanosine tetraphosphate) have been found to influence antibiotic production [57]. A γ-butyrolactone - factor A has been reported to affect both streptomycin production and morphological development in *Streptomyces griseus* [58]. There also exist certain regulators belonging to the SARP family of proteins in Streptomyces that act as transcription activators in a pathway specific manner [59]. In the pathogenic fungus *Aspergillus fumigatus,* an example of a pathway specific regulator is the GliZ gene (a zinc finger transcription factor) in the gliotoxin biosynthesis cluster [60]. The deletion of GliZ leads to the disappearance of gliotoxin synthesis. Global regulators like the LaeA and VeA genes in Apergilli have been reported to influence the biosynthesis of a number of secondary metabolites such as penicillin, lovastatin and hyphal pigments [61, 62]. Secondary metabolism in filamentous fungi is far more complex compared to bacteria and actinomycetes, involving a number of regulatory proteins working at tandem [63]. Further complexity arises with the possibility of cross-talks between two or more pathways indicating a combinatorial approach of biosynthesis of secondary metabolites. Although a lot of work has been done in this field especially in the post genomic era, the knowledge on secondary metabolite regulation is still in its infancy. A superior understanding of secondary metabolite regulation would be quite useful for the purpose of activation of cryptic gene clusters that could be responsible for the synthesis of a number of novel and biologically useful molecules.

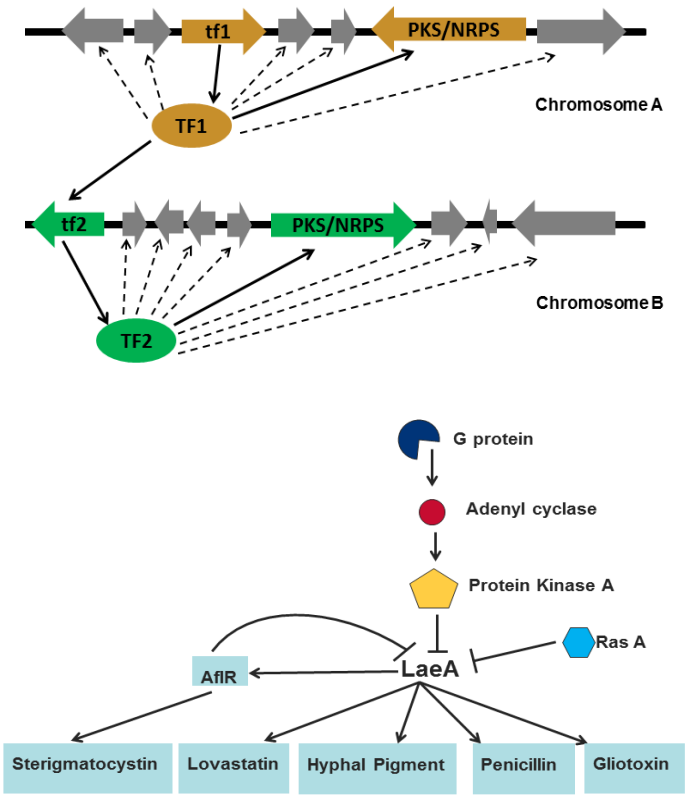


Figure 1.4: Regulation of secondary metabolism. (a) Cross-talk has often been observed between secondary metabolite clusters. In this example tf1 gene on Chromosome A is a transcription factor that activates not only activates its own cluster but is also seen to be tf2 gene on chromosome B, leading to the activation of another gene cluster. Illustration adapted from [63]. (b) The figure illustrates how the global regulator LaeA affects the production of more than one secondary metabolite in *Aspergillus sp*. Illustration adapted from [61].

Genomics guided approach to study secondary metabolism

For much of history, secondary metabolite discovery has been a process driven in large part by chance. In most cases, discovery of new secondary metabolites has been driven either by large scale screening that would involve bioactivity guided fractionation of crude fermentation broth extracts. As the inventory of natural products has grown rapidly in the past century, the long term success of secondary metabolite discovery campaigns have generally been determined by the degree to which this dependence on blind chance can be minimized [16]. Historically, the screening approach depended on a number of strategies to accomplish this task, including exploration of new ecologies, careful selection of strains [64, 65], and the development of better analytical methods with improved analytical separation and sensitivity [66]. Over the years there has been

stagnation in the field of secondary metabolite discovery accompanied by a rise in the rate of re-discovery of already known compounds.

The stagnation in new secondary metabolite discovery has had the biggest impact on the pharmaceutical and healthcare sector. Discovery of a pharmacologically relevant small molecule is an expensive affair and quite a long process. As a thumb rule, one in five-thousand leads reach the market as a drug, as most of them are not able to clear the clinical or pre-clinical trial stages [67]. Some estimates claim that it takes at least $ 1 billion and 5-7 years for the discovery of a single drug [67]. Hence, the screening and identification of as many bioactive compounds as possible is essential as only a few amongst them may finally qualify as therapeutic drugs. Traditionally, the approach for discovering novel microbial natural products would start with screening followed by fermentation/culturing and compound isolation. This method of screening for bioactivities is also known as the "compound-based" strategy. Pharmaceutical giants like Merck and Eli-Lilly have spent millions of dollars on screening thousands of microbial strains for new secondary metabolite discovery but have not met with much success in the last couple of decades. Apart from the marked increase in high re-discovery rate, the top down approach suffers from two major drawbacks – (a) the products of some of the gene clusters which remain cryptic or silent under laboratory conditions cannot be identified (and hence the metabolic capability of those clusters remain untapped), (b) the compounds which are produced in very low yields would go undetected and hence unidentified. As mentioned previously, this method heavily depends on a number of blind chances and hence other methods which are more high-throughput and robust in nature could be good alternatives.

Genome mining can be defined as the radical re-envisioning of the entire approach for looking at secondary metabolite discovery. Contrary to the traditional large scale screening mode of natural product discovery, the success of the genome mining method heavily depends on the extent to which the secondary metabolic gene clusters within a given genome can be unraveled [13]. Over the past few decades, genome sequencing technologies have become increasingly affordable for routine purposes. Bioinformatic platforms have also seen considerable development in this time frame [13]. This has led to the sequencing of whole genomes of plenty of microorganisms for the identification of the complete set of secondary metabolite gene clusters that these organisms code for. This strategy of novel secondary metabolite discovery is also called the "genomics-based"

approach. One of the major advantages of this "genomics-based" approach is that - using whole genome sequencing the entire set of secondary metabolite genes can be retrieved and hence nothing is missed out due the cryptic nature of expression of certain gene clusters. This strategy also reduces the possibility of re-discovery of already known compounds to a considerable extent. Thus, genome mining has opened up endless possibilities for further research in secondary metabolism. This approach will now allow investigators to explore the untapped potential of a huge reservoir of uncharacterized genes related to natural product biosynthesis [16].

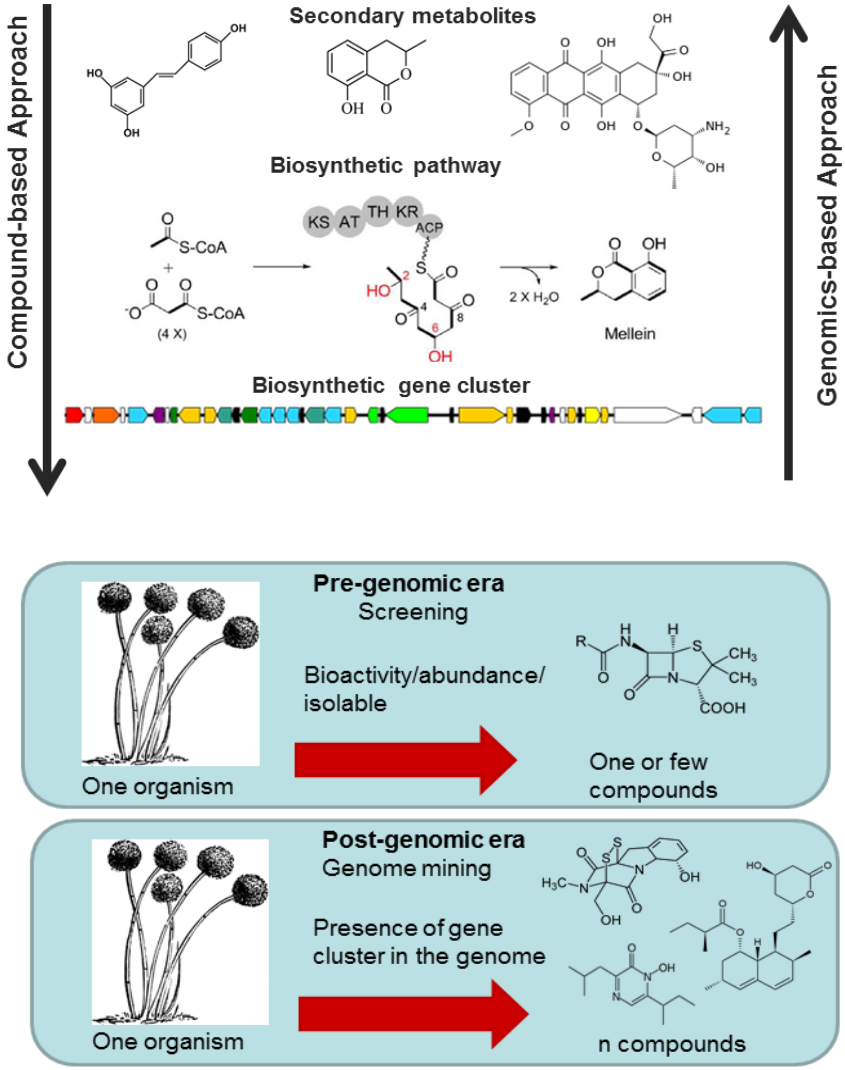


Figure 1.5: Strategies for natural product discovery. The first panel shows the difference between "compound-based" and "genomics-based" approach of secondary metabolite discovery. The second panel shows the flow of events during the pre-genomic secondary metabolite discovery era. Illustration adapted from [13].

Automated bioinformatics platforms now facilitate the semi-automated prediction of secondary metabolites encoded by secondary metabolic blueprints [68, 69].The sequenced genomes can further be annotated and gene cluster predictions can be done using advanced tools like antiSMASH (antibiotic And Secondary Metabolite SHell) [68], RAST (rapid annotation using subsystem technology) [70] and SMURF (Secondary Metabolite Unknown Regions Finder) [71]. For identification of genes responsible for the synthesis of secondary metabolite backbone, a number of other tools like SBSPKS (Structure Based Sequence Analysis of Polyketide Synthases) [72], NRPSPredictor [73], CLUster SEquence ANalyzer (CLUSEAN) [74], ClustScan [73] and Natural Product searcher (NP searcher) can be used [75]. After annotation and prediction of gene clusters there can be three categories of clusters – (1) the clusters which bear significant homology to certain known and characterized clusters, (2) clusters that do not have any significant homology to any characterized cluster (but have homologs in other species) and hence are of unknown function and (3) the cluster is completely novel and has not been found in any of the reported genomes so far. The gene clusters under category (2) and (3) can be further characterized using various molecular manipulation tools. The rapid advancements in recombinant DNA technology and molecular biology techniques towards the end of the 20^{th} century has also contributed to the success of the "bottom–up" strategy of secondary metabolite discovery immensely.

The search for entire gene clusters (pertaining to secondary metabolism) is also driven by the fact that the knowledge of all the enzymes (coded by the genes) involved in the building of a certain compound will only facilitate the elucidation of the entire biosynthetic pathway that the microorganism makes use of to synthesize the concerned molecule. The understanding of the intricate details pertaining to the mechanistics involved in small molecule manufacturing can pave the way for the synthetic or in-vitro synthesis of novel compounds using the well-studied enzymes. In fact, sometimes a few of the ancillary enzymes are capable of imparting extremely uncommon modifications in the small molecule backbone and the know-how of their functional behavior can be of immense use to the field of enzymology. For example, trioxacarcin A contains an unusual polycyclic ring fused to a spiro-epoxide [76]. The synthesis of this compound is accomplished by a type II PKS containing a novel starter unit and series of modifications involving more than 25 tailoring enzymes, many among which seem to be quite unique [76].

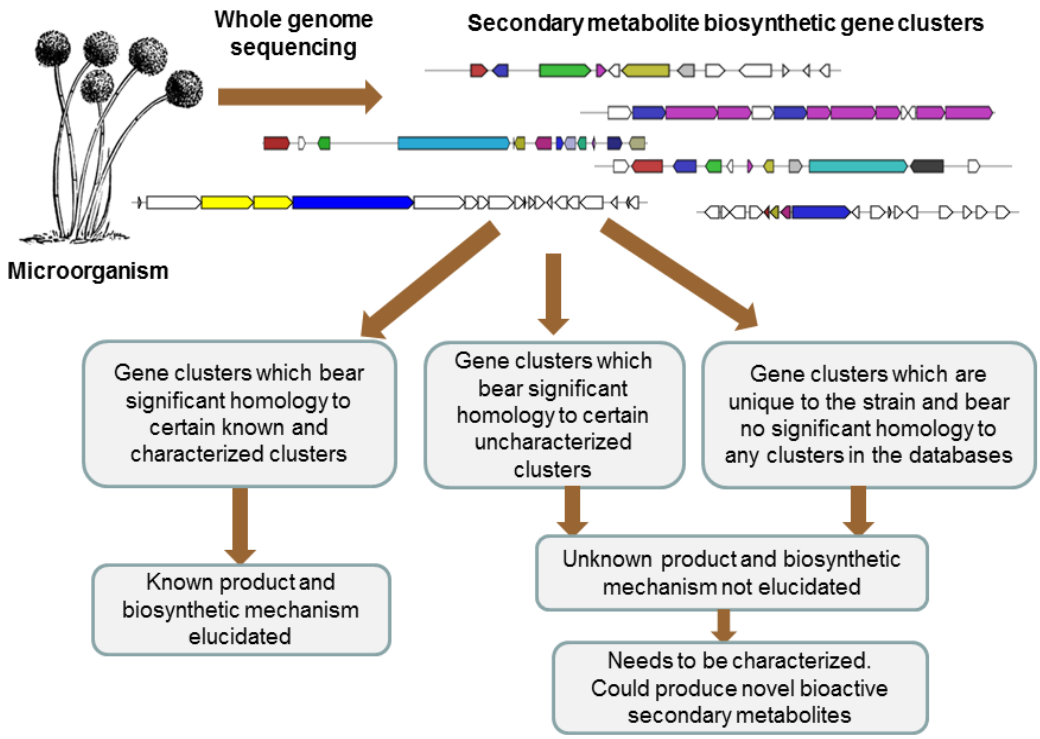


Figure 1.6: A schematic representation of the fate of all the secondary metabolite biosynthetic gene clusters after being identified by genome mining.

Apart from whole genome sequencing of microorganisms, metagenomics have also proved to be quite useful for secondary metabolism research. Metagenomics involves the study of genetic material from environmental samples. The genetic material isolated from a particular environment contains the collective genomes of all types of microbes inhabiting that particular habitat [77]. The DNA/RNA extracted from such samples can be used for the construction of libraries and can be sequenced. A number of PKS and NRPS genes have been isolated using metagenomics and have been characterized later. This approach could be quite handy for the isolation of genetic materials of unculturable microorganisms as well [78]. The 16S rRNA sequencing of these isolates has shown the presence of a number of species that have not been reported or cultured before in the labs. A significant number of small molecules including antibiotics have been discovered using this approach [79] [80]. The antibiotic turbomycin A and B have been isolated using metagenomic library screening made of soil isolates [81].

Which microbes to choose for genome-mining?

There is a debate as to which microorganisms would be the best source for discovery of novel secondary metabolites among scientists. While some believe that unculturable microorganisms might serve as the best reservoir of "untapped" resources for novel secondary metabolites discovery [82], others are of the opinion that exploration of new habitat could be a good way to look for novel metabolite producing microbes. On the other hand there are also some who opine that the taxa of microorganisms which have had a history of producing a variety of important small molecules should continue to be the ideal candidates for secondary metabolism research. With the advent of affordable sequencing platforms, it has become possible to explore the metabolic capacity of different groups of microorganisms simultaneously. In the past few years; marine microorganisms have become a very important source of novel secondary metabolite search. The void in the discovery of new secondary metabolites has inspired scientists to explore new terrains and habitats (from sub-Saharan desserts to the ice-packed Antarctica) that have not been explored so far. Microbes found associated to sponges and corals have been known to be prolific producers of secondary metabolites for quite some time now [83-85]. Oceans are home to an enormous diversity of microbes and hence, marine actinomycetes and fungi are increasingly becoming a popular subject of study for their ability to produce new bioactive compounds. The anti-quorum sensing compound N-*(2′*-phenylethyl)-isobutyramide and 3-methyl-*N*-(2′-phenylethyl)-butyramide have been isolated from the marine bacterium *Halobacillus salinus* [86].

A number of groups have tried to study the metabolic capabilities of "unculturable" microorganisms using diffusion chamber/microbial traps, metagenomics and isolation chips (ichips) [87] [88]. The latest antibiotic discovered in 2015, teixobactin has been found using the ichips technique [89]. "Unculturable organisms" are microbes that are almost impossible to culture under laboratory conditions and they need a medium that very closely mimics their natural environment for growth. In spite of the fact that this set of organisms are replete with the ability to produce a number of different metabolites, their culturing remains one of the bottle-necks for harnessing their secondary metabolite production capacity [87].

Interestingly, some studies have suggested that the presence of MbtH homolog in the genome of an actinomycete acts as an indicator of its metabolic potential [64]. MbtH

encoded by *Mycobacterium tuberculosis* offer unique cue for identification of "gifted microbes". This is a chaperone like protein which is known to stimulate adenylation reactions involving NRPS systems. Expression of MbtH-like protein is essential for the optimal production of secondary metabolites produced by mechanisms that employ NRPS enzymes [90]. Hence, the presence or absence of MbtH homolog can also be an indiactor for the selection of a particular microorganism for secondary metabolism studies.

Leveraging genomics with other techniques

Gene knockout study is one of the most widely used methods to validate the function of an uncharacterized gene clusters. The step by step deletion of genes in a secondary metabolite biosynthetic gene cluster can not only serve to determine gene cluster boundaries, it may also help in the identification of intermediates of the pathway hence leading to the elucidation of entire mechanism of biosynthesis [91].

Certain genes of the secondary metabolite cluster under study can be expressed in heterologous hosts like *E. coli*, yeast, Streptomyces and filamentous fungi for expression. The expression can be of two kinds: the first involves the over-expression and purification of protein in heterologous host. In this particular case, the enzyme is purified and is mixed with its predicted substrates and allowed to react for the identification of the product it synthesizes [92]. The second type involves the expression of a particular gene or the entire cluster in a heterologous host under the influence of an inducible promoter and the subsequent isolation of product from the culture broth extract [93] . *Streptomyces coelicolor* allows the expression of an entire gene cluster and thereby enables its characterization [94]. Likewise, the *A. nidulans* LO7020 strain [93] has been specially designed for the heterologous expression of secondary metabolite gene clusters followed by the identification and purification of the corresponding small molecule product.

The products encoded by some of the gene clusters have very low yield under standard laboratory conditions and hence it is difficult to identify and characterize them. Addition of an inducible promoter upstream of such genes both in homologous and/or heterologous host can facilitate an easier isolation for the identification and characterization of such products and elucidation of biosynthetic pathways. This strategy also holds good for the study of silent or cryptic gene clusters. The novel metabolite, Pyranonigrin E in *A. niger* have been identified using this technique [95].

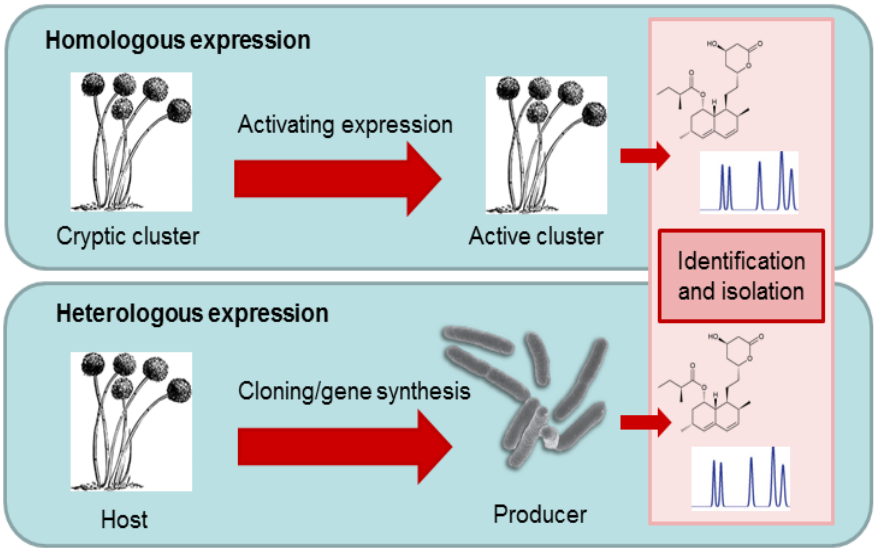


Figure 1.7: Two major approaches for investigating gene clusters identified by genome mining efforts. In heterologous expression, whole gene clusters are mobilized into expression strains and analyzed to identify new compounds. In homologous expression, endogenous transcriptional, translational, or metabolic elements are manipulated to activate secondary metabolite production. Both approaches involve identification of new peaks within a metabolome and isolation of these compounds for structural elucidation. Illustration adapted from [13].

Certain techniques involve a combination of genomics and addition of radio-labelled substrates to the culture medium followed by isotope-guided fractionation to isolate the products [96]. This method also known as "genomisotopic approach", is extremely useful for the characterization of orphan gene clusters (i.e. gene clusters that have not been characterized yet). This method can be used to identify compounds encoded by large gene clusters, which are either difficult to express in their entirety in a suitable host, or produce compounds which lack a predictable bioactivity. Since the genomisotopic approach does not require any prior information on the activity or structure of a secondary metabolite encoded by the gene cluster under study, it offers a powerful approach for the identification of novel compound classes encoded by orphan gene clusters [96].

A number of studies have indicated the effect of pH, light, type of carbon & nitrogen source present in the medium and temperature on the production of secondary metabolism [97]. Hence, altering these parameters while culturing microorganisms is important for the study of secondary metabolism. On the other hand, mimicking the natural environment; for example: addition of filtered sea water for culturing marine microorganisms can be useful in certain situations. In some cases co-culturing has been reported to enhance secondary metabolite production. It has been observed by some

groups that the addition of mycolic acid containing bacteria can increase the amount of heterologous secondary metabolite production in *Streptomyces lividans* [98]. Microorganisms are known to grow in structured communities and hence co-culturing of different microorganisms can also help mimic the natural environment resulting in the synthesis of the desired metabolite [99]. A study has shown that the addition of Gram-negative bacterial cell wall lipopolysaccharide might stimulate and modulate fungal secondary metabolite expression [100]. According to this study 15% of fungi tested responded positively to an optimal exposure of lipopolysaccharide resulting in either activation or enhancement in the production of secondary metabolites [100].

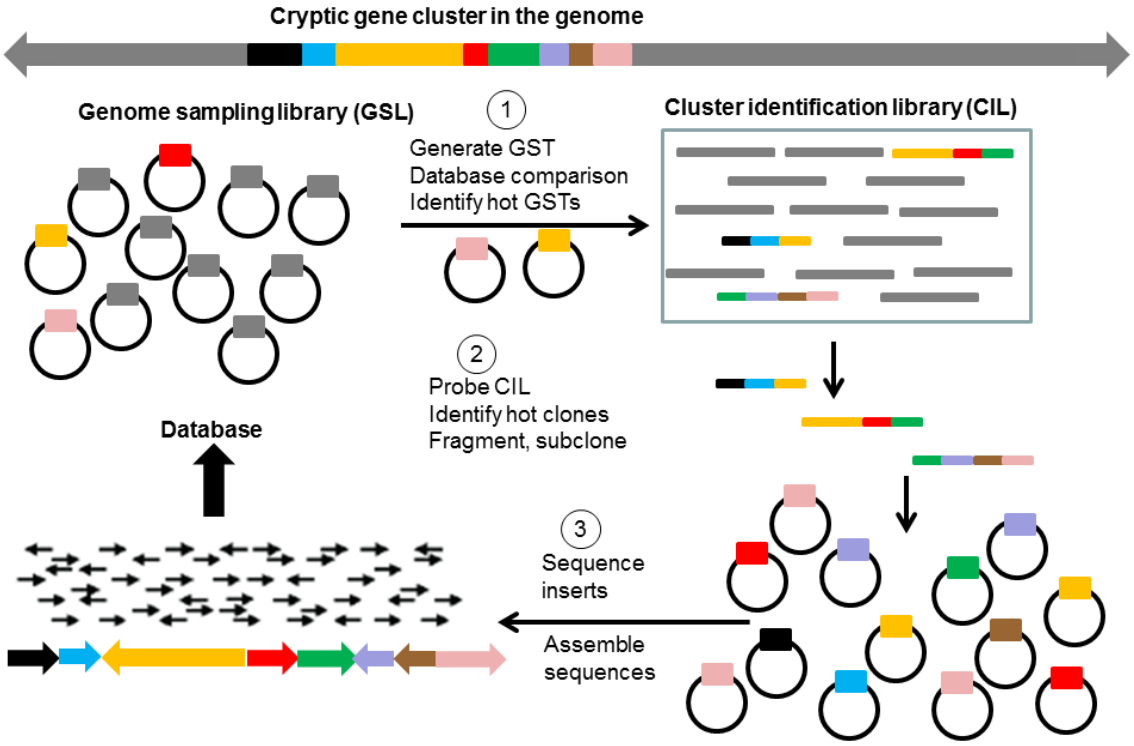


Figure 1.8: Genomics-guided approach for discovery and expression of cryptic metabolic pathways. Secondary metabolite biosynthetic genes are clustered in the microbial genome (only a single gene cluster is shown in the illustration). Genomic DNA is randomly fragmented and small fragments are used to prepare a genome sampling library (GSL) in a plasmid vector and the large fragments are used to prepare a cluster identification library (CIL) in a cosmid or BAC vector. Gene sequence tags (GSTs) are generated from the GSL clones using a universal primer located in the plasmid vector. The GSTs are compared to a database of secondary metabolite biosynthetic genes to identify tags derived from genes involved in secondary metabolite biosynthesis ("hot" GSTs, colored inserts; step 1). These genes are then used as probes to identify CIL clones containing the corresponding genes as well as their neighboring genes ("hot" CIL clones). Overlapping CIL clones may be identified by restriction fragment length mapping or during the subsequent sequencing step. The hot CIL clones are randomly fragmented and used to prepare a second plasmid library that provides templates for sequencing (step 2). Sequencing and assembly of the selected CIL clones result in a complete natural-product gene cluster that is then annotated and stored in the database (step 3). Illustration adapted from [101].

Some groups have reported the formulation of certain high-throughput genome scanning techniques that would enable the fast discovery of bioactive secondary metabolic loci independent of their expression (See Fig: 1.8) [101]. In this kind of approach, generally a shot-gun DNA library is created with short random genome sequence tags (GSTs). This method relies on the fact that all genes related to the biosynthesis of a particular secondary metabolite are clustered together. GSTs are derived from the genes that are likely to be involved in the biosynthesis of secondary metabolites. They are identified by sequence comparisons to a database of microbial gene clusters known to be involved in secondary metabolite biosynthesis [101]. Selected GSTs are then used to design screening probes to identify cloned subgenomic fragments (for example, cosmids or bacterial artificial chromosomes [BACs]) containing the genes of interest as well as the neighboring genes that together may constitute a biosynthetic gene cluster. Genome scanning method provides quite a comprehensive way to discover secondary metabolite gene clusters because the analysis of a relatively small number of GSTs provides reasonable assurance of full genome coverage [101].

A closer look at the uncharacterized gene clusters can sometimes hint at the possible nature of the final product they produce. A careful examination of the type of backbone synthesizing enzyme, prediction of substrate specificity of adenylation (A) domains in-case of NRPS containing clusters and the type of accessory genes present can be quite handy for an approximate product prediction. This kind of very preliminary prediction can help researchers to shortlist a set of gene clusters for further study amongst the vast array of clusters available in the databases. Analytical methodologies like TLC (thin layer chromatography), HPLC (high resolution liquid chromatography), High resolution LC-MS (liquid chromatography mass spectrometry) and NMR (nuclear magnetic resonance) are very useful and the most frequently used platforms for the isolation and structure elucidation of secondary metabolites in the culture broth of homologous or heterologous hosts. The compounds thus identified, if novel is thereafter subjected to various bioassays for the exploration of their biological properties [6, 102]. A considerable part of the success of secondary metabolism studies depends on the quality of separation and sensitivity of the aforementioned analytical tools.

Although there has been a significant shift towards genomics-guided techniques along with molecular biology methods for the study of secondary metabolism, on a number of occasions a combination of genome mining and classical activity-based screening for

natural product discovery, generation, and hyperproduction is used. This approach to combine the two methods of secondary metabolism study can help to maximize the possibilities of harnessing the metabolic capabilities of the microorganisms under study.

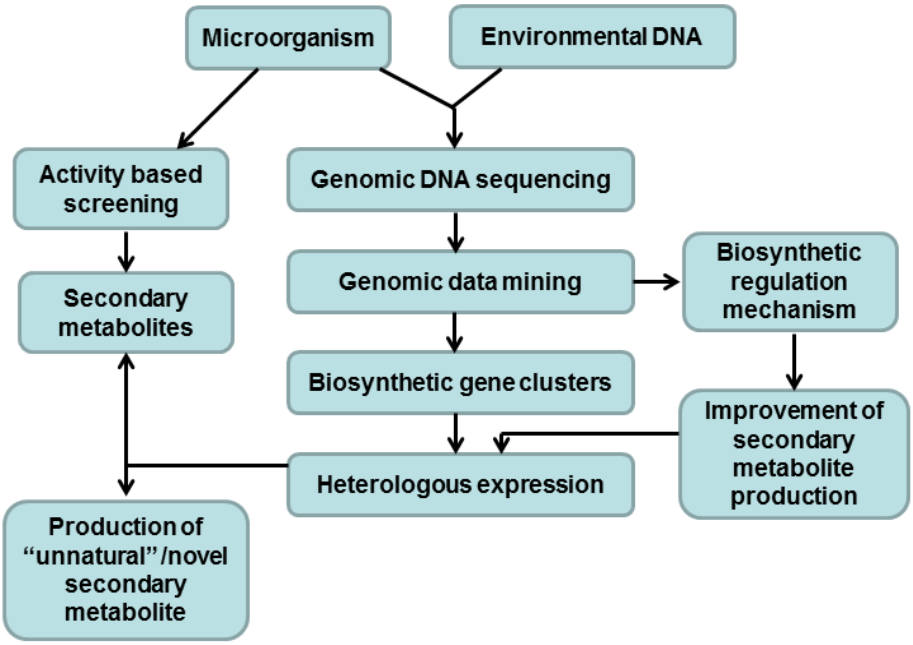


Figure 1.9: Combination of genome mining with classical activity-based screening for natural product discovery, generation, and hyperproduction. Illustration adapted from [103].

Indeed, genome mining now includes the full spectrum of the bioinformatic prediction of gene and pathway function, the control of gene expression and translation, and the identification and structural elucidation of new metabolites from within the metabolome of the producing microorganisms [13]. As already evident, genome mining studies have become more than just secondary metabolite discovery as they encompass the comprehensive understanding and manipulation of cellular systems at the molecular level [13].

Genomics and combinatorial biosynthesis of novel compounds

Another area that holds considerable promise for secondary metabolism research is enzyme engineering for the synthesis of novel secondary metabolite structures [16]. Using genetic engineering tools to modify some of the iPKS genes and subsequently study them using biochemical approaches is one way of synthesizing new secondary metabolites with novel bioactive properties [104]. Iterative PKSs are extremely

interesting for study from the biochemistry point of view as they use the same set of domains iteratively with the property of selective modification of substrates at different chain lengths. The genetic manipulation of these enzymes could include addition or deletion of domains, tuning substrate specificities by mutations in active sites or swapping of domains from another PKS gene [104]. This kind of study could serve a dual purpose: (a) be of relevance in understanding the detailed molecular mechanism in polyketide biosynthesis and probing the versatility of PKS; (b) generate a library of novel polyketide structural analogs by engineering the PKS. Although, NRPS functions in a modular fashion, the discovery of certain iterative NRPS have also added to the mystery of natural product biosynthesis [105, 106]. Additionally, compounds with much greater structural complexity are manufactured by the PKS-NRPS hybrid enzymes. These hybrid systems can be found both in modular as well as iterative forms. BaeJ gene of *Bacillus amyloliquefaciens* is one of the examples of bacterial modular PKS-NRPS hybrid [107]. Very few of the iterative PKS-NRPS enzymes that have been discovered have been characterized and very little knowledge is available on the programming involved in their functioning [108]. These are some of the questions that enzymologists are looking into. Enzyme engineering and combinatorial biosynthesis might help in answering some of these questions in future. Genetic manipulation of PKS and NRPS enzymes can be an attractive alternative for the generation of structural and functional diversity in secondary metabolites using semi-synthetic strategies [13]. Genome mining has provided a wealth of new NRPS and PKS sequences in order to facilitate the expansion of the secondary metabolite repertoire using synthetic biology approaches for combinatorial biosynthesis of novel NRPS and PKS pathways. The same strategy can be used to manipulate NRPS/PKS hybrid systems, and ancillary reactions. The study of mechanisms involved in the generation of complex secondary metabolite structures has been facilitated by the availability of enormous genomic data, advanced genetic manipulation, analytical methodologies and the various synthetic biology tools [13].

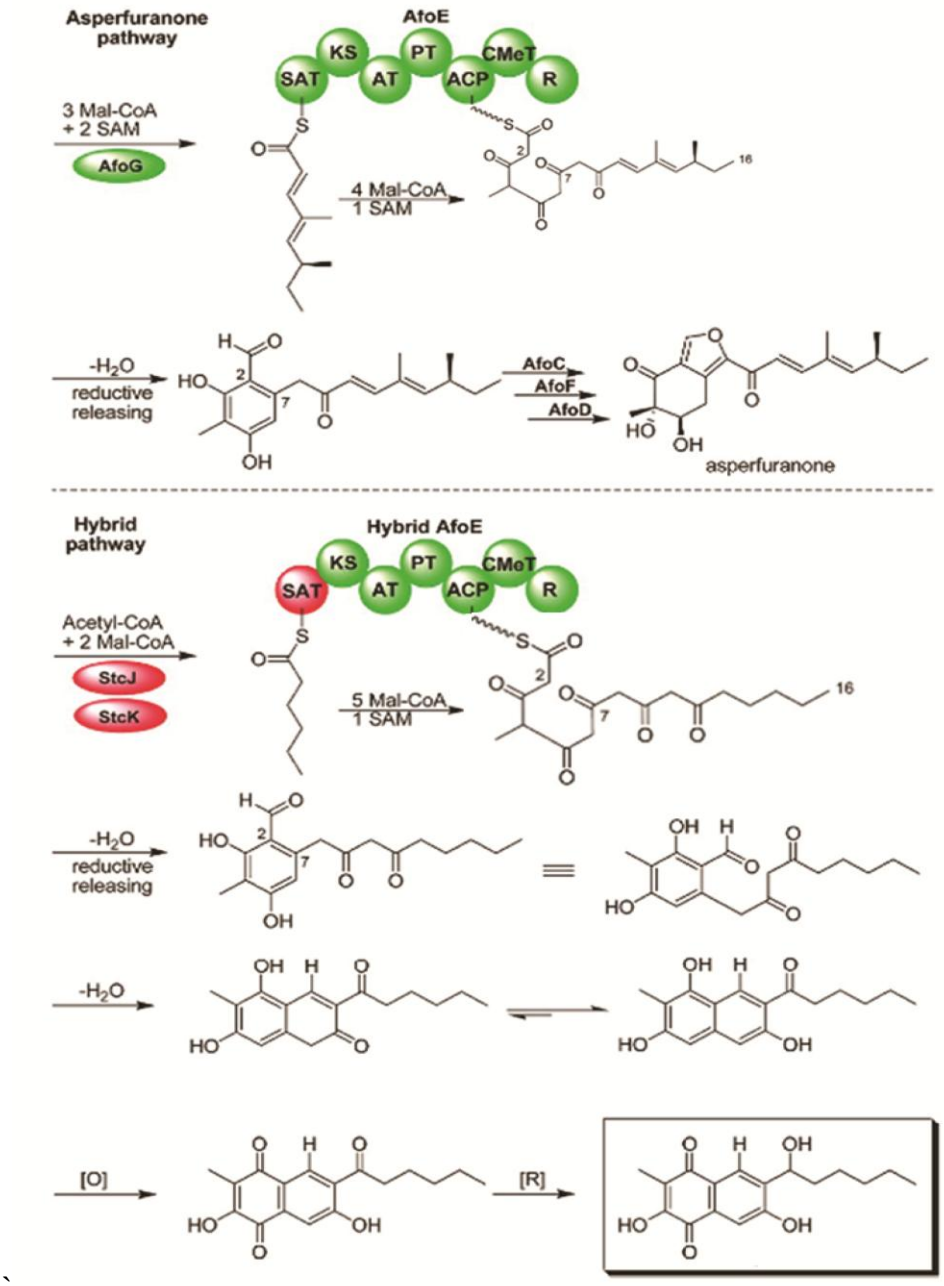


Figure 1.10: Asperfuran biosynthesis from the enzyme AfoE (on the top) and synthesis of a novel metabolite when the SAT module of AfoE is swapped with the SAT module of another PKS; StcA involved in sterigmatocystin biosynthesis (at the bottom) [109].

Synthetic biology, metabolomics and secondary metabolism

Synthetic biology, entails re-designing of biological systems for novel purposes and applications. It aims to enable the successful transfer of a secondary metabolite biosynthetic pathway from its organism of origin into heterologous hosts which is more amenable, and where the product of interest can be produced in desired quantities [16]. In

the last few years, there has been a quick emergence of synthetic biology as a platform for the reconstruction of biological systems (e.g., to assemble a novel biological regulon or to regulate gene expression in response to a specific input). In the industrial context, synthetic biology offers some tremendous scope for the creation of cell factories tailor-made for efficient production of molecules of interest [19]. Engineered strains of *S. cerevisiae* have been used for the production of stilbinoids (resveratrol), carotenoids, polketide (6-methylsalicylic acid), alkaloids (reticuline) and isoprenoids biosynthesis [110]. The future of synthetic biology with respect to secondary metabolism would involve engineering of a catalogue of production hosts that are engineered to supply enough amount of precursors (especially substrates like amino acids, acetyl-CoA and IPP, and of cofactors such as NADPH) to facilitate the overproduction of metabolites for introduced heterologous pathways [110]. These engineered production strains should also have the ability to allow efficient uptake of substrates, retention and compartmentalization of intermediates, and export of products. Knowledge gained from characterizing the function of native and introduced biosynthetic enzymes in yeast or other hosts will provide the parameters to enable the development of ever more advanced computational models of metabolic flux, thereby increasing the predictability of pathway engineering [110]. Thus, the tools emerging from synthetic biology have the exciting potential to advance the application of yeast or other organisms as a production host for the economically feasible and commercial-scale manufacturing of bioactive microbial secondary metabolites [110].

Metabolomics is the complete analysis of all metabolites present in a biological sample. As metabolomics is the final step in the "-omics" cascade, closest to the phenotype, it provides a direct snapshot of the physiological status of the cell at a certain time point and under specific conditions [111]. Advances in metabolomics have led to headways in analytical techniques in combination with development of suitable softwares for the interpretation of results [112]. Mass spectrometry has always been a favorite tool for metabolomics studies. The high mass accuracy and sensitivity to identify traces of compounds makes it a method of choice. MS is used in combination with liquid chromatography (LC-MS), with gas chromatography (GCMS) and capillary electrophoresis (ECMS) depending on the sample under study [113]. Metabolomics can be used for the identification and quantification of products of an orphan gene clusters expressed in an engineered strain. Comparative metabolomics can be useful for the

characterization of secondary metabolism gene clusters of unknown function [16]. By comparing the metabolic profile of a wild type strain and the corresponding mutants (in which a silent pathway has been activated), novel metabolites can be identified. Quite a few new small molecules and their isomers have been characterized as products of cryptic biosynthetic gene clusters using metabolomics in recent years [114-117]. Detection of novel metabolites using (high or ultra-performance) LC along with (ultra-) high resolution MS, subsequently followed by structural elucidation by NMR seems to have emerged as the most powerful strategy for compound discovery using metabolomics[110].

Metabolomics can have further applications in secondary metabolism. Identification of secondary metabolites by global metabolomics analysis is another strategy that can substantially contribute to the identification of metabolic bottlenecks in engineered biosynthetic pathways. Metabolic engineering has been applied to fine-tune the synthetic routes to obtain optimum product titers [19]. The commonly found bottlenecks in engineered biosynthetic pathways are: (1) the depletion of precursors, (2) excessive flux-shift towards by- product formation; (3) poor catalytic capacities of one or more enzyme(s), (4) incapacity to accommodate the high-volume flux to the targeted molecule; and (5) the accumulation of toxic intermediates [118]. The identification and quantification of by-products and precursors of a metabolic pathway could help biologists to understand the key reactions that would require further optimization [16].

Genomics and future of secondary metabolism research

The past decade has witnessed exceptional increase in genome sequencing information, and genome mining analyses which has unearthed the tremendous metabolic capability of microorganisms to produce bioactive secondary metabolites [103]. Studies on the biosynthetic mechanism and regulatory elements underlying secondary metabolite biosynthesis will maximize the exploration of secondary metabolites. Today, research on secondary metabolites has seen a drastic shift from pure chemical analysis to the combination of genetic manipulations and chemical synthesis, and the diversity of secondary metabolites can be further enhanced by genetic engineering of biosynthetic genes or gene clusters [103]. Experimental techniques for ways to discover the products of cryptic gene clusters continue to be developed and refined [119]. One of the important challenge that lies in front, is the development of general methods for the activation and

expression of silent or cryptic biosynthetic gene clusters [119]. Much of the biosynthetic chemistry that has been uncovered in the last 20 years has been possible because of the genomics-guided approach [13] [119]. The accelerated progress in this field of secondary metabolism can be largely attributed to the avalanche of genetic information available to us today. Therefore, it is needless to say that, a lot more novel, exciting and intriguing biosynthetic chemistry remains to be discovered in the future through the continued exploration and exploitation of information from genome sequencing projects [13].

Indeed, secondary metabolite research has progressed manifolds in the past century. Thanks to the sequencing revolution, today we have an enormous repertoire of uncharacterized secondary metabolite gene cluster sequences known to us. In spite of this, there lies a zone of the vast unknown in this field of secondary metabolite discovery that remains to be explored using a combination of all the approaches discussed so far.

Naturally occurring dihydroisocoumarins and their biosynthetic mechanism

Naturally occurring dihydroisocoumarins are derivatives of isocoumarin and have been found to exhibit a wide spectrum of biological activities such as cytotoxicity, antimicrobial, anti-diabetic, anti-tumor, antioxidant and anti-ulcerative properties [120]. The diverse biological activities of dihydroisocoumarins arise from the different substituents attached to the basic isocoumarin backbone [121]. The dihydroisocoumarin-containing ochratoxins A produced by fungal species exhibit carcinogenic, teratogenic and nephrotoxic properties [122-124]. Cladosporins which also contain the isocoumarin backbone, are fungal secondary metabolites having a wide spectrum of antimicrobial activity [125]. NM3 is an isocoumarin compound reported to have angiogenesis inhibiting properties [126]. *Xenorhabdus nematophila,* a bacterium that inhabits the gut of nematodes has been found to be the source of two important antimicrobial compounds xenocoumacin 1 and 2 [50]. And while the compound amicoumacin A exhibits anti-ulcerative, anti-inflammatory activity [127] and can also inhibit methicillin resistant *Staphylococcus aureus* (MRSA) [128], amicoumacin B has been implicated in increasing levels of bone morphogenetic factor 2 needed for bone development [50] and bacilosarcin A is reported to have antiplasmodial activity [129].

The dihydroisocoumarin ring is generated by different classes of PKS enzymes in different organisms. A partially reducing iterative PKS from *Saccharopolyspora erythraea* has been found to produce mellein [92], but in *Xenorhabdus nematophila* a

modular PKS synthesizes the isocoumarin ring of xenocoumacins [50]. Our lab has been involved in the study of the biosynthetic mechanism of dihydroisocoumarin-containing natural products (See Fig: 1.8). At the beginning of my Ph.D. project, we chose the organisms *Aspergillus westerdijkiae* [130, 131] and *Nocardia jinanensis* for our study because they were reported to produce the dihydroisocoumarin-containing compounds ochratoxin A and amicoumacin B [132]. The biosynthetic mechanisms of ochratoxin A and amicoumacin B remained unknown at that time.

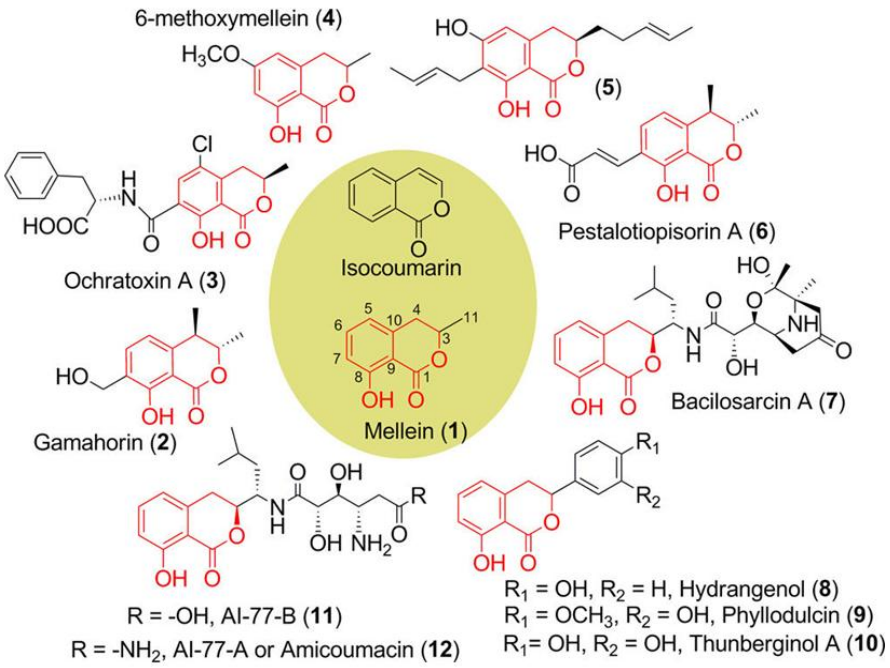


Figure 1.11: Isocoumarin and derivatives of naturally occurring dihydrocomarins. Mellein (1) exhibits anti-fungal, anti-bacterial and HCV protease-inhibiting properties; bacilosarcin A (7) exhibits potent anti-plasmodial activity; hydrangenol (8), phyllodulcin (9) and thunberginol A (10) promote adipogenesis of 3T3-L1 cells and exhibit anti-diabetic properties [92] (Illustration obtained using ChemDraw tool).

Objectives

The filamentous fungus *A. westerdijkiae CBS112803* is a major producer of the neurotoxic mycotoxin ochratoxin A (OTA) and potentially contains many other biosynthetic gene clusters. Although a huge amount of information is available on the toxicity, detection and occurrence of OTA in food and other agricultural products, the biosynthetic and regulatory mechanisms of OTA remain to be fully understood. *N. jinanensis* is a rare actinomycete that produces the bioactive compound amicoumacin B

according to a recent report [132]. In this study, my first objective was to identify the biosynthetic gene clusters responsible for the production of the two naturally occurring dihydroisocoumarins and elucidate the biosynthetic mechanisms. The approach involved the sequencing of the microbial genomes, bioinformatic analysis to identify the biosynthetic gene clusters and fast metabolite profiling. Aided by bioinformatic analysis, I planned to elucidate the biosynthetic mechanisms and function of the enzymes involved in the biosynthesis of the two secondary metabolites. Gene knockout and inactivation as well as in vitro enzymatic assays were employed to probe the function of the biosynthetic enzymes. My second objective was to elucidate the biosynthetic mechanism of some of the other secondary metabolites produced by *A. westerdijkiae* and *N. jinanensis*. While the identification of new compounds and elucidation of biosynthetic mechanism remained my main objective, I also planned to perform comparative analysis of the genomes of the *N. jinanensis* and *A. westerdijkiae* with other closely related species which would shed substantial light on the important factors underlying the physiology and pathogenicity of the two microorganisms.

Chapter 2 - The complete genome of *Aspergillus westerdijkiae* and secondary metabolism

Introduction

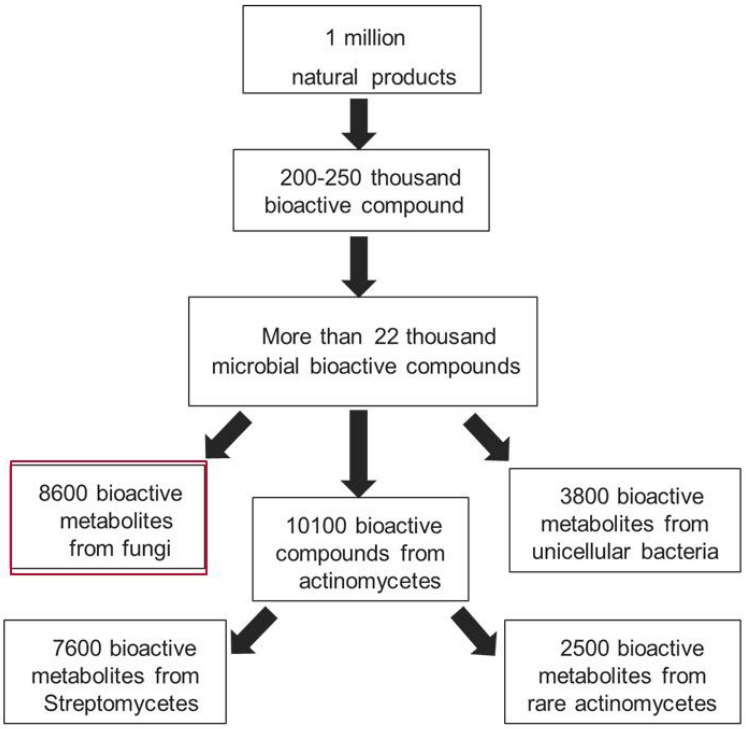


Figure 2.1: Distribution of known bioactive compounds according to their sources. Illustration adapted from [133].

Filamentous fungi are prolific producers of a wide range of secondary metabolites. Quite a large number of frequently prescribed medicines for humans are of fungal origin. According to a review published in 2005, 1500 fungal derived secondary metabolites have been isolated in between 1993-2001 alone [30]. The functional diversity of fungal secondary metabolites and their ability to bind to biological targets makes them crucial players in pharmaceutical research [30]. Fungi isolated from different sources have been investigated for secondary metabolite discovery. As evident from Fig: 2.1, fungi and actinomycetes account for a huge share in the bioactive metabolite production and hence have been a subject of intense study for scientists investigating secondary metabolism. Although, all fungi have the metabolic potential to produce an array of interesting

secondary metabolites, members of the class Ascomycetes have been the most prominent ones. So far, a huge repertoire of secondary metabolites of clinical relevance has been isolated from this class of filamentous fungi. The genus Aspergillus belongs to the class Ascomycetes and has a genome size of 30-40 megabases. *Aspergillus sp*. are ubiquitous in nature and are known to produce myriad bioactive molecules such as toxins, allergens and antimicrobials. Most of the Aspergilli are pathogenic to humans, animals and plants. Some of the members of the genus Aspergillus like *A. fumigatus* is known to cause aspergillosis of lungs and opportunistic infections in humans [134]. Aflatoxins and ochratoxins are some of the major fungal toxins produced by this group of Ascomycetes. The aflatoxin producing *A. flavus* and *A. parasiticus,* and ochratoxinogenic *A. niger, A. ochraceus* and *A. carbonarius* species are frequently encountered in agricultural products [135].

The genus Aspergillus is one of the major sources of pharmacologically relevant molecules and includes the cholesterol lowering lovastatin, the anti-cancer drug orsellinic acid, antibiotics like emericellamides, immunosupressants like cyclosporin and so on [136]. Lovastatin and its semi-synthetic derivatives under various brand names like zocor and lipitor constitute one of the best selling classes of prescription medication. Numerous other bioactive molecules have been isolated from Aspergilli and the list continues to grow with discovery of every new species and strain. Apart from being prolific producers of secondary metabolites, Aspergilli are also industrially exploited for the large scale production of various enzymes like lipases, amylases, cellulases and proteinases [137]. While some of these enzymes find uses in the food processing industry, a few others like the lipases and proteases are used as additives in commercial detergents used for laundry. Infact, *A. niger* is the biggest source of commercially available citric acid [138]. Some of the Aspergilli are also used in the fermentation industry for the production of fermented food products such as soyabean paste [139].

As a result of their commercial importance a greater knowledge of Aspergillus genomics can pave the way for (1) the discovery of new anti-fungal drug targets and virulence factors, (2) unraveling many unknown secondary metabolite gene clusters which remain silent under laboratory conditions and hence go undetected, (3) the identification of entire gene clusters could facilitate better understanding of the biosynthetic mechanism involved in the synthesis of important bioactive molecules, (4) genome comparisons between related species of fungi can be useful for the study evolution, (5) identification of

different regulatory elements that are involved in secondary metabolism which have not been known so far; and (6) identification of genes coding for unique enzymes which could be of industrial relevance.

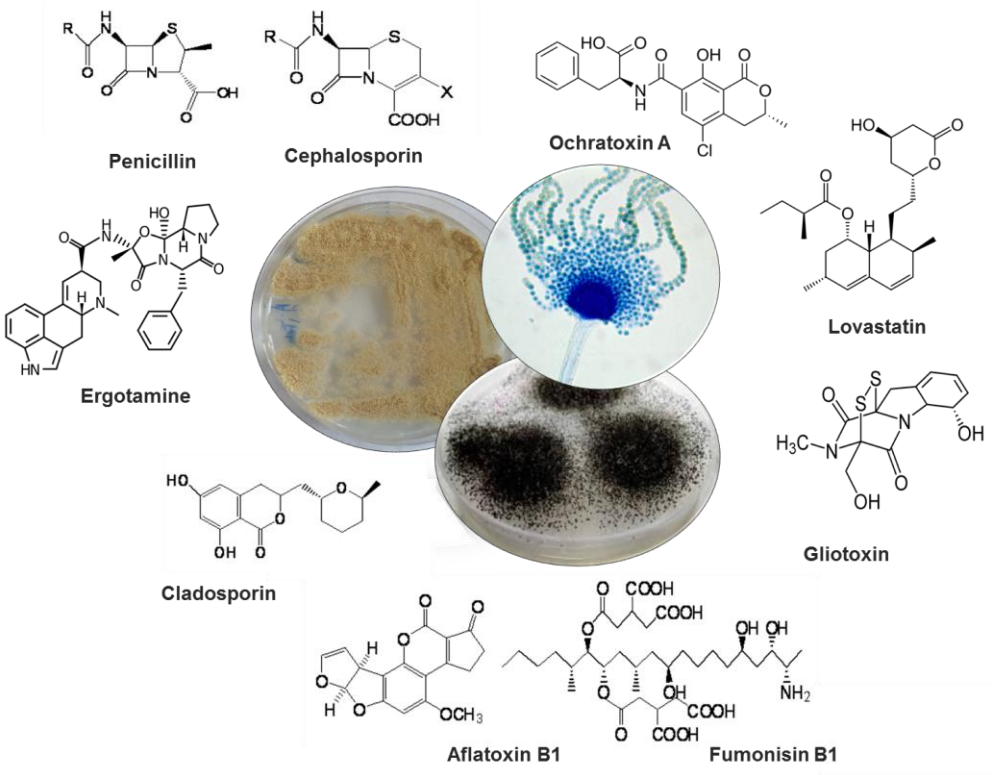


Figure 2.2:. Structures of some important secondary metabolites produced by fungi.

The current focus on fungal genomics coupled with advances in molecular manipulation techniques and better detection tools have resulted in the galvanization and revitalization of research on fungal secondary metabolism. Cutting edge genome sequencing tools are being used frequently these days for the detection of the complete set of secondary metabolism genes present in an organism. The whole genome approach has proved to be an extremely competent tool for the greater understanding of fungal secondary metabolism, the knowledge of which can be utilized in various pharmaceutical interests [140]. This technique has attracted importance for its ability to unravel cryptic secondary metabolite gene clusters which could have the potential to produce novel small molecules of relevance. Today, the search for new secondary metabolites has become more necessary due to the high re-discovery rate of previously known compounds. The

genomics guided approach has led to a considerably better understanding of the genetic basis of small molecule biosynthesis. The availability and accessibility of a large number of fungal genome sequences on different databases has made it possible to bypass the traditional screening approach [141] of secondary metabolite discovery. The usage of bioinformatic tools for secondary metabolite discovery is becoming more and more widespread. The convergence of genomics, microbiology and chemistry holds the promise of being a more comprehensive method for studying microbial secondary metabolism.

Perhaps because of being one of the most harmful as well as useful genera of fungi, in the past few decades a large number of Aspergillus genomes have been reported. The availability of genomic sequences of various members of the genus Aspergillus has laid the platform for the study of comparative genomics, evolution and functional studies of genes. *A. fumigatus* is a pathogen and produces gliotoxin [142]. The whole genome sequencing of this strain of Aspergillus has lead to a better knowledge of genetics involved in making a successful fungal pathogen. *A. nidulans* [143] is a model organism whose genomic study has enhanced understanding of different aspects of Aspergillus biology including growth and development. *A. oryzae* [144] and *A. niger* [145] are important for a variety of industrial processes and their genome analyses have shed considerable light on the role of fungi as cell factories replete with the ability synthesize commercially relevant molecules. Irrespective of pathogenicity or industrial relevance, one property that remains universal among all members of Aspergillus is their ability to produce a vast array of structurally and functionally diverse secondary metabolites. Multi-species comparative genomics has lead to the identification of orthologs and as well as species specific unique genes which has set the stage for a robust approach to study secondary metabolism. With more than a twenty species of Aspergillus genomes sequenced till date, a very large number of PKS, NRPS, PKS-NRPS hybrid and terpenoid synthesizing gene clusters have been added to the existing inventory of secondary metabolism gene clusters [2] . Most of these gene clusters remain uncharacterized and hence may hold the key to novel secondary metabolite discovery.

A. westerdijkiae is a filamentous fungal species that was initially believed to a member of the *A. ochraceus* taxon. However, unlike *A. ochraceus*, *A. westerdijkiae* is unable to grow at 37 ^{o}C and the white to cream coloured sclerotia produced by *A. westerdijkiae* differ from the pink to vinaceous purple sclerotia of *A. ochraceus* [130] . Also in contrast to the

A. ochraceus and other Aspergillus species such as *A. niger* and *A. melleus*, most isolates of *A. westerdijkiae* are able to produce consistently large amounts of OTA [146]. As mentioned earlier our initial focus was to study the biosynthetic mechanisms involved in dihydroisocoumarin compounds. Since OTA contains a dihydroisocoumarin moiety and *A. westerdijkiae* is a major producer of it, we decided to choose this organism for our study. *Aspergillus westerdijkiae* CBS112803 which was isolated originally from soil samples in South Africa was procured from CBS-KNAW (Centraalbureau voor Schimmelcultures (CBS) Fungal Biodiversity Centre) fungal collection for this study. Some antimicrobial activity assay done by us using the culture broth extract of *A. westerdijkiae* gave us promising results and it gave us a further reason to study secondary metabolism in this organism at length. In this study, we report the whole genome sequence of *A. westerdijkiae*. The genome of *A. westerdijkiae* have been found to encode for a larger number (>60) of biosynthetic gene clusters. Here we elucidate the complete set of predicted secondary metabolite gene clusters that have been found in this organism by AntiSMASH version 3.0 [68]. Further, we present the results of the culture broth extract analysis of *A. westerdijkiae* for the identification of compounds using compound-based approach. Lastly, we have done a comparative analysis of the genome of *A. westerdijikiae* viz.a.viz. the reported genomes of some of the other members of the genus Aspergillus for the better understanding of physiology, virulence and evolution of *A. westerdijikiae*.

Materials and methods

Culturing- *A. westerdijkiae* CBS112803 obtained from Centraalbureau Voor Schimmelculture, Netherlands (CBS) was cultured in LMM and SAM media as stationary cultures for 10-14 days. The composition of SAM medium is: 50 g glucose, 3 g NH_4NO_3 (ammonium nitrate), 26 g K_2HPO4 (di-potassium hydrogen phosphate), 1 g KCl (potassium chloride), 1 g $MgSO_4.7H_2O$ (magnesium sulfate heptahydrate), 10 ml mineral solution (composition per liter of distilled water): 70 mg, $Na_2B_4O_7.10H_2O$ (sodium tetraborate decahydrate), 50 mg $(NH_4)_6$ $Mo_7O_{24}.4H2O$ (sodium molybdate tetrahydrate), 1000 mg $FeSO_4.7H_2O$ (ferrous sulfate heptahydrate), 30 mg $CuSO_4.5H_2O$ (Copper sulphate pentahydrate), 11 mg $MnSO_4.H_2O$ (manganese sulfate), 1760 mg $ZnSO_4.7H_2O$ (zinc sulfate heptahydarte) [147]. LMM medium: 15g/L lactose, 6g/l $NaNO_3$ (sodium nitrate), 0.52g/KCl (potassium chloride), 0.52g/L MgSO4.$7H_2O$ (Magnesium sulphate heptahydrate), 1.52 g/L KH_2PO_4 (potassium dihydrogen phosphate) and 1ml/L trace elements solution [(composition per liter of distilled water) $ZnSO_4.7H_2O$ (zinc sulfate

heptahydrate) 2.2g or $ZnCl_2$ (zinc chloride), 1.04g H_3BO_3 (Boric acid), 1.1g$MnCl_2 \cdot 4H_2O$ (manganous chloride tetrahydrate), 0.5g $FeSO_4.7H_2O$ (ferrous sulphate heptahydrate), 0.5g $CoCl_2.6H_2O$ (Cobalt chloride), 0.16g $CuSO_4.5H_2O$ (Cupric sulphate pentahydrate) 0.16g or CuCl2 , 0.096g $(NH4)_6Mo_7O_{24}.4H2O$ (ammonium molybdate tetrahydrate), 0.11g Na4 EDTA.$4H_2O$ (EDTA, tetrasodium salt) 6.0g, pH6.5] with necessary supplements [148]. All the above chemicals were bought from Sigma, except for KH_2PO_4 (Merck) and $MnCl_2.4H_2O$ (Alfa aesar).

Isolation of genomic DNA - *A. westerdijkiae* was cultured in 50ml flasks containing LMM broth for 5 days at 25 ºC with shaking at 120 rpm. Fungal mycelial mat (2.0 g) was harvested and was ground with liquid nitrogen in a mortar into a fine powder. Buffers A (0.35M Sorbitol, 0.1M Tris-HCl pH 9, 5mM EDTA pH-8), B (0.2M Tris-HCl pH 9, 50mM EDTA pH 8, 2m NaCl, 2 % CTAB) and C (5 % sodium doedecyl sulfate) were prepared for the subsequent steps.17.5ml of lysis buffer (2.5 volume buffer A, 2.5 volume buffer B, 1 volume buffer C, 0.1 % of polyvinylpyrrolidone, 0.1mg proteinase K) was added to the powdered mycelia, mixed by vortexing and incubated at 65 ºC for 30 minutes. After incubation 0.33 volumes of 5M potassium acetate was added followed by mixing and incubated on ice for 30 minutes, proceeded by centrifugation for 20 minutes at 5000g at 4 ºC. One volume of chloroform: isoamyl alcohol (24:1) was added to the supernatant and centrifuged for 10 minutes at 4000g at 4 ºC. The aqueous phase was transferred into a 50ml falcon tube and 100µl of RNAse A (10mg/ml) was added and incubated for 90 minutes at 37ºC. One-tenth volume of 3M sodium acetate and one volume of isopropanol was added and incubated for 5 minutes at room temperature followed by centrifugation for 30 minutes at 4 ºC. The supernatant was discarded; the pellet was washed with 70 % ethanol and centrifuged for 10 minutes at 4ºC. The DNA pellet obtained was dried and resuspended in 500µl of TE buffer (10mM Tris-HCl pH-8, 1mM EDTA).The concentration of the DNA sample was measured using nanodrop and an agarose gel was run with the sample (to check for RNA contamination) before it was sent for whole genome sequencing to Macrogen.Inc, Korea (see Appendix XIV).

Antimicrobial activity assay

Disc diffusion assay- Sterilized filter paper discs were impregnated with 30-40µl of organic extract from LMM medium and left for drying at room temperature inside the laminar air-flow hood. The control discs were impregnated with 100 % methanol and

dried. The impregnated dried filter paper discs were then placed on LB (Luria-Bertanni) agar plates containing the following organisms: *Bacillus subtilis, Staphylococcus aureus, Escherichia coli and Streptococcus pyogenes,* and incubated at 37 °C overnight. After incubation the plates were observed for the detection of inhibition zones around the filter paper discs.

96-well Microtitre plate test- 10 µl of organic extract LMM medium was added on to designated wells and left for drying. 100 µl of bacterial culture (O.D 0.06) was added to assigned wells and incubated at 37 °C overnight with shaking at 150 rpm. The Tecan machine/plate reader was used to read the O.D of each well after incubation. For the antimicrobial activity assay, 14 day old cultures of *Aspergilli* were harvested and the broth was extracted with equal volumes of ethyl acetate after acidification. The organic extract was dried and reconstituted in 100% methanol for all assays.

Extraction of culture broth and LC-MS/NMR analysis - The culture broths (200 × 50 mL, total 10 L) of *Aspergilus westerdijkiae* were combined and centrifuged to separate the supernatant and the mycelia. The supernatant was separated using a C18 column (250 g, Phenomenex, Sepra C18-E bulk packing, 50um, 65A) with two isocratic conditions of 100% water and 100% methanol. The 100% methanol fraction was evaporated to dryness using rotary evaporation and separated by Agilent prep C18 (5u, 30 x 100 mm) reversed-phase preparative HPLC (gradient elution: solvent A (0.1% HCOOH/ H_2O): B (0.1% HCOOH/CH_3CN) 97:3 → 40:60 over 80 min, 40:60 → 0:100 over 15 min; flow rate 30 mL/min).

Whole genome sequencing and Bioinformatics - The whole genome was sequenced at Macrogen Inc., Korea using the Hiseq Illumina platform. The Whole genome sent for sequencing was first tested to check for quality of sample and then fragmented into small pieces for building a mate-paired library. This approach involves the cloning of random fragments of DNA with no information regarding the location of that fragment in the genome. Once the library was successfully prepared the sample was sequenced using MiSeq platform. The sequencing data was later assembled using SOAPdenovo tool and multiple long DNA sequences called scaffolds were obtained. Scaffold consists of a series of contigs that are in the right order but not necessarily connected in one continuous stretch of sequence. Contigs are small pieces of cloned DNA that have overlapping regions. SOAP or Short Oligonucleotide Analysis Package is especially suitable for

Illumina platform and is quite effective in assembling very large genomes.The assembled genome data in the form of scaffolds was further subjected to bioinformatic analysis at our end. The genome data was submitted to antiSMASH 3.0 (antibiotics & Secondary Metabolite Analysis Shell) for secondary metabolite gene cluster prediction [68] and Softberry FGENESH (http://www.softberry.com) [149] for the prediction of introns. The genes were further annotated using NCBI–BLAST [150].

Genome comparison - Genome and proteome sequences of other species presented in this study were all acquired from Aspergillus Comparative Database (3/29/2015). "Aspergillus Comparative Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)". A phylogenetic tree of *A. westerdijkiae* and other eight Aspergillus species was constructed [(*A. fumigatus* [142], *A. nidulans* [151], *A. niger* [145], *A. oryzae* [144], (*A. flavus* [152], *A. clavatus* [153], *A. terreus* [153], *Neosartorya fischeri* [153]]. Orthologous protein prediction was carried out using ProteinOrtho [154]. The maximum likelihood phylogenetic tree was created using the Dayhoff model in the TREE-PUZZLE [155] with 1,000 bootstrap replicates. The tree obtained was edited using Figtree. Carbohydrate-active enzymes and families of proteases were classified using CAZy database (http://www.cazy.org/) and MEROPS peptidase database (http://merops.sanger.ac.uk/) [156], respectively.

Results

General features of the *A. westerdijkiae* genome- The whole genome of *A. westerdijkiae* was sequenced using Hi-Seq illumina platform and was assembled using the Soap de novo assembly software at Macrogen Inc, Korea. The genome was found to have 47.73% GC content with 34,616,056 bases (34.6 MB) that spread over 3194 contigs (322 scaffolds). Comparison of the whole genome of *A. westerdijkiae* with other environmental and pathogenic Aspergillus strains revealed some interesting features of the sequenced genome (of *A. westerdijkiae*), which have been explained in the subsequent sections. Overall, the sequenced *A. westerdijkiae* genome is quite comparable in size to that of other filamentous fungi for which the genome sequence has been already reported (See Table 2.2). The *A. westerdijkiae* genome has been submitted to NCBI with accession numbers from LKBE01000001-LKBE01000239 (See Appendix XVII).

The genome of *A. westerdijkiae* contains a large number of secondary biosynthetic gene clusters- The genome sequencing data revealed the presence of a large number of secondary metabolite gene clusters in the *A. westerdijkiae* genome. By gene annotations and using the software AntiSMASH [157], we identified more than 60 secondary biosynthetic gene clusters (See Fig: 2.3) (See Appendix I and II). The majority of the biosynthetic gene clusters were found to be containing modular or iterative type I PKS and NRPS. The remaining ones could be classified under terpene and hybrid gene clusters. Interestingly, most of the gene clusters identified in *A. westerdijkiae* genome do not share significant homology or similarity to any of the characterized biosynthetic gene clusters, indicating that *A. westerdijkiae* could produce a large number of secondary metabolites with novel structure and bioactivity (See Appendix II for the domain organization of all the major PKS, NRPS and PKS-NRPS hybrids in *A. westerdijkiae*).

The genome of *A. westerdijkiae* contains at least 27 PKS (see Table: 2.2), with some of them sharing significant homology with PKS found in *A. niger*, *A. kawachii* and other species of Aspergillus. The number of iterative PKS genes found in the genome of *A. westerdijkiae* is remarkable considering that the number is comparable, if not greater than those of *A. niger, A. nidulans* and *A. fumigatus* (Appendix II and III). Based on the presence of KR domains, fungal PKSs can be further categorized into non-reducing, partially reducing and highly reducing PKSs. The products of majority of the PKS gene clusters remain to be determined. In addition, the *A. westerdijkiae* genome has been found to contain about 18 NRPS genes that do not contain non-canonical domain structures and the domain compositions of these predicted NRPSs are shown in Appendix II. Among the 20 predicted NRPSs, 14 were involved in putative secondary metabolism pathways, while the other may be involved in lipid biosynthesis. Based on the high homology shared with the genes of characterized biosynthetic gene clusters of other fungal species, we predict that *A. westerdijkiae* likely produces terrequinone, hexadehydroastechrome and notoamide and structurally similar compounds (See Appendix I). The biosynthetic gene cluster for the NRPS-produced dipeptide alkaloid notoamide B has been examined in 2010 [158]. A homologous gene cluster has been identified in the *A. westerdijkiae* genome (scaffold 27, cluster 48). By using LC-MS for analyzing the culture broth of *A. westerdijkiae*, we could detect both notoamide and the structurally related and Avrainvillamide. The two compounds are likely to be produced by the same biosynthetic gene cluster that contains a bi-modular NRPS. Based on our bioinformatic analysis, the *A.*

westerdijkiae genome also contains several NRPS gene clusters that are likely to be producing other di-peptide and tri-peptide alkaloids that contain the amino acid anthranilate [105] (discussed in chapter 4).

The genome of *A. westerdijkiae* also contains six PKS/NRPS hybrid gene clusters (see Appendix II and III) that could have the potential of producing unknown compounds. These PKS/NRPS hybrid gene clusters have been found to not bear any significant homology with any known characterized gene clusters. Although plants are the major producers of terpenes, some fungi such as Aspergillus can produce terpenes as well. The genome of *A. westerdijkiae* contains 11 gene clusters that are predicted to produce terpenes, including one terpene/PKS and one terpene/NRPS hybrid gene clusters. Refer Appendix I, II and III for the tables showing detailed analysis done by antiSMASH and the domain organization of all major PKS, NRPS, and PKS-NRPS hybrids found in the *A. westerdijkiae* genome.

Additionally, in the *A. westerdijkiae* genome we have identified an OTA biosynthetic gene cluster which we have attempted to characterize (refer chapter 3). In chapter 4 we have summarized our findings on peptidyl alkaloid synthesizing biosynthetic clusters that have been identified in *A. westerdijkiae* genome.

SOAPdenovo Assembly Summary

No. of contigs	Contig sum	N50	Longest contig	Shortest contig	Average length
3,194	34,616,056	28,533	179,756	6	10,837

No. of scaffolds	scaffolds sum	N50	Longest scaffolds	Shortest scaffolds	Average length
322	36,523,223	1,603,627	3,359,036	1,003	113,426

Contig/scaffold sum: total number of bases in contigs/scaffolds
N50: means half of all the bases reside in contigs/scaffolds of this size or longer.

GC content

No. of A	No. of T	No. of G	No. of C	No. of N	GC content
8,604,682	8,580,358	8,712,252	8,718,520	1,907,411	47.73%

No. of N: Number of ambiguous bases

Table 2.1: Summary of genome sequencing results.

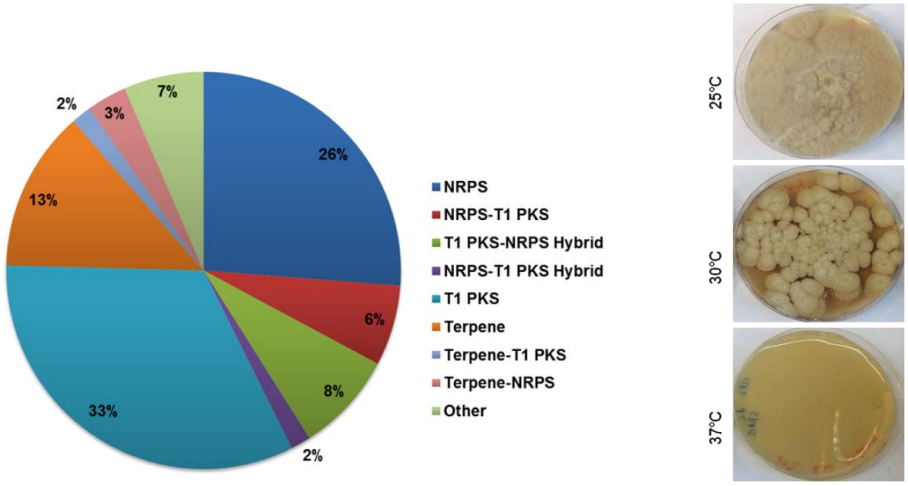


Figure 2.3: The left panel shows the summary of secondary metabolite gene clusters found in *A. westerdijkiae* genome as predicted by antiSMASH (Illustration has been made using SigmaPlot statistical analysis tool) and the right panel shows that *A. westerdijkiae* does not grow at 37°C and above.

	A. oryzae	*A. fumigatus*	*A. nidulans*	*A. westerdijkiae*
Genome size (Mb)	36.7	28.81	30.06	34.6
PKS	30	14	27	27
NRPS	18	14	14	21
Sesquineterpene cyclase	1	-	1	-
DMATS (Dimethylallyl tryptophan synthetase)	2	7	2	5
Secondary metabolite clusters	56	28	46	61

Table 2.2: Comparison of secondary metabolite biosynthesis genes in *A. westerdijkiae* and other *Aspergillus sp* [30].

Cross species genome comparison between *A. westerdijkiae* and other environmental and pathogenic Aspergillus strains revealed some interesting features about the *A. westerdijkiae* genome. Overall, the sequenced *A. westerdijkiae* genome is quite comparable in size to that of other filamentous *Aspergillus sp*. A phylogenetic tree constructed with an entire dataset of concatenated single-copy orthologous genes, indicated that *A. westerdijkiae* is close to *A. flavus* and *A. oryzae* phylogenetically rather than the pathogenic *A. fumigatus* (see Appendix XV). A set of orthologous gene clusters were identified using the Perl program Proteinoetho5, using the genomic data of nine members of the genus Aspergillus including *A. westerdijkiae*. The number of orthologs

found to be common between the species has been listed in Table 2.3, which indicates that *A. westerdijkiae* shares more orthologs with *A. flavus* and *A. oryzae* in comparison to other species.

Species	*A. westerdijkiae*	*A. clavatus*	*A. fumigatus*	*A. flavus*	*A. nidulans*	*A. niger*	*A. oryzae*	*A. terreus*	*N. fischeri*
A. westerdijkiae	3,969	-	-	-	-	-	-	-	-
A. clavatus	3,082	5,724	-	-	-	-	-	-	-
A. fumigatus	3,109	5,323	8,248	-	-	-	-	-	-
A. flavus	3,415	3,580	3,622	9,738	-	-	-	-	-
A. nidulans	2,723	3,015	3,011	3,016	3,569	-	-	-	-
A. niger	1,913	2,061	2,117	2,336	1,795	2,827	-	-	-
A. oryzae	3,337	3,465	3,525	9,217	2,917	2,270	9,553	-	-
A. terreus	3,161	3,424	3,509	3,728	2,962	2,125	3,588	4,477	-
N. fischeri	3,184	5,513	8,114	3,772	3,133	2,185	3,626	3,638	8,558

Table 2.3: Orthologous genes across the different species of Aspergillus**.**

Unlike the pathogenic *A. fumigatus, A. westerdijkiae* is unable to grow at temperature above 37 ^{o}C. Several gene products that include the O-mannosyltransferase (encoded by cgrA), the nuclear protein CgrA, and the heat shock chaperone protein Hsp1/Aspf12, contribute to the thermotolerance of *A. fumigatus*. A search of the genome of *A. westerdijkiae* suggests that the genes are absent in *A. westerdijkiae*. The lack of these proteins may account for the inability of *A. westerdijkiae* to grow at temperatures above 37°C. A study in 2005 indicates that *A. ochraceus* shows optimal growth in between 25 to 30°C, at 35°C the growth rate is retarded; while at 41°C a complete inhibition of growth can be observed [159]. Since, *A.westerdijkiae* was originally classified under *A. ochraceus* until 2004; it is expected to have similar growth patterns. To verify this, we grew *A. westerdijkiae* at 25, 30 and 37°C respectively and observed that *A.westerdijkiae* did not grow at 37°C (see Fig: 2.3), in accordance to earlier observations [159].

A. westerdijkiae is not considered to be a human pathogen unlike *A. fumigatus* which produces an array of toxins that are believed to contribute to the pathogenicity of the strain. Gliotoxin is the major and the most potent toxin produced by *A. fumigatus*. It belongs to the family of epipolythiodioxopiper-azines, which are characterized by a disulfide bridge across a piperazine ring which is essential for their toxicity. *A. westerdijkiae* is unlikely to produce gliotoxin as no similar gene cluster can be found in the genome. *A. westerdijikae* also lacks the gene clusters for the production of the other

toxins produced by *A. fumigatus*, which include aflatoxin B1 and G1, mitogillin, festuclavine, fumigaclavine A-C, fumitremorgin A-C, verruculogen, fumagillin, helvolic acid. The lack of the biosynthetic pathways for the production of gliotoxin and the other toxins are likely one of the major factors distinguishing *A. westerdijkiae* from the pathogenic *A. fumigatus*. We did not detect the common mycotoxins patulin and aflatoxin (by LC-MS) in the culture broth of *A. westerdijkiae*, which is consistent with the lack of the corresponding gene clusters in the genome and the previous reports that *A. westerdijkiae* is a non-producer of pataulin and aflatoxin [160, 161] .

Peptidases or proteolytic enzymes are essential for the survival of an organism and are also industrially relevant. Hence the knowledge of the kind of proteases and protease inhibitors an organism can code for can be useful [156]. A batch BLAST search against the MEROPS (the database of proteolytic enzymes, their substrates and inhibitors) [162], resulted in the identification of 378 protease coding genes, classified into five categories and distributed in 77 families (see Table 2.4). 72 of the predicted proteases were of the secreted kind. Using batch BLAST, we also identified 7 inhibitor coding genes, of which one was of the secreted type (see Table 2.4). All 378 genes coded 286 peptidases, 3 homologues and 7 inhibitors.

Genomes	Peptidases	Homologues	Inhibitors
Aspergillus westerdijkiae	286	3	7
Aspergillus clavatus	165	38	5
Neosartorya fischeri	185	41	5
Aspergillus flavus	198	56	5
Aspergillus fumigatus	236	109	12
Aspergillus nidulans	235	109	6
Aspergillus niger	254	159	11
Aspergillus oryzae	325	139	5
Aspergillus terreus	178	52	4
Average	229	78	7

Table 2.4: List of peptidases and protease inhibitors identified in different Aspergilli**.**

CAZymes or carbohydrates-active-enzymes is a database that contains information regarding all enzymes related to carbohydrate synthesis and metabolism. These enzymes are crucial for various processes needed for the survival of an organism. Usually the plant

pathogens require them to degrade polysaccharides of host tissues, such as cellulose, hemicellulose and pectin present on the plant cell wall. Localized degradation of cell wall is also necessary for penetrating and spreading across host tissues once an infection is established [163]. The genome of *A. westerdijkiae* was screened for genes coding for carbohydrate synthesis, metabolism and transport using the database CAZy annotation pipeline [164]. The results obtained were further compared to other species of Aspergillus whose sequences are available in the CAZymes database (http://www.cazy.org/). 633 putative CAZymes were identified in the *A. westerdijkiae* genome, which was more than those found in CAZY database for three other species of Aspergillus, namely, *A. nidulans* (430 genes), *A. niger* (478 genes) and *A. oryzae* (501 genes). Among the 633 CAZyme genes identified, 281, 95, 24, 130, 101 and 53 were assigned to classes of glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and carbohydrate-binding modules (CBMs), respectively.

A whole genome BLAST against pathogen-host interaction (PHI) database, revealed that 3124/10861 proteins (28.76%) in *A. westerdijkiae* genome shared homology with pathogenicity, virulence and effecter genes in PHI-Base, 223/3124 (0.72%) of which belonged to the potentially secreted protein class. The information with regard to the secreted proteases and CAZymes predicted in this study may serve as useful data for the better understanding of virulence and host-specificity of *A. westerdijkiae* (see Appendix XIV). Although, *A. westerdijkiae* does not grow at 37°C and hence it might have difficulty to cause systemic infections in mammals, the possibility of it causing opportunistic skin/hair infections or plant diseases cannot be ruled out at this point. A list of putative virulence and toxin related genes found in *A. westerdijkiae* genome have been listed in Appendix IV.

Antimicrobial activity assay of culture broth extracts- As Aspergilli are a source of quite a good number of antibacterial or bactericidal compounds, we performed antimicrobial activity assay with the organic culture broth extracts of *A. westerdijkiae* and another member of the genus Aspergillus; *A. alutaceus*. Clear zones of inhibition or clearance were observed for all the test organisms (*Staphylococcus aureus*, *Escherichia coli*, *Streptococcus pyogenes* and *Bacillus subtilis*). Results indicate that the extract from the *A. westerdijkiae* broth has a substantially more inhibitory effect on bacterial growth when compared to the *A. alutaceus* broth extracts. The results of the 96-well microtitre

plate test also reiterate the findings of the disc diffusion assay (Fig: 2.3). Note that the results shown in this section also indicate the antimicrobial activity of the culture broth extract of another species of Aspergillus i.e. *A. alutaceus*. Although at the initial stage we did try to look for antimicrobial activity in both the Aspergilli, the study on *A. alutaceus* was shelved and we decided to study *A. westerdijkiae* alone. One of the major reasons of this was the result of antimicrobial activity assay itself, which indicated that *A. westerdijkiae* produced a higher zone of inhibition against both gram positive and gram negative organisms when compared to *A. alutaceus*.

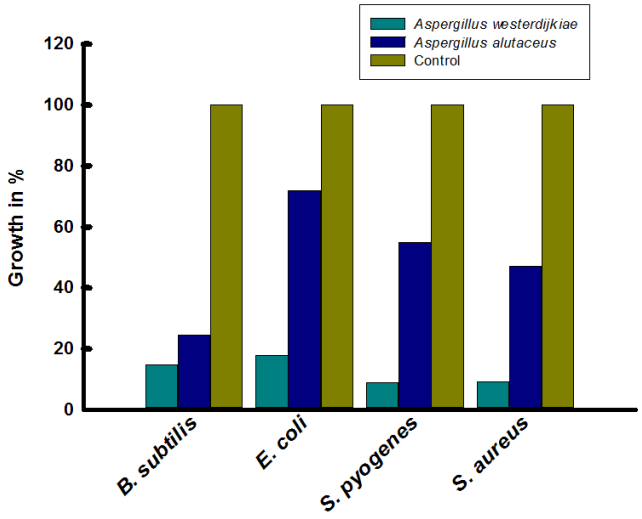


Figure 2.4: Graph indicates the trends of the 96-well plate antimicrobial activity assay done with organic extracts of *A. westerdijkiae* culture broth (Illustration has been made using SigmaPlot statistical analysis tool).

LC-MS/NMR- We used the liquid chromatography coupled mass spectrometry (LC-MS) to detect the compounds produced by *A. westerdijkiae* cultivated under several conditions. In addition to the mycotoxins OTA we could identify several compounds from the culture broth. The compounds whose identities could be established by high-resolution mass and NMR spectroscopy include circumdatin A, B [165] and F, notoamide I [166], avrainvillamide [167], coriolic (13-hydroxy-9, 11-octadecadienoic) acid [168], linolinic acid [169], neoaspergillic acid [170], neohydroxyaspergilic acid [171] and the siderophore desferrichrysin [172]. The LC chromatogram contains a large number of peaks that indicate the presence of many more compounds. However, the low abundance of the compounds prevented us from determining the structures of the compounds at this

moment (See Appendix X and XI for the NMR and LC-MS data of the compounds isolated).

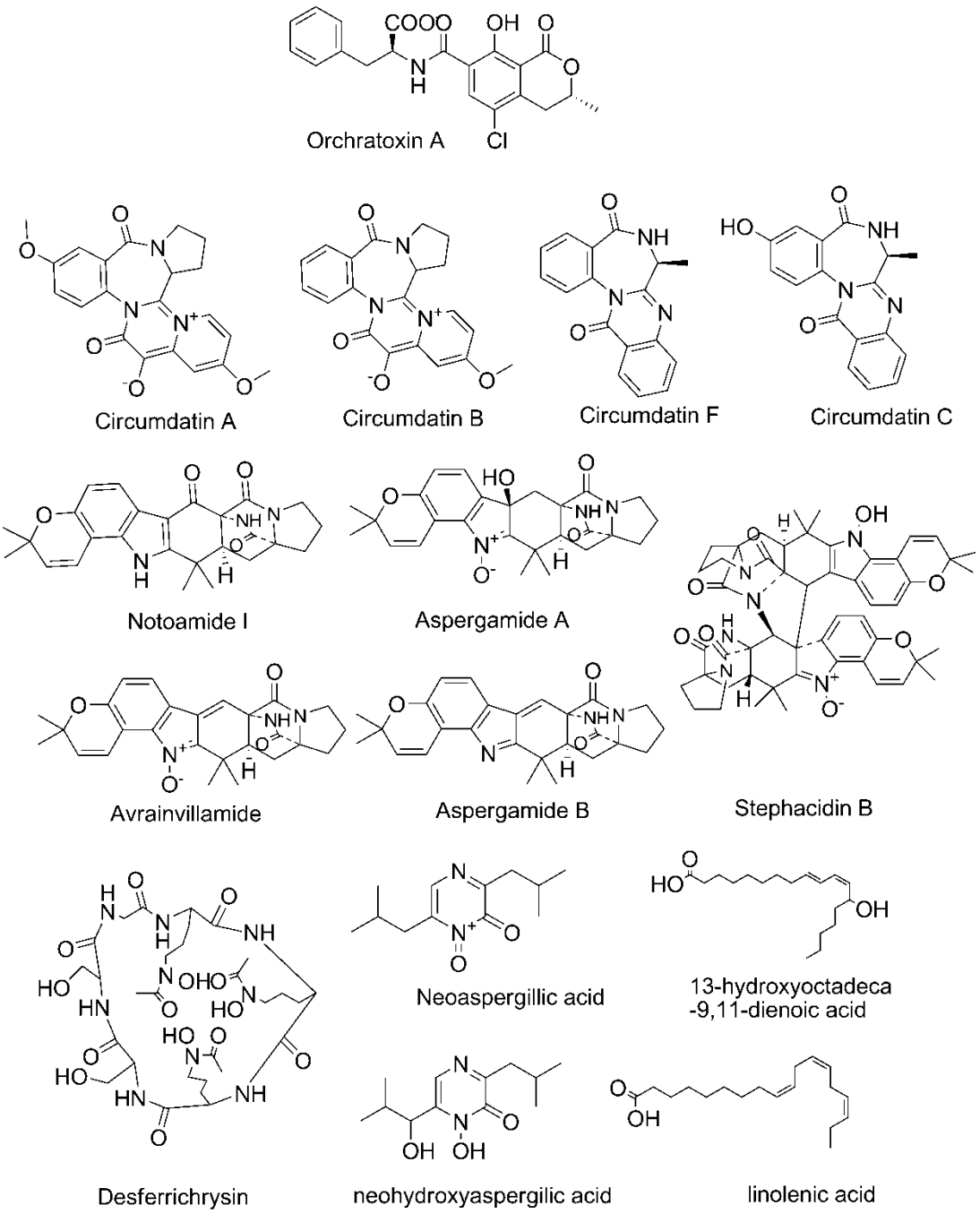


Figure 2.5: Structures of compounds identified in broth and biomass of *A. westerdijikiae* (Illustration obtained using ChemDraw tool)**.**

Discussion

The complete genome of *A. westerdijkiae* CBS112803 revealed interesting aspects about this important filamentous fungus, including its remarkable capability of producing secondary metabolites. Whole genome sequencing data suggests *A. westerdijkiae* CBS112803 contains comparable number of PKS and NRPS gene clusters as the other members of the genus Aspergillus that have been under intensive investigation. Amongst the more than 60 secondary metabolite gene clusters predicted by AntiSMASH, we have been able to identify an OTA gene clusters as well as clusters for notoamide and a hexahydroastechrome (HAS)-like molecule synthesis (refer chapters 3 and 4). A number of putative alkaloid synthesizing gene clusters have also been seen (refer chapter 4). Quite a number of the biosynthetic gene clusters found in *A. westerdijkiae* genome do not share significant similarity with any gene cluster from the already sequenced Aspergilli genomes, suggesting that they are likely to produce unknown compounds. Hence, they could be quite promising candidates for the exploration of novel small molecule production. Many of the biosynthetic pathways are likely to be silent or suppressed. Promoter replacement and co-culturing will be carried out in future to activate the silent biosynthetic pathways. In-vivo techniques like heterologous expression of an entire gene cluster or a particular backbone synthesizing enzyme can also be employed for the characterization of unknown/orphan gene clusters.

The genome comparison data highlights the evolutionary and phylogentic relationship of *A. westerdijkiae* with the other members of the genus Aspergillus. *A. westerdijkiae* has been observed to be much closer to *A. oryzae* and *A. flavus* compared to the pathogenic strain *A. fumigatus*. Apart from the different secondary metabolism gene clusters, a number of proteases and carbohydrate metabolism related genes have also been identified in *A. westerdijkiae* genome. The data on the type of proteases and carbohydrate metabolism related enzymes present in any particular species can be quite useful for the understanding of the lifestyle of the organism. While peptidases/proteases can be potential drug targets they can be very usefull commercially and find various industrial applications. Carbohydrate synthesis, transport and metabolism genes can shed light on the virulence and pathogenic aspects of a particular organism. The analysis of putative pathogenic genes suggests that *A. westerdijkiae* is most likely not a human pathogen. The inability to grow at 37°C and the phylogenetic distance with *A. fumigatus* adds more credibility to this hypothesis. Additionally, the large number of carbohydrate metabolism

related genes found in the genome points towards the possibility of *A. westerdijkiae* being a plant pathogen.

The culture broth of *A. westerdijkiae* has shown to possess antimicrobial activity but it has not been established if this activity is due to some already known compound, a novel compound or a combination of compounds. Further analysis of the culture broth i.e. the fractionation of the organic extracts into multiple sub-fractions along with their characterization and followed by antimicrobial assay with each fraction is required to be done in future.

The culture broth analysis using LC-MS/NMR has led to the identification of some of the already reported compounds. A scale-up of culture volume will be needed in future to overcome the yield issues and therefore detection of more number of compounds which might have gone undetected due to low yield. A variation in culture conditions and growth medium is also necessary for the identification of additional secondary metabolites in the broth. Composition of growth medium as well as other physical parameters like temperature, light and pH can affect the production of secondary metabolites in a huge way [97, 173]. As a matter of fact, certain pathways which normally remain silent might just happen to respond to any of the different culture conditions tried. Finding ways to increase the amount of biomass produced per litre of medium during the culturing period can also be a possible solution to the issue of low expression levels of compounds. We intend to set up a robust optimized regime in the lab to study the compounds that Aspergilli secrete into the culture medium in the future.

The whole genome sequencing of *A. westerdijkiae* has provided us with enormous genetic data, which could be harnessed in different ways to get a more sound understanding of Aspergillus biology. Our future endeavors in this respect would be centered around the characterization of unexplored gene clusters and elucidation of biosynthetic mechanisms of useful bioactive small molecules.

Chapter 3 - Biosynthesis of ochratoxin A in *A. westerdijkiae*

Introduction

Ochratoxins are a group of mycotoxins that are predominantly produced by fungi belonging to the genera Aspergillus and Penicillium. Ochratoxins are some of the most abundant mycotoxins known to contaminate agricultural stocks and are considered to be hazardous to public health [174]. Chemically, ochratoxins are polyketide derived secondary metabolites containing a dihydroisocoumarin moiety. In nature there are several forms of ochratoxins namely- ochratoxin A, B and C and some intermediates like ochratoxin α and ochratoxin β [174]. Ochratoxin A (OTA), first isolated in 1965 from the culture broth of *A. ochraceus*, is considered to be the most potent in terms of toxicity [174].

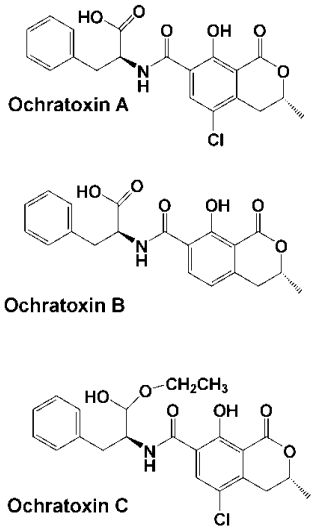


Figure 3.1: Structure of ochratoxins (Illustration obtained using ChemDraw tool).

OTA is reported to be nephrotoxic, immunogenic and even a potential carcinogen [122-124]. Studies in both animals and humans strongly point at the correlation between consumption of OTA and renal diseases [123]. There is evidence that OTA acts an immunosuppressant by reducing the size of the immune organs like thymus and spleen thereby affecting the quality of immune response [131]. The International Agency for Research on Cancer (IARC) classifies Ochratoxin A under the class 2B (possibly

carcinogenic to humans). Studies have also indicated that OTA can be teratogenic in mice [175] and chick embryo [176]. Certain reports have shown that OTA is capable of crossing the placental barrier and has also been detected in colostrum (breast milk) [177]. Research further highlights the relationship between exposure to OTA and oxidative damage to brain tissue [178]. Some studies have revealed that OTA is highly cytotoxic to nerve tissue [179] and may be a cause of neurodegenerative diseases as well.

Although a huge amount of information is available on the toxicity, detection and occurrence of OTA in food and other agriculture products, the biosynthetic and regulatory mechanisms of OTA production remain to be fully understood. In 1979 Huff and Hamilton suggested a model for the synthesis of OTA where mellein is the starting compound [180] (Fig: 3.2).

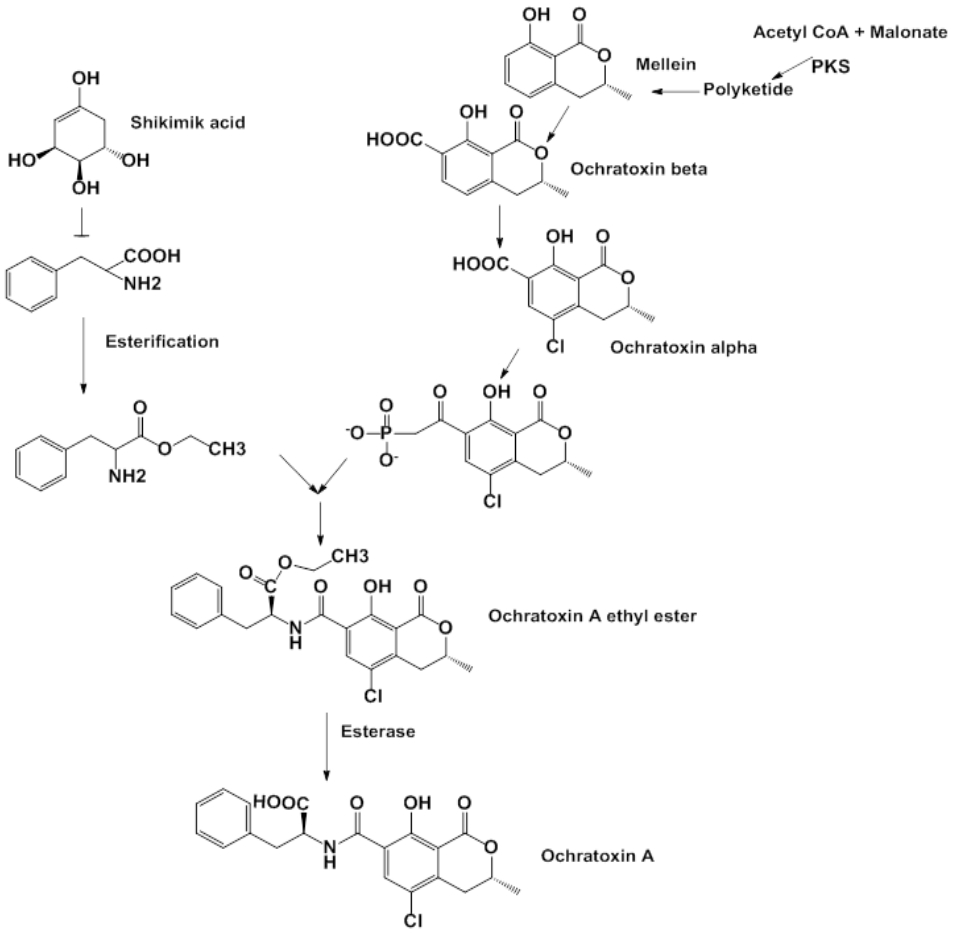


Figure 3.2: OTA biosynthetic mechanism proposed by Huff and Hamilton in 1979 [180] (Illustration obtained using ChemDraw tool)**.**

It was speculated for a long time that the pentaketide-derived mellein (which is also known as ochracin) is the precursor or intermediate in the OTA biosynthetic pathway [180]. A small portion of the PKS gene for OTA production in *A. ochraceus* was sequenced in 2003 (Gene Bank accession number: AY272043) [181]. The same group also observed a significant rise in the transcript levels of a few P450 genes during enhanced OTA production indicating a possible role in OTA biosynthesis. In 2009, a group knocked-out the PKS gene (Gene Bank Accession Number: AY583209) involved in OTA production in *A. westerdijkiae* NRRL 3174 strain and found that mellein synthesis was not affected, which suggests that mellein is not an biosynthetic intermediate for OTA [160, 174]. A group in 2012 reported the role of a NRPS (*AcOTAnrps)* in OTA biosynthesis in *A. carbonarius*. This study also proved that the chlorination step occurs after the addition of phenylalanine to the isocoumarin backbone and they proposed the following new scheme of OTA synthesis [174, 182].

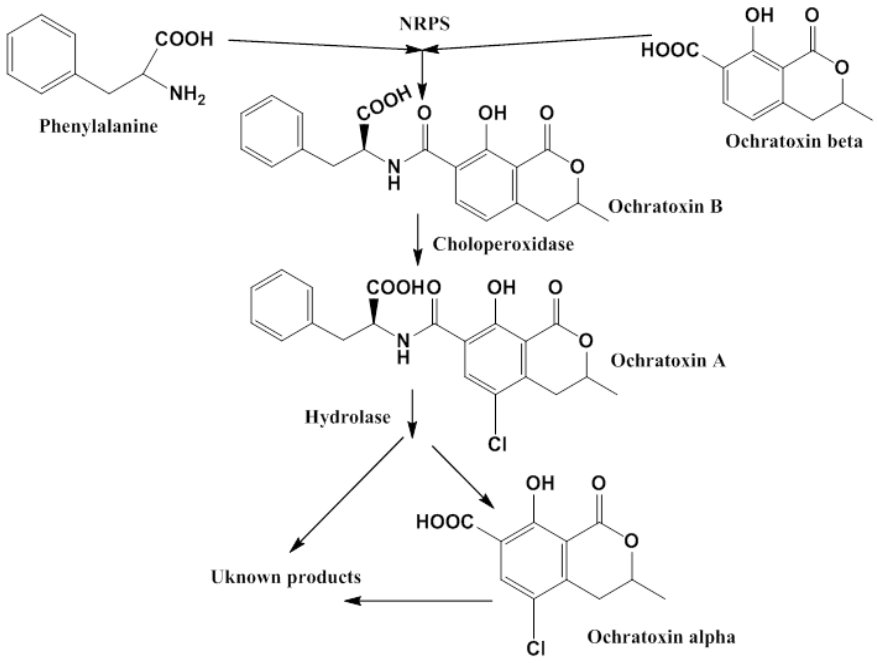


Figure 3.3: OTA biosynthetic mechanism proposed by Gallo et al in 2012 [182] (Illustration obtained using ChemDraw tool)**.**

Regulation of mycotoxins production often involves complex and multi-tiered signaling pathways. A number of secondary metabolite gene clusters remain cryptic under most of the conditions and it is difficult to identify what stimulates them. Since natural products are of great clinical relevance both as therapeutics and toxins, it is imperative for researcher to further investigate into the regulatory mechanisms that are involved in their synthesis. Secondary metabolite genes are often regulated at several levels (genetically) and are also influenced by certain environmental conditions/stimuli. While some of the secondary metabolite clusters function under a universal regulator like the LaeA protein in *A. nidulans* [61], there are other clusters that have specific regulators. It has been observed that the deletion of the LaeA regulator affects the production of penicillin, lovastatin and sterigmatocystin in a negative way while the overexpression of this regulator up-regulates the production of the aforementioned polyketides. On the other hand, instead of global regulation, the transcription factors associated with compactin or trichothecene synthesis gene clusters are specific in nature [183].

Interestingly, several studies have shown that regulators of secondary metabolism are also capable of affecting sporulation, virulence and pigmentation. VeA a global regulator that has been reported in several members of the genus Aspergillus, is a light dependent protein which is known to regulate both secondary metabolite production as well as sporulation [62]. The relationship between spore formation and toxin production in fungi has been an area which has been the subject of much study. According to reports, strains of *Claviceps purpurea* that are deficient in spore production produces less amount of toxic metabolites compared to the wild type [3]. A study done in *P. urticae* has shown a marked decrease in patulin synthesis when conidiation was blocked [3]. Similar correlation has been found between sporulation and fumonisin and aflatoxin production in species of Fusarium and Aspergillus respectively [3]. In certain cases fungal pigments are associated with spores and hence an alteration in spore production due to the deletion or over-expression of some secondary metabolite regulators that affect sporulation may also alter pigmentation. A group in 2008 has reported how secondary metabolism is related to growth/development in *A. nidulans* via a set of regulatory proteins (LaeA, VeA and VelB) with a global influence over the entire process in a light dependent manner [184] . In fact in 2013, [185] it was reported that the deletion of LaeA and VeA global regulators down-regulates the production of OTA in *A. carbonarius*.

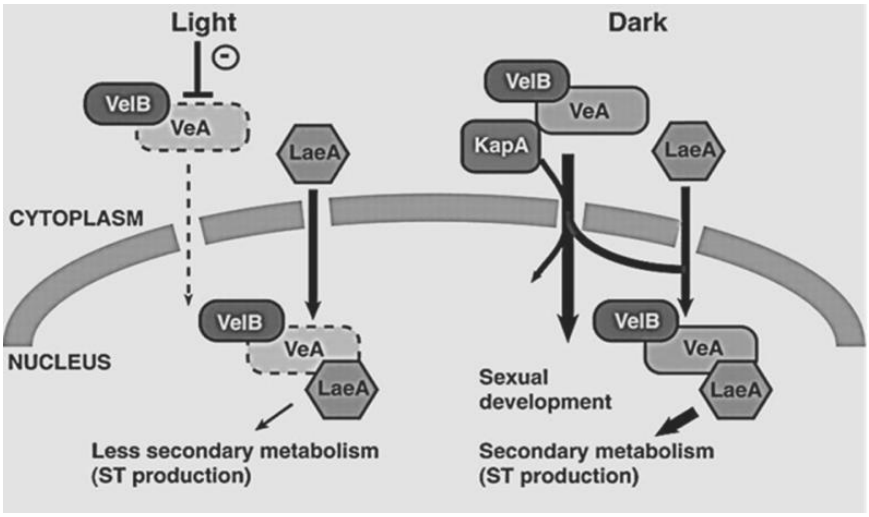


Figure3.4: Regulation of secondary metabolism in *A.nidulans*. The figure shows how the velB/Vae/Lae complex coordinates light and influences secondary metabolism in *A.nidulans* [184].

Recently, a new group of proteins which lack an intrinsic globular structure (also known as intrinsically unstructured proteins or IUPs) have gained prominence due to their involvement in a number of important regulatory processes. For example the HY5 bZIP (basic leucine zipper) transcription factor (TF) in the plant Arabidopsis, is an IUP and has been implicated to be involved in light regulation of transcription [186, 187]. The tumor suppressor protein, p53 which is also a TF is a partially disordered protein [188]. cAMP regulated transcription factors, CREB and certain heat shock TF in yeast are disordered in nature.[189]. The lack of structure sometimes can be advantageous, allowing the protein to interact with more than one binding partners. It is possible that the protein might fold into a proper ordered structure once it binds to its target. A mention of unstructured proteins in the context of OTA biosynthesis arises from the fact that we have identified a transcription factor (which acts as an activator of OTA synthesis) in the putative OTA cluster that is predicted to have a disordered N-terminal.

As for the regulation of OTA production, a study used the cDNA-amplified fragment length polymorphism method (cDNA-AFLP) to identify differentially expressed genes in OTA high- and low-producing *A. carbonarius* strains. Among the genes that were found to be up-regulated in OTA non-producing strains are a small number of transcriptional factors and proteins from the G-protein signalling and Ca^{2+}/calmodulin-dependent pathways [190]. Given the health concerns of the ochratoxins, understanding the

biosynthetic and regulatory mechanisms may lead to the discovery of novel solutions to suppress the production of the toxins during food production and storage.

Aspergillus westerdijkiae is a filamentous fungal species that was initially believed to be a member of the *A. ochraceus* taxon. However, in 2004, the *A. ochraceus* species was found to be further sub-divided into *A. westerdijkiae* and *A. ochraceus* [130]. As mentioned in chapter 1 and 2 *Aspergillus westerdijkiae* is one of the major producers of OTA and hence we sequenced the whole genome of this organism for further analysis. In this study, we have identified a putative OTA biosynthetic gene cluster in the genome of *A. westerdijkiae* CBS112803. We further validated the function of the putative OTA gene cluster with a series of gene knockout experiments. As mentioned previously, we have identified a bZIP transcription factor that functions as a transcriptional activator controlling OTA production, and affects other phenotypes such as spore formation and pigmentation. This information could facilitate the detection of ochratoxins-producing fungal species by PCR-based approaches as well as development of new methods for suppressing mycotoxins production.

Materials and methods

Culturing of *A. westerdijkiae*- *A. westerdijkiae* CBS112803 obtained from Centraalbureau Voor Schimmelculture, Netherlands (CBS) was cultured in LMM and SAM media as stationary cultures for 10-14 days. The composition of SAM medium is: 50 g glucose, 3 g NH_4NO_3 (ammonium nitrate), 26 g K_2HPO_4 (di-potassium hydrogen phosphate), 1 g KCl (potassium chloride), 1 g $MgSO_4.7H_2O$ (magnesium sulfate heptahydrate), 10 ml mineral solution (composition per liter of distilled water): 70 mg, $Na_2B4O_7\_10H_2O$ (sodium tetraborate decahydrate), 50 mg $(NH_4)_6$ $Mo_7O_{24}\_4H_2O$ (sodium molybdate tetrahydrate), 1000 mg $FeSO_4.7H_2O$ (ferrous sulfate heptahydrate), 30 mg $CuSO_4\_5H_2O$ (Copper sulphate pentahydrate), 11 mg $MnSO_4.H_2O$ (manganese sulfate), 1760 mg ZnSO4.7H2O (zinc sulfate heptahydarte) [147]. LMM medium: 15g/L lactose, 6g/l $NaNO_3$ (sodium nitrate), 0.52g/KCl (potassium chloride), 0.52g/L $MgSO_4.7H_2O$ (Magnesium sulphate heptahydrate), 1.52 g/L KH_2PO_4 (potassium dihydrogen phosphate) and 1ml/L trace elements solution [(composition per liter of distilled water) $ZnSO_4.7H_2O$ (zinc sulfate heptahydrate) 2.2g or $ZnCl_2$ (zinc chloride), 1.04g H_3BO_3 (Boric acid), 1.1g$MnCl_2$.4H2O (manganous chloride tetrahydrate), 0.5g FeSO4 · 7H2O (ferrous sulphate heptahydrate), 0.5g $CoCl_2.6H_2O$ (Cobalt chloride),

0.16g $CuSO_4.5H_2O$ (Cupric sulphate pentahydrate) 0.16g or $CuCl_2$, 0.096g (NH4)6Mo7O24 · 4H2O (Ammonium molybdate tetrahydrate) , 0.11g Na4 EDTA.$4H_2O$ (EDTA, tetrasodium salt) 6.0g, pH6.5] with necessary supplements [148]. All the above chemicals were brought from Sigma, except for KH_2PO_4 (merck) and $MnCl_2.4H_2O$ (alfa aesar).

Bioinformatics Analysis- The whole genome of *A. westerdijkiae* CBS112803 was sequenced at Macrogen Inc., Korea using the Hiseq Illumina platform. After sequencing the genome was assembled using the de novo assembly program. The assembled genome data in the form of scaffolds was further subjected to bioinformatic analysis in our lab. The genome data was submitted to the AntiSMASH 3.0 (antibiotics & Secondary Metabolite Analysis Shell) server for secondary metabolite gene cluster prediction [68] and Softberry FGENESH (http://www.softberry.com) [149] for the prediction of introns. The genes were further annotated using NCBI–BLAST [150]. The OTA cluster in *A. westerdijkiae* was compared to the homologous clusters from other Aspergillus species using BLAST and the Aspergillus genome database (http://www.aspergillusgenome.org). Structure/disorder prediction of TF gene was done using various online tools like DISOPRED2 [http://bioinf.cs.ucl.ac.uk/], DomPred (Protein Domain Prediction: http://cms.cs.ucl.ac.uk/, Foldindex [http://bip.weizmann.ac.il/fldbin/findex]. Predict protein [www.predictprotein.org] was used to predict disorder, molecular recognition sites, possible structure and motifs with respect to the bZIP transcription factor amino acid sequence. The NRPS A domain phylogenetic (maximum likelihood) tree –Bootstrap method (number of replications 1000) was built using the software Mega (mega evolutionary genetics analysis) V6.0 [191].

Transformation of *A. westerdijkiae* and generation of knockout strains- The plasmid pCB1535 (obtained from FGSC (Kansas City, Missouri USA) containing hygromycin (hph) resistant gene obtained from FGSC (Fungal genetic stock centre, Kansas City, Missouri USA) was used as a template and primers were designed to contain overlapping sequences with the hyd*Aw*ota/ tf*Aw*ota/ cp*Aw*ota/ pks*Aw*ota/ nrps*Aw*ota/ or*Aw*ota genes as well as the hph gene for PCR of the hydrolase, transcription factor, cytochrome P450 and PKS, oxidoreductase constructs respectively. The PCR fragment was cloned into the vector pET28b+ (Novagen). The vector containing the desired fragment was single digested and the linearized fragment was used for transforming the *A. westerdijkiae* CBS112803 strain. The nrps*Aw*ota construct was made using the plasmid pBC-phleo

(FGN 42:73) obtained from FGSC (Kansas City, Missouri USA) containing the phleomycin resistance gene. For all PCR reactions Kapa HIFI DNA polymerase (Kapa Biosystems) was used.

For transformation, 200 ml germinating media (0.5% yeast extract (Sigma or BD) and 2% glucose) is inoculated with 1×10^7 conidia/ml and incubated at 37°C in a shaker at 300 rpm for 5 hours. The harvested conidia was washed with ice cold water and resuspended in 25ml ice-cold pretreating buffer YED (1% yeast extract, 1% glucose and 20 mM HEPES adjusted to pH 8.0 with 100 mM Tris) and further incubated for 1 hour at 37°C at 100 rpm. The conidia were centrifuged at 1,500g for 10 minutes and resuspended in 2.5ml ice-cold electroporation buffer (10 mM Tris-HCl pH 7.5, 270 mM sucrose, 1 mM lithium acetate - Sigma). 1 µg of linearized plasmid (containing the relevant construct to be inserted into the fungal genome) was added to 50-75 µl of conidia suspension in a pre-chilled 2 cm cuvette (BioRad) and incubated in ice for 15 minutes, followed by electroporation at 1,000 Volts, 200 ohms resistance, 25 µF capacitance. 1 ml ice cold YED is added to the electroporated cuvettes and the contents were transferred into a 1.5ml tube and incubated at 30°C for 90 minutes at 100 rpm. The conidia were centrifuged, washed with water and plated on minimal media with relevant antibiotics and incubated at room temperature for 7-10 days. The antibiotics hygromycin and phleomycin were purchased from Invitrogen. The genomic DNA of *A. westerdijkiae* CBS112803 was used as template to amplify the fragment for tf*Aw*ota gene replacement. The primers used had overlapping regions with the tf*Aw*ota gene as well as the genomic DNA sequence adjoining the tf*Aw*ota gene in the genome. The PCR fragment was integrated into the vector pET28b+ and was used for transforming the *A. westerdijkiae* Δtf*Aw*ota strain to create a *A. westerdijkiae* tf*Aw*ota replaced strain.

Promoter replacement for the transcription regulator gene- The plasmid pMT-OVE obtained from FGSC (Kansas City, Missouri USA) containing the alcohol dehydrogase promoter [alcA] was used as a template for the synthesis of the PCR product which contained the promoter gene with overlapping regions with the genomic DNA of *A.westerdijkiae* adjoining the transcription factor gene. The PCR product was integrated into a pET28b vector. The resultant vector was single digested and transformed into *A. westerdijkiae CBS112803.* The positive clones (p-TF or promoter-TF) were later selected using small scale genomic DNA extraction followed by PCR.

Isolation and detection of OTA in culture broth- Fourteen day old culture was used for the extraction of secondary metabolites. The culture broth was initially acidified to pH 3 using HCl and then extracted with equal volume of ethyl acetate and dried. The dried extract was then re-dissolved in 100% methanol for further analysis.

Thin layer chromatography (TLC) technique was used to detect the presence of OTA and other compounds in the organic extract. 10x10cm Silica gel coated glass plates were used to spot samples (OTA standard and organic extract from fungal broth). The TLC was run using chloroform: diethyl ether: acetic acid (17:5:1) as mobile phase (while check for the presence or absence of OTA in the crude extract). The plates were observed under UV light for the detection of secondary metabolites. The organic extracts of the culture broth of the *A. westerdijkiae* wild type and the knockout strains were analyzed using HPLC using the Agilent LC1200 system with the XDB-C18column (5 um, 4.6 x 150 mm) using mobile phase (57% acetonitrile, 41% water and 2% acetic acid, 1 mL/min flow rate).

PKS*Awota* protein expression in *E. coli*- The pks gene of the OTA cluster (2556 amino acid long after removal of introns) was codon optimized for expression in *E.coli* and the gene was synthesized and integrated into pUC57 by Genscript. The Polyketide synthase gene was cleaved out of from pUC57 and ligated into pET28b+ vector (containing both N-terminal and C-terminal 6xHis tag) and co-transformed with pCDf2 plasmid (containing 4'-phosphopantetheinyl transferase or PPTase or Sfp gene of *Bacillus subtilis* [Sfp plays an essential role in priming PKS by covalently converting serine residues in acyl carrier proteins within the enzyme from inactive apo-form to active holo-form] [192] into *E. coli BL21 DE3.* LB medium with 10% glycerol was supplemented with Kanamycin and Streptomycin (final concentration 50mg/ml for each) was inoculated with the *E. coli* BL21 DE3 carrying the vector with the gene of interest, and was allowed to grow at 37°C, 180 rpm to OD_{600} ~0.6-0.8. The culture was induced with 0.8mM of Isopropyl-B-D-thiogalactopyranoside (IPTG) and grown at 25 °C, 180 rpm, for 16 hours. The expression level of this protein was in very small amounts and the protein could not be purified as it did not bind to the Ni-NTA (Qiagen) or Co-NTA beads (NTA beads were regenerated using – Sheng lab protocol for column regeneration, York University (in this case Cobalt (II) sulfate heptahydrate was used instead of Nickel (II) sulfate heptahydrate for re-charging the resin).

PKS*Aw*ota protein expression in Yeast- The two multiple cloning sites of the yeast shuttle vector pESC-ura (obtained from A/P Mathew Chang's lab, National University of Singapore) were ligated with the pks*Aw*ota gene and sfp gene respectively for the expression of protein in yeast expression strain Sachharomyces cerevisiae INVSc01 (Invitrogen). pESC-ura contains the Gal1 and Gal 10 promoter which can be induced by the addition of galactose (sigma) into the culture medium. A polyhistidine tag was introduced at the C-terminal of the OTA_PKS for purification. The transformed yeast cells were grown overnight in YPD medium (1% yeast extract, 2% peptone and 2% dextrose-all chemicals from sigma) at 30°C overnight and shaken at 175 rpm. The overnight culture was used to inoculate 10-25 ml of Stage I medium (0.67% yeast nitrogen base , 1% glucose (sigma), 1% raffinose (sigma) and ura- 1X DO supplement) and incubated at 30°C, 175 rpm for 18-24 hours. 500ml Stage II medium (Induction Medium -0.67% yeast nitrogen base, 2 % galactose and ura- 1X DO supplement) with inoculated with stage I culture and incubated at 30°C, 175 rpm for 2-3 days. Thereafter the cells were harvested, centrifuged, resuspended in lysis buffer [50 mM Tris-HCL pH 8.0, 500 mM NaCl, 1 mM DTT, 10 % Glycerol, 10mM imidazole and PMSF] and sonicated. The lysed cell suspension was filtered through a 0.45µ filter and run through both Ni-NTA and Co-NTA column for purification. As seen in case of E.coli, the expression level of this protein was in very small amounts and the protein could not be purified as it did not bind to the Ni-NTA or Co-NTA beads.

Yeast transformation protocol- Yeast expression strain *Sachharomyces cerevisiae INVSc01* (Invitrogen) were grown in 10ml YPD medium overnight. 200ml YPD was inoculated with the overnight culture and allowed to grow for 4-5 hours until the O.D reaches 0.3. The cells were harvested and centrifuged at 4000 rpm for 5 minutes. The pellet was resuspended in water and centrifuged again at 4000rpm for 5 minutes. The pellet was washed once with 25ml electroporation buffer (1M Sorbitol, 1mM CaCl2) and again resuspended resupended in 10-20ml 0.1mLiAc, 5mMDTT and incubated at 30°C for 30 minutes. The cells were centrifuged and washed with Electroporation buffer once before being resuspended in 500 µl to 1ml of electroporation buffer. 10-20 µl of plasmid DNA and 100 µl of competent cells were added to a 2cm cuvvette (pre-chilled) and incubated for 15 minutes in ice before electroporation at 1500V, 25 µF capacitance and 200 ohms resistance. Immediately 500µlof 1M sorbitol and 500µl of YPD medium were added to the electroporated cells and kept for recovery overnight at 30°C in a 1.5ml tube.

Post recovery the cells were centrifuged and washed in water once before being resuspended in 200µl of water and plated on selective media (0.67% Yeast Nitrogen Base, 1X Drop-out mix Ura- ,2% Glucose and 2% agar.) and incubated for 5-7 days at 30°C.

Heterologous expression of PKS*Awota* in *A. nidulans LO7020*- The *A. nidulans* LO7020 strain [93] was gifted to us by the Oakley lab at the Department of Molecular Biology, University of Kansas. This strain of *A. nidulans* has been specially designed for the heterologous expression of secondary metabolite genes followed by easy identification and purification of small molecule product. The genes pyrG (encodes for orotide-5'-phosphate decarboxylase- synthesis of uridine and uracil), riboB (necessary for the synthesis of riboflavin) and pyroA (necessary for the synthesis of pyridoxine have been mutated and rendered dysfunctional in the strain LO7020. Hence, the homologs of these genes in *A. fumigatus* (AfpyrG, AfriboB & AfpyroA) have been used as selectable markers for the insertion of secondary metabolite genes for heterologous expression in *A. nidulans* LO7020 strain. The alcohol dehydrogenase promoter (alcA), from *A. nidulans* which is repressed by glucose and induced by alcohols was used for the expression of *pksAw*ota gene in *A. nidulans* LO7020 strain. Six different PCR fragments were generated with overlapping regions with the preceding and succeeding fragments (primers were designed with overlaps). For the synthesis of fragment 1, 2, 5 and 6 the templates used were plasmids ANIp7 (Plasmid 53:191-204), pMT-OVE (Toews et al, 2004), pTN2 (Genetics 172:1557-1566) respectively and all obtained from FGSC (Kansas City, Missouri USA). Fragment 4 was the OTA_PKS gene synthesized by Genscript and Fragment 3 was amplified using Fragment 4 as template. For transformation in *A. nidulans* LO7020, the six fragments were divided into two groups (three fragments in each group) and assembled using a Gibson assembly kit (New England Biolabs). The two Gibson assembly reactions were co-transformed into the LO7020 strain. For all PCR reactions Kapa HIFI DNA polymerase (Kapa Biosystems) was used.

Results

Identification of OTA gene cluster by bioinformatic analysis and gene knockout- Upon further analysis of the biosynthetic gene clusters identified by AntiSMASH (see Chapter 2), we identified a potential gene cluster for OTA synthesis. This potential gene cluster consists of 7-10 genes and is located in Scaffold 14. The most important genes

from this gene cluster include a type I iterative PKS (*pksAw*ota), a NRPS (*nrpsAw*ota), a flavin-dependent oxidoreductase (*orAw*ota), a cytochrome P450 monooxygenase (*cpAw*ota), a flavin-dependent halogenase (*hlgAw*ota), a putative hydrolase (*hydAw*ota) and a bZIP transcription factor (*tfAw*ota). A portion of the sequence for the PKS gene responsible for OTA biosynthesis in *A. ochraceus* LC35-12 has been reported previously [193]. The 5'-portion of the OTA PKS gene from *A. westerdijkiae* is 98% identical to the reported sequence. The cytochrome P450 gene of this putative cluster also bears 99% identity with the *A. ochraceus* P450 monoxygenase implicated in OTA biosynthesis [194]. In addition, we also found similar gene clusters from the genomes of *A. niger CBS513.88* [195, 196] and *A. carbonarius ITEM5010* [197] (see Fig: 3.6 and Appendix XII and XIII). Besides the seven genes within the putative gene cluster, the genes flanking the seven genes do not share significant homology with the genes in *A. niger CBS513.88* [195, 196] and *A. carbonarius ITEM5010* [197]. These observations led us to propose that the seven genes constitute the core of the OTA biosynthetic gene cluster. Considering the multi-tier regulatory mechanisms usually involved in the regulation of secondary metabolism in fungi, it is likely that other regulatory genes located elsewhere in the genome can also control OTA synthesis directly or indirectly.

To confirm the involvement of the genes in OTA biosynthesis, we disrupted the genes encoding the putative *oxidoreductase, PKS, NRPS, cytochrome P450* and the putative *hydrolase* individually and observed the effect of gene disruption on OTA production by using HPLC and TLC. The production of OTA was abolished in the absence of the any of the biosynthetic genes (see Fig: 3.5 and 3.7), supporting the essential role of the genes in OTA biosynthesis. The gene encoding the halogenase was not knocked out in our experiment because its role in OTA biosynthesis has been confirmed previously [198].

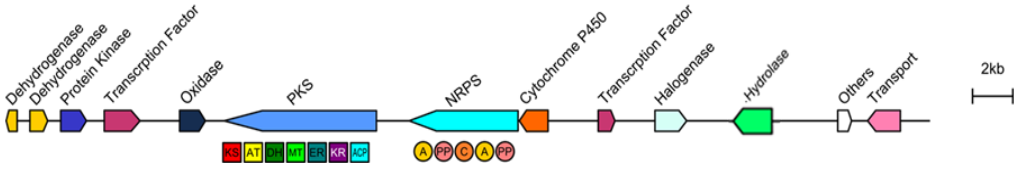


Figure 3.5: Putative OTA biosynthetic gene cluster in *A. westerdijkiae* (Illustration made using the Ink-scape tool)**.**

TF*Awota* is a transcriptional regulator that activates OTA gene expression- In addition to the genes that encode for the six biosynthetic enzymes, the OTA gene cluster contains a putative transcriptional regulator *tfAwota* gene located between the *cytP450* and *halogenase* genes. The TF*Awota* protein is predicted to contain a disordered N-terminal domain (~100 residues) and a C-terminal basic leucine-zipper (bZIP) domain. The full length TF*Awota* of the OTA cluster in *A. westerdijkiae* has an approximately 38% identity with Atf21 like proteins (Activating transcription factor) in *A. oryzae* and *A. flavus* and 14% identity with CRE1 binding protein (cAMP responsive element) of *A. parasiticus* [199]. The C-terminal bZIP domain shares ~50% identity with the homologous genes An15g07890 and Acar5010_007821 from A. *niger CBS513.88* and *A. carbonarius ITEM5010* respectively.

The NCBI conserved domain search also indicates that the DNA binding region of TF*Awota* has significant homology with the DNA binding region of Atf-2 (Activating transcription factor) like proteins. Meanwhile, the N-terminal domains (~100 residues) of TF*Awota* and its homologues from A. *niger CBS513.88* and *A. carbonarius* do not share significant homology with any characterized protein domains. Secondary structure analysis suggested that the domain has a high degree of disordered region with little α-helix and β-strand elements.

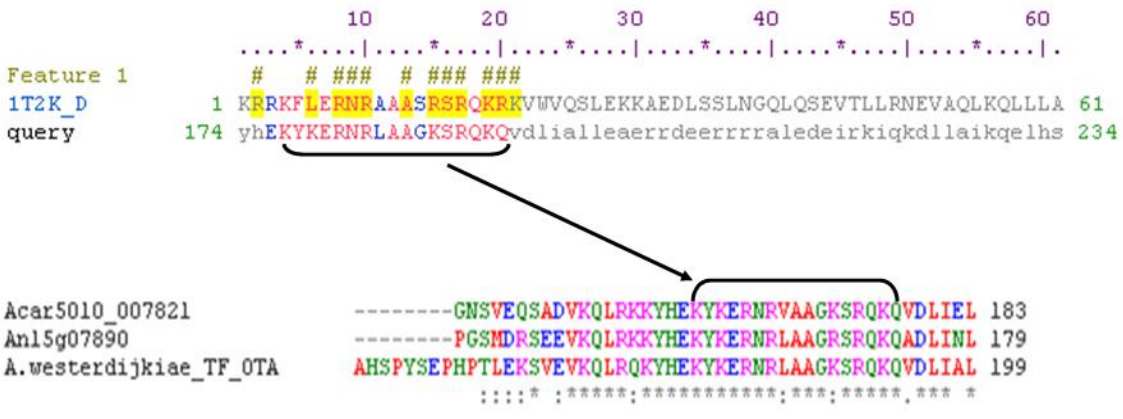


Figure 3.6: Evidence of similarity between the DNA binding sites of bZIP domain of TF*Awota* (and its homologs and An15g07890 and Acar5010_007821 from A. *niger CBS513.88* and *A. carbonarius ITEM5010* respectively) and 1T2K (Human ATF-2/AP1 (c-Jun) bZIP domain.

When we disrupted the *tfAwota* gene, OTA production was completely eliminated as shown in the HPLC analysis (see Fig: 3.8). Gene replacement was able to recover OTA production, suggesting that that gene product is an important for OTA production most likely by acting as a transcriptional activator. We further found that incorporation of a

strong promoter upstream of *tfAwota* to further enhance gene expression did not increase OTA production significantly. Production of OTA and the lack of significant increase in OTA synthesis upon promoter insertion indicate that there are other transcriptional regulators that might act in conjuction with the transcription activator identified by us. Interestingly, a comparison between the colony appearance of the wild type, Δ*tfAwota* and *tfAwota* replaced strains suggested that the deletion of the *tfAwota* gene not only affects OTA synthesis but also affects other phenotypes. The Δ*tfAwota* strain grew at a slower rate, formed less spores and the colony had a paler appearance (see Fig: 3.8 and Appendix XIV); whereas the wild type and the replaced strains exhibited similar appearance, growth rate and spore formation. The other deletion mutants that include Δ*orAwota*, Δ*pksAwota*, Δ*nrpsAwota*, Δ*cpAwota* and Δ*hydAwota* did not exhibit significant differences in colony appearance, growth rate and sporulation when compared to the wild type strain. These observations suggest that the phenotypic differences caused by the deletion of the transcriptional regulator gene are unlikely to be caused by the absence of OTA. Instead, the results indicate that the transcription regulator may regulate the expression of other genes outside of the OTA gene cluster.

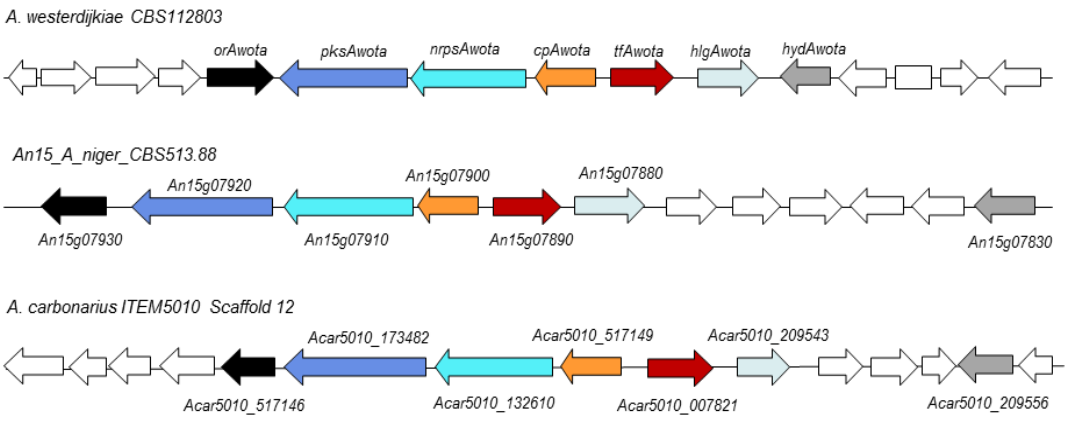


Figure 3.7: Comparison of OTA gene clusters in *A. westerdijkieae, A. niger* and *A. carbonarious.*

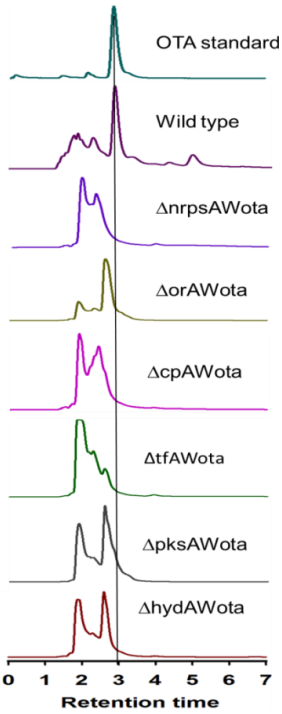


Figure 3.8: HPLC profiles of organic extracts of culture broth of *A. westerdijkiae* wild type, OTA standard and all the knockout strains viz. Δ*nrpsA*wota, Δ*orA*wota, Δ*cpA*wota, Δ*tfA*wota, Δ*pksA*wota and Δ*hydA*wota.

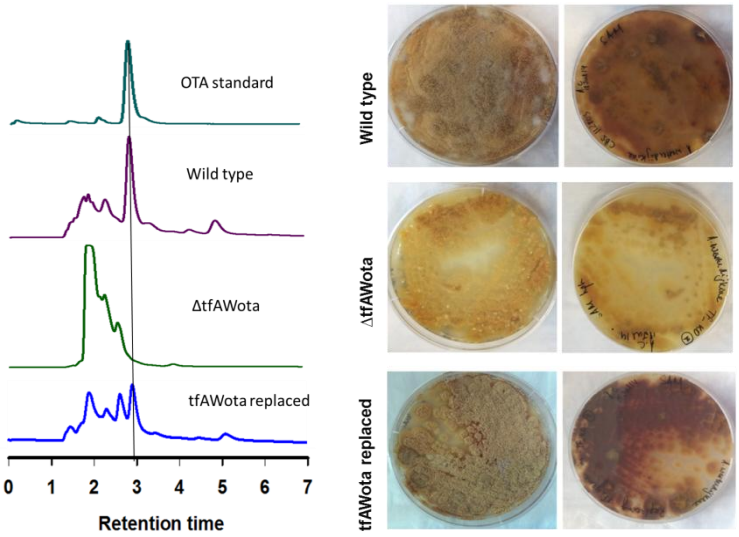


Figure 3.9: The TF knockout strain does not produce OTA as seen in the (a) HPLC profile and (b) shows paler appearance and less spore production when compared to the wild type. Replacement of the TF gene results in both OTA production and recovery of appearance phenotype.

Discussion

The gene knockout studies of the putative OTA biosynthetic gene cluster confirmed its role in OTA biosynthesis. The presence of homologous clusters in other Aspergillus species confirms that a similar pathway is likely to exist in the fungi for the production of orchratoxins. Based on the identified biosynthetic genes and existing information, we propose a biosynthetic mechanism for OTA (Fig: 3.9). The dihydroisocoumarin core, or the 7-methylmellein moiety, is assembled by the highly reducing iterative PKS. In contrast to early suggestions that the methyl group is installed by a stand-alone methyltransferase enzyme, observation of the methyltransferase (MT) domain in the PKS supports that the methyl group is likely to be incorporated during the assembly of the pentaketide product. Dihydroisocoumarin has been known to be synthesized by partially reducing PKS and modular PKS [50, 92]. Confirmation of the role of the highly reducing PKS of the OTA cluster would further underscore the diverse ways of constructing the pentaketide-derived dihydroisocoumarin moeity. The cytochrome P450 is mostly like to catalyze the oxidation reaction at carbon 7 (C7) and adds the hydroxyl group to the 7-methylmellein. The 7-methylmellein is converted to the biosynthetic intermediate ochratoxin β following hydroxylation catalyzed by the cytochrome P450 monooxygenase and oxidation by the flavin-dependent oxidoreductase (Ochratoxin β has long been known as an intermediate in the OTA biosynthetic pathway). A group in 2001 did a study with labelled precursors and observed that ochratoxin β was biotransformed very efficiently into both ochratoxins A and B production [200]. This group also suggested that mellein was not found to have an intermediary role in OTA synthesis. For the next step, ochratoxin β is activated by adenylation and covalently tethered to one of the PCP domains of the NRPS. Upon activation of the phenylalanine by the same NRPS, the ochratoxin β moiety is ligated to the phenylalanine moiety through the formation of a new amide bond. The biosynthetic intermediate tethered to the NRPS is released by the hydrolase to give the intermediate ochratoxin B. As the last step, the halogenase installs the chorine atom onto ochratoxin B to yield the final product OTA. The production of ochratoxin B upon the deletion of the halogenase supports that halogenation is the last step in the biosynthetic pathway. [182]. The proposed biosynthetic mechanism requires the NRPS to activate the carboxylic substrate ochratoxin β for amide bond formation. Considering the NRPS usually accept amino acids as substrates, one of the adenylation (A) domains must be able to recognize and activate the carboxylate. A phylogenetic

(maximum likelihood) tree of the A domains of the NRPS using Bootstrap method (number of replications 1000) was built with the software Mega (mega evolutionary genetics analysis) V6.0 [191]. The two A domains of NRPS*Aw*ota were found to be quite distant from each other in the phylogenetic tree. Our results also suggest that the A1 domain of NRPS*Aw*ota is close to A1 domains of Pc16g0490 of *P. crysogenum* which is predicted to be specific for phenylalanine, [201], A1 domain of AN0016.2 of *A. nidulans* and NRPS Pes1 of *A. fumigatus*. In contrast, the A2 domain of NRPS*Aw*ota is close to the A domain of LPS2 (lysergyl peptide synthetase 2) protein which is known to be responsible for the activation of lysergic acid in the biosynthesis of the D-lysergyl-tripeptide lactams [202] {A previous study has noted that the A2 domains of the NRPS of the putative OTA clusters in *A. niger* and *A. carbonarius* are quite unique and were closer to the ergot alkaloid synthesizing A domains [203]}. The LPS2 adnylation domain is known to be quite unique i.e. it recruits a lysergic acid group and then transfers it to the adjoining A domain specific for alanine in the synthesis of D-lysegyl-tripeptide lactams in fungi [202]. Based on the phylogenetic analysis and sequence comparison, the A2 domain of the NRPS is mostly likely the A domain that recognizes and activates the ochratoxin β to form the amide bond.

It is important to note that, for the identification of the intermediates of OTA biosynthesis pathway, we attempted to (1) express and purify the PKS*Aw*ota protein in *E. coli* and *S. cerevisiae* but the protein was expressed in extremely low amounts and did not bind to the Ni-NTA column for purification, (2) we then tried to express the PKS*Aw*ota in *A. nidulans* LO7020 strain (see Appendix XX). Although we successfully transformed the *pksAw*ota gene into the heterologous host we could not isolate the PKS product of the OTA cluster and (3) we also tried to detect the intermediates from the knockout strains but that too did not meet with any success, probably due to the low yield of the compounds in the culture broth and biomass.

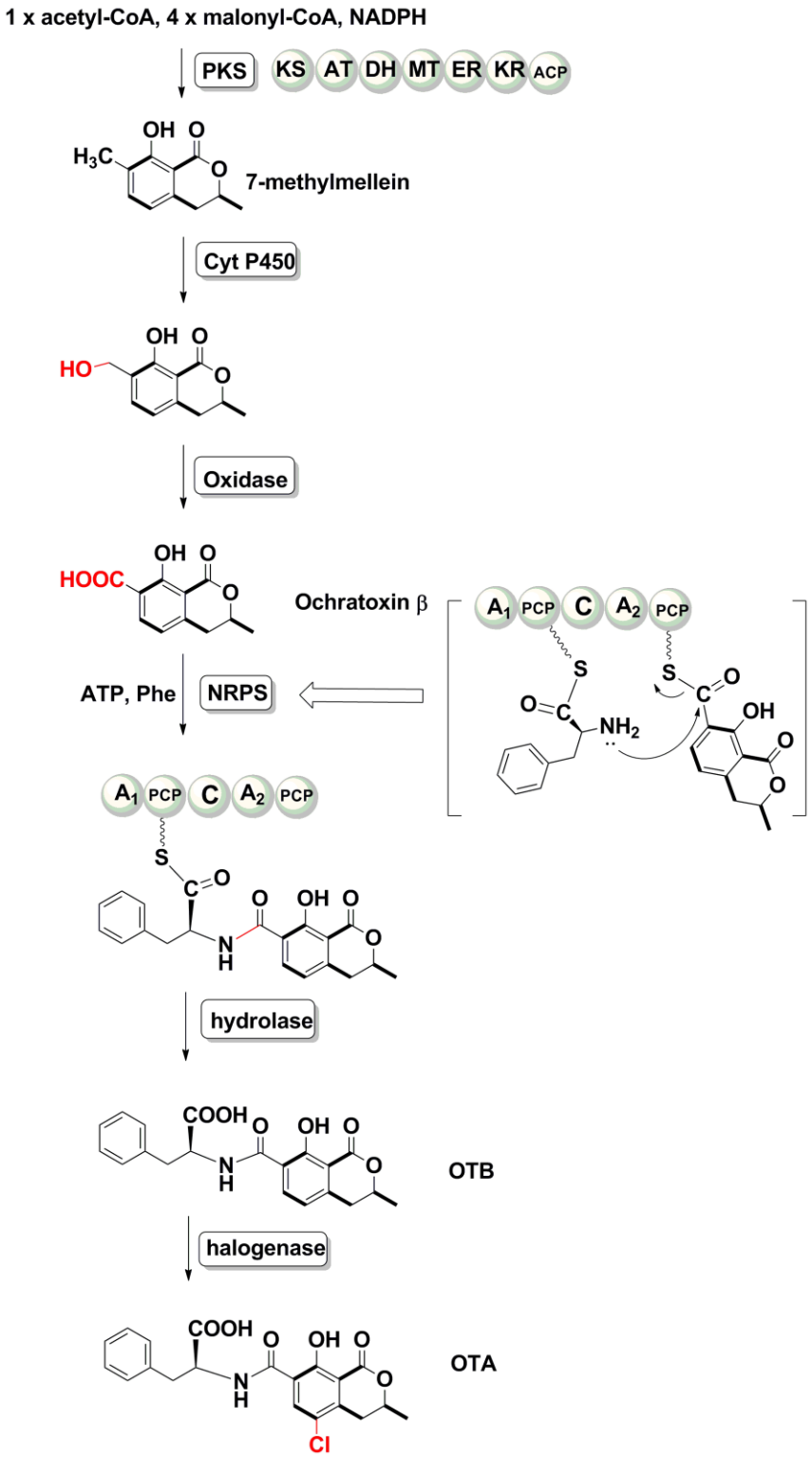


Figure 3.10: Proposed OTA biosynthetic mechanism (Illustration was drawn using the ChemDraw tool).

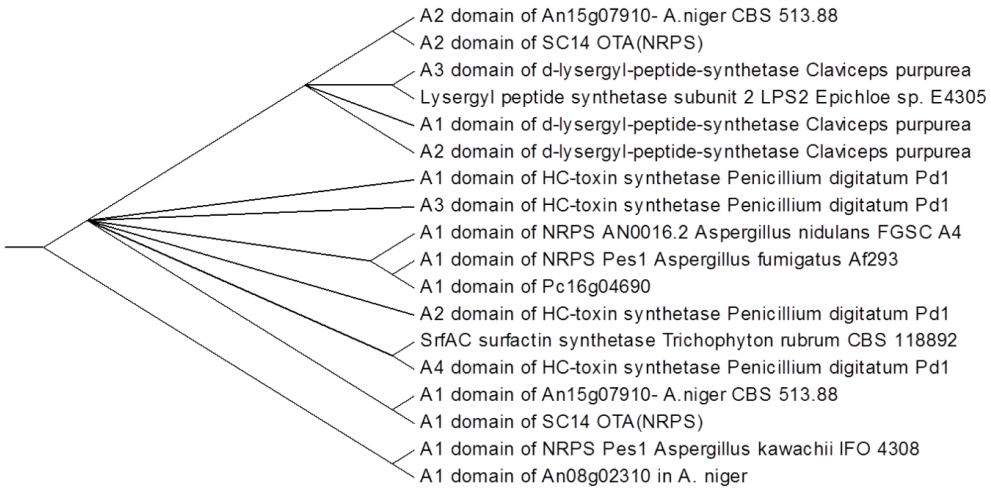


Figure 3.11: Phylogenetic tree of the two A domains found in the NRPS. While the A1 domain is closer to adenylation domains that prefers phenylalanine, the A2 domain has been found to be closer to the A domains that have preference for carboxylic acid substrates. The NRPS found in the OTA biosynthesis cluster has been indicated as Sc14 OTA (NRPS) in the phylogenetic tree (since this gene cluster has been found in scaffold 14 of *A. westerdijkiae* genome). The NRPS A domain phylogenetic (maximum likelihood) tree was drawn using the MEGA V6.0 tool.

The above study also clearly demonstrates the pivotal role of *tf*Awota in the regulation of OTA synthesis. Since this *tf*Awota protein acts as an activator, and is key in the regulation of OTA production, this might be a potential drug target to tackle the problems related to OTA contamination in the food industry. Blocking this *tf*Awota could block the synthesis of OTA. Our observations also indicate that the insertion of a promoter upstream of the *tf*Awota enhances sporulation to some extent but there is not a significant change in the amount of OTA production. The role of *tf*Awota as a transcription activator is in accordance with the bioinformatics results, wherein the conserved domain analysis shows a significant homology in the DNA binding domain of our transcription factor with the Atf2 (activating transcription factors) subfamily of TFs. It has been reported already that CREB/Atf2 family TFs bind to DNA palindrome sequence "TGACGTCA". From our knockout and gene replacement study it is quite evident that the *tfAw*ota is an activator of the pathway. However, it is still not clear whether there are one or more protein binding partners that join this TF protein to form an activation complex to trigger OTA production. It is likely a few more proteins participate in this pathway since secondary metabolite regulation is highly complex and intertwined. This theory is further supported by our experiments which have shown that inserting a promoter upstream of the TF gene does not increase the levels of OTA significantly. Hence, there must be other

binding partners whose expression levels also need to be upregulated in order to enhance OTA production. In 2013 it has been reported that the deletion of VeA and LaeA global regulators leads to drastic decrease in OTA synthesis in *A. carbonarius* [185]. Since, *A. westerdijkiae* OTA cluster has a homolog in *A. carbonarius*; it is most likely that the deletion of the global regulators would affect the OTA production in *A. westerdijkiae* as well. Hence, there might be a role of VeA and LaeA homologs in this whole cascade. It is also worthy to note that there is a presence of a zinc-finger transcription factor next to the oxidoreductase gene in the *A. westerdijkiae* OTA cluster (which we did not investigate). According to a study in *A. carbonarius,* a certain degree of differential expression has been seen in levels of zinc finger TFs in OTA producing and non-producing strains [198]. Hence it might also happen that this zinc-finger transcription factor next to the oxidoreductase gene could have a role in controlling OTA synthesis. A protein kinase gene is also present next to the zinc-finger TF and could also be a part of this regulatory process.

We have seen in the earlier section that the regulation of secondary metabolism is closely related to the rate of sporulation, pigmentation, virulence etc. Hence, it is not surprising that the presence or absence of the tf*Aw*ota has been found to affect both sporulation and pigmentation. Our observation also indicates that this kind of regulation might be present/conserved in the other homologous putative clusters of OTA found in *A. niger CBS513.88* (An15g07890) and *A. carbonarius ITEM5010* (Acar5010_007821) as well i.e. across species of Aspergilli. Since, the tf*Aw*ota protein has a predicted disordered/unstructured N-terminal, we might surmise that this disorderliness confers a certain degree of flexibility of binding to a number of second messenger molecules/proteins partners rather than a specific one. Further investigation is needed to identify possible ligands/binding partners of this *tfAw*ota protein for the better understanding of the regulation involved in OTA synthesis or the other phenotypes that get affected by its deletion.

Chapter 4 - Biosynthesis of circumdatins and other peptidyl alkaloids in *A. westerdijkiae*

Introduction

Naturally occurring alkaloids are nitrogen containing cyclic organic compounds with many of them exhibiting potent bioactivity [204]. Alkaloids can be divided into several large classes according to their biosynthetic origin. Peptidyl alkaloids are structurally diverse and their backbone or skeleton is synthesized by NRPS [205]. Peptidyl alkaloids contain heterocyclic nitrogen and the peptide backbone is usually derived from phenylalanine, proline, lysine, tryptophan and ornithine [204]. Many among this class of secondary metabolites are synthesized by NRPS using non-proteinogenic amino acid building blocks [205]. An increasing number of peptide based alkaloids have been isolated from filamentous fungi in recent years. Fungi in general and Aspergilli in particular are prolific producers of peptidyl alkaloids which are often known for their toxicity toward mammalian tissues [205]. A characteristic feature of many of these metabolites is the multi-cyclic, constrained architectures that can lead to high affinity for biological targets [205]. Several fungal alkaloids have been studied at the genetic level, but by far the ergot alkaloids and the indole diterpene group of alkaloids remain the most studied ones [46]. In this chapter, we will focus on the study of fungal alkaloids that involve NRPS for their biosynthesis.

Ergot alkaloids (originally isolated from *Claviceps purpurea*) have been isolated from the human pathogen *A. fumigatus* [206].Ergot alkaloids are indole alkaloids which contain a characteristic tetracyclic ergoline ring system and constitutes one of the most widely studied groups of peptide derived alkaloids [207]. Indole alkaloids are usually derived from tryptophan and dimethylallyl pyrophosphate[208]. Ergot alkaloids can be classified into broadly three types depending on their structures viz. clavines, ergoamides and ergopeptines. Clavines contain just the tetracyclic ergoline ring system or a tricyclic precursor of it [209]. Ergoamides and ergopeptines are derivatives of D-lysergic acid [209] and also contain the ergoline ring [209]. Clavine alkaloids are produced by many other fungi other than the members of Claviceps including Penicillum and Aspergillus [209]. On the other hand, ergopeptines and ergoamides are mainly produced by members

of the genera Claviceps and Epichloe [209]. Some of the ergot alkaloids are used as therapeutics for headaches and migraines due to their property of inhibiting tryptaminergic receptors [207]. These small molecules also interact with serotonin, dopamine and adrenergic receptors of the central nervous system as well as the adrenergic receptors present on blood vessels and hence find various uses as pharmaceutical drugs such as treatment of parkinson's disease [210]. The difference in the bioactivities of these molecules is because of the different functional groups attached to the D-lysergic acid moiety [209]. The peptidyl alkaloids fumigaclavine C, festuclavine, fumigaclavine A, and fumigaclavine B which also belong to the ergot alkaloid group have been isolated from *A. fumigatus* [211]. *A. fumigatus* is also known for the production of the prenylated indole alkaloid fumitremorgin, which is a toxin and is capable of inducing convulsion [212]. Fumitremorgens are derived from proline, tryptophan and dimethylallyl pyrophosphate [208].

Indole alkaloids in general can be of several categories depending on their biosynthetic origin. They could be synthesized by a bi-modular (notoamide) or a trimodular NRPS (asperlicin and ardeemins) or could be a product of a NRPS-PKS hybrid (cyclopiazonic acid) or a diterpenoid (paxillin) [208]. Recently, several new prenylated indole diketopiperazine alkaloids have been reported from *A. fumigatus* (isolated from sea-cucumber), including novel molecules like, spirotryprostatins, some derivatives of fumitremorgin and oxoverruculogen as well as a number of known compounds [213]. Some of the peptidyl alkaloid compounds like fumitremorgin, oxoverruculogen paspalitrem, penitrem, and tryptoquivaline are tremorgenic in nature i.e. they are capable of inducing neurological disorders [214]. The number of piperazinedione ring containing indole alkaloids is expanding and are gaining prominence due to many of their interesting biological activities. Infact, many of the well studied bi-modular NRPS systems have been associated with the production of diketopiperazine alkaloids in fungus [205]. The gene cluster responsible for the synthesis of another set of prenylated indole alkaloids notoamide and stephacidin has been reported and characterized in the marine derived *Aspergillus sp* MF297-2 in 2010 [215]. The notoamide/stephacidin cluster has been found to be associated with a bi-modular NRPS, along with a number of FAD- binding oxidoreductases and prenyl transferases. The synthesis of peptide based alkaloids that contain isoprene groups usually involve enzymes such as aromatic prenyltransferases like

dimethylallyltryptophan synthases in addition to the NRPS and other tailoring enzymes [205].

A large number of fungal peptidyl alkaloids are synthesized using the non-proteinogenic amino acid anthranilate which is a derivative of tryptophan. Benzodiazepine, quinoline and acridine group of peptide based alkaloids use anthranilate for their biosynthesis. Since fungi use tryptophan as a building block for many of the signature indole alkaloids, clearly some of the tryptophan is diverted for the synthesis of anthranilate [205]. The anthranilate-derived frameworks are usually generated using bi- and trimodular NRPS systems and subsequently released as bicyclic 6,7-benzodiazepinediones, 6,6,6-tricyclic pyrazinoquinazolinediones, or as 6,6,7,6-diazepinone-containing cyclic scaffolds [205]. They are then further processed by four kinds of tailoring enzymes. These enzymes include – a set of FAD-dependent oxygen-utilizing enzymes, a set of FAD-dependent indole epoxygenases [205]. The second kind of enzyme is responsible for the oxidizing the pyrazinone ring to a cyclic imine that undergoes intramolecular addition reactions by N, and O nucleophiles. A set of acylation enzyme catalysts that transfer acyl groups from acyl thioester substrates and the final set of enzymes consists of prenyl transferases adding dimethylallyl groups to the beta position of the indole ring of the framework and thereby facilitates heterocyclization [205]. Aszonalenin (*N. fischeri*), cyclopenin (*A. nidulans*), circumdatin (*A. ochraceus*), fumiquinazolines (*A. fumigatus*) all belong to this class of peptidyl alkaloids [205]. Ardeemins isolated from Aspergillus which is capable of blocking the efflux of anticancer drugs from the tumor cells [15] are synthesized by a tri-modular NRPS systems using anthranilate, alanine and tryptophan. Interestingly, the same set of of amino acids when recruited by a NRPS system in a different order (i.e. anthranilate, tryptophan and alanine) leads to the synthesis of fumiquinazolin [216]. The mycotoxin asperlicin reported to be produced by *A. alliaceus*, also uses the non-proteinogenic amino acid anthranilate as a building block and is synthesized by a tri-modular NRPS system [105].

Genome mining and subsequent characterization of biosynthetic gene clusters has led to the discovery of a series of new peptidyl alkaloids from fungi. Fungal peptidyl alkaloids form only a subset of a huge repertoire of fungal secondary metabolites and a very few among them have been discussed in this section. Although a significant amount of progress has been made with regard to knowledge of microbial peptidyl alkaloid

biosynthesis, with every new genome that is being reported and with more and more compounds identified, a lot still remains to be explored and understood.

A. westerdijkiae was formerly known as *A. ochraceus* till the bifurcation of *A. ochraceus* into *A. westerdijkiae* and *A. ochraceus* in 2004 [130]. The peptidyl alkaloid circumdatin A, B, C, D, E, F, G and H have been reported from *A. ochraceus* previously [85, 217-220]. In 2013, two new benzodiazepine alkaloids, circumdatins K and L, two new prenylated indole alkaloids, 5-chlorosclerotiamide and 10-epi-sclerotiamide have been reported from the deep-sea-derived fungus *A. westerdijkiae* DFFSCS013 [221]. The indole alkaloid stephacidin A has also been reported from *A. ochraceus* previously in 2002 [222]. Our whole genome sequencing of *A. westerdijkiae* revealed a number of NRPS gene clusters that are potentially involved in the biosynthesis of peptidyl alkaloids. In this study, we isolated several peptidyl alkaloids from *A. westerdijkiae* culture broth and biomass that includes notoamides, avrainvillamide, stephacidin B, circumdatin A, B and F. The homologous biosynthetic gene clusters for notoamides, avrainvillamide, stephacidin B and hexadehydroastechrome (HAS) have been identified in the *A. westerdijkiae* genome. The biosynthetic gene cluster for the circumdatins has not been reported so far. Here we report the identification of a NRPS gene cluster potentially responsible for the biosynthesis circumdatins along with a proposed biosynthetic mechanism for circumdatin production.

Materials and method

The protocols for cultivation of *A. westerdijkiae*, identification of biosynthetic gene clusters by AntiSMASH, other bioinformatic analysis along with isolation and identification of circumdatins from *A. westerdijkiae* can be found in chapter 2. For the detailed protocol for knockout generation and fungal transformation refer chapter 3.

Results

Isolation of peptidyl alkaloids from *A. westerdijkiae*- Under our cultivation conditions, a number of compounds could be detected readily by TLC and HPLC (RP-C18 column) from the culture broth and biomass of *A. westerdijkiae*. The identity of the compounds was first probed by using liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) (with the culture broth and biomass obtained from 10 L liquid medium, preparative HPLC was used for separation and isolation of the secondary metabolites for further structure elucidation). As discussed previously, we isolated a

dozen compounds (range from 0.1 to 5 mg) from the culture broth and biomass. Among the compounds whose identities were established by high-resolution mass spectrometry and NMR spectroscopy are several peptidyl alkaloids, which includes notoamide I (m/z 446.2083) [166], avrainvillamide (m/z 446.2080) [167], stephacidin B (m/z 891.4071) [223], circumdatin A, B, and F (m/z 394.1395, 364.1313 and 292.1090 respectively) [165, 224] (see Appendix X and XI). Notoamide I, avarinvillamide and stephacidin B are structurally related and were likely to be produced by the same biosynthetic pathway. Hexadehydroastechrome (HAS) is a peptidyl-alkaloid based siderophore that has been known to be involved in iron acquisition and contribute to the pathogenicity of *A. fumigatus* [225]. Circumdatins belong to a class of benzodiazepine alkaloids formed by the fusion of benzene and a diazephine ring. Isolation of circumdatin A, B, C, D, E, F and G from *A. ochraceus* have previously been reported [220] [217] [218]. Circumdatin H, an inhibitor of mitochondrial NADH oxidase, was isolated from *A. ochraceus* in 2005 [219]. Circumdatin I, which has an ultraviolet-A protecting property, was isolated from the marine fungus *Exophiala sp* more recently [85].

Identification of the biosynthetic gene clusters of peptidyl alkaloids from the *A. westerdijkiae* genome- The complete genome of *A. westerdijkiae* revealed a large number of PKS, NRPS, PKS/NRPS hybrid and terpene biosynthetic gene clusters. A close inspection of the annotated NRPS gene clusters reveals a number of clusters potentially involved in the biosynthesis of peptidyl alkaloids. The potential gene clusters for peptidyl alkaloid biosynthesis include clusters 10, 11, 12, 31, 48, 55, 68 and 72 (See Table: 4.3 and Appendix III). The prediction is based on the presence of the genes that code for di- or tripeptide-synthesizing NRPS and other ancillary enzymes such as prenyltransferase, dimethylallyl tryptophan synthase cytochrome P450 and a FAD-containing mono-oxygenase.

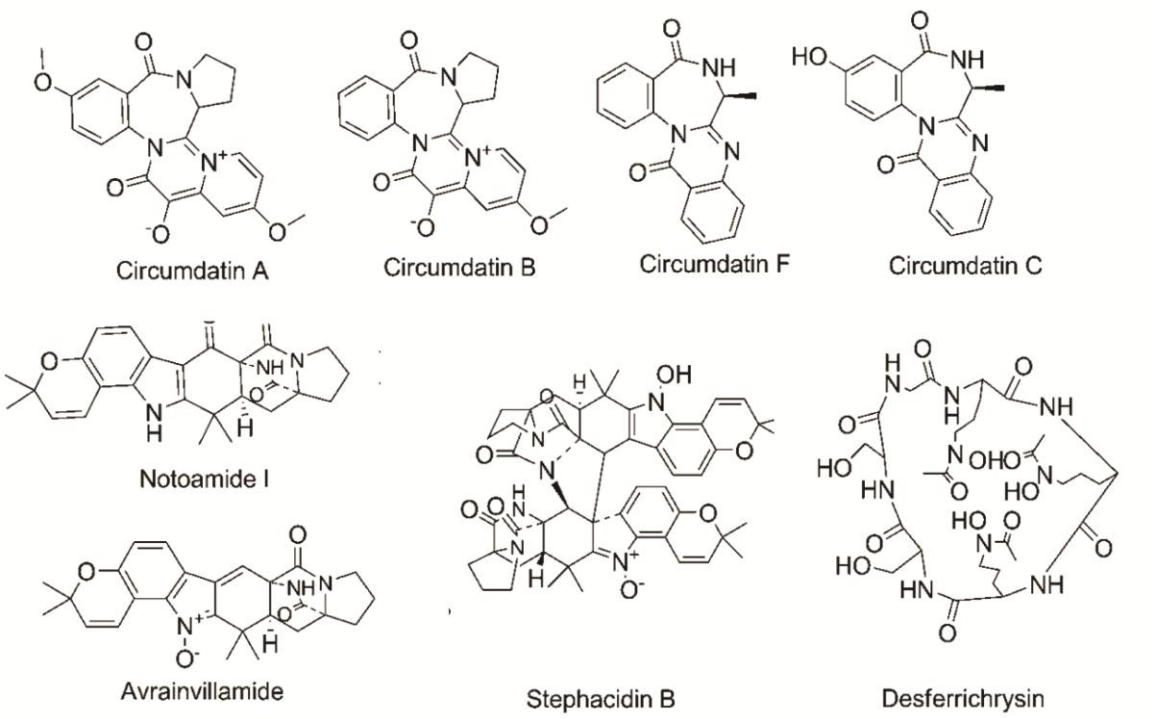


Figure 4.1: The structures of the alkaloids isolated from *A. westerdijkiae* culture broth and biomass (Illustration made using the ChemDraw tool).

Notoamides are prenylated indole alkaloids whose biosynthetic mechanism was studied by Ding et al. [226]. Several structurally related alkaloids that include stephacidin A and avrainvillamide were also shown to be produced by the notoamide pathway [215] (Fig: 4.2), with stephacidin B formed from the dimerization of avrainvillamide [227]. One of the NRPS gene clusters (cluster 48) from *A. westerdijkiae* genome shows high homology to the notoamide biosynthetic gene cluster from *Aspergillus sp.* MF297-2 with the six-domain-containing NRPS and the ancillary prenyltransferase, cytochrome P450 and FAD-containing oxidoreductases conserved (see Fig: 4.2, Table 4.1). Cluster 48 is most likely responsible for notoamide biosynthesis in *A. westerdijkiae* [215]. Stephacidin B and avrainvillamide are also like the products of the same biosynthetic pathway.

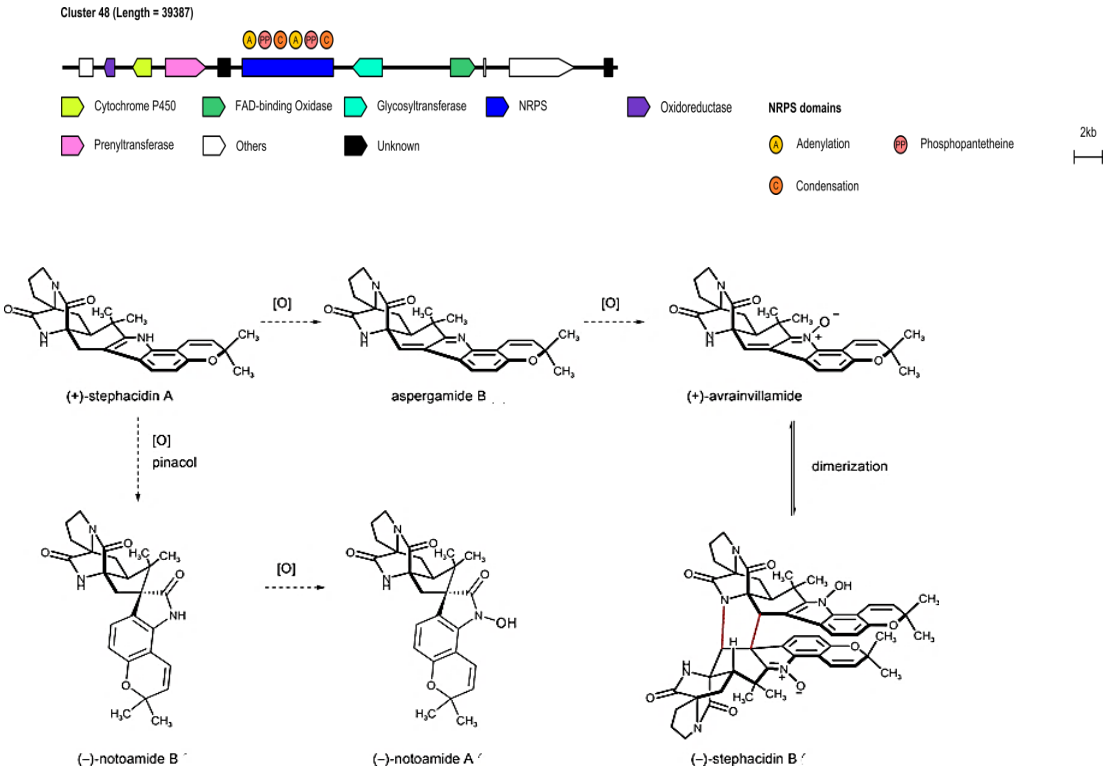


Figure 4.2: Notoamide biosynthetic gene cluster and mechanism. A. Notoamide biosynthetic gene cluster. B. Proposed biosynthetic mechanism of Notoamide, stephacidin B and avrainvillamide [223].

Gene in *A. westerdijkiae*	*Aspergillus sp.* MF297-2 notoamide biosynthetic gene locus	Identity
Gene 1 Hypothetical protein	ADM34143.1(Hypothetical protein)	60%
Gene 2 FAD monooxygenase	ADM34142.1(FAD binding protein)	68%
Gene 3 Cytochrome p450	ADM34141.1(Cytochrome p450)	58%
Gene 4 Cytochrome p450	ADM34140.1(Cytochrome p450)	75%
Gene 5 Aromatic prenyltransferases	ADM34139.1(Aromatic prenyl transferase)	69%
Gene 6 NRPS	ADM34138.1 (NRPS)	61%
Gene 8 FAD linked oxidoreductase	ADM34137.1 (Oxidoreductase)	69%
Gene 5 Aromatic prenyltransferases	ADM34136.1 (Aromatic prenyl transferase)	32%
Gene 2 FAD monooxygenase	ADM34135.1 (FAD binding protein)	45%
Gene 7 NmrA family protein	ADM34134.1 (NmrA)	69%

Table 4.1: Comparison of the putative notoamide biosynthesis gene cluster of *A. westerdijkiae* genome with the the notoamide cluster reported from *Aspergillus sp.* MF297-2.

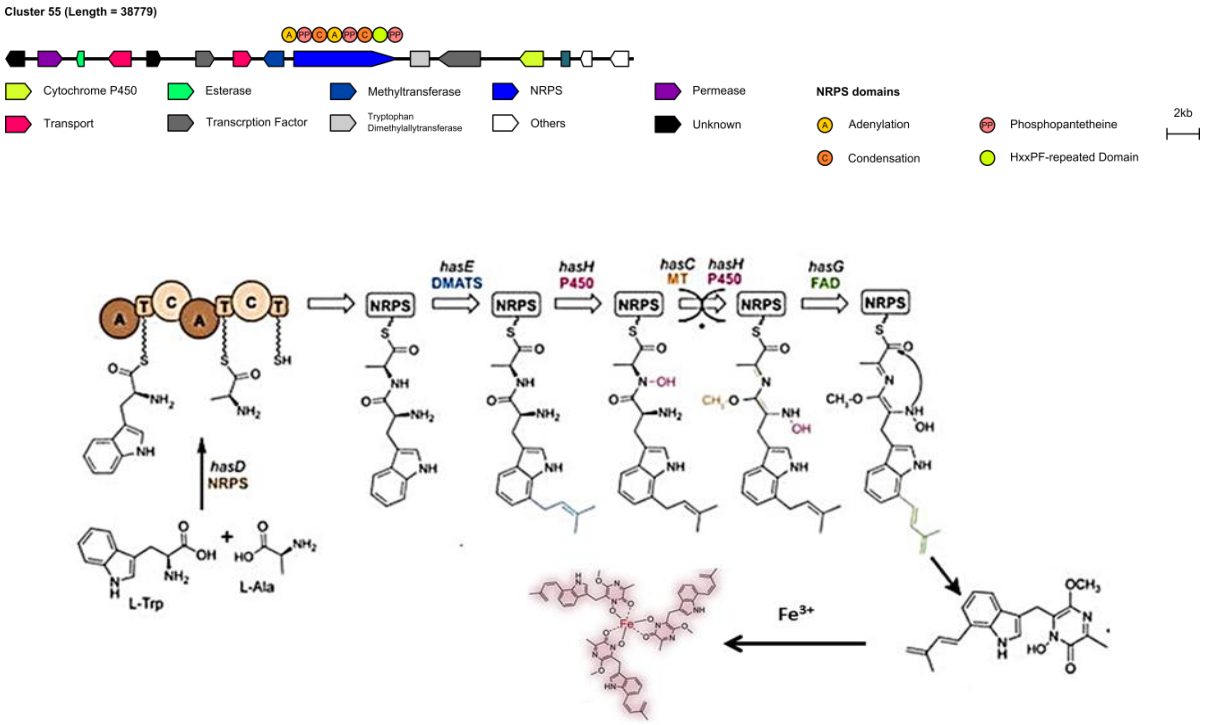


Figure 4.3: Hexadehydroastechrome (HAS) biosynthesis in Aspergillus. A. Hexadehydroastechrome biosynthetic gene cluster. B. Proposed biosynthetic mechanism of hexadehydroastechrome [225].

Gene in *A. westerdijkiae*	*Aspergillus fumigatus*-Hexadehydro-astechrome (HAS) biosynthetic gene cluster	Identity
Gene 6 Transcription factor	AFUA_3G12890 (Transcription factor)	56%
Gene 7 Major facilitator protein	AFUA_3G12900 (MFS transporter)	76%
Gene 8 Methyltransferase	AFUA_3G129010 (O-methyltransferase)	82%
Gene 9 NRPS	AFUA_3G12920 (NRPS)	67%
Gene 10 Tryptophan dimethylallyltransferase	AFUA_3G12930 (Dimethylallyl tryptophan synthase)	75%
Gene 12 Transcription factor	AFUA_3G12940 (Transcription factor)	68%
Gene 11 FAD binding protein	AFUA_3G12950 (FAD binding protein)	73%
Gene 13 Cytochrome p450	AFUA_3G12960 (cytochrome p450)	68%

Table 4.2: Comparison of the putative hexadehydroasteochrome biosynthesis gene cluster in *A. westerdijkiae* genome with the hexadehydroasteochrome cluster reported from *Aspergillus fumigatus.*

The peptidyl alkaloid based siderophore hexadehydro-astechrome was isolated from a cryptic gene cluster in *A. fumigatus*. *A. fumigatus* which is known to cause aspergillosis of lungs has been reported to have shown a sharp increase in virulence upon the activation of the hexadehydroastechrome cluster [179]. The hexadehydro-astechrome biosynthetic gene cluster from *A. fumigatus* has been characterized to reveal the involvement of a NRPS and several ancillary proteins [179]. The hexadehydro-astechrome biosynthetic gene cluster can be found in a large number of Aspergillus and

other fungal species, indicating the wide use of this siderophore for iron acquisition. Gene cluster 55 (identified by AntiSMASH) of *A. westerdijkiae* genome share high similarity with the hexadehydroastechrome biosynthesis gene cluster in *A. fumigatus* [179] (see Fig: 4.3, Table 4.2), with the NPRS and the ancillary enzymes conserved in *A. westerdijkiae*.

As discussed earlier, circumdatins have been isolated from *A. ochraceus* and other fungal strains but the corresponding gene cluster remains unknown. Based on the biosynthetic mechanism for other structurally related di- and tripeptidyl alkaloids [205], circumdatins are most likely produced by a di- or tri-modular NRPS systems that incorporate two anthranilate moieties into the structure. If the mechanism is conserved, a cytochrome P450 and a methyltransferase enzyme will be necessary to synthesize circumdatins. A number of fungal peptidyl alkaloids have been identified that make use of the non-proteinogenic amino acid anthranilate as a building block to construct bi-, tri- or tetracyclic nitrogen-containing alkaloid backbones [205]. Circumdatins are structurally similar to cyclopenin and asperlicin despite the use of different amino acid building blocks. The peptidyl alkaloid cyclopenin is synthesized using anthranilate and phenylalanine; whereas the toxin asperlicin is formed by anthranilate and tryptophan [205]. Based on the protein sequences homology, we attempted to identify the anthranilate-specific A-domains in *A. westerdijkiae* genome. By sequence alignment, we identified several A-domains that share significant homology with the anthranilate-specific A-domains. Phylogenetic analysis shows that the A1 domains of the NRPS genes of cluster 10, cluster 12, cluster 49 and the A3 domain of cluster 10 genes exhibit the highest homology to the anthranilate-specific A-domains - Afua_6g12080(A1), NFIA_057960(A1) and NFIA_058030(A1), NFIA_043670(A1), ACLA_017890 (A1), ACLA_076770(A1) and ACLA_095980(A1) [228]. Among gene clusters 10, 12, and 49, cluster 49 can be ruled out for the synthesis of the tri-peptidyl circumdatin because it contains four A domains. Cluster 10 is the most likely candidate for circumdatins for a few reasons. First, it contains two NRPS genes that code for three A domains in total and two of them have predicted specificity for anthranilate (see Fig: 4.5). Second, the domain organizations of the two NRPS module are highly similar to that of the asperlicin gene cluster [105]. Third, the gene cluster contains gene encoding the cytochrome P450 and methyltransferase required for circumdatin synthesis. Cluster 12 contains a single NRPS gene that codes for a NRPS with two A domains, one of which is anthranilate specific (see Fig: 4.5). However, the gene cluster has a cytochrome P450 gene but does not have a

methyltransferase gene. Based on these considerations, we postulate that cluster 10 is responsible for the biosynthesis of circumdatins by *A. westerdijkiae*.

Gene knockout to inactivate the predicted circumdatin biosynthetic gene cluster- In order to confirm the involvement of the genes in circumdatin biosynthesis, we disrupted the *nrps*, *cytochrome P450* and *methyltransferase (MT)* genes of cluster 10 individually by using the same protocol discussed in previous chapter 3. Next the effect of gene disruption on circumdatin production will be examined by using HPLC and LC-HRMS. The abolition of circumdatin production due to *nrps* and *cytochrome P450* gene disruption will confirm the role of cluster 10 in circumdatin biosynthesis (see Table 4.3 for the list of knockouts generated).

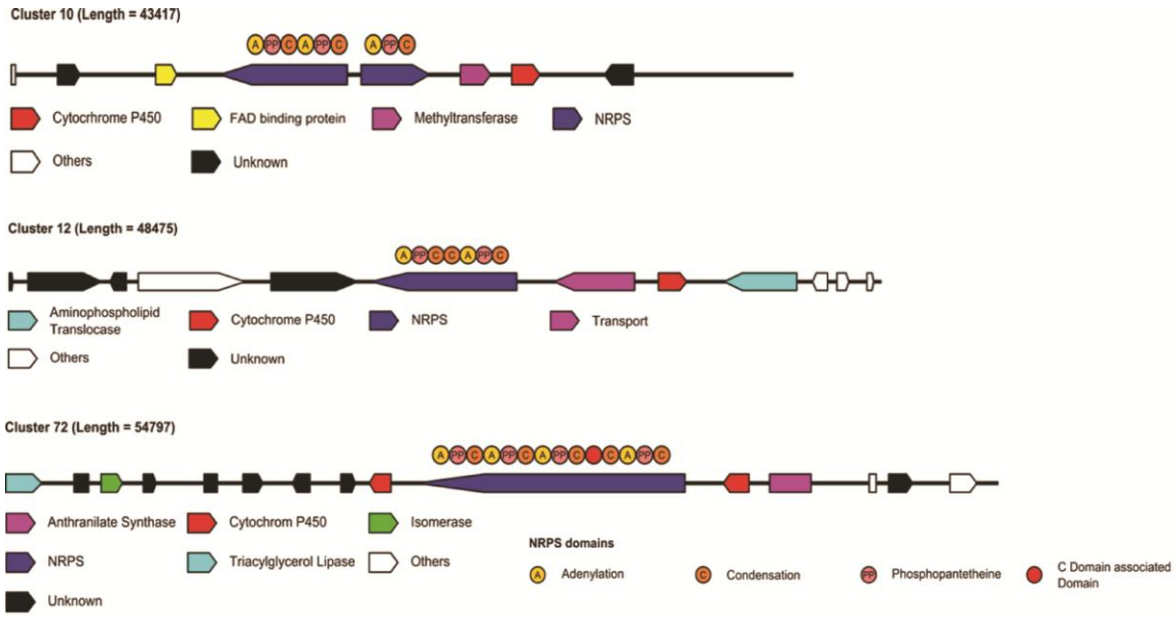


Figure 4.4: Three gene clusters predicted to contain anthranilate-specific A domains (Illustration drawn using Ink-scape tool).

Gene cluster	Gene knocked out				Comments
10	NRPS I	NRPS II	P450	MT	Contains two NRPS genes with two and one "A" domains respectively, a cytochrome P450, a methyltransferase.
11	NRPS I	NRPS II	P450	-	Contains two NRPS genes with two "A" domains each, an elymoclavin monooxygenase. Antismash predicts similarity with ergot alkaloid cluster.
12	NRPS	-	P450	-	Contains NRPS genes with two "A" domains and a cytochrome p450. Antismash predicts similarity with ergot alkaloid cluster.
31	NRPS	-	-	-	Contains a NRPS (with two "A" domains) and a methyltransferase and a FAD binding protein.
68	NRPS	-	-	-	Contains a NRPS (with three "A" domains) and a methyltransferase.
72	NRPS	-	-	-	Contains a NRPS (with four "A" domains), a cytochrome P450, a methyltransferase and an anthranilate synthase.

Table 4.3: List of all knockouts that have been generated to identify the circumdatin biosynthesis gene cluster.

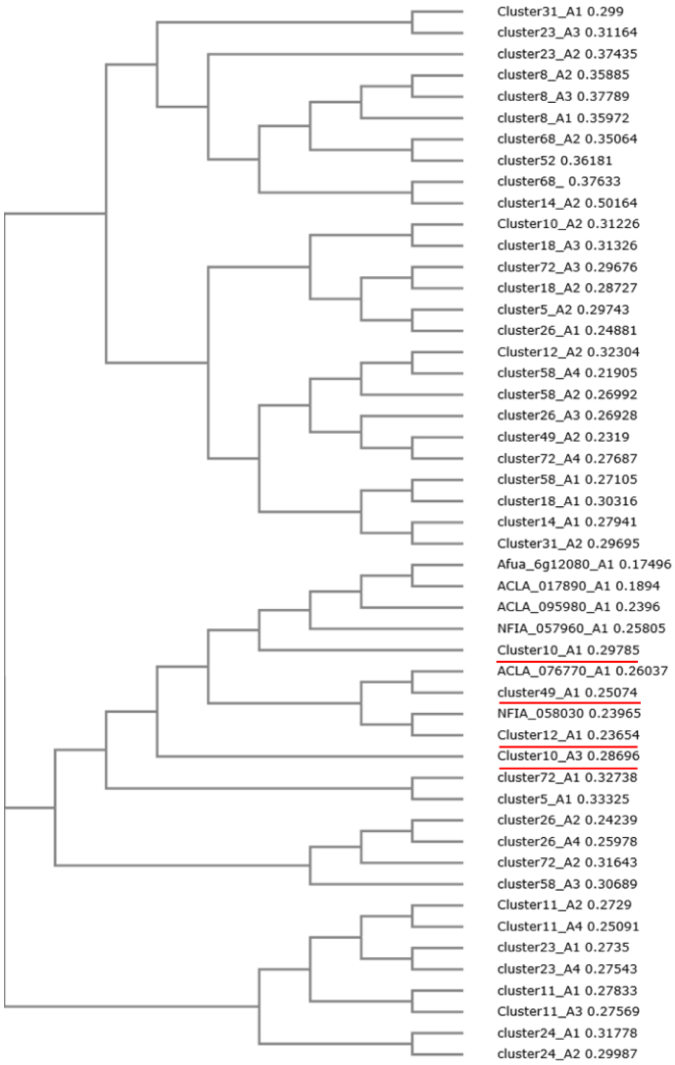


Figure 4.5: Phylogenetic analysis of all the adenylation (A) domains found in the *A. westerdijkiae* genome. Several known anthranilate-specific A domains from other species are included here for the identification of anthranilate-specific A domains from *A. westerdijkiae* (Illustration obtained using Clustal omega platform).

Discussion

Sequencing the whole genome of *A. westerdijkiae* revealed a large number of biosynthetic gene clusters, with several NRPS gene clusters potentially involved in the biosynthesis of peptidyl alkaloids. We isolated several peptidyl alkaloids from the culture broth and biomass of *A. westerdijkiae* that include notoamide and the related stephacidin B and avrainvillamide and circumdatins A, B and F. The gene clusters for notoamide and hexadehydroastechrome biosynthesis from other species have been characterized [215, 225]. We could readily identify the homologs of notoamide and hexadehydroastechrome clusters in *A. westerdijkiae* genome based on high sequence homology (see Table: 4.1 and 4.2). Although the tri-peptidyl circumdatins were first reported from *A. ochraceus* in late 1990 [217, 220], the biosynthetic gene cluster of the circumdatins has not been identified so far. By identifying the potential anthranilate-specific A domains and analyzing the domain organization of the NRPS modules and ancillary enzymes, we identified cluster 10 as the potential circumdatin biosynthetic gene cluster. We propose a biosynthetic mechanism for the circumdatins as illustrated in Fig: 4.6. Two molecules of anthranilate and one molecule of proline/alanine are condensed together by the three NRPS modules. In agreement with the function of the NRPS, the active site of one of the A domains (in the case of cluster 10, it is the second A domain of NRPS I) is predicted to have similarity to A domains that have preference for proline and alanine, whereas the other two are closely related to anthranilate-specific A domains (i.e. the first A domain of NRPS I and the A domain of NRPS II) (see Fig: 4.5). The condensation of proline/alanine with the two anthranilate moieties is accompanied by a cyclodehydration reaction catalyzed by the condensation (C) domains. Upon the formation of the cyclic intermediate, hydroxylation and methylation will afford the circumdatins F, H and J. Circumdatin F is likely to result from the accidental loading of alanine instead of proline by one of the A domain. Circumdatins A and B are likely artifacts produced by the non-enzymatic oxidation of circumdatins H and J by air. The timing of the hydroxylation catalyzed the cytochrome P450 in the biosynthetic mechanism remains to be determined. We have prepared the *cytochrome P450* knockout mutant strain to detect the intermediates of the biosynthetic pathway encoded by cluster 10. Analysis of the products of the mutant strain will reveal whether the hydroxylation occurs at the early stage (i.e. hydroxylation of anthranilate) or late stage (hydroxylation of the cyclic intermediate) of the biosynthesis.

To provide experimental evidence for the role of cluster 10 in circumdatin biosynthesis, we have knocked out the *nrps* I and II, *cyt P450* and *methyltransferase* genes in some of the potential peptidyl alkaloid synthesizing clusters. We are in the process of cultivating the mutant strains and analyzing the culture broth extract of the knockout strain. To fully establish the biosynthetic mechanism, we also plan to express the biosynthetic enzymes in heterologous hosts for in vitro enzyme assays. The role of the NRPS can be further established by (a) the expression of the A domains in *E. coli* host and subsequently identify their substrate specificity or (b) express the entire NRPS protein and identify its product by in vitro enzymatic assay.

And lastly, given the discrepancy between the large number of NRPS gene clusters found in the *A. westerdijkiae* genome and the relatively small number of peptide-based compounds identified in the culture broth/biomass, some of the gene clusters are likely silent or suppressed. Based on bioinformatic analysis, we predict that *A. westerdijikae* is capable of producing several other peptidyl alkaloids in addition to the ones discussed above. Some of the gene clusters do not share significant similarity or homology with any other characterized gene cluster and are likely to produce compounds with novel structures. Co-cultivation and genetic manipulation are currently being explored in our lab to activate or awaken these gene clusters.

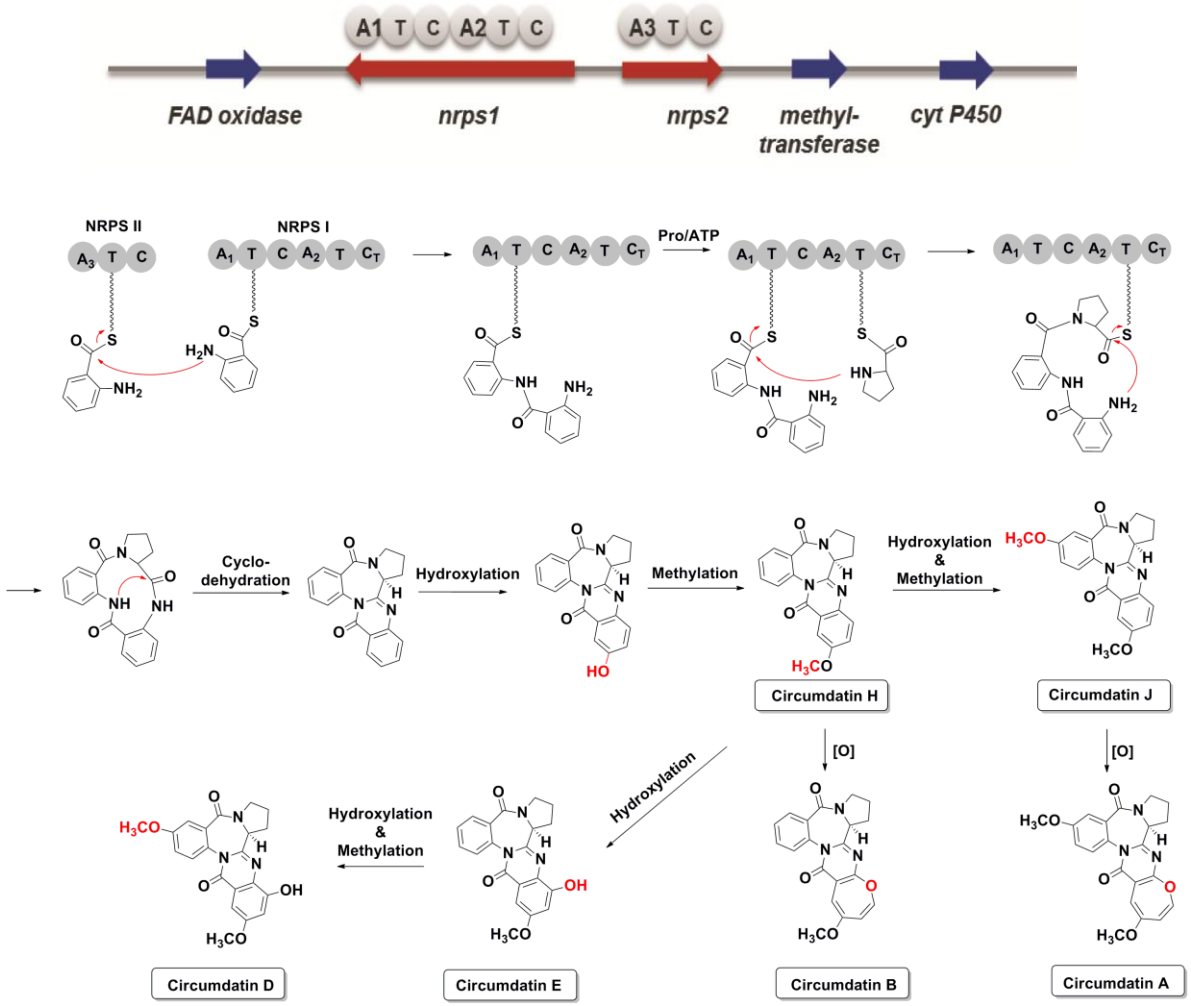


Figure 4.6: Proposed mechanism of circumdatin synthesis in *A. westerdijkiae* [by cluster 10 found in the scaffold 2 of the *A. westerdijkiae* genome] (Illustration made using the ChemDraw tool)

Chapter 5 - Genome sequence and biosynthetic potential of *Nocardia jinanensis*

Introduction

Actinomycetes have been known to produce a bounty of pharmacologically important secondary metabolites. Majority of the bioactive small molecules reported till date come from this group of organisms. Actinomycetes are mainly soil dweller, exhibit filamentous morphology and possess high genomic G+C content [191]. In fact, the characteristic smell of soil that is perceived everytime it rains is due to the secretion of two small molecules – geosmin and methylisoborneol (MIB) by actinomycetes [229]. Apart from soil, actinomycetes have been isolated from various habitats like marine sediments, coral and sponges for the discovery of novel secondary metabolites [229] [230]. Among the actinomycetes, the genus Streptomyces is the most important in terms of secondary metabolite synthesis. The better part of the actinomycete derived biologically important drugs in the market are from the genus Streptomyces (nystatin, vancomycin, noviobiocin, kanamycin, tetracycline, neomycin, streptomycin are just a few examples) [231]. The other genera belonging to actinomycetes - Saccharopolyspora, Amycolatopsis, Micromonospora, Nocardia, Actinoplanes and Mycobaceria are also known to produce a number of important secondary metabolites but the numbers are lower compared to Streptomyces [133].

More than 70% of the versatile bioactive microbial compounds are derived from actinomycetes, which comprises of the genus Streptomyces (68%) and the rare actinomycetes (32%) [232]. Among the rare actinomycetes with potential for production of bioactive compounds, the genus Nocardia is the most predominant one (see Fig: 2.1, Chapter 2) [232]. Distinct from the more common genera of Actinomycetes such as Streptomyces, Nocardia is is partially acid-fast in nature. Acid-fast organisms contain mycolic acid in their cell wall and hence have to be stained using acid-fast staining techniques. In this respect, the genus Nocardia is close to *Mycobacterium sp* as both of them are non-spore formers, catalase positive and have a mycolic acid rich cell wall [233].

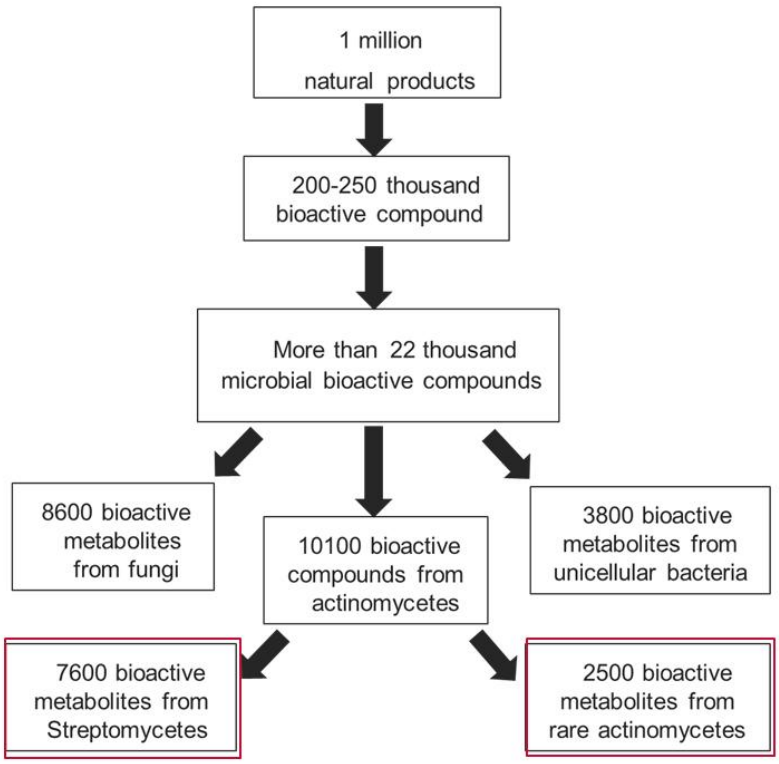


Figure 5.1: Distribution of known bioactive compounds according to their sources. Illustration adapted from [133].

Various Nocardia strains have been isolated from both aquatic and terrestrial habitats as well as from the tissue of infected patients. Several Nocardia strains, such as *N. farcinica, N. brasiliensis* and *N. cyriacigeorgica*, are of biomedical importance because they are known as opportunistic human pathogens that cause nocardiosis. According to the public health datasheet of the Canadian health agency the genus Nocardia contains 70 species of which 25 have been reported to be pathogenic to humans. Among the 25, *N. vertana* [234], *N. farcinica* [235], *N. cyriacigeorgica* [236], *N. brasiliensis* [237], *N. nova* [235] and *N. otitidiscaviarum* [238] are the best known pathogenic Nocardia strains. The infection caused by Nocardia can be airborne, and some of the infections such as the ones caused by *N. farcinica,* can be fatal. Nocardia strains are also capable of causing pulmonary diseases as well as brain and skin infections. Nocardiosis is a condition caused by nocardial infections with a wide variety of clinical manifestations in patients. Nocardiosis has been found to affect more than one part of the body as the infection spreads via the bloodstream mostly affecting the brain, forming lesions, or the skin forming abscesses in the dermis or epidermis [239].

The genus Nocardia has also attracted attention in the last two decades due to their versatility in secondary metabolite production [240] and their ability to breakdown complex hydrocarbons and polymers [241]. Several Nocardia strains are capable of desulfurizing crude oil and thereby facilitating the bio-refining of diesel and other derivatives of crude oil [242]. The strain *N. amarae* was reported to possess the ability of demulsifying oil-field emulsions [243]. Given their ecological, biomedical and biotechnical relevance, the complete or draft genomes of a number of *Nocardia* strains have been sequenced in recent years [244].The most notable and best annotated genomes include *N. farcinica* [245], *N. brasilensis* [246], *N. cyriacigeorgica* [247] and *N. nova* [248].

A number of Nocardia strains are endowed with the ability to produce secondary metabolites with novel chemical structure and bioactivity. *N. pseudobrasiliensis* produces the antimicrobial compounds nocardicyclin A and B [249], which are capable of inhibiting the growth of both Mycobacterium and Nocardia. One of the Nocardia strains is known to synthesize the type II fatty acid synthase inhibitor thiolactomycin [250]. The macrolide antibiotic tubelactomycin A which shows activity even against drug resistant variety of Mycobacterium has been isolated from a *Nocardia sp* [251]. *N. mediterranei* and *Nocardia sp* CS682 are known to produce the antibiotics rifamycin B [252] and nargenicin [253] respectively. The siderophore nocobactin, isolated from *N. farcinica*, is known to contribute to pathogenicity [254] of the microorganism. The pathogenic *N. brasiliensis*, of which the whole genome has been sequenced, produces the terpenoid compound brasilicardin A that exhibits immunosupressive activity [255] [256]. *N. brasiliensis* also produces the antifungal macrolide brasilinolide A and B [257] and indole alkaloid brasilidine A [258]. Nocardithiocin, which was isolated from the pathogenic strain *N. pseudobrasiliensis*, has been found to be effective against acid-fast bacilli like Mycobacterium [259]. A number of antitumor compounds such as amanistatin A & B [260], nocardichelin A & B [261] and the potent beta-lactam antibiotic norcardicin A have been isolated from various members of the genus Nocardia. [262]. The free radical scavenging molecule formobactin [263] and the thiopeptide antibiotic nocardithiocin [264] were isolated from *Nocardia sp*. Antibacterial lipopeptides have been reported to be produced by a marine Nocardia strain [265]. Considering the fact that all the compounds mentioned here were discovered from a relatively small number (~80) of recognized

Nocardia strains that have been isolated and cultivated, the genus of Nocardia seems to represent a rich yet underexplored source of secondary metabolites.

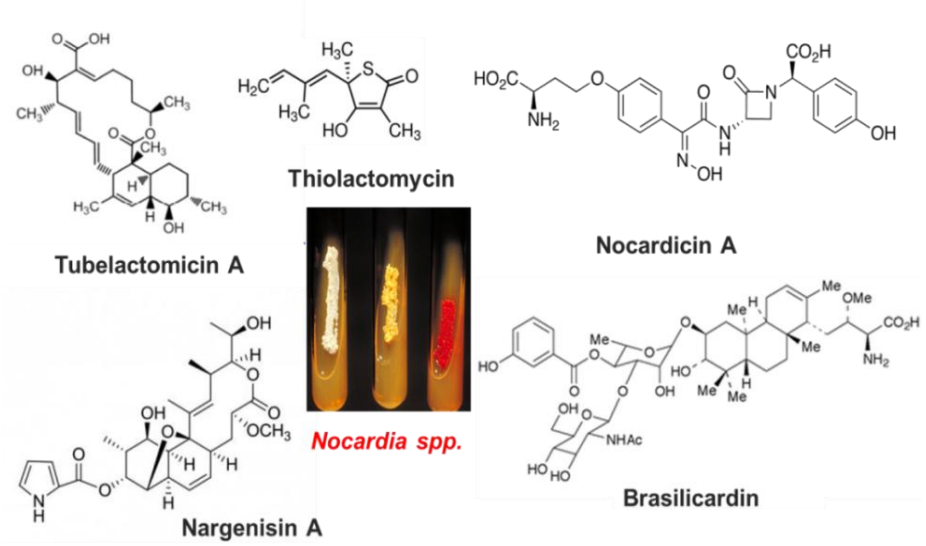


Figure 5.2: Some of the bioactive secondary metabolites isolated from *Nocardia sp*.

Research on secondary metabolites associated with Nocardia has been somewhat overshadowed due to their pathogenic nature. However the application of advanced genomics-guided methodologies along with genetic manipulation and metabolic engineering techniques can substantially help in exploring the metabolic capabilities of *Nocardia sp* as robust microbial cell factories [256]. The rapid increase in the number of Nocardia genomes being sequenced, could pave the way for a deeper understanding of the molecular biology underlying the metabolic versatility in this group of organisms [240].

The *N. jinanensis* strain NBRC 108249/ (type strain- 04-5195$^{T)}$ [132] was isolated from soil sample in China and is considered to be a pathogenic strain that can cause cellulitis [266]. In addition, *N. jinanensis* was also reported to produce the anti-osteoporotic compound amicoumacin B [132]. Amicoumacin B can increase the expression of the bone morphogenetic factor-2 gene that is important for bone development [132]. The biosynthetic origin of amicoumacin B, which contains a dihydroisocoumarin moiety (the same backbone that can also be found in mellein and orchratoxins), was unknown at the time when this study was initiated. With the initial aims of elucidating the amicoumacin B biosynthetic pathway and exploring the biosynthetic potential of *N. jinanensis*, and at

the same time gaining a deeper understanding of the pathogenicity of *N. jinanensis* by identifying virulence factors and secondary metabolites, we set out to sequence the genome of *N. jinanensis.* Further, we carried out bioinformatics analysis of the secondary biosynthetic gene clusters, bioactivity assay and metabolite profiling by using LC-MS and other analytic methods. The results reveal that *N. jinanensis* contains a large number of putative virulence factors shared with other pathogenic Nocardia strains. Our study also reveals the great biosynthetic potential of *N. jinanensis* as evidenced by a large number of secondary biosynthetic gene clusters found in the genome, which surprisingly lacks the amicoumacin B gene cluster.

Materials and method

Culturing of *N. jinanensis* - The *N. jinanensis* strain NBRC 108249 was grown at 28 ºC with shaking (180 rpm) for 14 days in ISP2 medium or Yeast Malt medium (Yeast extract 0.4 %, Malt extract 1 %, Glucose 0.4 %, pH 7.3) [132].and ISP3 medium (oatmeal 20g/l, 1ml trace element solution/L [0.1g $FeSO_4.7H_2O$,0.1g $MnCl_2.4H_2O$ and 0.1g $ZnSO_4.7H_2O$ per 100ml]), Stress medium (0.1 % Corn starch, 0.3 % Soy bean meal, 0.1 % $CaCO_3$, 0.2 % $MgSO_4$, 0.1 % Molasses in 1L of water [267].

Genomic DNA isolation and genome sequencing- *N. jinanenesis* was cultured in a 50ml flask containing Yeast Malt (ISP2) medium (yeast extract 0.4 %, malt extract 1 %, glucose 0.4 %, pH-7.3), at 28 ºC for 4 days with shaking. D-cycloserine (1.2 mg/ml) was added to the culture flask and the culture was further incubated at 28ºC for 48 hrs. Cells were harvested and the pellet was resuspended in lysis buffer (15 % sucrose, 0.05M EDTA, 0.05M Tris pH-8, lysozyme 3mg/ml) and incubated for 30 minutes at 37 ºC. The mixture was sonicated and further processed using the Tiangen Bacteria DNA kit. The eluted DNA was incubated with 1µl of RNAse A (10mg/ml) at 37 ºC for 90 minutes before centrifugation to collect the supernatant. The concentration of the eluted DNA sample was measured using NanoDrop and agarose gel electrophoresis was performed to check the quality of DNA before it was frozen and shipped for genome sequencing. The draft genome of *N. jinanensis* was obtained using whole-genome shotgun technique (Illumina platform) with the DNA fragments assembled using the SOAPdenovo tool at Macrogen Inc. (South Korea).

Gene annotation and gene cluster prediction- The *N. jinanensis* genome was annotated using BLASTNR and BLASTGO (functional annotation). The secondary metabolite gene

clusters were predicted using the online tool AntiSMASH (antibiotics & Secondary Metabolite Analysis Shell version 3.0).

Comparative genome analysis- A comparative analysis of *N. jinanensis* genome sequence was done with the genome of *N. brasiliensis* and *N. farcinica* using the web tools Mauve and WebACT (http://www.webact.org/WebACT/generate) and RAST (Rapid Annotation using Subsystem Technology). The RAST annotation files of *N. jinanensis, N. brasilensis, N. farcinica* and *N. cyriacigeorgica* were also submitted to COG (cluster of orthologous groups) for the identification of orthologs across species and genes which are unique to *N. jinanensis*. The COG database classifies genes into different functional categories and contains information on the orthologous genes of both prokaryotes and eukaryotes. This is a tool for the analysis of protein function and evolution among different organisms [268] [269] . RNAmmer 1.2 server [270] was used for 16s rRNA sequence prediction and phylogenetic tree was drawn using the MEGA (mega evolutionary genetics analysis) V6.0 tool [191].

Analysis of genes related to potential pathogenicity- We have also tried to shortlist the possible gene in *N. jinanensis* which might be involved in cell-wall synthesis and virulence/pathogenicity. For determining the putative genes involved in virulence and pathogenicity we used the list of virulence genes present in *N. farcinica* as a reference (we found the information from the genome view of *N. farcinica* on the *Nocardia farcinica* Genome Project Page; http://nocardia.nih.go.jp. We gathered from this list that the genes belonging to the genes belonging to the Mce (mammalian cell entry) and YrbE family, superoxide dismutases, alkylhydroperoxidases, mycolyltransferases, certain NRPSs, LSR2 proteins, some lysine-N-oxygenases etc. are usually involved in pathogenicity of an organism and hence we searched for the corresponding genes in the annotated orfs of *N. jinanensis.*

Antimicrobial activity assay- For the antimicrobial activity assay, 14 day old culture of *N. jinanensis* was harvested and the broth was extracted with equal volumes of ethyl acetate. The organic extract was dried and reconstituted in 100% methanol for all assays. Sterilized filter paper discs were impregnated with 30-40µl of organic extract from ISP2 and Stress medium and left for drying at room temperature inside the laminar air-flow hood. The control discs were impregnated with 100 % methanol and dried. The impregnated dried filter paper discs were then placed on LB (Luria-Bertanni) agar plates

containing the following organisms: *Bacillus subtilis*, *Staphylococcus aureus*, *Escherichia coli* and *Streptococcus pyogenes*, and incubated at 37 °C overnight. After incubation the plates were observed for the detection of inhibition zones around the filter paper discs. For the 96-well microtitre plate test, 10 µl of organic extract ISP2 and Stress medium was added on to designated wells and left for drying. 100 µl of bacterial culture (O.D 0.06) was added to assigned wells and incubated at 37 °C overnight with shaking at 150 rpm. The Tecan machine/plate reader was used to read the O.D of each well after incubation.

Extraction and analysis of culture broth- The culture broths (200 × 50 mL, total 10 L) of *N. jinanensis* were combined and centrifuged to separate the supernatant and the biomass. The supernatant was separated using a C18 column (250 g, Phenomenex, Sepra C18-E bulk packing, 50um, 65A) with two isocratic conditions of 100% water and 100% methanol. The 100% methanol fraction was evaporated to dryness using rotary evaporation and separated by Agilent prep C18 (5u, 30 x 100 mm) reversed-phase preparative HPLC (gradient elution: solvent A (0.1% HCOOH/ H_2O): B (0.1% HCOOH/CH3CN) 97:3 → 40:60 over 80 min, 40:60 → 0:100 over 15 min; flow rate 30 mL/min) to give cyclo(isoleucyl-prolyl) [271, 272], cyclo(leucyl-prolyl) [271, 272] and cyclo(phenylalanyl-prolyl) [271]. The freeze-dried biomass of *N. jinanensis* was extracted two times with methanol. The methanol extract was evaporated to dryness using rotary evaporation and separated by Agilent prep C18 (5u, 30 x 100 mm) reversed-phase preparative HPLC (gradient elution: solvent A (0.1% HCOOH/ H_2O): B (0.1% HCOOH/CH_3CN) 95:5 → 40:60 over 80 min, 40:60 → 0:100 over 15 min, flow rate 30 mL/min).

Results

General features of the *N. jinanensis* genome- The whole genome of *N. jinanensis* NBRC 108249/(type strain- 04-5195T) was sequenced by Macrogen Inc. (South Korea) using the HiSeq Illumina platform. After obtaining the DNA reads, the genome was assembled into scaffolds using *de novo* assembly programs. The assembled scaffolds were further subjected to bioinformatic analysis such as gene annotation and gene cluster mining in our lab. Overall, the draft genome of *N. jinanensis* has a GC content of 67.3% and consists of 161 scaffolds and a total of 7,957,899 bases. The complete genomes of several Nocardia strains have been sequenced prior to our study. The size of the *N. jinanensis* genome is larger than that of the *N. farcinica* (6.29 Mb) [245] and *N.*

cyriacigeorgica (6.19 Mb) [247], but smaller than that of *N. brasiliensis* (9.44 Mb) [246]. The detailed analysis of cluster of orthologs (COG) has been attached in Appendix VIII. The *Nocardia jinanensis* genome has been submitted to NCBI with accession numbers from LNDA01000001-LNDA01000107 (See appendix XVIII). *N. jinanensis* 16S rRNA sequence was retrived from the whole genome using the RNAmmer platform and a phylogenetic tree was generated (Fig: 5.3) using the MEGA 6.0 software to establish the evolutionary relationship between *N. jinanensis* and other members of the genus Nocardia. Analysis of 16S rRNA sequences indicates that (1) the 16S rRNA sequence of our strain shows 100% identity to the *N. jinanensis* strain X0315 16S rRNA sequence (GenBank ID: JX857479.1) and (2) *N. jinanensis* is evolutionarily closer to *N. carnea*, *N. brevicatena* and *N. cyriacigeorgica* compared to other species like *N. farcinica*, *N. brasiliensis* and *N. absseccus* and *N. asteroids, N. uniformis and N. otitidiscavarium.*

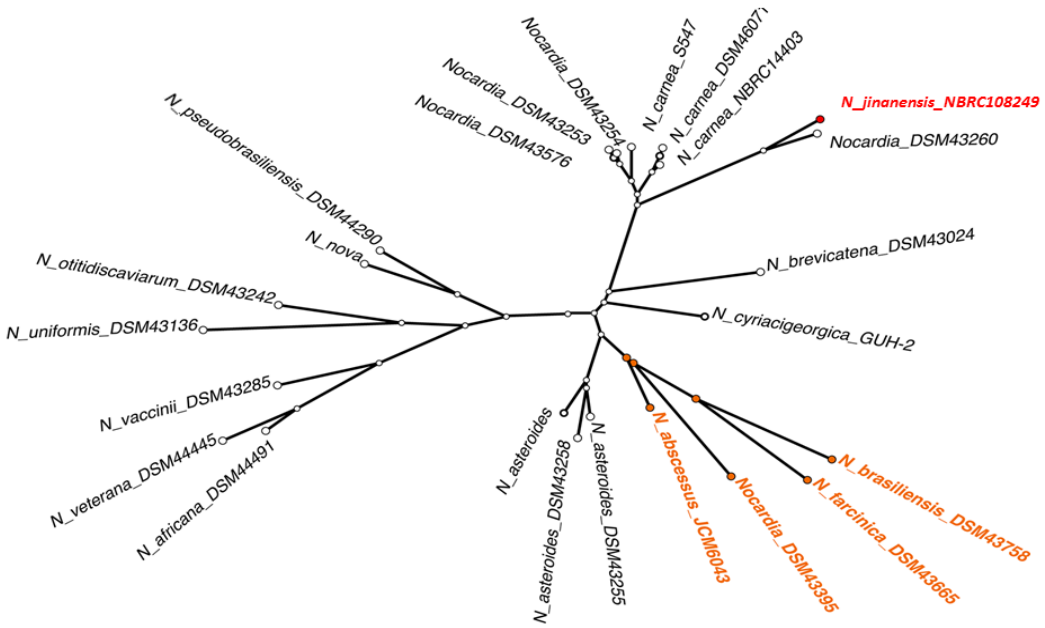


Figure 5.3: Phylogenetic tree indicating the evolutionary distance between diferent members of the genus Nocardia w.r.t to *N. jianensis*. The phylogenetic (maximum likelihood) tree was drawn using the MEGA V6.0 tool.

N. jinanensis has been isolated from soil samples and has recently been found to be a pathogenic strain that can cause cellulitis [266]. *N. jinanensis* mostly infects the skin and has not been found to cause pulmonary infections like in the case of the more common pathogenic Nocardia strains; *N. farcinica* and *N. brasiliensis*. The putative pathogenic genes have been identified in *N. farcinica* and all related information is available at the

Nocardia farcinica Genome Project Page (http://nocardia.nih.go.jp). By searching the *N. jinanensis* genome and comparing the data with the information available on *Nocardia farcinica* Genome Project Page, we found a large number of genes in *N. jinanensis* that are homologous to the pathogenic genes from *N. farcinica* (see Fig:5.3 and Appendix IX). The homologous pathogenic genes include the Mce family and YbrE family virulence factors used for mammalian cell invasion and infection, superoxide dismutases, antigenic proteins and transporters, esterases, hemolysin, PKS and NRPSs. Mce or mammalian cell entry proteins are an important group of protein found in Mycobacterium and related species [273]. While LRS2 is a transcription factor related to virulence mainly in Mycobacterium [274], invasins are a group of proteins on the surface of microorganisms that facilitate the entry of the pathogen into the host cell [275]. Superoxide dismutases are capable of breaking down superoxides into peroxides and molecular oxygen thereby neutralizing some of the reactive oxygen species generated by the macrophages of the host immune system [276]. The presence of the potential virulence genes further supports that *N. jinanensis* can be an opportunistic pathogen under suitable condition or environment. It was found that *N. jinanensis* is located in the same clade as the pathogenic *N. farcinica* and *N. cyriacigeorgica* in the phylogenetic tree constructed based on the 16S RNA sequences for 78 Nocardia species [277]. Hence, sharing of the pathogenic genes is not very surprising given the close evolutionary relationship between *N. jinanensis* and *N. farcinica*.

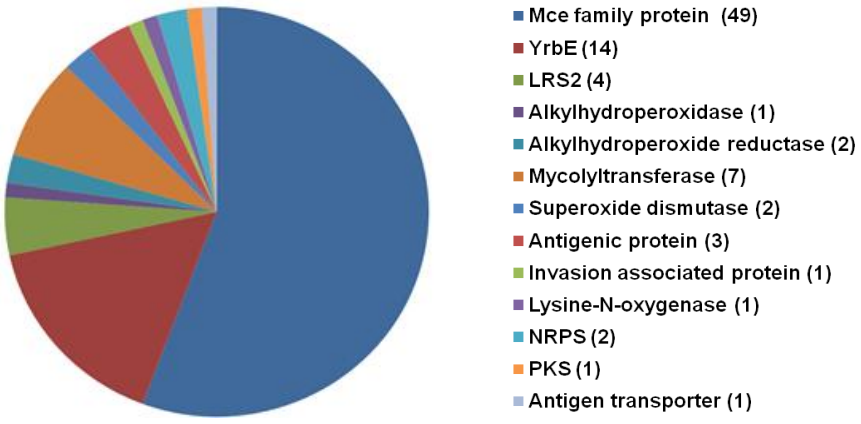


Figure 5.4: Summary of potential virulence genes found in *N. jinanensis*. A total of 88 genes were found which could putatively play a role in pathogenesis (Illustration has been made using SigmaPlot statistical analysis tool).

Bioinformatic analysis of the secondary biosynthetic gene clusters- Genome sequencing of *N. farcinica, N. brasiliensis, N. cyriacigeorgica* have revealed the remarkable biosynthetic capacity of the Nocardia species (see Appendix III, V, VI and VII). Search of biosynthetic gene clusters by using AntiSMASH 3.0 suggests the presence of more than 30 putative secondary biosynthetic gene clusters (including fatty acid and saccharide gene clusters) in the genome of *N. jinanensis* (see Appendix V and VI). The large number of secondary metabolite biosynthetic gene clusters found in the *N. jinanensis* genome once again confirms that the genus of Nocardia includes many highly prolific producers of secondary metabolites. Apart from the common ectoine biosynthetic gene cluster that can be found in many bacterial genomes and small DNA fragments that encode orphan NRPS and PKS domains, quite a number of large and seemingly intact PKS, NRPS and PKS-NRPS hybrid gene clusters were readily identified. We noted that *N. jinanensis* does not seem to contain gene clusters for the production of terpenes, lantibiotic or ribosome-derived peptide that can be found in many actinomycetes strains. *N. jinanensis* also codes for some strain-specific biosynthetic gene clusters that cannot be found in other Nocardia strains, indicating the ability of *N. jinanensis* to produce novel secondary metabolites. The *N. jinanensis* strain was previously reported to produce the dihydroisocoumarin-containing antibiotic amicoumacin B. The amicoumacin gene cluster that contains several PKS and NRPS modules has been recently identified in *Bacillus subtilis* [278], together with the similar gene cluster for xenocoumacin biosynthesis in *Xenorhabdus nematophila* [50]. To our surprise, no homologous amicoumacin gene cluster could be found in the *N. jinanensis* genome. The lack of amicoumacin gene cluster is consistent with our LC-HRMS studies that indicate the absence of amicoumacin B in the culture broth and biomass. Considering that no experimental details were given in the previous report on the isolation and characterization of amicoumacin B from *N. jinanensis* [132], we suspect that a different strain might have been used in that study.

Despite the lack of the amicoumacin B gene cluster, *N. jinanensis* contains a number of PKS, NRPS and PKS/NRPS hybrid gene clusters (see Fig: 5.4 and Appendix III, V and VII). By using AntiSMASH and other bioinformatic tools for gene annotation and based on the prediction of the substrate specificity of the A domain, we set out to determine whether the gene clusters are related to other characterized gene clusters with known products. Most of the gene clusters from the *N. jinanensis* genome are NRPS containing gene clusters. We could identify eight NRPS gene clusters, with six of them appearing to

be intact (the intact NRPS clusters include cluster number: 6, 20, 22, 26, 35 and 104). Interestingly, we could not find a corresponding homologous gene clusters for cluster 6, which is a large modular NRPS that potentially produces a high-molecular weight linear or cyclic peptide. Cluster 20 shares sequence homology with the biosynthetic gene cluster involved in the production of mirubactin, an unusual siderophore that contains a hydroxamic acid ester group [279]. The domain organization of the NRPS encoded by cluster 20 versus that of the mirubactin gene cluster are not totally identical and the some of the ancillary enzymes for mirubactin biosynthesis are missing. These observations indicate that cluster 20 is probably involved in the production of a peptide-derived siderophore that bears some structural similarity to mirubactin. Cluster 22 has a single NRPS that contains five adenylation domains (A domains) and several free-standing type II fatty acid synthesizing enzymes, including a putative desaturase. The A domains are predicted to have preference for hydrophobic amino acids such as Trp, Tyr, Leu and Phe. Hence, cluster 22 is likely to produce a small linear or cyclic lipopeptide that is decorated with an unsaturated fatty acid moiety. Cluster 26 contains a single NRPS with four A domains and a number of ancillary enzymes that include oxidoreductases, methyltransferase and glycosyltransferase. Cluster 26 is most likely to produce a peptide-derived compound that is heavily modified by the ancillary enzymes. Cluster 35 contains a single NRPS enzyme with several ancillary enzymes and the gene clusters does not share significant homology with any other known biosynthetic gene clusters. Because the cluster is found at the end of scaffold 104, it is possible that some genes from this gene cluster are missing. Notably, the gene cluster for nocardicin [280], one of the best known NRPS-derived secondary metabolites found in *N. uniformis* subsp. *tsuyamanensis*, was not found here in the *N. jinanensis* genome. The gene cluster for the lantibiotic nocathiacin found in *Nocardia sp.* ATCC 202099 was not found in the genome either.

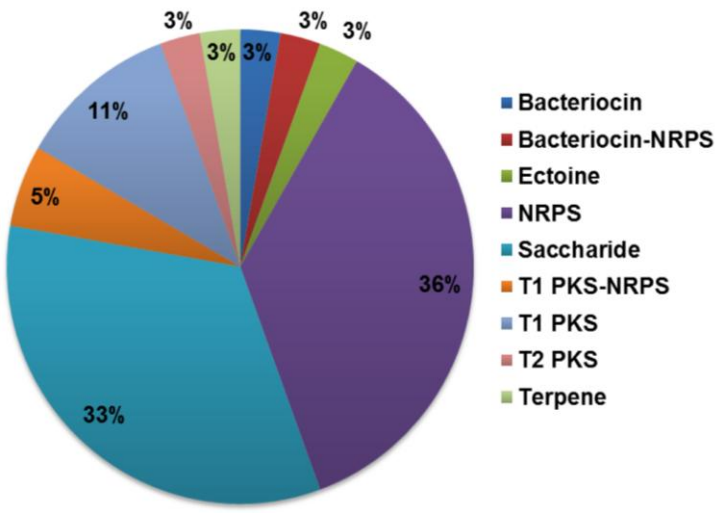


Figure 5.5: Summary of secondary metabolism gene clusters predicted by AntiSMASH. As evident from the pie-chart *N. jinanensis* genome carries a large number of NRPS based gene clusters (Illustration has been made using SigmaPlot statistical analysis tool).

Many Actinomycetes contain type I modular PKS systems for the production of structurally complex polyketides, polyenes or polyether compounds. However, no such large modular PKS (mPKS) gene clusters could be found in the *N. jinanensis* genome. The lack of mPKS systems in *N. jinanensis* suggests that this strain is unlikely to produce nargenicin A1[281], mycocerosic acid [282] and other mPKS-derived secondary metabolites isolated from other Nocardia strains. Instead, three gene clusters (clusters 12, 24 and 27) that seem to contain type I iterative PKS could be identified in the genome (see Appendix III and V). Nocardia produces a variety of lipids derived from palmitic acid, mycolic acid, 14-methyl-pentadecanoic acid and other iso- and anteiso-fatty acids [283, 284]. The cell wall envelope of Nocardia strains usually contain the α-branched hydroxylated fatty acids known as mycolic acids, which contains 36 to 66 carbon atoms and are important components of the cell wall of Nocardia species. Clusters 12 (see Appendix III and VII) shares significant homology with the gene clusters involved in the biosynthesis of mycolic acid in *M. tuberculosis*. Importantly, cluster 12 contains all the genes that encode the PKS13, AMP dependent-ligase, carboxyl transferase and several other conserved enzymes required for mycoloic acid biosynthesis. Hence, cluster 12 is most likely the gene cluster involved in the biosynthesis of mycolic acid. No homologous gene clusters could be found for cluster 24 and cluster 27, suggesting that they may be involved in the synthesis of lipids or secondary metabolites with uncharacterized structures.

N. jinanensis has two PKS-NRPS hybrid gene clusters (clusters 33 and 37) (see Appendix III and V). Cluster 37 has two modules that contain PKS catalytic domains (AT-KR-ACP-MT, KS-TE) and four NRPS modules (see Appendix III and V). This gene cluster is conserved in Nocardia species like *N. cyriacigeorgica* GUH-2 and *N. farcinica*. This cluster bears resemblance to the nocobactin cluster of *N. farcinica* [254]. Nocobactin is a siderophore which has been implicated in iron uptake and virulence. Nocobactin is also structurally related to the siderophore mycobactin produced by *Mycobacterium sp*. So not surprisingly, cluster 37 also shares some similarity with the mycobactin gene cluster. Mycobactins are capable of iron acquisition even when inside the macrophage cells and hence are essential for the survival of the pathogen [285]. Cluster 37 is not totally identical to the nocobactin and mycobactin gene clusters because of the differences in domain number and domain organization and the lack of salicylate synthase and ligase. We speculate that cluster 37 produces a structurally related siderophore that may contribute to the pathogenicity of *N. jinanensis*. Refer Appendix VII for the detailed domain organization of the major PKS and NRPSs in *N. jinanensis* genome.

Fermentation and metabolite profiling- Analysis of the genome and identification of the gene clusters suggested that *N. jinanensis* has the potential to produce a number of novel secondary metabolites. We cultivated *N. jinanensis* in liquid culture medium in an effort to identify the secondary metabolites. Analysis of the metabolites contained in the culture broth and biomass of *N. jinanensis* was performed by using various analytical techniques such as LC-HRMS and NMR spectroscopy. By using LC-HRMS, we could detect several small cyclic dipeptides that include 2-phenyl ethyl amine, cyclo (isoleucyl-prolyl), cyclo (leucyl-prolyl), and cyclo (phenylalanyl-prolyl) in the culture broth (Appendix X and XI). We also detected the highly fluorescent alkaloid flazine methyl ester (m/z 322.32) in the culture broth (see Fig: 5.8). This is surprising because flazine has only been known to be produced by plants and animals. We also detected a highly hydrophobic unknown lipopeptide molecule (m/z 962.5866) that is rich in hydrophobic residues and a few other compounds with likely novel structures. The low yield of the peptide prevented us from determining its structure at this moment. Based on the information about the biosynthetic gene clusters, the most likely gene cluster responsible for the production of this high hydrophobic lipopeptide is gene cluster 22. From the 10 L fermentation, only small quantities (<0.2 mg) of these compounds could be obtained, which prevented us from determining the structures by NMR spectroscopy. Considering

the much greater number of secondary biosynthetic gene clusters found in the genome, when compared to the metabolites identified using LC-MS, it is most likely that *N. jinanensis* is capable of producing other secondary metabolites that were not detected, likely due to the cryptic nature of genes. In particular, the genome contains a several NRPS gene clusters that are predicted to produce large linear or cyclic peptides. The absence of the peptides seems to indicate the suppression of the biosynthetic pathways.

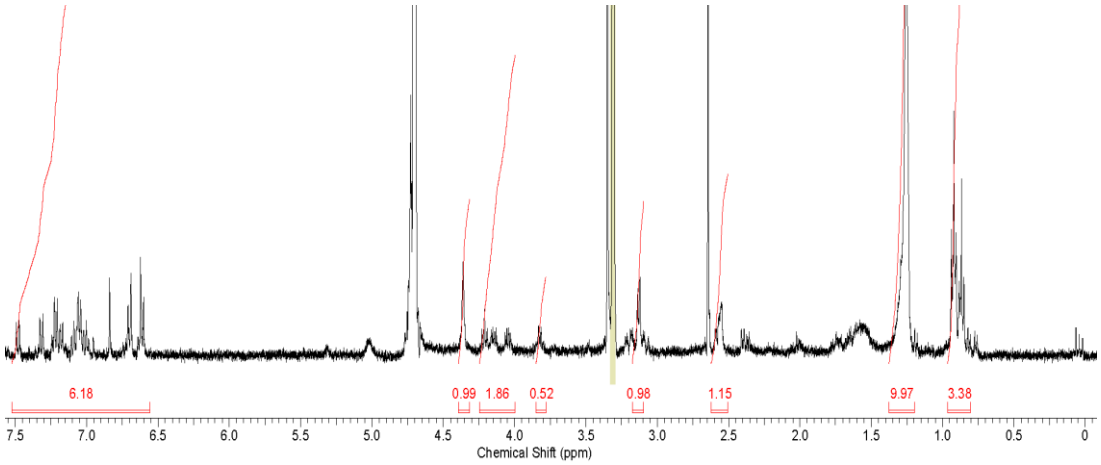


Figure 5.6: NMR spectrum of the hydrophobic lipopeptide isolated from the biomass of *N. jinanensis*.

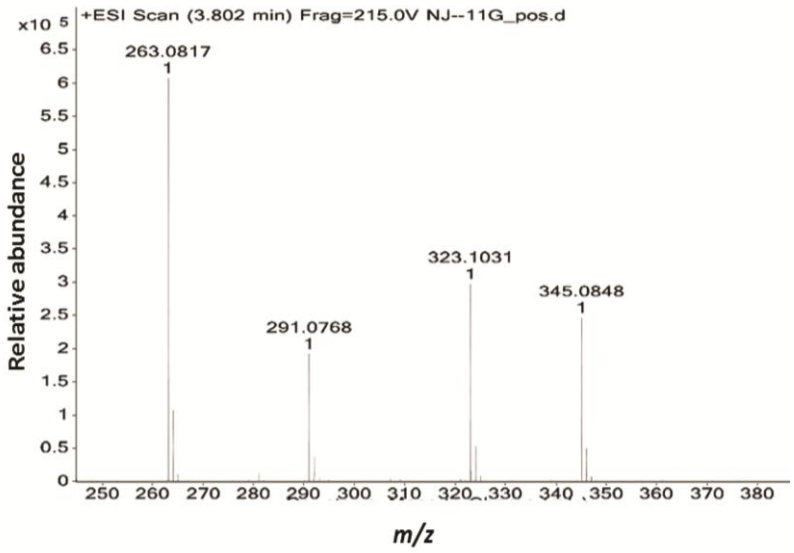


Figure 5.7: Mass spectrum of the putative flazine methyl ester isolated from the biomass of *N. jinanensis*.

***N. jinanensis* produces antimicrobial compounds in stress medium-** Since several Nocardia species are known to produce compounds that inhibit the growth of either gram-positive or gram-negative bacteria, we set out to test the antimicrobial activity of the organic extract of *N. jinanensis* culture broth. The antimicrobial activity assays also serves as a more sensitive way of detecting the secondary metabolites that are produced by fermentation. Although the standard disc-diffusion assay showed that the organic extract from the ISP2 medium exhibited little inhibitory effect on the growth of the testing strains that include *Escherichia coli*, *Bacillus subtilis*, *Staphylococcus aureus* and *Streptococcus pyogenes*, more significant inhibitory effect was observed when *N. jinanensis* was cultured in stress medium. Further testing by using the more robust 96-well microtitre plate assay method also showed similar results, with greater effect observed for the stress medium extract (see Fig: 5.6). This is not a very surprising observation since the composition of culture medium is known to affect secondary metabolite production, which is why different media are usually explored for the production of secondary metabolites [286]. These observations indicate the production of one or more antimicrobial compounds in the stress medium, presumably as the result of the stimulation or activation of the corresponding biosynthetic pathways under the stress medium conditions. We did not proceed to identify the antimicrobial compound and establish its structure because the antimicrobial activity is still considered to be moderate.

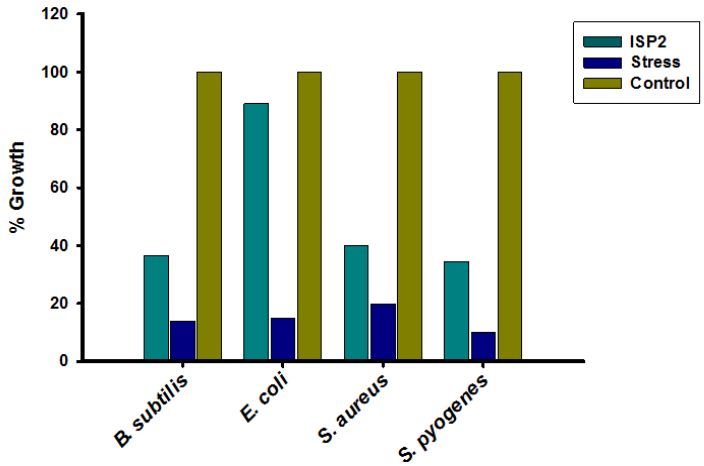


Figure 5.8: Graph indicates the trends of the 96-well plate antimicrobial activity assay done with organic extracts of *N. jinanensis* culture broth (Illustration has been made using SigmaPlot statistical analysis tool).

Discussion

Our initial aim of this study was to uncover and elucidate the biosynthetic pathway and mechanism for the dihydroisocoumarin-containing antibiotic amicoumacin B. So it was a surprise that the genome does not contain a gene cluster that is homologous to the amicoumacin B or xenocoumcin synthesizing gene cluster found in other bacterial species [132]. The lack of the amicoumacin B gene cluster is fully in accordance with the absence of amicoumacin B in the culture broth and biomass. Considering the likelihood for *N. jinanensis* to use a completely different amicoumacin B gene cluster is rather low, we do not believe *N. jinanensis* is capable of producing amicoumacin B. Despite the aforementioned claims, if in case of an unlikely event *N. jinanensis* is found to contain a amicoumacin cluster (homologous to the reported cluster in Bacillus [278]) in future, it could be attributed to two possible explanations: (1) we missed it due to gaps in sequencing and (2) the absence of amicoumacin B in the culture broth and biomass is due to the cryptic nature of the cluster.

Apart from the lack of amicoumacin B gene cluster, the genome sequences and metabolite producing have revealed the biosynthetic potential and pathogenicity of this emerging pathogenic Nocardia species. The surprising discovery of flazine methyl ester [287] (m/z 322.32) from the culture broth indicates that *N. jinanensis* contains enzymes that can construct the beta-carboline scaffold. Although flazine and other beta-carboline natural products are mostly known to be produced by plants and animals, flazine has been found in a marine Streptomyces strain. More recently, a novel microbial biosynthetic enzyme (McbB) that catalyzes the beta-carboline skeleton formation through a Pictet-Spengler reaction was discovered in the marine bacterium *Marinoactinospora thermotolerans* SCSIO 00652 [288]. The homolog of McbB can be identified in several Nocardia strains, including *N. concava, N. brevicatena* and *N. otitidiscaviarum*. We could not identify any ORF(open reading frame) from the *N. jinanensis* genome that shares significant homology with McbB. Based on the literature available on the biosynthetic mechanism involved in marinacarbolines production [288], we propose a biosynthetic mechanism for flazine as illustrated in Fig: 5.9. Although, flazine methyl ester was detected in the culture broth, we could not identify a McbB homolog in in *N. jinanensis* genome. The failure to identify a potential homolog of McbB [288] could either be due to (1) the fact that we have missed the gene as a result of gaps present in the draft genome or (2) because of a completely novel gene that is used by *N. jinanensis* to synthesize flazine.

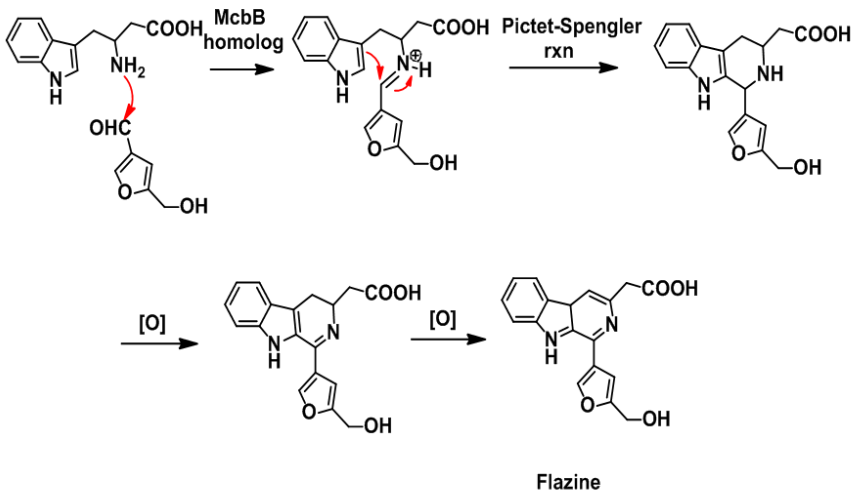


Figure 5.9: A potential microbial biosynthetic pathway for flazine. (Illustration was drawn using the ChemDraw tool).

The genome sequence of *N. jinanensis* reveals that the strain contains a significant number of secondary metabolite genes clusters that are comparable to other Nocardia strains. While a few of these gene clusters bear significant homology to known clusters in synthesizing siderophores and lipids, many of them seem to be strain-specific gene clusters and hence could be potential candidates for novel secondary metabolite discovery (see Appendix III, V, VI and VII). A detailed look at the gene organization of these gene clusters, indicate the possibility of the synthesis of some interesting structures and could thereby enrich the repertoire of natural products. From our effort to isolate and identify secondary metabolites produced by *N. jinanensis*, we could only detect a highly hydrophobic lipopeptide in the culture broth. This seems to suggest that most of the PKS, NRPS and PKS-NRPS hybrid gene clusters and pathways are suppressed, if not totally silent. To uncover the secondary metabolites in future, co-cultivation and other generic approaches can be explored to activate the biosynthetic gene clusters and pathways. As a simple proof-of-principle experiment, our antimicrobial activity assay showed that *N. jinanensis*, when cultivated using stress medium, started to produce compounds that can inhibit the growth of both gram-positive and gram-negative bacteria. *N. jinanensis* has been established recently as a human pathogenic bacterium. In consistence with this, more than 80 genes that potentially contribute to virulence and pathogenicity were

identified by us upon the genome-wide search and genome comparison. The genes uncovered based on the significant homology they share with the pathogenic genes of *N. farcinica* and *N. cyriacigeorgica,* two of the most studied pathogenic species of Nocardia [289] [236] (see Appendix IX). Future experiments and a deeper analysis of the bioinformatics data will validate the role and function of these genes and their protein products. We plan to do some knockout studies in *N. jinanensis* for the characterization of some of the gene clusters. In the past we have made attempts to express and purify some of the PKS genes from *N. jinanensis* in *E. coli* which has been unsuccessful as the proteins were majorly insoluble and remained in the pellet. Hence, we would also consider the usage of alternative heterologous hosts like Streptomyces and yeast expression systems for the purification of PKS and NRPS proteins from Nocardia in future in order to unravel the secondary metabolite backbone or frameworks these enzymes could synthesize. Hosts like *Streptomyces coelicolor*, allow the expression of an entire gene cluster and thereby enables the identification of the secondary metabolite product [94]. Hence, heterologous expression is a strategy that can be used to explore the identity of the products of the various biosynthetic gene clusters found in *N. jinanensis* genome.

Reference

1. Demain, A.L. and A. Fang, *The natural functions of secondary metabolites.* Adv Biochem Eng Biotechnol, 2000. 69: p. 1-39.
2. Pi Borui, Yu Dongliaang, Dai Fangwei,Song Xiaoming, Zhu Congyi, Li Hongye, Yu Yunsong, *A Genomics Based Discovery of Secondary Metabolite Biosynthetic Gene Clusters in Aspergillus ustus.* PLoS ONE, 2015. 10(2): p. e0116089.
3. Calvo, A.M.,Wilson, Richard A., Bok, Jin Yu., Keller, Nancy P., *Relationship between Secondary Metabolism and Fungal Development.* Microbiol. Mol. Biol. Rev., 2002. 66(3): p. 447-459.
4. Adams, T.H. and J.H. Yu, *Coordinate control of secondary metabolite production and asexual sporulation in Aspergillus nidulans.* Curr Opin Microbiol, 1998. 1(6): p. 674-7.
5. O'Brien, J. and G.D. Wright, *An ecological perspective of microbial secondary metabolism.* Curr Opin Biotechnol., 2011. 22(4): p. 552-558.
6. Marinelli, F. and G.L. Marcone, *3.26 - Microbial Secondary Metabolites*, in *Comprehensive Biotechnology (Second Edition)*, M. Moo-Young, Editor. 2011, Academic Press: Burlington. p. 285-297.
7. Marinelli, F., *Chapter 2. From microbial products to novel drugs that target a multitude of disease indications.* Methods Enzymol, 2009. 458: p. 29-58.
8. Malik, V.S., *Microbial secondary metabolism.* Trends Biochem Sci., 1980. 5(3): p. 68-72.
9. Tansey, T., *Book Review.* N Engl J Med, 2004. 351(16): p. 1697-1698.
10. Silver, L.L., *Challenges of Antibacterial Discovery.* Clin Microbiol Rev., 2011. 24(1): p. 71-109.
11. Davies, J. and D. Davies, *Origins and Evolution of Antibiotic Resistance.* Microbiol. Mol. Biol. Rev.: MMBR, 2010. 74(3): p. 417-433.
12. Khosla, C. and J.D. Keasling, *Metabolic engineering for drug discovery and development.* Nat Rev Drug Discov, 2003. 2(12): p. 1019-1025.
13. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?* J Ind Microbiol Biotechnol., 2014. 41(2): p. 175-184.

14. Fleischmann, R. D., Adams, M. D.,White, O.Clayton, R. A.Kirkness, E. F.Kerlavage, A. R.Bult, C. J.Tomb, J. F.Dougherty, B. A.Merrick, J. M.et al.,, *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. 269(5223): p. 496-512.
15. Balibar, C.J. and C.T. Walsh, *GliP, a Multimodular Nonribosomal Peptide Synthetase in Aspergillus fumigatus, Makes the Diketopiperazine Scaffold of Gliotoxin†.* Biochemistry, 2006. 45(50): p. 15029-15038.
16. Nguyen, Q.-T.,Merlo, M.E.,Medema, M.H., Jankevics, A., Breitling, R., Takano, E., et al., *Metabolomics methods for the synthetic biology of secondary metabolism.* FEBS Letters, 2012. 586(15): p. 2177-2183.
17. Chaudhary, A.K., D. Dhakal, and J.K. Sohng, *An Insight into the "-Omics" Based Engineering of Streptomycetes for Secondary Metabolite Overproduction.* Biomed Res Int , 2013. 2013: p. 15.
18. Hopwood, D.A., *Soil to genomics: the Streptomyces chromosome.* Annu Rev Genet, 2006. 40: p. 1-23.
19. Nielsen, J. and J.D. Keasling, *Synergies between synthetic biology and metabolic engineering.* Nat Biotech, 2011. 29(8): p. 693-695.
20. Li, J.W. and J.C. Vederas, *Drug discovery and natural products: end of an era or an endless frontier?* Science, 2009. 325(5937): p. 161-5.
21. Gomes, E.S., V. Schuch, and E.G. de Macedo Lemos, *Biotechnology of polyketides: New breath of life for the novel antibiotic genetic pathways discovery through metagenomics.* Braz J Microbiol., 2013. 44(4): p. 1007-1034.
22. Klopries, S., U. Sundermann, and F. Schulz, *Quantification of N-acetylcysteamine activated methylmalonate incorporation into polyketide biosynthesis.* Beilstein J Org Chem, 2013. 9: p. 664-74.
23. Malla, S., Prasad, N.N., Singh, B., Liou, K.,Kyung, S.J., *Limitations in doxorubicin production from Streptomyces peucetius.* Microbiological Research, 2010. 165(5): p. 427-435.
24. Lussier, F.-X., Colatriano, D., Wiltshire, Z., Page, J.E., Martin, V.J.J., *Engineering Microbes for Plant Polyketide Biosynthesis.* Comput Struct Biotechnol J., 2012. 3: p. e201210020.
25. Wang, C.C.C., *Recent advances in genome mining of secondary metabolites in Aspergillus terreus.* Front Microbiol , 2014. 5.

26. Mor, A., *Peptide-based antibiotics: A potential answer to raging antimicrobial resistance.* Drug Develop Res., 2000. 50(3-4): p. 440-447.
27. Fritz Lipmann1, W.G., Horst Kleinkauf2, Robert Roskoski Jr1, *Polypeptide Synthesis on Protein Templates: The Enzymatic Synthesis of Gramicidin S And Tyrocidine.* Adv Enzymol Relat Areas Mol Biol, , 22 NOV 2006. Volume 35.
28. Griffiths, G.L., Sigel, S.P., Payne, S.M., Neilands, J.B., *Vibriobactin, a siderophore from Vibrio cholerae.* Journal of Biological Chemistry, 1984. 259(1): p. 383-385.
29. Finking, R. and M.A. Marahiel, *Biosynthesis of Nonribosomal Peptides.* Annu Rev Microbiol, 2004. 58(1): p. 453-488.
30. Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism - from biochemistry to genomics.* Nat Rev Micro, 2005. 3(12): p. 937-947.
31. Hooper, I.R., *The Naturally Occurring Aminoglycoside Antibiotics*, in *Aminoglycoside Antibiotics*, H. Umezawa and I. Hooper, Editors. 1982, Springer Berlin Heidelberg. p. 1-35.
32. Crawford, J.M. and C.A. Townsend, *New insights into the formation of fungal aromatic polyketides.* Nat Rev Micro, 2010. 8(12): p. 879-889.
33. Crawford, J.M., Vagstad, A.L., Ehrlich, K.C., Townsend, C.A., *Starter unit specificity directs genome mining of polyketide synthase pathways in fungi.* Bioorg. Chem. , 2008. 36(1): p. 16-22.
34. Boettger, D. and C. Hertweck, *Molecular diversity sculpted by fungal PKS-NRPS hybrids.* Chembiochem, 2013. 14(1): p. 28-42.
35. Tang, L., Shah, S., Chung, L., Carney, J., KAtz, L., Khosla, C., Julien, B., *Cloning and heterologous expression of the epothilone gene cluster.* Science, 2000. 287(5453): p. 640-2.
36. Pelludat, C., Rakin, A., Jacobi, C.A., Schubert, S., Heesemann, J., *The yersiniabactin biosynthetic gene cluster of Yersinia enterocolitica: organization and siderophore-dependent regulation.* J Bacteriol, 1998. 180(3): p. 538-46.
37. Khaw, L.E., Böhm, G.A., Metcalfe, S., Staunton, J., Leadlay, P.F., *Mutational Biosynthesis of Novel Rapamycins by a Strain of Streptomyces hygroscopicus NRRL 5491 Disrupted in rapL, Encoding a Putative Lysine Cyclodeaminase.* J Bacteriol, 1998. 180(4): p. 809-814.
38. Tokuoka, M., Seshime, Y., Fujii, I., Kitamoto, K., Takahashi, T., Koyama, Y., *Identification of a novel polyketide synthase-nonribosomal peptide synthetase*

(PKS-NRPS) gene required for the biosynthesis of cyclopiazonic acid in Aspergillus oryzae. Fungal Genet Biol, 2008. 45(12): p. 1608-15.

39. Gressler, M., Zaehle, C., Scherlach, K., Hertweck, C., Brock, M., *Multifactorial induction of an orphan PKS-NRPS gene cluster in Aspergillus terreus.* Chem Biol, 2011. 18(2): p. 198-209.
40. Scott, R.B. and M.R. Agnes, *A Plethora of Polyketides: Structures, Biological Activities, and Enzymes,* in *Polyketides*. 2007, American Chemical Society. p. 2-14.
41. Chan, Y.A., Podevels, A.M., Kevany, B.M., Thomas, M.G., *Biosynthesis of Polyketide Synthase Extender Units.* Nat Prod Rep, 2009. 26(1): p. 90-114.
42. A. Argüelles Arias, M.C., P. Fickers, *Gram-positive antibiotic biosynthetic clusters: a review* Science against Microbial Pathogens: Communicating Current Research and Technological Advances, 2011(26-Sept-2011).
43. Campbell, C.D. and J.C. Vederas, *Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes.* Biopolymers, 2010. 93(9): p. 755-63.
44. Yu, J., Chang, P-K., Ehrlich, K.C., Cary, J.W., Bhatanagar, D., Cleveland, T.E., Payne, G.A., Linz, J.E., Woloshuk, C.P., Bennette, J.W., *Clustered Pathway Genes in Aflatoxin Biosynthesis.* Appl Environ Microbiol, 2004. 70(3): p. 1253-1262.
45. Hertweck, C., Luzhetskyy, A., Rebets, Y., Bechthold, A., *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Nat Prod Rep, 2007. 24(1): p. 162-90.
46. *Comprehensive Natural Products II: Chemistry and Biology, Volumes 1−10.* J. Am. Chem. Soc, 2010. 132(28): p. 9929-9929.
47. Yu, D., Xu, F., Zeng, J., Zhan, J., *Type III polyketide synthases in natural product biosynthesis.* IUBMB Life, 2012. 64(4): p. 285-295.
48. Funa, N., T. Awakawa, and S. Horinouchi, *Pentaketide resorcylic acid synthesis by type III polyketide synthase from Neurospora crassa.* J Biol Chem, 2007. 282(19): p. 14476-81.
49. Schmartz, P.C., Zerbe, K., Abou-Hadeed, K., Robinson, J. A., *Bis-chlorination of a hexapeptide-PCP conjugate by the halogenase involved in vancomycin biosynthesis.* Org Biomol Chem, 2014. 12(30): p. 5574-7.

50. Park, D., Ciezki, K., Van Der Hoeven, R., Singh, S., Reimer, D., Bode, H.B., Frost, S., *Genetic analysis of xenocoumacin antibiotic production in the mutualistic bacterium Xenorhabdus nematophila.* Mol Microbiol, 2009. 73(5): p. 938-949.
51. Tholl, D., *Terpene synthases and the regulation, diversity and biological roles of terpene metabolism.* Curr Opin Plant Biol, 2006. 9(3): p. 297-304.
52. Caruthers, J.M., Kang, I., Rynkiewicz, M. J., Cane, D.E., Christianson, D. W., *Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti.* J Biol Chem, 2000. 275(33): p. 25533-9.
53. Rynkiewicz, M.J., D.E. Cane, and D.W. Christianson, *Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade.* Proc Natl Acad Sci U S A, 2001. 98(24): p. 13543-8.
54. Tudzynski, B., Hedden, P., Carrera, E., Gaskin, P., *The P450-4 gene of Gibberella fujikuroi encodes ent-kaurene oxidase in the gibberellin biosynthesis pathway.* Appl Environ Microbiol, 2001. 67(8): p. 3514-22.
55. Goossens, A., Hakkinen, S. T., Laakso, I., Oksman-Caldentey, K. M., Inze, D., *Secretion of secondary metabolites by ATP-binding cassette transporters in plant cell suspension cultures.* Plant Physiol, 2003. 131(3): p. 1161-4.
56. Bibb, M.J., *Regulation of secondary metabolism in streptomycetes.* Curr Opin Microbiol., 2005. 8(2): p. 208-215.
57. Chakraburtty, R. and M. Bibb, *The ppGpp synthetase gene (relA) of Streptomyces coelicolor A3(2) plays a conditional role in antibiotic production and morphological differentiation.* J Bacteriol., 1997. 179(18): p. 5854-5861.
58. Choi, S.U., Lee, C. K., Hwang, Y. I., Kinosita, H., Nihira, T., *γ-Butyrolactone autoregulators and receptor proteins in non-Streptomyces actinomycetes producing commercially important secondary metabolites.* Archives of Microbiology, 2003. 180(4): p. 303-307.
59. Wietzorrek, A. and M. Bibb, *A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold [3].* Mol Microbiol, 1997. 25(6): p. 1181-1184.
60. Bok, J.W., Chung, D., Balajee, S. A., Marr, K. A., Andes, D., Nielsen, K. F., Frisvad, J. C., Kirby, K. A., Keller, N. P., *GliZ, a transcriptional regulator of gliotoxin biosynthesis, contributes to Aspergillus fumigatus virulence.* Infect Immun, 2006. 74(12): p. 6761-8.

61. Bok, J.W. and N.P. Keller, *LaeA, a Regulator of Secondary Metabolism in Aspergillus spp.* Eukaryotic Cell, 2004. 3(2): p. 527-535.
62. Shaaban, M.I., Bok, J.W., Lauer, C., keller, N.P., *Suppressor Mutagenesis Identifies a Velvet Complex Remediator of Aspergillus nidulans Secondary Metabolism.* Eukaryotic Cell, 2010. 9(12): p. 1816-1824.
63. Brakhage, A.A., *Regulation of fungal secondary metabolism.* Nat Rev Micro, 2013. 11(1): p. 21-32.
64. Baltz, R.H., *MbtH homology codes to identify gifted microbes for genome mining.* J Ind Microbiol Biotechnol, 2014. 41(2): p. 357-69.
65. Jensen, P.R., Mincer, T.J., Williams, P.G., Fenical, W., *Marine actinomycete diversity and natural product discovery.* Antonie Van Leeuwenhoek, 2005. 87(1): p. 43-8.
66. Molinski, T.F., *Microscale methodology for structure elucidation of natural products.* Curr Opin Biotechnol, 2010. 21(6): p. 819-26.
67. Perrior, T., *Overcoming Bottlenecks in drug discovery.* Drug Discovery World, 2010.
68. Blin, K., Medema, M.H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E.,Weber, T., *antiSMASH 2.0- "a versatile platform for genome mining of secondary metabolite producers*. Nucleic Acids Res., 2013.
69. Boddy, C.N., *Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides.* J Ind Microbiol Biotechnol, 2014. 41(2): p. 443-50.
70. Aziz, R.K., Bartels, D., Best, A. A., , DeJongh, M. Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L.et al., *The RAST Server: rapid annotations using subsystems technology.* BMC Genomics, 2008. 9: p. 75.
71. Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., Fedorova, N. D., *SMURF: genomic mapping of fungal secondary metabolite clusters.* Fungal Genet Biol : FG & B, 2010. 47(9): p. 736-741.
72. Anand, S., Prasad, M. V. R., Yadav, G., Kumar, N., Shehara, J., Ansari, Md Z., Mohanty, D., SBSPKS*: structure based sequence analysis of polyketide synthases.* Nucleic Acids Res., 2010. 38(suppl 2): p. W487-W496.
73. Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., Kohlbacher, O., *NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity.* Nucleic Acids Res., 2011. 39(Web Server issue): p. W362-W367.

74. Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D. H., Wohlleben, W., *CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters.* J Biotechnol, 2009. 140(1–2): p. 13-17.
75. Michael HT Li, P.M.U., James Zajkowski, Sylvie Garneau-Tsodikova and David H Sherman*, *Automated genome mining for natural products.* BMC Bioinfomatics, 2009.
76. Zhang, M., Hou, X-F., Qi, L-H., Yin, Y., Li, Q., Pan, H-X., Chen, X-Y., Tang, G-L., *Biosynthesis of trioxacarcin revealing a different starter unit and complex tailoring steps for type II polyketide synthase.* Chem. Sci., 2015. 6(6): p. 3440-3447.
77. Nováková, J. and M. Farkašovský, *Bioprospecting microbial metagenome for natural products.* Biologia, 2013. 68(6): p. 1079-1086.
78. Handelsman, J., *Metagenomics: Application of Genomics to Uncultured Microorganisms.* Microbiol. Mol. Biol. Rev., 2004. 68(4): p. 669-685.
79. MacNeil, I.A., et al., *Expression and isolation of antimicrobial small molecules from soil DNA libraries.* J Mol Microbiol Biotechnol, 2001. 3(2): p. 301-8.
80. Courtois, S., Cappellano, C. M., Ball, M., Francou, F-X., Normand, P., Helynck, G., Martinez, A., Kolvek, S. J., Hopke, J., Osburne, M. S., August, P. R., Nalin, R., Guérineau, M., Jeannin, P., Simonet, P., Pernodet, J-L., *Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products.* Appl Environ Microbiol, 2003. 69(1): p. 49-55.
81. Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., Clardy, J., Goodman, R.M., Handelsman, J., *Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA.* Appl Environ Microbiol, 2002. 68(9): p. 4301-4306.
82. Iqbal, H.A., Z. Feng, and S.F. Brady, *Biocatalysts and small molecule products from metagenomic studies.* Curr Opin Chem Biol, 2012. 16(1-2): p. 109-16.
83. Azumi, M., Ogawa, K-I., Fujita, T., Takeshita, M., Yoshida, R., Furumai, T., Igarashi, Y., *Bacilosarcins A and B, novel bioactive isocoumarins with unusual heterocyclic cores from the marine-derived bacterium Bacillus subtilis.* Tetrahedron, 2008. 64(27): p. 6420-6425.

84. Huang, Y-F., Li, L-H., Tian, L., Qiao, L., Hua, H-M., Pei, Y-H., *Sg17-1-4, a Novel Isocoumarin from a Marine Fungus Alternaria tenuis Sg17-1.* J Antibiot, 2006. 59(6): p. 355-357.

85. Zhang, D., Yang, X., Kang, J. S., Choi, H. D., Son, B. W., *Circumdatin I, a new ultraviolet-A protecting benzodiazepine alkaloid from a marine isolate of the fungus Exophiala.* J Antibiot (Tokyo), 2008. 61(1): p. 40-2.

86. Teasdale, M. E., Liu, J., Wallace, J., Akhlaghi, F., Rowley, D. C., *Secondary Metabolites Produced by the Marine Bacterium Halobacillus salinus That Inhibit Quorum Sensing-Controlled Phenotypes in Gram-Negative Bacteria.* Appl Environ Microbiol., 2009. 75(3): p. 567-572.

87. Lewis, K., Epstein, S., D'Onofrio, A., Ling, L.L., *Uncultured microorganisms as a source of secondary metabolites.* J Antibiot, 2010. 63(8): p. 468-476.

88. Bollmann, A., K. Lewis, and S.S. Epstein, *Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates.* Appl Environ Microbiol, 2007. 73(20): p. 6386-90.

89. Ling, L. L., Schneider, T., Peoples, Aaron J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schaberle, T.F., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. A., Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C.,Lewis, K., *A new antibiotic kills pathogens without detectable resistance.* Nature, 2015. 517(7535): p. 455-459.

90. Baltz, R.H., *Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery.* J Ind Microbiol Biotechnol, 2011. 38(11): p. 1747-60.

91. Chiang, Y-M.,Szewczyk, E., Nayak, T., Davidson, A. D., Sanchez, J. F.,Lo, H-C., Wen-Yueh, H.,Simityan, H., Kuo, E., Praseuth, A., Watanabe, K., Oakley, B. R., Wang, C. C. C., *Molecular genetic mining of the Aspergillus secondary metabolome: Discovery of the emericellamide biosynthetic pathway.* Chemistry & biology, 2008. 15(6): p. 527-532.

92. Sun, H.,Ho, C. L., Ding, F., Soehano, I., Liu, X-W., Liang, Z-X., *Synthesis of (R)-Mellein by a Partially Reducing Iterative Polyketide Synthase.* J. Am. Chem. Soc, 2012. 134(29): p. 11924-11927.

93. Chiang, Y-M., Oakley, C. E., Ahuja, M., Entwistle, R., Schultz, A., Chang, S-L., Sung, C. T., Wang, C. C. C., Oakley, B. R., *An Efficient System for Heterologous*

Expression of Secondary Metabolite Genes in Aspergillus nidulans. J. Am. Chem. Soc., 2013. 135(20): p. 7720-7731.

94. Gomez-Escribano, J.P. and M.J. Bibb, *Streptomyces coelicolor as an expression host for heterologous gene clusters.* Methods Enzymol, 2012. 517: p. 279-300.
95. Awakawa, T., Yang, X-L., Wakimoto, K., Abe, I., *Pyranonigrin E: A PKS-NRPS Hybrid Metabolite from Aspergillus niger Identified by Genome Mining.* ChemBioChem, 2013. 14(16): p. 2095-2099.
96. Gross, H., Stockwell, V.O., Henkels, M.D., Nowak-Thompson, B.,Loper, J.E., Gerwick, W.H ., *The Genomisotopic Approach: A Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters.* Chemistry & Biology, 2007. 14(1): p. 53-63.
97. Zong, Y., B. Li, and S. Tian, *Effects of carbon, nitrogen and ambient pH on patulin production and related gene expression in Penicillium expansum.* Int J Food Microbiol, 2015. 206(0): p. 102-108.
98. Onaka, H., Ozaki, T., Mori, Y., Izawa, M., Hayashi, S., Asamizu, S., *Mycolic acid-containing bacteria activate heterologous secondary metabolite expression in Streptomyces lividans.* J Antibiot, 2015.
99. Netzker, T., Fischer, J., Weber, J., Mattern, D.J., König, C. C., Valiante, V., Schroeckh, V., Brakhage A.A., *Microbial communication leading to the activation of silent fungal secondary metabolite gene clusters.* Front Microbiol, 2015. 6.
100. Khalil, Z.G., P. Kalansuriya, and R.J. Capon, *Lipopolysaccharide (LPS) stimulation of fungal secondary metabolism.* Mycology, 2014. 5(3): p. 168-178.
101. Zazopoulos, E., Huang, K., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B., Farnet, C. M., *A genomics-guided approach for discovering and expressing cryptic metabolic pathways.* Nat Biotech, 2003. 21(2): p. 187-190.
102. Gram, L., *Silent clusters – speak up!* Microbial Biotechnology, 2015. 8(1): p. 13-14.
103. Zhao, X.-Q., *Genome-Based Studies of Marine Microorganisms to Maximize the Diversity of Natural Products Discovery for Medical Treatments.* Evidence-Based Complementary and Alternative Medicine, 2011. 2011: p. 11.
104. Katz, L., Kennedy, J., Mutka, S.C., Carney, J.R., MacMillan, K.S., Murli, S., *Novel Polyketides from Genetic Engineering (... and Lessons We Have Learned*

from Making Them), in *Polyketides*. 2007, American Chemical Society. p. 200-216.

105. Haynes, S.W., Gao, X., Tang, Y., Walsh, C.T., *Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two-enzyme pathway generates heptacyclic scaffold complexity in asperlicin E.* J Am Chem Soc, 2012. 134(42): p. 17444-7.
106. Felnagle, E.A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., Thomas, M. G., *Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products.* Mol. Pharm., 2008. 5(2): p. 191-211.
107. Chen, X.-H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A. W., Gottschalk, G., Süssmuth, R. D., Borriss, R., *Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in Bacillus amyloliquefaciens FZB 42.* J Bacteriol., 2006. 188(11): p. 4024-4036.
108. Fisch, K.M., *Biosynthesis of natural products by microbial iterative hybrid PKS-NRPS.* RSC Advances, 2013. 3(40): p. 18228-18247.
109. Liu, T., Chiang, Y-M., Somoza, A. D., Oakley, B. R.,Wang, C. C. C., *Engineering of an "Unnatural" Natural Product by Swapping Polyketide Synthase Domains in Aspergillus nidulans.* Journal of the American Chemical Society, 2011. 133(34): p. 13314-13316.
110. Siddiqui, M.S., Thodey, K., Trenchard, I., Smolke, C.D., *Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools.* FEMS Yeast Research, 2012. 12(2): p. 144-170.
111. Nielsen, J. and S. Oliver, *The next wave in metabolome analysis.* Trends Biotechnol, 2005. 23(11): p. 544-6.
112. Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., Van Ommen, B., Pujos-Guillot, E., Verheij, E., Wishart, D., Wopereis, S., *Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research.* Metabolomics, 2009. 5(4): p. 435-458.
113. Garcia, D.E., Baidoo, E. E., Benke, P. I., Pingitore, F., Tang, Y. J.,Villa, S., Keasling, J. D., *Separation and mass spectrometry in microbial metabolomics.* Curr Opin Microbiol, 2008. 11(3): p. 233-9.

114. Sandmann, A., Dickschat, J., Jenke-Kodama, H., Kunze, B., Dittmann, E., Muller, R., *A Type II polyketide synthase from the gram-negative Bacterium Stigmatella aurantiaca is involved in Aurachin alkaloid biosynthesis.* Angew Chem Int Ed Engl, 2007. 46(15): p. 2712-6.

115. Krug, D., Zurek, G., Revermann, O., Vos, M., Velicer, G. J., Müller, R., *Discovering the Hidden Secondary Metabolome of Myxococcus xanthus: a Study of Intraspecific Diversity.* App Environ Microbiol, 2008. 74(10): p. 3058-3068.

116. Bunet, R., Song, L., Mendes, M. V., Corre, C., Hotel, L., Rouhier, N., Framboisier, X., Leblond, P., Challis, G. L., Aigle, Be., *Characterization and Manipulation of the Pathway-Specific Late Regulator AlpW Reveals Streptomyces ambofaciens as a New Producer of Kinamycins.* J Bacteriol, 2011. 193(5): p. 1142-1153.

117. Jansen, R., Gerth, K., Steinmetz, H., Reinecke, S., Kessler, W., Kirschning, A., Müller, R., *Elansolid A3, a Unique p-Quinone Methide Antibiotic from Chitinophaga sancti.* Chemistry – A European Journal, 2011. 17(28): p. 7739-7744.

118. Weeks, A.M. and M.C. Chang, *Constructing de novo biosynthetic pathways for chemical synthesis inside living cells.* Biochemistry, 2011. 50(24): p. 5404-18.

119. Challis, G.L., *Mining microbial genomes for new natural products and biosynthetic pathways.* Microbiol, 2008. 154(Pt 6): p. 1555-69.

120. Napolitano, E., *The Synthesis Of Isocoumarins Over The Last Decade. A Review.* Org Prep Proced Int, 1997. 29(6): p. 631-664.

121. Chapter Author: Alessandra Braca, A.B., T. Nunziatina De, and A.-u.-R. Book Author/Editor, *Chapter 7: Plant and Fungi 3,4-Dihydroisocoumarins: Structures, Biological Activity, and Taxonomic Relationships*. 2012, Amsterdam: ELSEVIER.

122. Varga, J., Kevei, E., Rinyu, E., Teren, J., Kozakiewicz, Z., *Ochratoxin production by Aspergillus species.* Appl Environ Microbiol, 1996. 62(12): p. 4461-4464.

123. Hope, J.H., Hope, B.E., *A Review of the Diagnosis and Treatment of Ochratoxin A Inhalational Exposure Associated with Human Illness and Kidney Disease including Focal Segmental Glomerulosclerosis*. 2012 [cited 2012; Available from: http://dx.doi.org/10.1155/2012/835059

124. Clark, H.A. and S.M. Snedeker, *Ochratoxin a: Its Cancer Risk and Potential for Exposure.* J Toxicol Environ Health., Part B, 2006. 9(3): p. 265-296.

125. Anke, H. and H. Zahner, *Metabolic products of microorganisms. 170. On the antibiotic activity of cladosporin.* Arch Microbiol, 1978. 116(3): p. 253-7.
126. Reimer, C.L., Agata, N., Tammam, J. G., Bamberg, M., Dickerson, W. M., Kamphaus, G. D., Rook, S. L., Milhollen, M., Fram, R., Kalluri, R., Kufe, D., Kharbanda, S., *Antineoplastic Effects of Chemotherapeutic Agents Are Potentiated by NM-3, an Inhibitor of Angiogenesis.* Cancer Research, 2002. 62(3): p. 789-795.
127. Itoh, J., Shomura, T., Omoto, S., Miyado, S., Yuda, Y., Shibata, U., Inouye, S., *Isolation, Physicochemical Properties and Biological Activities of Amicoumacins Produced by Bacillus pumilus.* Agric Biol Chem, 1982. 46(5): p. 1255-1259.
128. Lama, A., Pané-Farré, J., Chon, T., Wiersma, A. M., Sit, C.S., Vederas, J. C., Hecker, M., Nakano, M. M., *Response of Methicillin-Resistant Staphylococcus aureus to Amicoumacin A.* PLoS ONE, 2012. 7(3): p. e34037.
129. Boya, C.A., Herrera, L., Guzman, H. M., Gutierrez, M., *Antiplasmodial activity of bacilosarcin A isolated from the octocoral-associated bacterium Bacillus sp. collected in Panama.* J. Pharm. Bioall. Sci, 2012. 4(1): p. 66-69.
130. Frisvad, J.C.F., J.M.; Houbraken, J.A.M.P.; Kuijpers, A.F.A.; Samson, R.A, *New ochratoxin A producing species of Aspergillus section Circumdati.* . Stud Mycol., 2004. 50(1): p. 23-44.
131. Creppy, E.E., *Human Ochratoxicosis.* Toxin Reviews, 1999. 18(3-4): p. 277-293.
132. Sun, W., Zhang, Y. Q., Huang, Y., Zhang, Y. Q., Yang, Z. Y., Liu, Z. H., *Nocardia jinanensis sp. nov., an amicoumacin B-producing actinomycete.* Int J Syst Evol Microbiol, 2009. 59(Pt 2): p. 417-20.
133. Solecka, J., Zajko, J., Postek, M., Rajnisz, A., *Biologically active secondary metabolites from Actinomycetes.* Cent Eur J Biol., 2012. 7(3): p. 373-390.
134. Ben-Ami, R. and D. Kontoyiannis, *Pathogenesis of Invasive Pulmonary Aspergillosis*, in *Aspergillosis: From Diagnosis to Prevention*, A. Comarú Pasqualotto, Editor. 2010, Springer Netherlands. p. 345-379.
135. Perrone, G., Perrone, G., Susca, A., Cozzi, G., Ehrlich, K., Varga, J., Frisvad, J. C., Meijer, M., Noonim, P., Mahakarnchanakul, W., Samson, R. A., *Biodiversity of Aspergillus species in some important agricultural products.* Stud Mycol., 2007. 59: p. 53-66.

136. Bhetariya, P.J., Madan, T., Basir, S. F., Varma, A., Usha, S. P., *Allergens/Antigens, Toxins and Polyketides of Important Aspergillus Species.* Indian J Clin Biochem., 2011. 26(2): p. 104-119.

137. Singh, S., Sczakas, G., Soccol, C.R., Pandey, A., *Production of Enzymes by Solid-state Fermentation*, in *Current Developments in Solid-state Fermentation*, A. Pandey, C. Soccol, and C. Larroche, Editors. 2008, Springer New York. p. 183-204.

138. Roukas, T., *Biotechnology of Citric acid production.* Food Biotechnology, 2006. 2nd Edition.

139. Hong, K.J., C.H. Lee, and S.W. Kim, *Aspergillus oryzae GB-107 fermentation improves nutritional quality of food soybeans and feed soybean meals.* J Med Food, 2004. 7(4): p. 430-5.

140. Fox, E.M. and B.J. Howlett, *Secondary metabolism: regulation and role in fungal biology.* Curr Opin Microbiol, 2008. 11(6): p. 481-7.

141. Cacho, R.A., Y. Tang, and Y.-H. Chooi, *Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi.* Front Microbiol, 2014. 5: p. 774.

142. Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R, .Kim, H. S., Arroyo, J., Berriman, M.et al., *Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus.* Nature, 2005. 438(7071): p. 1151-6.

143. Galagan, J.E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., Clutterbuck, J., Kapitonov, V., Jurka, J., et al., *Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae.* Nature, 2005. 438(7071): p. 1105-1115.

144. Machida, M., et al., *Genome sequencing and analysis of Aspergillus oryzae.* Nature, 2005. 438(7071): p. 1157-61.

145. Pel, H.J., de Winde, J. H., Archer, D. B., Dyer, P. S., Hofmann, G., Schaap, P. J., Turner, G., de Vries, R. P., Albang, R., Albermann, K., et al., *Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88.* Nat Biotechnol, 2007. 25(2): p. 221-31.

146. Morello, L.G., Sartori, D., de Oliveira M., A. L., Vieira, M. L., Taniwaki, M. H., Fungaro, M. H., *Detection and quantification of Aspergillus westerdijkiae in coffee beans based on selective amplification of beta-tubulin gene by using real-time PCR.* Int J Food Microbiol, 2007. 119(3): p. 270-6.

147. Bacha, N., Dao, H. P., Atoui, A., Mathieu, F., O'Callaghan, J., Puel, O., Liboz, T., Dobson, A.D. W., Lebrihi, A., *Cloning and characterization of novel methylsalicylic acid synthase gene involved in the biosynthesis of isoasperlactone and asperlactone in Aspergillus westerdijkiae.* Fungal Genet Biol, 2009. 46(10): p. 742-749.

148. Yaegashi, J., B.R. Oakley, and C.C. Wang, *Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of heterologous expression systems in Aspergillus nidulans.* J Ind Microbiol Biotechnol, 2014. 41(2): p. 433-42.

149. Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., *Automatic annotation of eukaryotic genes, pseudogenes and promoters.* Genome Biol, 2006. 7(Suppl 1): p. S10-S10.

150. Boratyn, G.M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S.D., Merezhuk, Y, Raytselis, Y, Sayers, E.W., Tao, T., Ye, J., Zaretskaya, I., *BLAST: a more efficient report with usability improvements.* Nucleic Acids Res., 2013. 41(W1): p. W29-W33.

151. Galagan, J.E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., et al., *Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae.* Nature, 2005. 438(7071): p. 1105-15.

152. Payne, G.A., Nierman, W. C., Wortman, J. R., Pritchard, B. L., Brown, D., Dean, R. A., Bhatnagar, D., Cleveland, T. E., Machida, M., Yu, J., *Whole genome comparison of Aspergillus flavus and A. oryzae.* Med Mycol, 2006. 44(s1): p. 9-11.

153. Wortman, J.R., Fedorova, N., Crabtree, J., Joardar, V., Maiti, R., Haas, B. J., Amedeo, P., Lee, E., Angiuoli, S. V., Jiang, B., Anderson, M. J., Denning, D. W., White, O. R., Nierman, W. C., *Whole genome comparison of the A. fumigatus family.* Med Mycol, 2006. 44(s1): p. 3-7.

154. Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., Prohaska, S. J., *Proteinortho: detection of (co-)orthologs in large-scale analysis.* BMC Bioinformatics, 2011. 12: p. 124.

155. Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., *TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.* Bioinformatics, 2002. 18(3): p. 502-4.

156. Rawlings, N.D., Morton, F. R., Kok, C. Y., Kong, J., Barrett, A. J., *MEROPS: the peptidase database.* Nucleic Acids Res, 2008. 36(Database issue): p. D320-5.
157. Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., Breitling, R., *antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.* Nucleic Acids Res.
158. Abad, A., Fernandez-Molina, J. V., Bikandi, J., Ramirez, A., Margareto, J., Sendino, J., Hernando, F. L., Ponton, J., Garaizar, J., Rementeria, A., *What makes Aspergillus fumigatus a successful pathogen? Genes and molecules involved in invasive aspergillosis.* Rev Iberoam Micol, 2010. 27(4): p. 155-82.
159. Palacios-Cabrera, H., Taniwaki, M. H., Hashimoto, J. M., Menezes, H. C.de, *Growth of Aspergillus ochraceus, A. carbonarius and A. niger on culture media at different water activities and temperatures.* Braz. J. Microbiol., 2005. 36: p. 24-28.
160. Bacha, N., Atoui, A., Mathieu, F., Liboz, T., Lebrihi, A., *Aspergillus westerdijkiae polyketide synthase gene "aoks1" is involved in the biosynthesis of ochratoxin A.* Fungal Genet Biol, 2009. 46(1): p. 77-84.
161. R. Latha, H.K. Manonmani , E.R. Rati, *Multiplex PCR Assay for the Detection of Aflatoxigenic and Non-Aflatoxigenic Aspergilli.* Res. J Microbiol., 2008. 3(3): p. 136.
162. Rawlings, N.D. and F.R. Morton, *The MEROPS batch BLAST: a tool to detect peptidases and their non-peptidase homologues in a genome.* Biochimie, 2008. 90(2): p. 243-59.
163. Zhao, Z., Liu, H., Wang, C., Xu, J. R., *Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi.* BMC Genomics, 2013. 14: p. 274.
164. Cantarel, B.L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.* Nucleic Acids Res, 2009. 37(Database issue): p. D233-8.
165. Ookura, R., Kito, K., Ooi, T., Namikoshi, M.,Kusumi, T., *Structure revision of circumdatins A and B, benzodiazepine alkaloids produced by marine fungus Aspergillus ostianus, by X-ray crystallography.* J Org Chem, 2008. 73(11): p. 4245-7.

166. Tsukamoto, S., Kato, H., Samizo, M., Nojiri, Y., Onuki, H., Hirota, H., Ohta, T., *Notoamides F-K, prenylated indole alkaloids isolated from a marine-derived Aspergillus sp.* J Nat Prod, 2008. 71(12): p. 2064-7.
167. William Fenical W, J.P., and Cheng XC, *US Pat., 6 066 635*. 2000.
168. Babudri, F., Fiandanese, V., Marchese, G., Punzi, A., *Novel Synthetic Approach to (S)-Coriolic Acid.* Tetrahedron, 2000. 56(2): p. 327-331.
169. Behnke, C.P.a.J., *Aldrich Library of 13C and 1H FT-NMR Spectra.* 1993.
170. Assante, G., Camarda, L., Locci, R., Merlini, L., Nasini, G., Papadopoulos, E., *Isolation and structure of red pigments from Aspergillus flavus and related species, grown on a differential medium.* J Agric Food Chem, 1981. 29(4): p. 785-787.
171. Liu, J.F., Kaselj, M., Isome, Y., Chapnick, J., Zhang, B., Bi, G., Yohannes, D., Yu, L.,Baldino, C. M., *Microwave-assisted concise total syntheses of quinazolinobenzodiazepine alkaloids.* J Org Chem, 2005. 70(25): p. 10488-93.
172. Tino G, V.P.a.E.V., *US Patent. 3342795A.* 1967.
173. Wang, Y., Fang, X., An, F., Wang, G., Zhang, X., *Improvement of antibiotic activity of Xenorhabdus bovienii by medium optimization using response surface methodology.* Microb Cell Fact., 2011. 10(1): p. 98.
174. El Khoury, A. and A. Atoui, *Ochratoxin A: General Overview and Actual Molecular Status.* Toxins, 2010. 2(4): p. 461-493.
175. O'Brien, E., A. Prietz, and D.R. Dietrich, *Investigation of the teratogenic potential of ochratoxin A and B using the FETAX system.* Birth Defects Research Part B: Developmental and Reproductive Toxicology, 2005. 74(5): p. 417-423.
176. Gilani, S.H., Bancroft, J., and Reily, M., *Teratogenisity of ochratoxin A in chick embryos.* Toxicol Appl Pharmacol, 1978. 46(6).
177. Munoz, K., Campos, V., Blaszkewicz, M., Vega, M., Alvarez, A., Neira, J., Degen, G. H., *Exposure of neonates to ochratoxin A: first biomonitoring results in human milk (colostrum) from Chile.* Mycotoxin Res, 2010. 26(2): p. 59-67.
178. Brown, M.H., G.M. Szczech, and B.P. Purmalis, *Teratogenic and toxic effects of ochratoxin A in rats.* Toxicol Appl Pharm, 1976. 37(2): p. 331-338.
179. Zhang, X., Boesch-Saadatmandi, C., Lou, Y., Wolffram, S., Huebbe, P., Rimbach, Gerald *Ochratoxin A induces apoptosis in neuronal cells.* Genes & Nutrition, 2009. 4(1): p. 41-48.

180. Huff, W. and P. Hamilton, *Mycotoxins-Their Biosynthesis in Fungi-Ochratoxins-Metabolites of Combined Pathways.* J Food Prot., 1979. 42(10): p. 815-820.
181. O'Callaghan, J., M.X. Caddick, and A.D. Dobson, *A polyketide synthase gene required for ochratoxin A biosynthesis in Aspergillus ochraceus.* Microbiol., 2003. 149(Pt 12): p. 3485-91.
182. Gallo, A., Bruno, K. S., Solfrizzo, M., Perrone, G., Mule, G., Visconti, A.,Baker, S. E., *New insight into the ochratoxin A biosynthetic pathway through deletion of a nonribosomal peptide synthetase gene in Aspergillus carbonarius.* Appl Environ Microbiol, 2012. 78(23): p. 8208-18.
183. Yu, J.-H. and N. Keller, *Regulation of Secondary Metabolism in Filamentous Fungi.* Annu Rev Phytopathol., 2005. 43(1): p. 437-458.
184. Bayram, O., Krappmann, S., Ni, M., Bok, J. W., Helmstaedt, K., Valerius, O., Braus-Stromeyer, S., Kwon, N. J., Keller, N. P., Yu, J. H., Braus, G. H., *VelB/VeA/LaeA complex coordinates light signal with fungal development and secondary metabolism.* Science, 2008. 320(5882): p. 1504-6.
185. Crespo-Sempere, A., MarÃn, S., Sanchis, V., Ramos, A. J., *VeA and LaeA transcriptional factors regulate ochratoxin A biosynthesis in Aspergillus carbonarius.* Int J Food Microbiol., 2013. 166(3): p. 479-486.
186. Ang, L.-H., Chattopadhyay, S., Wei, N., Oyama, T., Okada, K., Batschauer, A., Deng, X-W., *Molecular Interaction between COP1 and HY5 Defines a Regulatory Switch for Light Control of Arabidopsis Development.* Mol.Cell, 1998. 1(2): p. 213-222.
187. Yoon, M.-K., Shin, J., Choi, G., Choi, B-S., *Intrinsically unstructured N-terminal domain of bZIP transcription factor HY5.* Proteins: Struct., Funct., Bioinf., 2006. 65(4): p. 856-866.
188. Anne-Sophie Huart, T.R.H., *Evolution of Conformational Disorder & Diversity of the P53 Interactome.* Biodiscovery, 2013. 5(8).
189. Ma, B. and R. Nussinov, *Regulating highly dynamic unstructured proteins and their coding mRNAs.* Genome Biol., 2009. 10(1): p. 204-204.
190. Botton, A., Ferrigo, D., Scopel, C., Causin, R., Bonghi, C., Ramina, A., *A cDNA-AFLP approach to study ochratoxin A production in Aspergillus carbonarius.* Int J Food Microbiol, 2008. 127(1-2): p. 105-15.

191. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., *MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.* Mol Biol Evol, 2013. 30(12): p. 2725-2729.

192. Walsh, C.T., Gehring, A. M., Weinreb, P.H., Quadri, L. E. N., Flugel, R. S. *Post-translational modification of polyketide and nonribosomal peptide synthases.* Curr. Opin Chem Biol., 1997. 1(3): p. 309-315.

193. Dao, H.P., F. Mathieu, and A. Lebrihi, *Two primer pairs to detect OTA producers by PCR method.* Int J Food Microbiol, 2005. 104(1): p. 61-67.

194. O'Callaghan, J., P.C. Stapleton, and A.D.W. Dobson, *Ochratoxin A biosynthetic genes in Aspergillus ochraceus are differentially regulated by pH and nutritional stimuli.* Fungal Genet Biol., 2006. 43(4): p. 213-221.

195. Yamada, O., Takara, R., Hamada, R., Hayashi, R., Tsukahara, M., Mikami, S., *Molecular biological researches of Kuro-Koji molds, their classification and safety.* J Biosci Bioeng., 2011. 112(3): p. 233-237.

196. Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., Fedorova, N. D., *SMURF: Genomic mapping of fungal secondary metabolite clusters.* Fungal Genet Biol., 2010. 47(9): p. 736-741.

197. Cerqueira, G.C., Arnaud, M. B., Inglis, D. O., Skrzypek, M. S., Binkley, G., Simison, M., Miyasato, S. R., Binkley, J., Orvis, J., Shah, P., Wymore, F., Sherlock, G., Wortman, J. R., *The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations.* Nucleic Acid Res., 2013. p. D705-D710.

198. Gallo, A., Bruno, K. S., Solfrizzo, M., Perrone, G., Mule, G., Visconti, A., Baker, S. E., *New Insight into the Ochratoxin A Biosynthetic Pathway through Deletion of a Nonribosomal Peptide Synthetase Gene in Aspergillus carbonarius.* Appl Environ Microbiol, 2012. 78(23): p. 8208-8218.

199. Roze, L.V., Miller, M. J., Rarick, M., Mahanti, N., Linz, J. E., *A Novel cAMP-response Element, CRE1, Modulates Expression of nor-1 in Aspergillus parasiticus*. J Biol Chem., 2004. p. 27428-27439.

200. Harris, J.P. and P.G. Mantle, *Biosynthesis of ochratoxins by Aspergillus ochraceus.* Phytochem, 2001. 58(5): p. 709-716.

201. Ali, H., Ries, M. I., Lankhorst, P. P., van der Hoeven, R. A. M., Schouten, O. L., Noga, MarekHankemeier, T., van Peij, N. N. M. E., Bovenberg, R. A. L., Vreeken, Rob J., Driessen, Arnold J. M., *A Non-Canonical NRPS Is Involved in*

the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in Penicillium chrysogenum. PLoS ONE, 2014. 9(6): p. e98212.

202. Keller, U. and F. Schauwecker, *Nonribosomal biosynthesis of microbial chromopeptides*, Progress in Nucleic Acid Research and Molecular Biology. 2001, Academic Press. p. 233-289.
203. Gallo, A., M. Ferrara, and G. Perrone, *Phylogenetic Study of Polyketide Synthases and Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Mycotoxins.* Toxins, 2013. 5(4): p. 717-742.
204. Warnhoff, E.W., *Peptide Alkaloids*, in *Fortschritte der Chemie Organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products*, W. Herz, H. Grisebach, and A.I. Scott, Editors. 1970, Springer Vienna. p. 162-203.
205. Walsh, C.T., Haynes, S. W., Ames, B. D., Gao, X., Tang, Y., *Short Pathways to Complexity Generation: Fungal Peptidyl Alkaloid Multicyclic Scaffolds from Anthranilate Building Blocks.* Am Chem Biol., 2013. 8(7): p. 1366-1382.
206. Coyle, C.M. and D.G. Panaccione, *An Ergot Alkaloid Biosynthesis Gene and Clustered Hypothetical Genes from Aspergillus fumigatus.* Appl EnvironMicrobiol, 2005. 71(6): p. 3112-3118.
207. Schiff, P.L., *Ergot and Its Alkaloids.* Am J Pharm Educ., 2006. 70(5): p. 98.
208. Xu, W., D.J. Gavia, and Y. Tang, *Biosynthesis of fungal indole alkaloids.* Nat Prod Rep., 2014. 31(10): p. 1474-1487.
209. Wallwey, C. and S.M. Li, *Ergot alkaloids: structure diversity, biosynthetic gene clusters and functional proof of biosynthetic genes.* Nat Prod Rep, 2011. 28(3): p. 496-510.
210. Gerhards, N., Neubauer, L., Tudzynski, P., Li, S-M., *Biosynthetic Pathways of Ergot Alkaloids.* Toxins, 2014. 6(12): p. 3281-3295.
211. Panaccione, D.G. and C.M. Coyle, *Abundant Respirable Ergot Alkaloids from the Common Airborne Fungus Aspergillus fumigatus.* Appl Environ Microbiol., 2005. 71(6): p. 3106-3111.
212. Tsunematsu, Y., shikawa, N., Wakana, D., Goda, Y., Noguchi, H., Moriya, H., Hotta, K., Watanabe, K., *Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk.* Nat Chem Biol, 2013. 9(12): p. 818-825.
213. Wang, F., Fang, Y., Zhu, T., Zhang, M., Lin, A., Gu, Q., Zhu, W., *Seven new prenylated indole diketopiperazine alkaloids from holothurian-derived fungus Aspergillus fumigatus.* Tetrahedron, 2008. 64(34): p. 7986-7991.

214. Knaus, H.-G., McManus, O. B., Lee, S. H., Schmalhofer, W. A., Garcia-Calvo, M., Helms, L. M. H., Sanchez, M., Giangiacomo, K., Reuben, J. P. *Tremorgenic Indole Alkaloids Potently Inhibit Smooth Muscle High-Conductance Calcium-Activated Potassium Channels.* Biochem., 1994. 33(19): p. 5819-5828.

215. Ding, Y., de Wet, J. R., Cavalcoli, J., Li, S., Greshock, T. J., Miller, K. A., Finefield, J. M., Sunderhaus, J. D., McAfoos, T. J., Tsukamoto, S., Williams, R. M., Sherman, D. H., *Genome-based characterization of two prenylation steps in the assembly of the stephacidin and notoamide anticancer agents in a marine-derived Aspergillus sp.* J Am Chem Soc, 2010. 132(36): p. 12733-40.

216. Haynes, S.W., Gao, X., Tang, Y., Walsh, C. T., *Complexity generation in fungal peptidyl alkaloid biosynthesis: a two-enzyme pathway to the hexacyclic MDR export pump inhibitor ardeemin.* Am Chem Biol, 2013. 8(4): p. 741-748.

217. Rahbaek, L. and J. Breinholt, *Circumdatins D, E, and F: further fungal benzodiazepine analogues from aspergillus ochraceus.* J Nat Prod, 1999. 62(6): p. 904-5.

218. Dai, J., Carte, B. K., Sidebottom, P. J., Sek Yew, A. L., Ng, S., Huang, Y., Butler, M. S., *Circumdatin G, a new alkaloid from the fungus Aspergillus ochraceus.* J Nat Prod, 2001. 64(1): p. 125-6.

219. Lopez-Gresa, M.P., Gonzalez, M. C., Primo, J., Moya, P., Romero, V., Estornell, E., *Circumdatin H, a new inhibitor of mitochondrial NADH oxidase, from Aspergillus ochraceus.* J Antibiot (Tokyo), 2005. 58(6): p. 416-9.

220. Rahbaek, L., Breinholt, J., Frisvad, J. C., Christophersen, C., *Circumdatin A, B, and C: Three New Benzodiazepine Alkaloids Isolated from a Culture of the Fungus Aspergillus ochraceus.* J Org Chem, 1999. 64(5): p. 1689-1692.

221. Peng, J., Zhang, X. Y., Tu, Z. C., Xu, X. Y., Qi, S. H., *Alkaloids from the deep-sea-derived fungus Aspergillus westerdijkiae DFFSCS013.* J Nat Prod, 2013. 76(5): p. 983-7.

222. Qian-Cutrone, J., Huang, S., Shu, Y-Z., Vyas, D., Fairchild, C., Menendez, A., Krampitz, K., Dalterio, R., Klohr, S. E., Gao, Q., *Stephacidin A and B: Two Structurally Novel, Selective Inhibitors of the Testosterone-Dependent Prostate LNCaP Cells.* Journal of the Am Chem Soc., 2002. 124(49): p. 14556-14557.

223. Artman, G.D., A.W. Grubbs, and R.M. Williams, *Concise, Asymmetric, Stereocontrolled Total Synthesis of Stephacidins A, B and Notoamide B.* Journal of the Am Chem Soc. 2007. 129(19): p. 6336-6342.

224. Tseng, M.-C., H.-Y. Yang, and Y.-H. Chu, *Total synthesis of asperlicin C, circumdatin F, demethylbenzomalvin A, demethoxycircumdatin H, sclerotigenin, and other fused quinazolinones.* Org Biomol Chem., 2010. 8(2): p. 419-427.

225. Yin, W.B., Baccile, J. A., Bok, J. W., Chen, Y., Keller, N. P., Schroeder, F. C., *A nonribosomal peptide synthetase-derived iron(III) complex from the pathogenic fungus Aspergillus fumigatus.* J Am Chem Soc, 2013. 135(6): p. 2064-7.

226. Ding, Y., de Wet, J. R., Cavalcoli, J., Li, S., Greshock, T. J., Miller, K. A., Finefield, J. M., Sunderhaus, J. D., McAfoos, T. J., Tsukamoto, S., Williams, R. M., Sherman, D. H., *Genome-Based Characterization of Two Prenylation Steps in the Assembly of the Stephacidin and Notoamide Anticancer Agents in a Marine-Derived Aspergillus sp.* J Am ChemSoc., 2010. 132(36): p. 12733-12740.

227. Herzon, S.B. and A.G. Myers, *Enantioselective Synthesis of Stephacidin B.* Journal of the Am Chem Soc., 2005. 127(15): p. 5342-5344.

228. Ames, B.D. and C.T. Walsh, *Anthranilate-Activating Modules from Fungal Nonribosomal Peptide Assembly Lines.* Biochem, 2010. 49(15): p. 3351-3365.

229. Bentley, R. and R. Meganathan, *Geosmin and methylisoborneol biosynthesis in streptomycetes: Evidence for an isoprenoid pathway and its absence in non-differentiating isolates.* FEBS Letters, 1981. 125(2): p. 220-222.

230. Manivasagan, P., Venkatesan, J., Sivakumar, K., Kim, S-K., *Marine actinobacterial metabolites: Current status and future perspectives.* Microbiol Res., 2013. 168(6): p. 311-332.

231. de Lima Procópio, R.E., da Silva, I.R., Martins, M. K., de Azevedo, J. L., de Araújo, J.M., *Antibiotics produced by Streptomyces.* The Braz J Infect Dis., 2012. 16(5): p. 466-471.

232. Miyadoh, S., *Research on Antibiotic Screening in Japan over the Last Decade: A Producing Microorganism Approach.* Actinomycetologica, 1993. 7(2): p. 100-106.

233. McMurray, D.N., *Mycobacteria and Nocardia.* Medical Microbiology, 1993. 4th edition, chapter 33.

234. Pottumarthy, S., Limaye, A.P., Prentice, J. L., Houze, Y. B., Swanzy, S. R., Cookson, B. T., *Nocardia veterana, a New Emerging Pathogen.* J Clin Microbiol, 2003. 41(4): p. 1705-1709.

235. Unzaga, M.J., A. Gaafar, and R. Cisterna, *Pulmonary Infection Due to Nocardia nova.* Archivos de Bronconeumología (English Version), 2003. 39(10): p. 478-478.

236. Schlaberg, R., R.C. Huard, and P. Della-Latta, *Nocardia cyriacigeorgica, an Emerging Pathogen in the United States.* J Clin Microbiol., 2008. 46(1): p. 265-273.

237. Smego, R.A., Jr. and H.A. Gallis, *The clinical spectrum of Nocardia brasiliensis infection in the United States.* Rev Infect Dis, 1984. 6(2): p. 164-80.

238. Betrán, A., Villuendas, M. C., Rezusta, A., Moles, B., Rubio, M. C., Revillo, M. J., Boiron, P., Bello, S., Rodríguez-Nava, V., *Cavitary pneumonia caused by Nocardia otitidiscaviarum.* Braz J Microbiol, 2010. 41: p. 329-332.

239. Beaman, B.L., *Mechanisms for the virulence of Nocardia.* Molecular Mechanisms of Bacterial Virulence, 1994.

240. Luo, Q., S. Hiessl, and A. Steinbuchel, *Functional diversity of Nocardia in metabolism.* Environ Microbiol, 2014. 16(1): p. 29-48.

241. Van Hamme, J.D., A. Singh, and O.P. Ward, *Recent Advances in Petroleum Microbiology.* Microbiol. Mol. Biol. Rev., 2003. 67(4): p. 503-549.

242. Chang, J.H., Rhee, S. K., Chang, Y. K., Chang, H. N., *Desulfurization of diesel oils by a newly isolated dibenzothiophene-degrading Nocardia sp. strain CYKS2.* Biotechnol Prog, 1998. 14(6): p. 851-5.

243. Cairns, W.L., Cairns, William L., Cooper, D. G., Zajic, J. E., Wood, J. M., Kosaric, N., *Characterization of Nocardia amarae as a Potent Biological Coalescing Agent of Water-Oil Emulsions.* App Environ Microbiol., 1982. 43(2): p. 362-366.

244. Tamura, T., Matsuzawa, T., Oji, S., Ichikawa, N., Hosoyama, A., Katsumata, H., Yamazoe, A., Hamada, M., Suzuki, K., Gonoi, T., Fujita, N., *A genome sequence-based approach to taxonomy of the genus Nocardia.* Antonie Van Leeuwenhoek, 2012. 102(3): p. 481-91.

245. Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T., Hattori, M., *The complete genomic sequence of Nocardia farcinica IFM 10152.* Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(41): p. 14925-14930.

246. Vera-Cabrera, L., Ortiz-Lopez, R.,Elizondo-Gonzalez, R., Perez-Maya, A. A., Ocampo-Candiani, J., *Complete Genome Sequence of Nocardia brasiliensis HUJEG-1.* J Bacteriol, 2012. 194(10): p. 2761-2762.

247. Zoropogui, A., Pujic, P., Normand, P., Barbe, V., Beaman, B., Beaman, L., Boiron, P., Colinon, C., Deredjian, A., Graindorge, A., Mangenot, S., Nazaret, S., Neto, M., Petit, S., Roche, D., Vallenet, D., Rodríguez-Nava, V., Richard, Y., Cournoyer, B., Blaha, D., *Genome Sequence of the Human- and Animal-Pathogenic Strain Nocardia cyriacigeorgica GUH-2.* J Bacteriol, 2012. 194(8): p. 2098-2099.

248. Luo, Q., Hiessl, S., Poehlein, A., Daniel, R., Steinbuchel, A., *Insights into the microbial degradation of rubber and gutta-percha by analysis of the complete genome of Nocardia nova SH22a.* Appl Environ Microbiol, 2014. 80(13): p. 3895-907.

249. Tanaka, Y., Grafe, U., Yazawa, K., Mikami, Y., Ritzau, M., *Nocardicyclins A and B: new anthracycline antibiotics produced by Nocardia pseudobrasiliensis.* J Antibiot (Tokyo), 1997. 50(10): p. 822-7.

250. Brown, M.S., Akopiants, K., Resceck, D. M., McArthur, H. A., McCormick, E., Reynolds, K. A., *Biosynthetic origins of the natural product, thiolactomycin: a unique and selective inhibitor of type II dissociated fatty acid synthases.* J Am Chem Soc, 2003. 125(34): p. 10166-7.

251. Igarashi, M., Nakamura, H., Naganawa, H., Takeuchi, T., *Tubelactomicin A, a novel 16-membered lactone antibiotic, from Nocardia sp. II. Structure elucidation.* J Antibiot (Tokyo), 2000. 53(10): p. 1102-7.

252. Vastrad, B.M., Neelagund, S. E., Iiger, S. R., Godbole, A. M., Kulkarni, V., *Improved Rifamycin B Production by Nocardia mediterranei MTCC 14 under Solid-State Fermentation through Process Optimization.* Biochem Res Int., 2014. 2014: p. 13.

253. Sohng, J., Yamaguchi, T., Seong, C-N., Baik, K-S., Park, S-C., Lee, H-J., Jang, S-Y., Simkhada, J. R., Yoo, J. C., *Production, isolation and biological activity of nargenicin from Nocardia sp. CS682.* Arch. Pharmacal Res, 2008. 31(10): p. 1339-1345.

254. Hoshino, Y., Chiba, K., Ishino, K., Fukai, T., Igarashi, Y., Yazawa, K., Mikami, Y., Ishikawa, J., *Identification of Nocobactin NA Biosynthetic Gene Clusters in Nocardia farcinica.* J Bacteriol., 2011. 193(2): p. 441-448.

255. Shigemori, H., Komaki, H., Yazawa, K., Mikami, Y.Nemoto, A., Tanaka, Y., Sasaki, T., In, Y., Ishida, T., Kobayashi, J., *Brasilicardin A. A Novel Tricyclic Metabolite with Potent Immunosuppressive Activity from Actinomycete Nocardiabrasiliensis.* J Org Chem, 1998. 63(20): p. 6900-6904.

256. Dipesh Dhakal, A.K.C., Anaya Raj Pokhrel, Amit Kumar Jha, and B.S. Sumangala Darsandhari, and Jae Kyung Sohng*, *Underpinning the secondary metabolites from Nocardia spp.* J. Biomolecule Reconstruction. Vol. 10(No. 1): p. 9-17.

257. Mikami, Y., Komaki, H., Imai, T., Yazawa, K., Nemoto, A., Tanaka, Y., Graefe, U., *A new antifungal macrolide component, brasilinolide B, produced by Nocardia brasiliensis.* J Antibiot (Tokyo), 2000. 53(1): p. 70-4.

258. Kobayashi, J., Tsuda, M., Nemoto, A., Tanaka, Y., Yazawa, K., Mikami, Y., *Brasilidine A, a new cytotoxic isonitrile-containing indole alkaloid from the actinomycete Nocardia brasiliensis.* J Nat Prod, 1997. 60(7): p. 719-20.

259. Mukai, A., Fukai, T., Hoshino, Y., Yazawa, K., Harada, K., Mikami, Y., *Nocardithiocin, a novel thiopeptide antibiotic, produced by pathogenic Nocardia pseudobrasiliensis IFM 0757.* J Antibiot, 2009. 62(11): p. 613-619.

260. Fennell, K.A., U. Möllmann, and M.J. Miller, *Syntheses and Biological Activity of Amamistatin B and Analogs.* J. Org. Chem, 2008. 73(3): p. 1018-1024.

261. Schneider, K., Rose, I., Vikineswary, S., Jones, A. L., Goodfellow, M., Nicholson, G., Beil, W., Sussmuth, R. D., Fiedler, H. P., *Nocardichelins A and B, siderophores from Nocardia strain acta 3026.* J Nat Prod, 2007. 70(6): p. 932-5.

262. Nishida, M., Mine, Y., Nonoyama, S., Kojo, H., *Nocardicin A, a new monocyclic beta-lactam antibiotic III. In vitro evaluation.* J Antibiot (Tokyo), 1977. 30(11): p. 917-25.

263. Murakami, Y., Kato, S., Nakajima, M., Matsuoka, M., Kawai, H., Shin-Ya, K., Seto, H., *Formobactin, a novel free radical scavenging and neuronal cell protecting substance from Nocardia sp.* J Antibiot (Tokyo), 1996. 49(9): p. 839-45.

264. Mukai, A., Fukai, T., Hoshino, Y., Yazawa, K., Harada, K., Mikami, Y., *Nocardithiocin, a novel thiopeptide antibiotic, produced by pathogenic Nocardia pseudobrasiliensis IFM 0757.* J Antibiot (Tokyo), 2009. 62(11): p. 613-9.

265. Wyche, T.P., Hou, Y., Vazquez-Rivera, E., Braun, D., Bugni, T. S., *Peptidolipins B-F, Antibacterial Lipopeptides from an Ascidian-derived Nocardia sp.* J Nat Prod., 2012. 75(4): p. 735-740.

266. Tan, C.K., Lai, C. C., Lin, S. H., Liao, C. H., Chou, C. H., Hsu, H. L., Huang, Y. T., Hsueh, P. R., *Clinical and microbiological characteristics of Nocardiosis including those caused by emerging Nocardia species in Taiwan, 1998–2008.* Clin Microbiol Inf., 2010. 16(7): p. 966-972.

267. Wang, Y., Fang, X., An, F., Wang, G., Zhang, X., *Improvement of antibiotic activity of Xenorhabdus bovienii by medium optimization using response surface methodology.* Microb Cell Fact, 2011. 10: p. 98.

268. Tatusov, R.L., Galperin, M. Y., Natale, D. A., Koonin, E. V., *The COG database: a tool for genome-scale analysis of protein functions and evolution.* Nucleic Acids Res., 2000. 28(1): p. 33-36.

269. Tatusov, R.L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al., *The COG database: an updated version includes eukaryotes.* BMC Bioinformatics, 2003. 4: p. 41.

270. Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H-H., Rognes., T., Ussery, D. W., *RNAmmer: consistent and rapid annotation of ribosomal RNA genes.* Nucleic Acids Res., 2007. 35(9): p. 3100-3108.

271. Adamczeski, M., A.R. Reed, and P. Crews, *New and Known Diketopiperazines from the Caribbean Sponge, Calyx cf. podatypa.* Journal of Natural Products, 1995. 58(2): p. 201-208.

272. Yang, B., et al., *Proline-Containing Dipeptides from a Marine Sponge of a Callyspongia Species.* Helvetica Chimica Acta, 2009. 92(6): p. 1112-1117.

273. Zhang, F. and J.P. Xie, *Mammalian cell entry gene family of Mycobacterium tuberculosis.* Mol Cell Biochem, 2011. 352(1-2): p. 1-10.

274. Bartek, I.L., Woolhiser, L. K., Baughn, A. D., Basaraba, R. J., Jacobs, W. R., Jr., Lenaerts, A. J., Voskuil, M. I., *Mycobacterium tuberculosis Lsr2 is a global transcriptional regulator required for adaptation to changing oxygen levels and virulence.* MBio, 2014. 5(3): p. e01106-14.

275. Niemann, H.H., W.D. Schubert, and D.W. Heinz, *Adhesins and invasins of pathogenic bacteria: a structural view.* Microbes Infect, 2004. 6(1): p. 101-12.

276. Cox, G.M., Harrison, T. S., McDade, H. C., Taborda, C. P., Heinrich, G., Casadevall, A., Perfect, J. R., *Superoxide dismutase influences the virulence of*

Cryptococcus neoformans by affecting growth within macrophages. Infect Immun, 2003. 71(1): p. 173-80.

277. Komaki, H., Ichikawa, N., Hosoyama, A., Takahashi-Nakaguchi, A., Matsuzawa, T., Suzuki, K., Fujita, N., Gonoi, T., *Genome based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in seven strains of five representative Nocardia species.* BMC Genomics, 2014. 15: p. 323.
278. Li, Y., Li, Z., Yamanaka, K., Xu, Y., Zhang, W., Vlamakis, H ., Kolter, R., Moore, B. S., Qian, P-Y., *Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis.* Sci. Rep., 2015. 5.
279. Giessen, T.W., Franke, K. B., Knappe, T. A., Kraas, F. I., Bosello, M., Xie, X., Linne, U., Marahiel, M. A., *Isolation, structure elucidation, and biosynthesis of an unusual hydroxamic acid ester-containing siderophore from Actinosynnema mirum.* J Nat Prod, 2012. 75(5): p. 905-14.
280. Gunsior, M., Breazeale, S. D., Lind, A. J., Ravel, J., Janc, J. W., Townsend, C. A., *The biosynthetic gene cluster for a monocyclic beta-lactam antibiotic, nocardicin A.* Chem Biol, 2004. 11(7): p. 927-38.
281. Cane, D.E. and G. Luo, *Biosynthesis of polyketide antibiotics. Incorporation of a pentaketide chain elongation intermediate into nargenicin.* J Am Chem Soc., 1995. 117(24): p. 6633-6634.
282. Mathur, M. and P.E. Kolattukudy, *Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.* J Biol Chem., 1992. 267(27): p. 19388-95.
283. Yano, I., Y. Furukawa, and M. Kusunose, *Phospholipids of Nocardia coeliaca.* J Bacteriol, 1969. 98(1): p. 124-130.
284. Yano, I., Tomiyasu, I., Kaneda, K., Kato, Y., Sumi, Y., Kurano, S., Sugimoto, N., Sawai, H., *Isolation of mycolic acid-containing glycolipids in Nocardia rubra and their granuloma forming activity in mice.* J Pharmacobiodyn, 1987. 10(3): p. 113-23.
285. De Voss, J.J., Rutter, K., Schroeder, B. G., Su, H., Zhu, Y., Barry, C. E., 3rd., *The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages.* Proc Natl Acad Sci U S A, 2000. 97(3): p. 1252-7.

286. VanderMolen, K.M., Raja, H. A., El-Elimat, T., Oberlies, N. H., *Evaluation of culture media for the production of secondary metabolites in a natural products screening program.* AMB Express, 2013. 3: p. 71-71.
287. Nakatsuka, S.-i., Feng, B-N.,Goto, T.,Kihara, K., *Structures of flazin and YS, highly fluorescent compounds isolated from japanese soy sauce.* Tetrahedron Lett., 1986. 27(29): p. 3399-3402.
288. Chen, Q., Ji, C., Song, Y., Huang, H., Ma, J., Tian, X., Ju, J., *Discovery of McbB, an Enzyme Catalyzing the β-Carboline Skeleton Construction in the Marinacarboline Biosynthetic Pathway.* Angewandte Chemie, 2013. 125(38): p. 10164-10168.
289. Kumar, V.A., Augustine, D., Panikar, D., Nandakumar, A., Dinesh, K. R., Karim, S., Philip, R., *Nocardia farcinica brain abscess: epidemiology, pathophysiology, and literature review.* Surg Infect (Larchmt), 2014. 15(5): p. 640-6.

Appendix

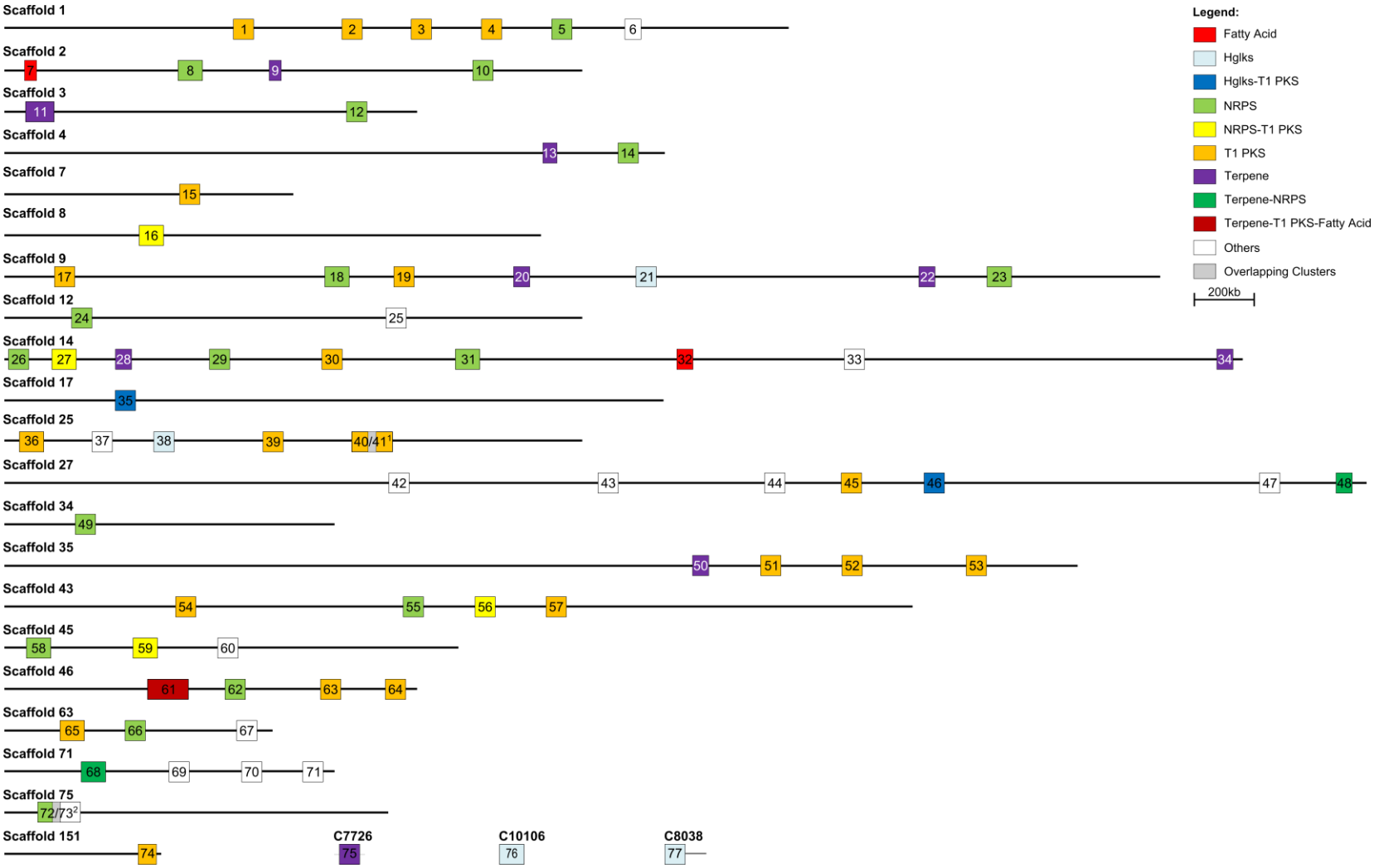
Appendix I: Summary of antiSMASH predictions for *A. westerdijkiae* genome.

Scaffold	Cluster	Type	From	To	Most similar known cluster
Scaffold 1	Cluster 1	T1 PKS	663381	709920	
(length=1959216)	Cluster 2	T1 PKS	944398	991485	
	Cluster 3	T1 PKS	1066421	1114207	
	Cluster 4	T1 PKS	1163123	1208823	
	Cluster 5	NRPS	1218455	1261943	
	Cluster 6	Other	1342841	1385490	
Scaffold 2	Cluster 7	Fatty Acid	28474	54233	
(length=1444596)	Cluster 8	NRPS	397890	457272	
	Cluster 9	Terpene	614896	636361	
	Cluster 10	NRPS	1125546	1168962	
Scaffold 3	Cluster 11	NRPS	39269	103575	
(length=993355)	Cluster 12	NRPS	801452	849926	
Scaffold 4	Cluster 13	Terpene	1434686	1457238	Putative Botrydial biosynthetic gene cluster (40% of genes show homology)
(length=1646026)					
	Cluster 14	NRPS	1551655	1597405	
Scaffold 7	Cluster 15	T1 PKS	407679	455891	
(length=736687)					
Scaffold 8	Cluster 16	NRPS-T1 PKS	396102	446569	Interesting Cluster (PKS-NRPS hybrid)
(length=1383725)					
Scaffold 9	Cluster 17	T1 PKS	118489	166395	
(length=2817998)	Cluster 18	NRPS	852372	899756	
	Cluster 19	T1 PKS	923376	971803	Interesting Cluster (Highly reducing PKS with stanadalone ER domain)
	Cluster 20	Terpene	1279521	1300711	
	Cluster 21	Hglks	1587552	1638366	
	Cluster 22	Terpene	2230725	2251900	
	Cluster 23	NRPS	2306669	2364194	
Scaffold 12	Cluster 24	NRPS	146483	190206	
(length=1469756)	Cluster 25	Other	960372	1003473	
Scaffold 14	Cluster 26	NRPS	1	48143	
(length=3084807)	Cluster 27	NRPS-T1 PKS	70927	123131	Interesting Cluster (PKS-NRPS hybrid with a number of oxidoreductase)
	Cluster 28	Terpene	145161	166886	
	Cluster 29	NRPS	399784	445613	Putative Ochratoxin biosynthetic gene cluster
	Cluster 30	T1 PKS	667073	714983	
	Cluster 31	NRPS	994755	1047136	
	Cluster 32	Fatty Acid	1609701	1633758	
	Cluster 33	Other	1966819	2010943	
	Cluster 34	Terpene	3060930	3084807	
Scaffold 17	Cluster 35	Hglks-T1 PKS	245099	293526	
(length=1603627)					
Scaffold 25	Cluster 36	T1 PKS	27941	80567	
(length=1430477)	Cluster 37	Other	194652	237736	
	Cluster 38	Hglks	311112	351738	
	Cluster 39	T1 PKS	597233	642544	
	Cluster 40	T1 PKS	936142	981904	
	Cluster 41	T1 PKS	975038	1027565	

Contd…

Scaffold 27	Cluster 42	Other	899246	943981	
(length=3359036)	Cluster 43	Other	1377024	1423592	
	Cluster 44	Other	1847570	1890965	
	Cluster 45	T1 PKS	2046709	2095139	
	Cluster 46	Hglks-T1 PKS	2175686	2223246	
	Cluster 47	Other	3148859	3192138	
	Cluster 48	Terpene-NRPS	3226288	3265674	Putative Notoamide biosynthetic gene cluster (31% of genes show homology)
Scaffold 34	Cluster 49	NRPS	149241	196718	
(length=822893)					
Scaffold 35	Cluster 50	Terpene	1764874	1786256	
(length=2604827)	Cluster 51	T1 PKS	1857580	1905602	Interesting Cluster (2 highly reducing iterative PKS)
	Cluster 52	T1 PKS	1932543	1982801	
	Cluster 53	T1 PKS	2332751	2378107	Putative Asperlactone biosynthetic gene cluster
Scaffold 43	Cluster 54	T1 PKS	391302	438207	
(length=2212477)	Cluster 55	NRPS	952492	996915	Putative Hexadehydro-astechrome (HAS) biosynthetic gene cluster (50% of genes show homology)
	Cluster 56	NRPS-T1 PKS	1131446	1185875	Interesting Cluster (PKS-NRPS hybrid)
	Cluster 57	T1 PKS	1262987	1310804	
Scaffold 45	Cluster 58	NRPS	18363	71494	
(length=1089895)	Cluster 59	NRPS-T1 PKS	281377	334968	Interesting Cluster (PKS-NRPS hybrid)
	Cluster 60	Other	539506	584502	
Scaffold 46	Cluster 61	Terpene-T1 PKS-Fa	357224	451187	
(length=1025085)	Cluster 62	NRPS	494571	543302	
	Cluster 63	T1 PKS	807483	856392	
	Cluster 64	T1 PKS	942254	989595	
Scaffold 63	Cluster 65	T1 PKS	109714	168657	
(length=654656)	Cluster 66	NRPS	269717	317894	
	Cluster 67	Other	587478	630064	
Scaffold 71	Cluster 68	Terpene-NRPS	187501	239853	
(length=796945)	Cluster 69	Other	412340	454079	Putative Terrequinone biosynthetic gene cluster (40% similarity)
	Cluster 70	Other	597896	641361	
	Cluster 71	Other	648657	691355	
Scaffold 75	Cluster 72	NRPS	144080	198876	
(length=929178)	Cluster 73	Other	194266	239270	
Scaffold 151	Cluster 74	T1 PKS	350017	379913	
(length=379913)					
C7726	Cluster 75	Terpene	1	4215	
(length = 4215)					
C8038	Cluster 76	Hglks	1	5360	
(length =5360)					
C10106	Cluster 77	Hglks	1	24683	
(length=27120)					

Arrangement of secondary metabolism clusters in scaffolds (*A. westerdijkiae*).



Appendix II: Summary of NRPS, PKS and PKS-NRPS hybrid genes from *A. westerdijkiae* genome.

Polyketide synthases

Location	Domain architecture	Closest BLAST hit
Sc1_Contig1_orf182	KS-AT-DH-ACP-ACP-TE	55% identity to *Trichophyton tonsurans* CBS 112818 - Protein EGD94040.1
Sc1_Contig1_orf285	KS-AT-DH-CMET-KR-ACP	59% identity to *Aspergillus niger* ATCC 1015 - Protein EHA20663.1
Sc7_Contig7_orf113	KS-AT-DH-CMET-ER-PP	43% identity to [*Aspergillus flavus* NRRL3357] Protein XP_002382732.1 40% identity to *Pyrenophora tritici-repentis* Pt-1C-BFP] Protein XP_001939440.1
Sc9_Contig9_orf207	KS-AT-DH-CMET-ER	87% identity to *Aspergillus niger* CBS 513.88 Protein XP_001396752.1
Sc14_Contig14_orf112	KS-AT-DH-CMET-ER-KR-ACP	77% identity to *polyketide synthase [Aspergillus steynii*] - Protein AHZ61902.1
Sc14_Contig14_orf182	KS-AT-DH-CMET- KR	34% identity to *[Glarea lozoyensis* ATCC 20868] Protein XP_008078705.1
Sc16_Contig16_orf58	KS-AT-DH-ER-KR-ACP	65% identity to *Aspergillus ustus*] Protein KIA75490.1
Sc23_Contig23_orf12	KS-AT-DH	48% identity to PKS/NRPS enzyme EqiS-like protein of [*Aspergillus fumigatus* A1163]- protein EDP53404.1
Sc23_Contig23_orf188	KS-AT-DH-ACP-TE	45% identity to PKS16 of [*Cladonia grayi*] - protein ADM79459.1
Sc23_Contig23_orf301	KS-AT-DH-KR-ACP	63% identity to [*Aspergillus niger* CBS 513.88] - protein XP_001399961.2
Sc23_Contig23_orf288	KS-AT-DH-CMET	99% identity to [*Aspergillus ochraceus*] –protein AAS98204.1
Sc25_Contig25_orf618	KS-AT-DH-CMET-ER-KR-ACP	45% identify to *Aspergillus kawachii* IFO 4308] - protein GAA84851.1
Sc25_Contig25_orf581	KS-AT-DH-ER-KR-ACP	61% identity to *Stachybotrys chartarum* IBT 7711] - protein KEY68601.01
Sc30_Contig30_orf640	KS-AT-DH-KR-ACP	99% identity to the MSAS-type polyketide synthase [*Aspergillus ochraceus]* -protein AAS98200.1
Sc30_Contig30_orf517	DH-CMET-ER-KR-SDR	67% identity to *Aspergillus kawachii* IFO 4308] - protein GAA89183.1
Sc30_Contig30_orf520	KS-AT-DH-CMET-ER-KR-ACP	66% identity to [*Aspergillus kawachii* IFO 4308] - protein GAA90346.1
Sc30_Contig30_orf539	A-PP-KS-AT-KR-ACP-SDR	52% identity to [*Eutypa lata* UCREL1] -protein XP_007798173.1
Sc37_Contig37_orf112	KS-AT-ACP-ACP-TE	48% identify to *Byssochlamys spectabilis* No. 5] - protein GAD98443.1
Sc37_Contig37_orf347	KS-AT-DH-CMET-ER-KR	73% identity to *Aspergillus clavatus* NRRL 1] - protein XP_001274957.1
Sc37_Contig37_orf319	KS-AT-DH-CMET-ER	63% identity to *Metarhizium acridum* CQMa 102] –protein XP_0078112233.1
Sc40_Contig40_orf205	KS-AT-DH-ER-KR	36% identity to putative Phthiocerol synthesis polyketide synthase type I ACPsC [*Glarea lozoyensis* 74030] –protein EHK99373.1
Sc40_Contig40_orf239	KS-AT-DH-ER-KR	56% identity to *Talaromyces stipitatus* ATCC 10500] -protein XP_002480264.1
Sc40_Contig40_102	KS-AT-DH-ACP-CMET	66% identity to *Aspergillus terreus* NIH2624] - protein XP_001217248.1
Sc40_Contig40_106	FAS-AT	57% identity to *Aspergillus terreus* NIH2624] - protein XP_001217253.1
Sc40_Contig40_107	KR-KS-ACP	61% identity to 3-oxoacyl-[acyl-carrier-protein] synthase [*Aspergillus oryzae* RIB40] -protein XP_001827151.2
Sc53_Contig53_orf42	KR-KS-AT-DH-TE	52% identity to *Botryotinia fuckeliana T4]* -protein CCD51172.1
Sc122_Contig122_orf90	KS-AT-DH-CMET-ER-KR	51% identity to Endocarpon pusillum Z07020] - protein XP_007804269.1

Non-ribosomal peptide synthetases

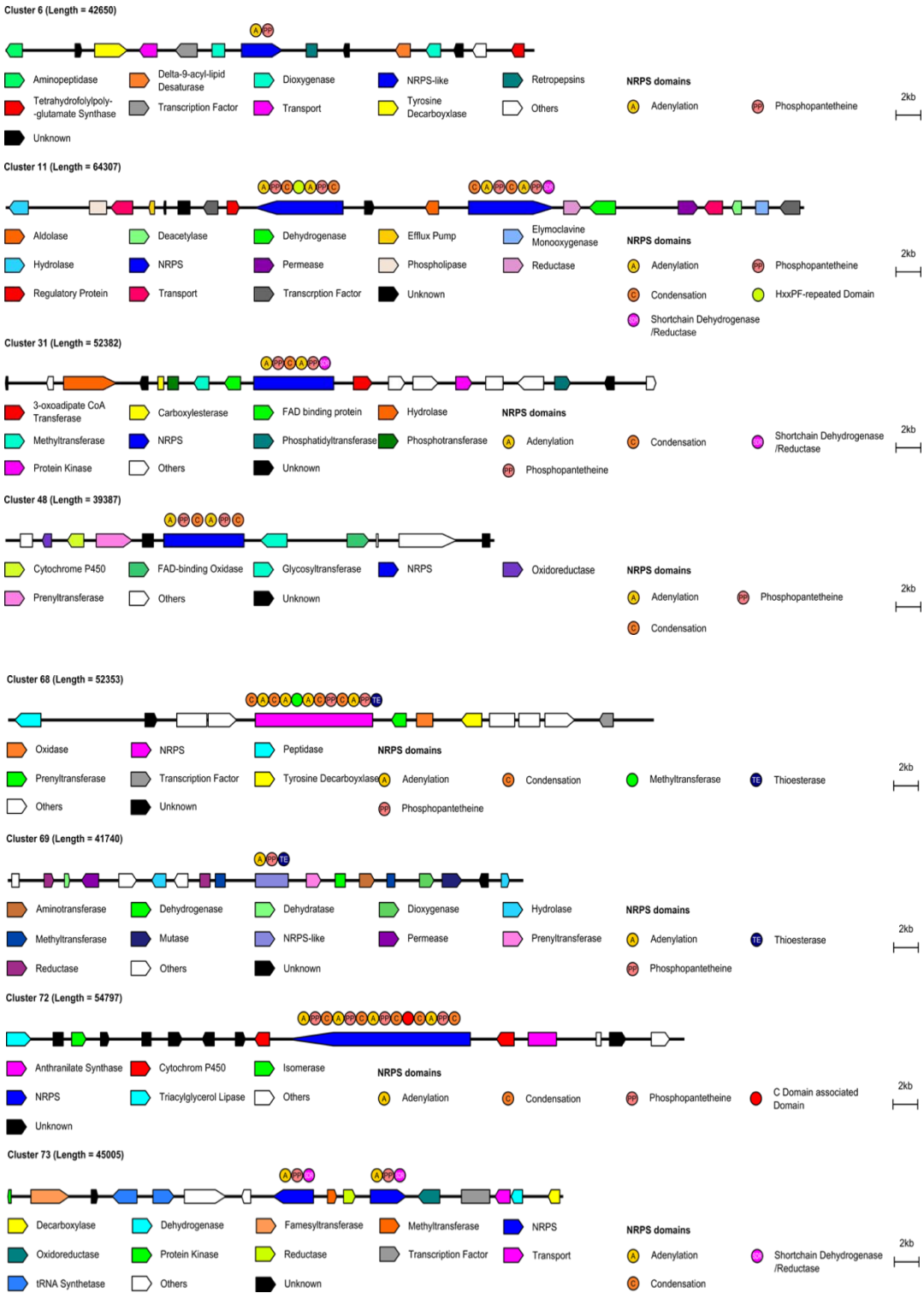
Location	Domain architechture	Closest BLAST hit
Sc1_contig1_orf327	A-P-C-A-P-C	56% identity to AMP-dependent synthetase ligase [*Penicillium solitum*] –protein KJJ27368.1
Sc2_contig2_orf117	A-P-C-A-P-C-A-P-C-P-C-P-C	73% identity to Nonribosomal siderophore peptide synthase SidC [*Aspergillus flavus* NRRL3357] –protein XP_002376812.1
Sc2_contig2_orf308	A-P-C-A-P-C	36% identity to nonribosomal peptide synthase, putative [*Neosartorya fischeri* NRRL 181] –protein-KFY94737.1
Sc2_contig2_orf310	A-P-C	35% identity to nonribosomal peptide synthase, putative [*Aspergillus ustus*] –protein-KIA75377.1
Sc3_contig3_orf14	A-P-C-HxxP-A-P-C	70% identity to Predicted protein [*Aspergillus terreus* NIH2624] –protein XP_001217686.1
Sc3_contig3_orf18	C-A-P-C-A-P-TE	74% identity to *predicted protein [Aspergillus terreus* NIH2624] –protein XP_001217690.1
Sc3_contig3_orf218	A-P-C-C-A-P-C	48% identity to Nonribosomal peptide synthase, putative [*Neosartorya fischeri* NRRL 181] – protein XP_001258350.1
Sc4_contig4_orf418	A-P-C	66% identity to Nonribosomal siderophore peptide synthase Sid2 [*Aspergillus oryzae* RIB40]] – protein XP_001826763.2
Sc9_contig9_orf191	C-A-P-C-C-A-P-C	50% identity to AMP-dependent synthetase/ligase [*Penicillium camemberti*] –protein CRL20855.1
Sc9_contig9_orf584	A-P-C-C-A-C-A-P-C-A-P-C-C-P-C-P	58% identity to nonribosomal peptide synthetase 1 [*Neosartorya udagawae*] –protein GAO88772.1
Sc12_contig12_orf31	A-P-C-A-P-C-C	50% identity to nonribosomal peptide synthase [*Arthroderma otae* CBS 113480] –protein XP_002843835.1
Sc14_contig14_orf0002	A-P-C-A-P-C-C-A-P-C-A-P-C	47% identity to hypothetical protein HIM_11088 [*Hirsutella minnesotensis* 3608] –protein KJZ69526.1
Sc14_contig14_orf112	A-P-C-A-P	72% identity to non-ribosomal peptide synthetase [*Aspergillus steynii]* –protein AHZ61901.1
Sc14_contig14_orf273	A-P-C-A-P-TE	64% identity to AMP-dependent synthetase/ligase [*Penicillium camemberti*] –protein CRL24942.1
Sc27_contig25_orf903	A-P-C-A-P-C	61% identity to non-ribosomal peptide synthetase [*Aspergillus sp. MF297-2*] –protein ADM34138.1
Sc34_contig29_orf50	A-P-C-C-A-P-C-C	38% identity to hypothetical protein UCREL1_1598 [*Eutypa lata* UCREL1] –protein XP_007789545.1
Sc43_contig37_orf262	A-P-C-A-P-C- HxxP-P	66% identity to Nonribosomal peptide synthase GliP2 [*Neosartorya fischeri* NRRL 181] –protein XP_001263173.1
Sc45_contig39_orf18	A-P-C-A-P-C-A-P-C-A-P-C	47% identity to AMP-dependent synthetase ligase [*Penicillium solitum*] –protein KJJ26924.1
Sc46_contig40_orf137	P-C-A-P	40% identity to HC-toxin synthetase [*Pyrenophora tritici-repentis Pt-1C-BFP]* – protein XP_001940764.1
Sc63_contig53_orf86	A-P-C-P-C-P-C	57% identity to AMP-binding enzyme [*Aspergillus parasiticus* SU-1] –protein KJK66558.1
Sc71_contig60_orf56	C-A-CMET-P-C-A-P-TE	46% identity to non-ribosomal peptide synthetase module [*Aspergillus oryzae* 3.042] –protein EIT81213.1
Sc75_contig64_orf41	A-P-C-A-P-C-A-P-C-A-P-C	36% identity to Nonribosomal peptide synthetase, serinocyclin synthetase NPS1 [*Metarhizium robertsii*]–protein EXU96269.1

PKS-NRPS hybrids

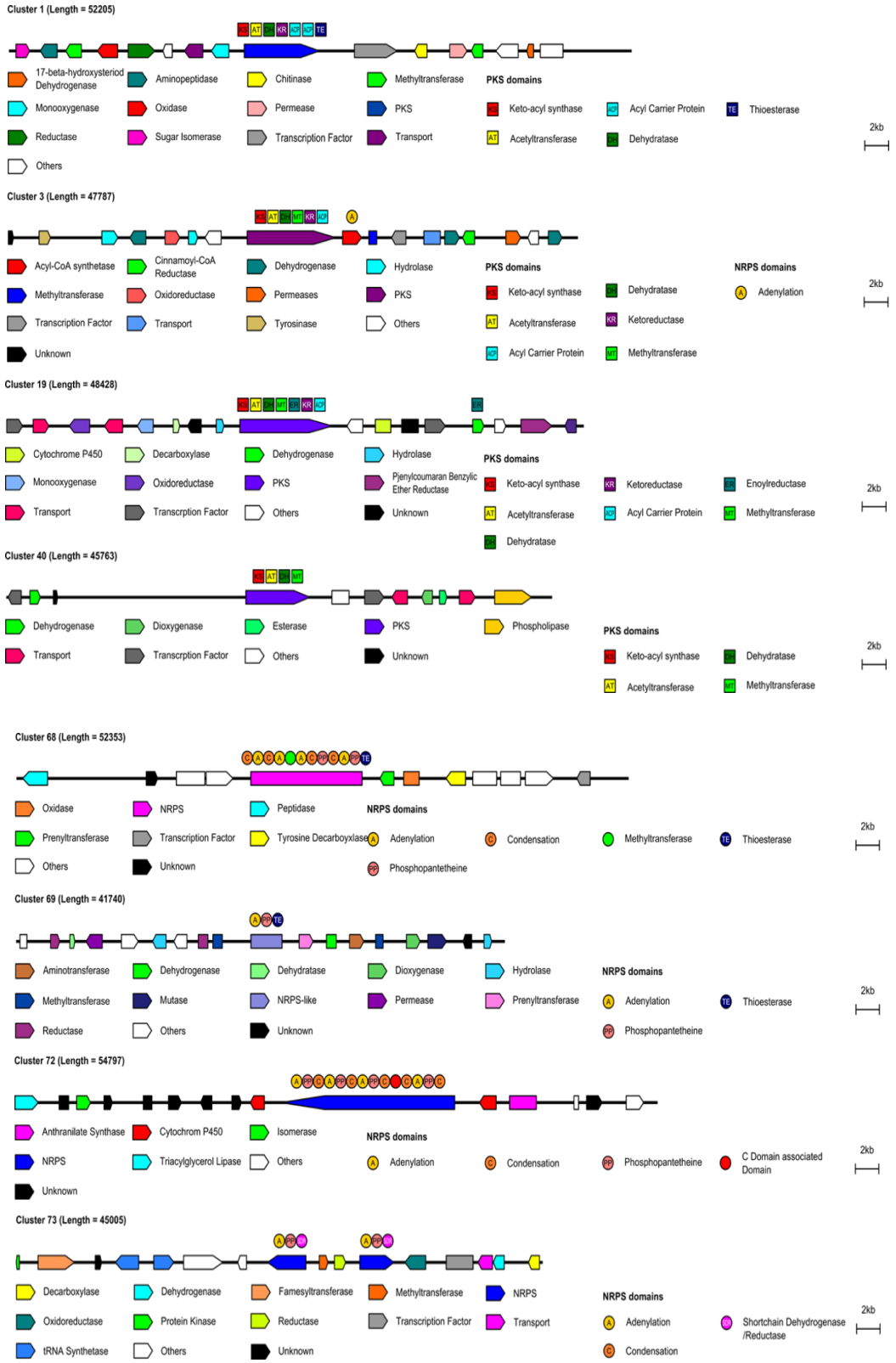
Location	Domain architecture	Closest BLAST hit
Sc8_Contig8_orf129	KS-AT-DH- KR-ACP-C-A-P	69% identity to polyketide synthase 3 [*Xylaria sp.* BCC 1067] - Protein AAS46233.1
Sc14_Contig14_orf26	KS-AT-DH-ER- KR-P-C-A-P-TE	54% identity to lovastatin nonaketide synthase [*Microsporum gypseum* CBS 118893] - protein XP_003176265.1
Sc25_Contig23_orf10	KS-AT-DH-CMET-KR-P-C-A-P-TE	44% identity to hybrid PKS/NRPS enzyme EqiS-like, putative [*Aspergillus fumigatus* A1163] -protein EDP53404.1
Sc25_Contig23_orf301	KS-AT-DH-CMET-KR-P-C-A-P-TE	51% identity to unnamed protein product [*Aspergillus niger*] -protein CAK96944.1
Sc45_Contig39_orf87	KS-AT-DH-CMET-KR-P-C-A-P	46% identity to Nonribosomal peptide synthetase 14 [*Talaromyces islandicus*] –protein CRG87308.1
Sc35_Contig30_orf539	A-P-KS-AT-KR-P-TE	53% identity to Nonribosomal peptide synthetase 14 [*Talaromyces islandicus*] -protein CRG87308.1

Appendix III: Diagrams of gene clusters from *A. westerdijkiae* and *N. jinanensis*

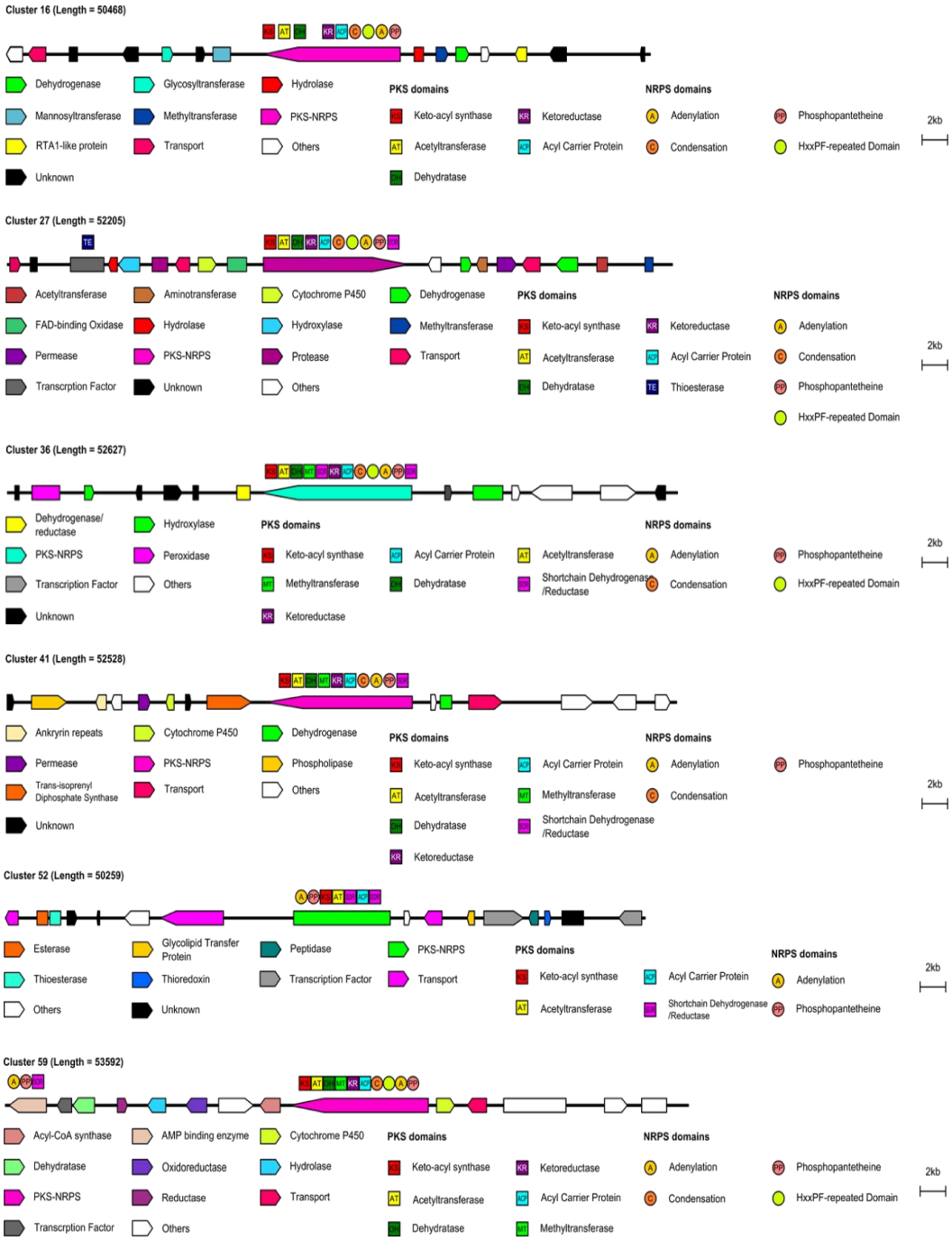
NRPS gene clusters from *A. westerdijkiae*.



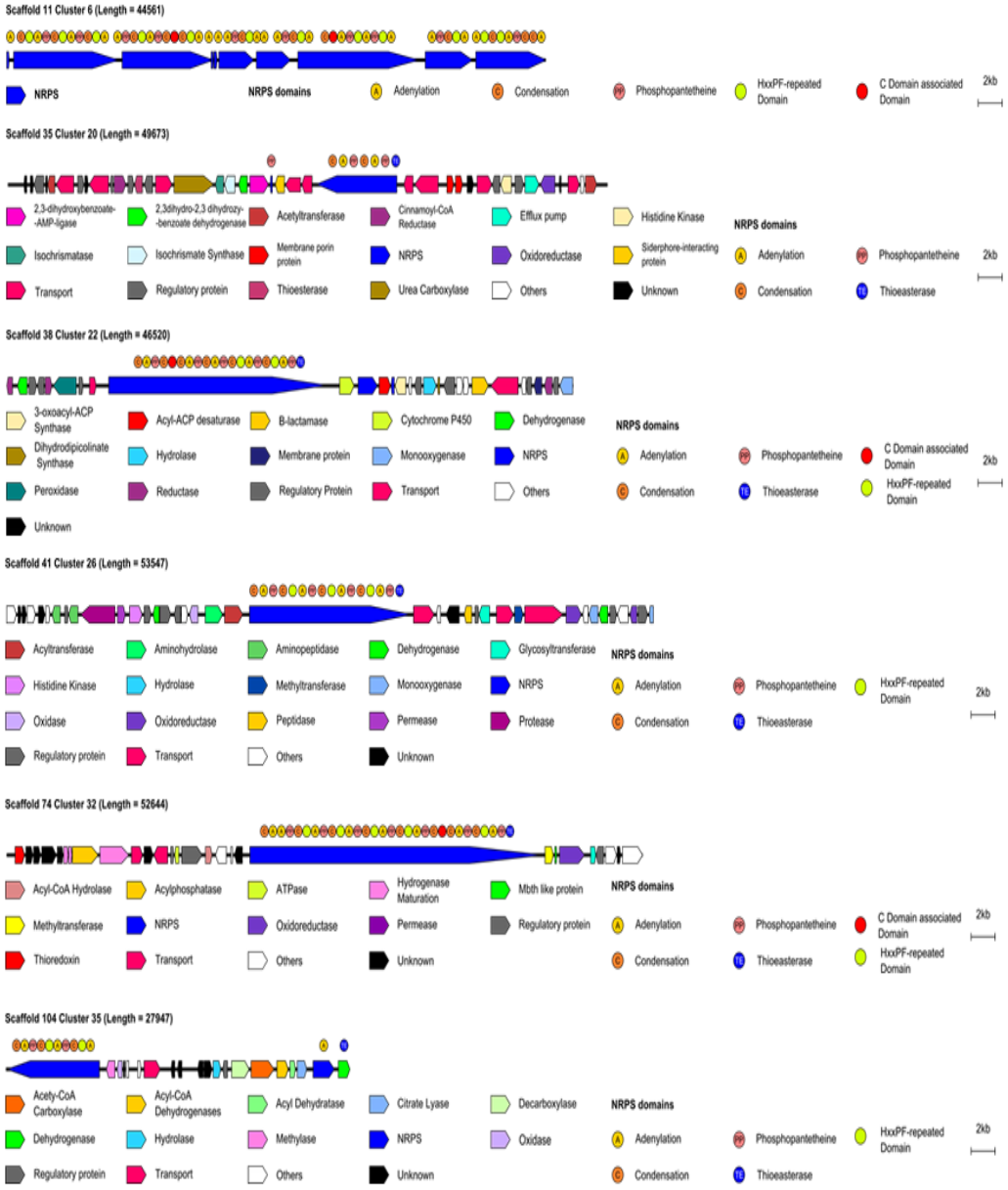
PKS gene clusters from *A. westerdijkiae*.



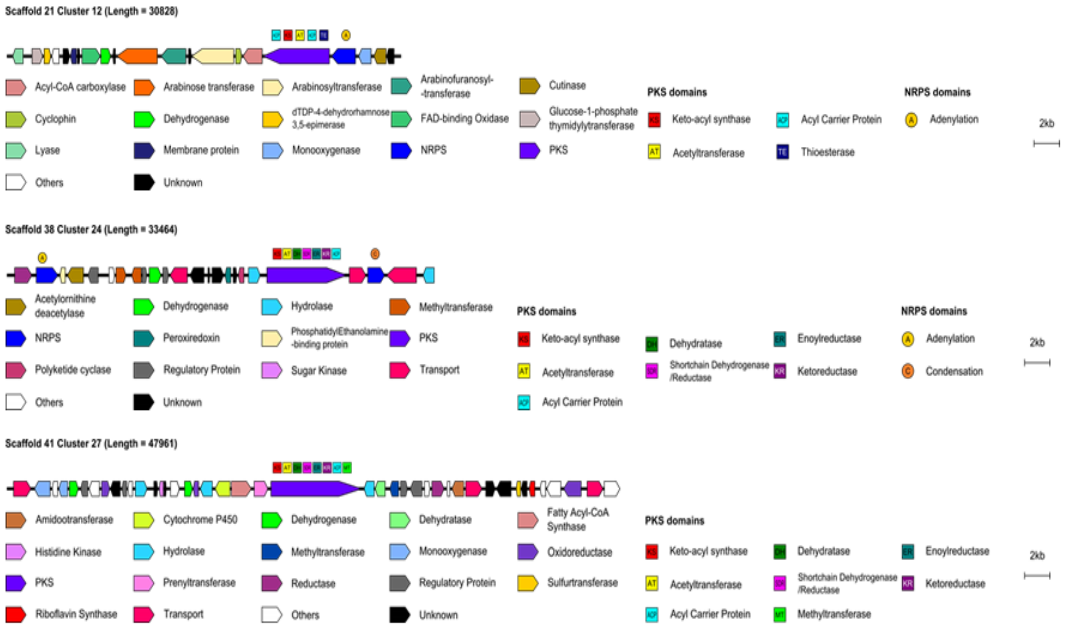
PKS-NRPS hybrid gene clusters from *A. westerdijkiae*.



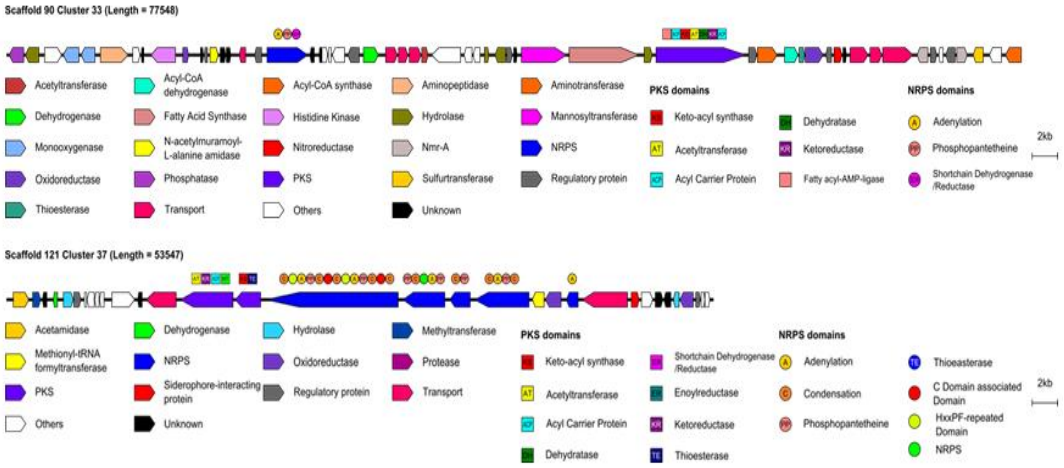
NRPS gene clusters from *N. jinanensis*



PKS gene clusters from *N. jinanensis*



PKS-NRPS hybrid gene clusters from *N. jinanensis*



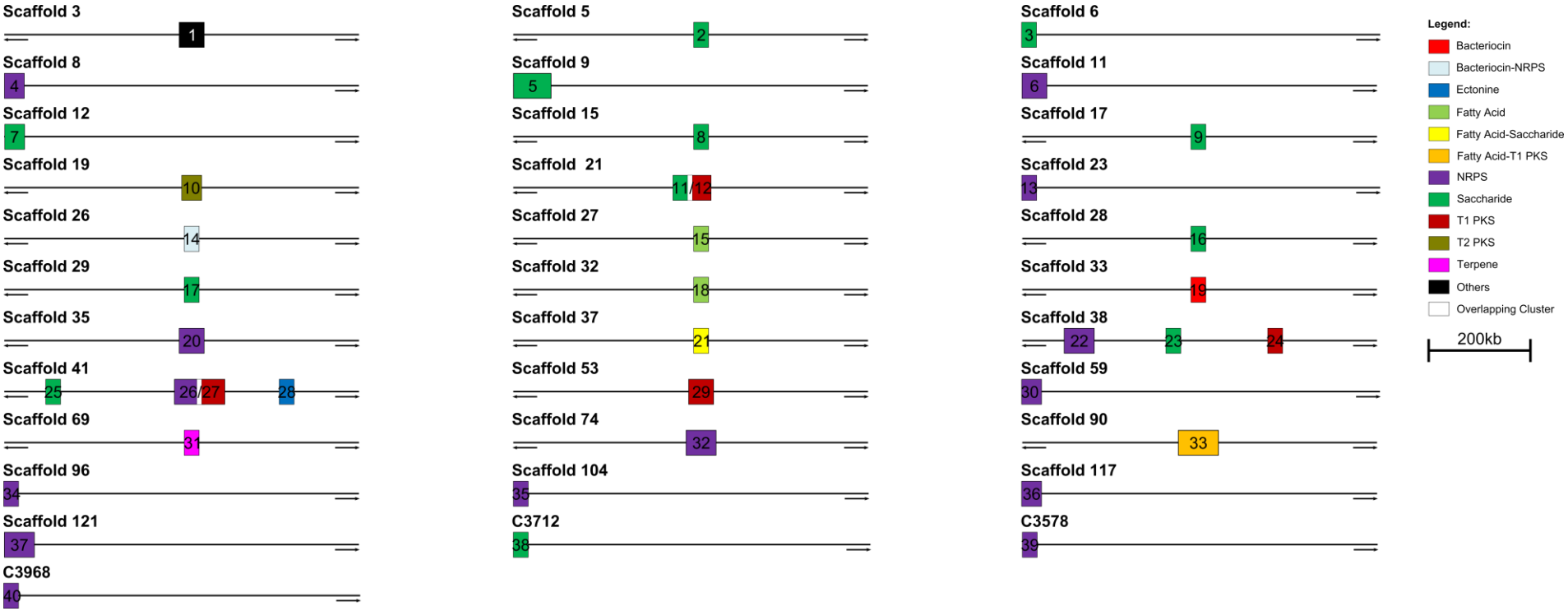
Appendix IV: Putative virulence and toxin related genes in *A. westerdijkiae* genome.

Scaffold/gene cluster	Type of putative virulence/allergen/toxin gene present
Scaffold 1 /Cluster 2	Glycoside hydrolase
Scaffold 3 /Cluster 11	Phospholipase, Arrestin (S-antigen)
Scaffold 4 /Cluster 13	This cluster has 40% homology to botrydial toxin biosynthesis cluster
Scaffold 8/Cluster 16	Mannosyltransferase
Scaffold 12/Cluster 25	Glycoside hydrolase
Scaffold 14/Cluster 27	Aspartic protease
Scaffold 14/Cluster 29	OTA biosynthesis cluster
Scaffold 14/Cluster 31	Phospholipase
Scaffold 14/Cluster 33	Glycoside hydrolase/cellulase, Aspartyl protease
Scaffold 25/Cluster 36	Glycoside hydrolase
Scaffold 25/Cluster 37	Mycosin like serine protease
Scaffold 34/Cluster 49	Glycoside hydrolase, Mannosyltransferase
Scaffold 43/Cluster 55	This cluster has 50% homology to Hexadehydro-astechrome biosynthetic cluster
Scaffold 43/Cluster 56	Glycoside hydrolase
Scaffold 45/Cluster 58	Glycoside hydrolase, Manganese superoxide dismutase
Scaffold 63/Cluster 67	Glycoside hydrolase, Necrosis inducing protein
Scaffold 71/Cluster 70	Glycoside hydrolase, Cytokine induced anti-apoptosis inhibitor
Scaffold 71/Cluster 71	Number of Glycoside hydrolase
Scaffold 151/Cluster 74	Glycoside hydrolase

Appendix V: Summary of antiSMASH predictions for *N. jinanensis* genome.

Scaffold	Cluster	Type	From	To	
Scaffold 3	Cluster 1	Other	190266	234276	Putative Phosphonoglycans biosynthetic gene cluster (3% of genes show homology)
Scaffold 5	Cluster 2	Saccharide	220821	241540	
Scaffold 6	Cluster 3	Saccharide	1	17442	Putative Amychelin biosynthetic gene cluster (12% of genes show homology)
Scaffold 8	Cluster 4	NRPS	1	37465	
Scaffold 9	Cluster 5	Saccharide	1	15480	
Scaffold 11	Cluster 6	NRPS	1	44562	Putative Daptomycin biosynthetic gene cluster (4% of genes show homology)
Scaffold 12	Cluster 7	Saccharide	1	28104	
Scaffold 15	Cluster 8	Saccharide	259306	280265	
Scaffold 17	Cluster 9	Saccharide	17761	38888	
Scaffold 19	Cluster 10	T2 PKS	3603	39005	Putative Auricin biosynthetic gene cluster (20% of genes show homology)
Scaffold 21	Cluster 11	Saccharide	29790	56891	Putative Dihydrochalcomycin biosynthetic gene cluster (9% of genes show homology)
	Cluster 12	T1 PKS	52055	82893	Putative Clorobiocin biosynthetic gene cluster (5% of genes show homology)
Scaffold 23	Cluster 13	NRPS	1	2532	Putative Laspartomycin biosynthetic gene cluster (6% of genes show homology)
Scaffold 26	Cluster 14	Bacteriocin-NRPS	8180	35795	
Scaffold 27	Cluster 15	Fatty Acid	5514	34447	
Scaffold 28	Cluster 16	Saccharide	5276	26529	Putative Macrotetrolide biosynthetic gene cluster (33% of genes show homology)
Scaffold 29	Cluster 17	Saccharide	7491	28708	
Scaffold 32	Cluster 18	Fatty Acid	58237	82175	
Scaffold 33	Cluster 19	Bacteriocin	29592	52531	
Scaffold 35	Cluster 20	NRPS	2386	51069	Putative Mirubactin biosynthetic gene cluster (50% of genes show homology)
Scaffold 37	Cluster 21	Fatty Acid -Saccharide	11774	37167	
Scaffold 38	Cluster 22	NRPS	101273	159527	
	Cluster 23	Saccharide	168051	189310	
	Cluster 24	T1 PKS	202164	235628	
Scaffold 41	Cluster 25	Saccharide	17909	39234	
	Cluster 26	NRPS	132189	185736	
	Cluster 27	T1 PKS	176146	224107	
	Cluster 28	Ectoine	266319	276714	Putative Ectoine biosynthetic gene cluster (75% of genes show homology)
Scaffold 53	Cluster 29	T1 PKS	11106	54216	Putative Epothilone biosynthetic gene cluster (40% of genes show homology)
Scaffold 59	Cluster 30	NRPS	1	3666	Putative Calcium-dependent antibiotic biosynthetic gene cluster (7% of genes show homology)
Scaffold 69	Cluster 31	Terpene	3866	15427	
Scaffold 74	Cluster 32	NRPS	20364	73008	Putative Herboxidiene biosynthetic gene cluster (10% of genes show homology)
Scaffold 90	Cluster 33	Fatty Acid -T1 PKS	55508	133056	Putative Laidlomycin biosynthetic gene cluster (18% of genes show homology)
Scaffold 96	Cluster 34	NRPS	1	25119	
Scaffold 104	Cluster 35	NRPS	1	27947	Putative Plipastatin biosynthetic gene cluster (15% of genes show homology)
Scaffold 117	Cluster 36	NRPS	1	30659	Putative Skyllamycin biosynthetic gene cluster (4% of genes show homology)
Scaffold 121	Cluster 37	NRPS	15546	70709	Putative Nocobactin NA biosynthetic gene cluster (87% of genes show homology)
C3712	Cluster 38	Saccharide	1	7358	
C3578	Cluster 39	NRPS	1	5592	
C3968	Cluster 40	NRPS	1	13089	

Arrangement of secondary metabolism clusters in scaffolds (*N. jinanensis*).



Appendix VI: Summary some of gene clusters identified in *N. jinanensis*.

Scaffold	Cluster	Type	Comments
Overlap	**(Putative)**		
Scaffold 21	Cluster 11	Saccharide	The NAD dependent epimerase, glucose-1-phosphate adenylytransferase and the dehydrohamnose 3 epimerase genes have 50% homology to the noviose sugar genes of Chlorobiocin/Novobiocin synthesis cluster. The cluster shares an overall 9% homology with the Dihydrochalcomycin, Chlorobiocin and the Novobiocin biosynthesis gene cluster.
Scaffold 21	Cluster 12	T1 PKS	This cluster seems to have a lot of homologs in different Actinomycete species. The glucose-1-phosphate adenylytransferase and the dehydrohamnose 3 epimerase genes have 33% homology to the noviose sugar genes of Chlorobiocin synthesis cluster. The cluster has overall 5% homology with the chlorobiocin and novobiocin synthesis cluster. Might be responsible for the production of a chlorobiocin like molecule.
Putative			
Scaffold 35	Cluster 20	NRPS	This cluster has 50% homology to the Mirubactin biosynthesis cluster. This gene cluster has three adenylation domains predicted to be specific for 2, 3 dihydroxybenzoate, arginine and 4-hydoxy-phenyl-glycine respectively. The amino acid specificities of the first two A domain s is the same as the mirucbactin NRPS [the Mirubactin cluster has three A domain specific for 2, 3 dihydroxybenzoate, arginine and hydroxyornithine respectively].This cluster also contains isochorismate synthase, permease, oxidoreductase and other transport proteins like the Mirubactin cluster.
Scaffold 121	Cluster 37	NRPS	This cluster has 87% homology to the Nocobactin biosynthesis gene cluster. The genes nfa7610 to nfa7680 of the Nocobactin cluster have their corresponding homologs in this cluster of *Nocardia jinanensis*.
Unknown	**yet**	**note**	**worthy**
Scaffold 11	Cluster 6	NRPS	4% homology to Daptomycin cluster. This NRPS cluster has atleast 18 adenylation domains, which might be responsible for the synthesis of a complex peptide based natural product.
Scaffold 38	Cluster 22	NRPS	The cluster contains a NRPS with five adenylation domains, a ACP-acyl desaturase, a cytochrome p450, a monooxygenase and peroxidase.
Scaffold 38	Cluster 24	T1 PKS	The cluster has an iterative PKS and a NRPS along with an acetylornithine deacetylase and an esterase gene among others. The PKS and NRPS genes together could generate an interesting secondary metabolite structure. It might be worthy to note that there is a NRPS and a saccharide gene cluster in close proximity of this particular cluster.
Scaffold 41	Cluster 26-27	NRPS and T1PKS (overlap)	These NRPS and iterative PKS containing overlapping clusters along with other accessory genes might be responsible for the synthesis of structurally unique and biologically relevant small molecule.
Scaffold 74	Cluster 32	NRPS	This cluster has more than five adenylation domains of NRPS and hence could be a source of a small molecule worthy of attention.
Scaffold 90	Cluster 33	Fatty acid – T1 PKS	This cluster has a PKS as well as a NRPS gene. The cluster has 18% homology to the Laidlomycin cluster. Unlike the Laidlomycin cluster, this particular gene cluster has a NRPS gene. Hence, this cluster has the possibility of synthesizing an interesting secondary metabolite.
Scaffold 104	Cluster 35	NRPS	This NRPS cluster has a stand-alone ER domain. This cluster has 15% homology to the Plipastatin biosynthesis cluster. Hence might be of interest for further investigation.

Appendix VII: Summary of NRPS and PKS genes in *N. jinanensis* genome.

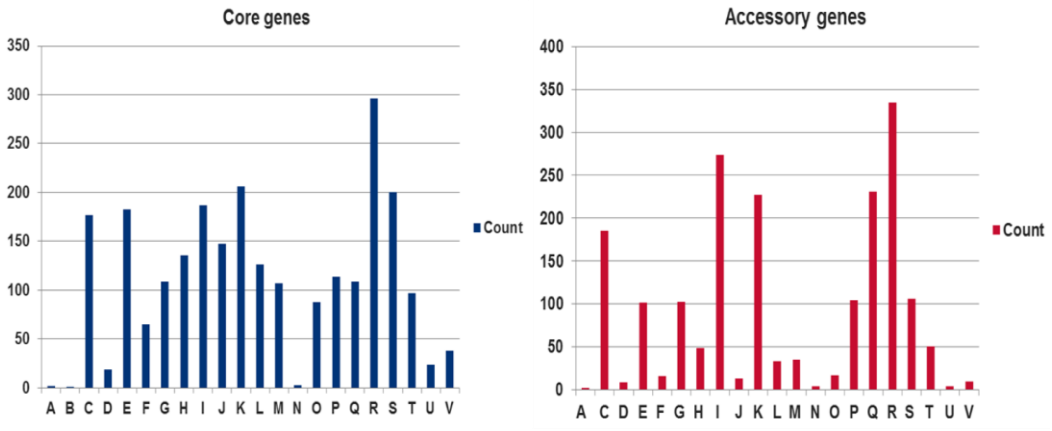
Non-ribosomal peptide synthetases

Scaffold No.	Predicted Domain Organization	Nearest homology
Scaffold 8	C-A-PP-TE	50% identity to *Nocardia tenerifensis* protein WP_040737662.1
Scaffold 11	A-C-A-P-C-A-P-C-A-P-C-A-P-E-C-A-P-C-A-P-C-E-C-A-P-C-A-P-C-A-P-C-A-P-C-A-P-C	88% identity to *Nocardia testacea* protein WP_039826763.1
Scaffold 23	C-HxxPF_rpt-A	61% identity to *Nocardia paucivorans* protein WP_040792650.1
Scaffold 35	C-A-PP-C-A	88% identity to *Nocardia rhamnosiphila* protein WP_030524090.1
Scaffold 38	C-A-P-C	86% identity to *Nocardia rhamnosiphila* protein WP_030519936.1
Scaffold 41	C-A-P-C-A-P-C-A-P-C-A-P-TE	89% identity to *Nocardia paucivorans* protein WP_030519037.1
Scaffold 59	C-A-P-C	81% identity to *Nocardia paucivorans* protein WP_036516119.1
Scaffold 74	A-P-C-HxxP-A-P-C-HxxP	77% identity to *Nocardia paucivorans* protein WP_040792356.1
Scaffold 96	C—HxxP-A-P-C-C	69% identity to *Nocardia brevicatena* protein WP_040840939.1
Scaffold 104	C-A-P-C-HxxP-A	81% identity to *Nocardia paucivorans* protein WP_036516119.1
Scaffold 117	A-P-C-C-A-P-C	75% identity to *Nocardia testacea* protein WP_039822393.1
Scaffold 121	C-HxxP-A-C-C	82% identity to *Nocardia rhamnosiphila* protein WP_030522672.1

Polyketide synthases

Scaffold No:	Gene Type	Predicted Domain Organization	Nearest Homology
Scaffold 21	PKS	ACP-KS-AT-ACP-TE	92% identity to *Nocardia rhamnosiphila* - protein [WP_030524332.1]
Scaffold 38	PKS	KS-AT-DH-ER-KR-ACP	76% identity to *Nocardia paucivorans* - protein [WP_040790680.1]
Scaffold 41	PKS	KS-AT-DH-MT-ER-KR-ACP	73% identity to *Nocardia paucivorans* - protein [WP_040789102.1] *42% identity to the PKS uncultured bacterium N27-1E – protein [AHZ46196.1]
Scaffold 53	PKS	KS-AT	74% identity to *Nocardia paucivorans* - protein [WP_040793265.1]
Scaffold 90	PKS	AT AT FAAL-KS-AT-DH-KR-ACP	57% identity to *Nocardia nova* – protein [WP_025353457.1]
Scaffold 121	PKS	AT-KR-ACP-MT KS-TE	83% identity to *Nocardia rhamnosiphila* - protein [WP_036514611.1]

Appendix VIII: COG analysis of the *N. jinanensis* genome



COG analysis of the *N. jinanensis* genes when compared the genomes of *N. brasilensis, N. farcinica* and *N. cyriacigeorgica*. While core genes indicate the number of genes that are shared by all the species of Nocardia mentioned above, accessory genes are those which are unique to *N. jinanensis*. The alphabetic names of the functional categories of genes as per COG have been listed below.

List of functional categories of genes in the above COG analysis.

COG class	Description
A	RNA processing and modification
C	Energy production and conversion
D	Cell cycle control
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
G	Carbohydrate transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
J	Translation
K	Transcription
L	Replication
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modifications
P	Inorganic ion transport and metabolism
Q	Secondary metabolite biosynthesis
R	General function prediction only
S	Function unknown
T	Signal transduction mechanism
U	intracellular trafficking
V	Defense mechanism

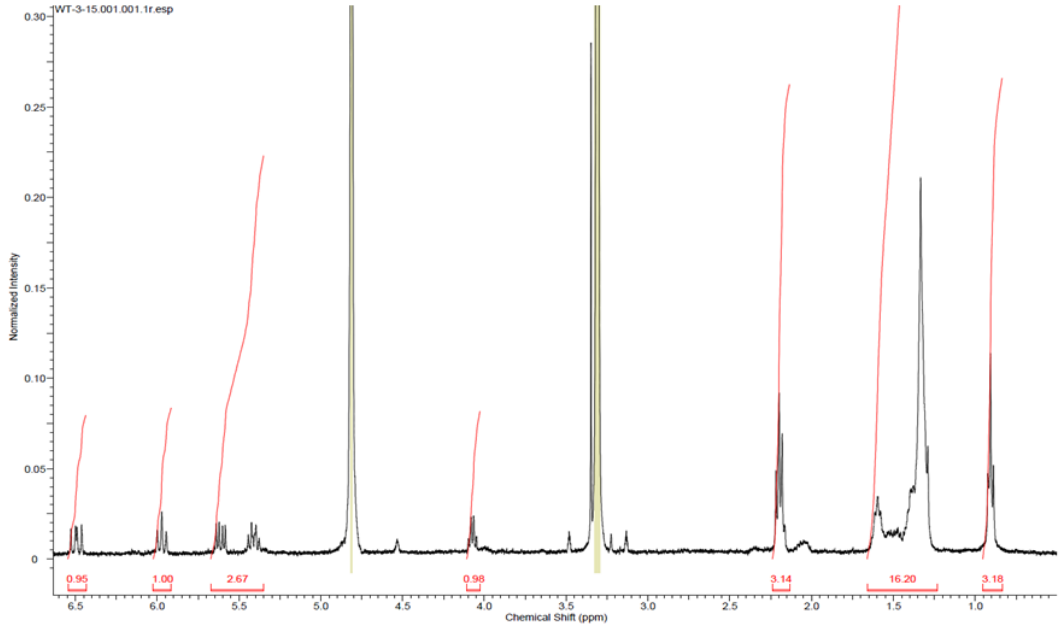
Appendix IX: List of putative genes related to cell wall synthesis and virulence (the ones homologous to *N. farcinica* are highlighted) in *N. jinanensis*.

Name	Length	Start	End	Nearest homology/Description
Cell wall synthesis				
scaffold102_orf00001	360	1	359	UDP-N-acetylmuramoylalanine--D-glutamate ligase (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid-adding enzyme) [Nocardia cyriacigeorgica GUH-2]
scaffold102_orf00003	100	1	60	MURD_NOCFA RecName: Full=UDP-N-acetylmuramoylalanine--D-glutamate ligase; AltName: Full=D-glutamic acid-adding enzyme; AltName: Full=UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
scaffold102_orf00005	359	1	358	putative phospho-N-acetylmuramoyl-pentapeptide- transferase [Nocardia farcinica IFM 10152]
scaffold102_orf00006	517	1	490	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanine ligase [Nocardia cyriacigeorgica GUH-2]
scaffold102_orf00008	539	1	538	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase [Nocardia cyriacigeorgica GUH-2]
scaffold102_orf00009	598	12	597	putative peptidoglycan synthetase ftsI precursor (Peptidoglycan glycosyltransferase 3) (Penicillin-binding protein 3) (PBP-3) [Nocardia cyriacigeorgica GUH-2]
Virulence/ Pathogenicity				
scaffold104_orf00037	394	1	393	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold104_orf00038	370	1	360	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold104_orf00040	367	1	366	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold104_orf00041	338	1	335	putative Mce family protein [Nocardia farcinica IFM 10152]
scaffold104_orf00042	55	2	51	putative Mce family protein [Nocardia farcinica IFM 10152]
scaffold11_orf00029	78	15	77	antigenic cell wall protein MP2 [Rhodococcus erythropolis SK121]
scaffold120_orf00003	387	1	386	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00004	354	1	352	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00006	367	9	366	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00007	332	1	329	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00008	344	4	343	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00009	518	1	516	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold120_orf00011	288	1	287	putative YrbE family protein [Nocardia farcinica IFM 10152]
scaffold120_orf00012	247	1	246	YrbE family protein [Nocardia cyriacigeorgica GUH-2]
scaffold121_orf00050	439	7	437	putative polyketide synthase MbtC [Mycobacterium abscessus subsp. bolletii BD]
scaffold121_orf00062	1493	1	1468	non-ribosomal peptide synthetase (MBTE peptide synthetase) [Nocardia cyriacigeorgica GUH-2]
scaffold121_orf00064	430	14	428	lysine-N-oxygenase [Nocardia cyriacigeorgica GUH-2]
scaffold121_orf00065	568	18	346	non-ribosomal peptide synthetase (MBTE peptide synthetase) [Nocardia cyriacigeorgica GUH-2]
scaffold122_orf00025	334	1	333	virulence factor Mce family protein [Nocardia brasiliensis ATCC 700358]
scaffold15_orf00002	310	1	309	mycolyltransferase [Nocardia brasiliensis ATCC 700358]
scaffold15_orf00050	323	5	320	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00051	336	1	335	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00052	367	38	359	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00054	343	18	342	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00055	338	24	336	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00057	344	2	317	putative Mce family protein [Rhodococcus opacus B4]
scaffold15_orf00058	294	3	293	putative YrbE family protein [Rhodococcus opacus B4]
scaffold15_orf00059	236	7	235	putative YrbE family protein [Rhodococcus opacus B4]
scaffold22_orf00011	208	1	207	superoxide dismutase, Mn [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00081	317	21	316	Conserved Mce associated membrane protein [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00083	489	1	488	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00086	424	13	410	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00087	440	3	425	Mce family protein [Nocardia brasiliensis ATCC 700358]
scaffold3_orf00088	392	21	337	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00091	342	1	341	Mce family protein [Nocardia brasiliensis ATCC 700358]
scaffold3_orf00092	377	6	376	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold3_orf00094	286	6	285	YrbE family protein [Nocardia brasiliensis ATCC 700358]
scaffold3_orf00095	255	1	254	putative YrbE family protein [Nocardia farcinica IFM 10152]
scaffold3_orf00279	156	37	155	LSR2 protein [Nocardia brasiliensis ATCC 700358]
scaffold31_orf00040	321	1	318	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold31_orf00041	356	12	355	putative Mce family protein [Nocardia farcinica IFM 10152]
scaffold31_orf00042	356	1	355	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold31_orf00043	353	15	352	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold31_orf00045	335	1	333	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold31_orf00046	326	32	323	Mce family protein [Nocardia cyriacigeorgica GUH-2]
scaffold31_orf00050	319	49	318	putative YrbE family protein [Nocardia farcinica IFM 10152]
scaffold37_orf00016	219	1	216	Superoxide dismutase [Cu-Zn] [Nocardia cyriacigeorgica GUH-2]
scaffold38_orf00273	109	1	104	putative LSR2 protein [Nocardia farcinica IFM 10152]
scaffold42_orf00024	62	1	59	myosin-crossreactive antigen [Rhodopseudomonas palustris CGA009]
scaffold42_orf00098	296	1	293	mycolyltransferase [Nocardia brasiliensis ATCC 700358]
scaffold44_orf00023	306	2	305	mycolyltransferase [Nocardia brasiliensis ATCC 700358]
scaffold51_orf00284	491	57	482	MOP superfamily O-antigen transporter [Nocardia brasiliensis ATCC 700358]
scaffold51_orf00407	349	33	348	putative mycolyltransferase [Nocardia cyriacigeorgica GUH-2]
scaffold51_orf00408	111	1	108	LSR2 protein [Nocardia brasiliensis ATCC 700358]
scaffold64_orf00052	182	1	181	putative invasion-associated protein [Nocardia cyriacigeorgica GUH-2]
scaffold68_orf00001	484	2	482	MCE-family protein [Mycobacterium marinum M]
scaffold68_orf00002	374	1	318	virulence factor mce family protein [Mycobacterium thermoresistibile ATCC 19527]
scaffold68_orf00003	416	1	402	virulence factor mce family protein [Mycobacterium smegmatis str. MC2 155]
scaffold68_orf00004	344	1	335	virulence factor Mce family protein [Mycobacterium thermoresistibile ATCC 19527]
scaffold68_orf00006	339	8	337	MCE-family protein [Mycobacterium kansasii ATCC 12478]
scaffold68_orf00007	350	1	342	virulence factor Mce [Mycobacterium colombiense CECT 3035]

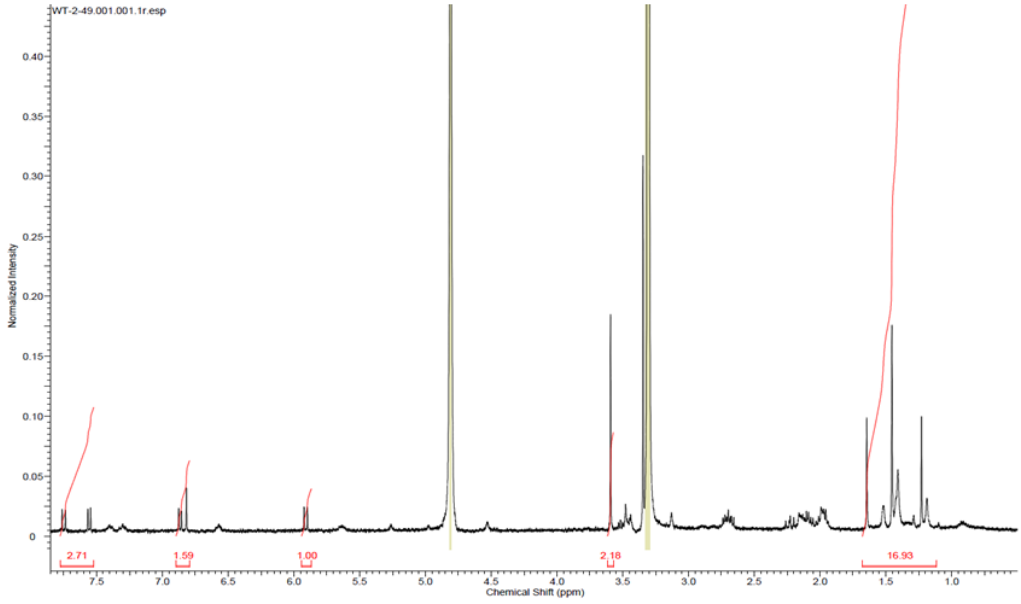
scaffold72_orf00168	255	5	250	putative YrbE family protein [Rhodococcus opacus B4]
scaffold72_orf00169	272	5	271	putative YrbE family protein [Rhodococcus opacus B4]
scaffold72_orf00171	406	6	401	putative Mce family protein [Rhodococcus opacus B4]
scaffold72_orf00172	341	1	338	putative Mce family protein [Rhodococcus opacus B4]
scaffold72_orf00174	402	10	381	putative Mce family protein [Rhodococcus opacus B4]
scaffold72_orf00176	416	27	410	putative Mce family protein [Rhodococcus opacus B4]
scaffold72_orf00177	381	14	376	putative Mce family protein [Rhodococcus opacus B4]
scaffold72_orf00179	484	1	463	putative Mce family protein [Rhodococcus opacus B4]
scaffold76_orf00042	278	11	277	YrbE family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00044	292	16	291	YrbE family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00045	342	2	330	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00046	318	1	314	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00047	314	2	313	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00049	362	3	359	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00050	339	8	338	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold76_orf00051	320	1	315	Mce family protein [Gordonia polyisoprenivorans NBRC 16320]
scaffold82_orf00146	49	5	42	alkylhydroperoxidase AhpD family core domain protein [Saccharomonospora glauca K62]
scaffold89_orf00025	196	1	195	alkylhydroperoxide reductase [Nocardia brasiliensis ATCC 700358]
scaffold89_orf00026	180	1	179	Alkyl hydroperoxide reductase D protein AhpD [Nocardia cyriacigeorgica GUH-2]
scaffold90_orf00057	318	1	101	putative LSR2 protein [Nocardia farcinica IFM 10152]
scaffold94_orf00142	263	19	258	putative YrbE family protein [Rhodococcus opacus B4]
scaffold94_orf00143	262	1	261	putative YrbE family protein [Rhodococcus opacus B4]
C3952_orf00017	144	4	143	18 kDa antigen 2 [Nocardia brasiliensis ATCC 700358]
C4008_orf00021	265	16	261	YrbE family protein [Gordonia amarae NBRC 15530]
C4132_orf00017	338	1	337	mycolyltransferase [Nocardia brasiliensis ATCC 700358]
C4132_orf00019	330	7	329	SECRETED ANTIGEN 85-B FBPB (85B) (ANTIGEN 85 COMPLEX B) (MYCOLYL TRANSFERASE 85B) (FIBRONECTIN-BINDING PROTEIN B) (EXTRACELLULAR ALPHA-ANTIGEN) [Nocardia cyriacigeorgica GUH-2]
C4132_orf00021	309	9	308	mycolyltransferase [Nocardia brasiliensis ATCC 700358]

Appendix X: NMR data

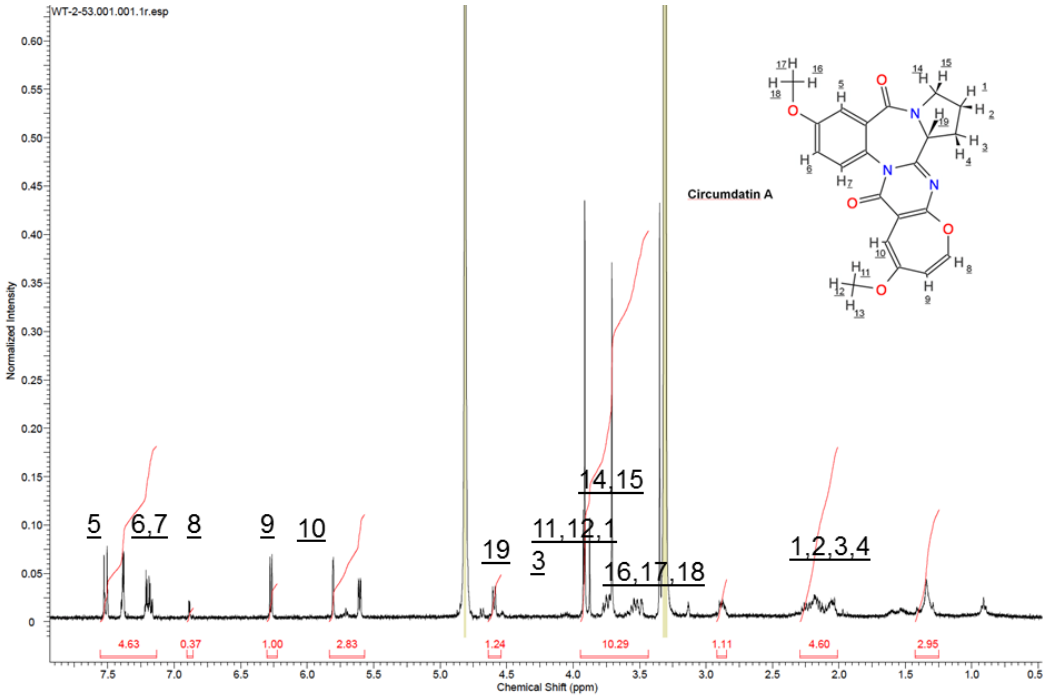
Aw- 13-hydroxyoctadeca-9,11-dienoic acid



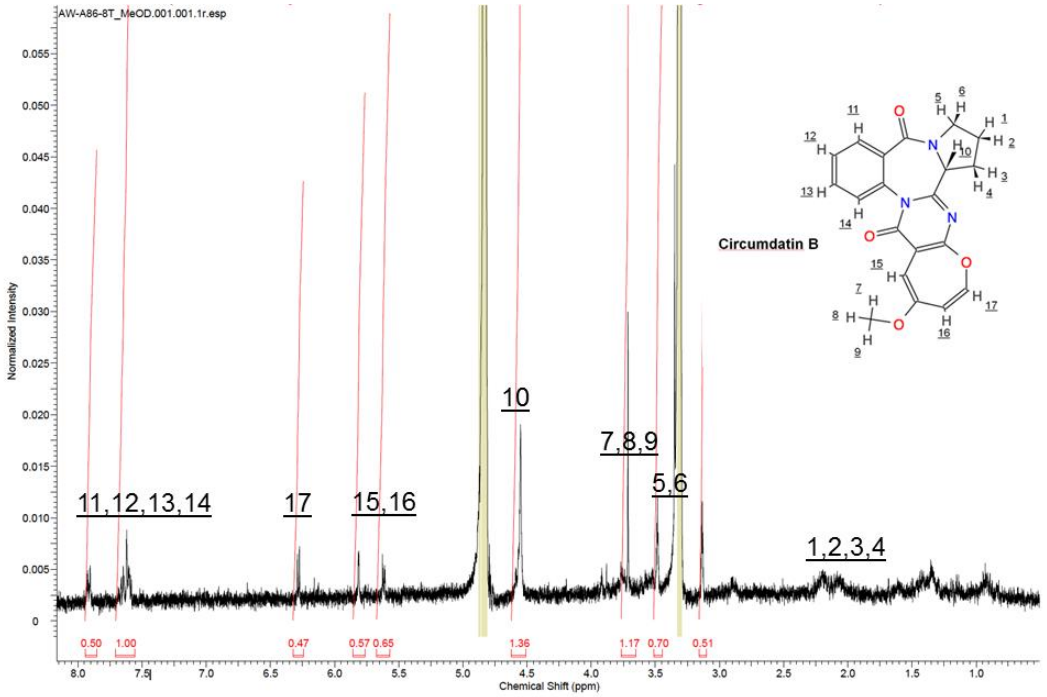
Aw- Avrainvillamide



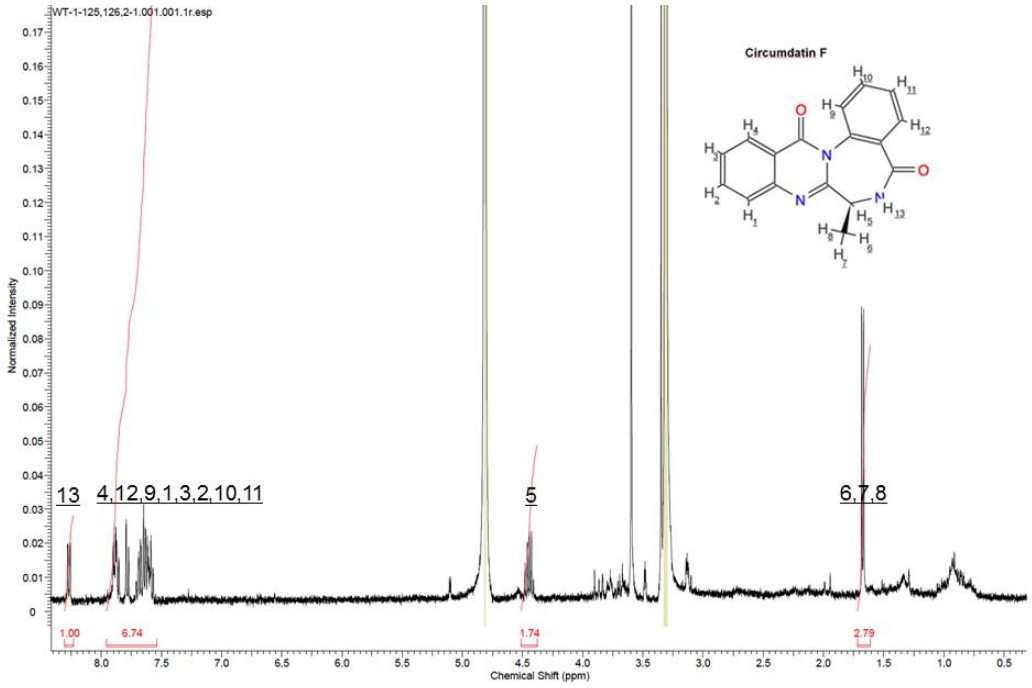
Aw- Circumdatin A



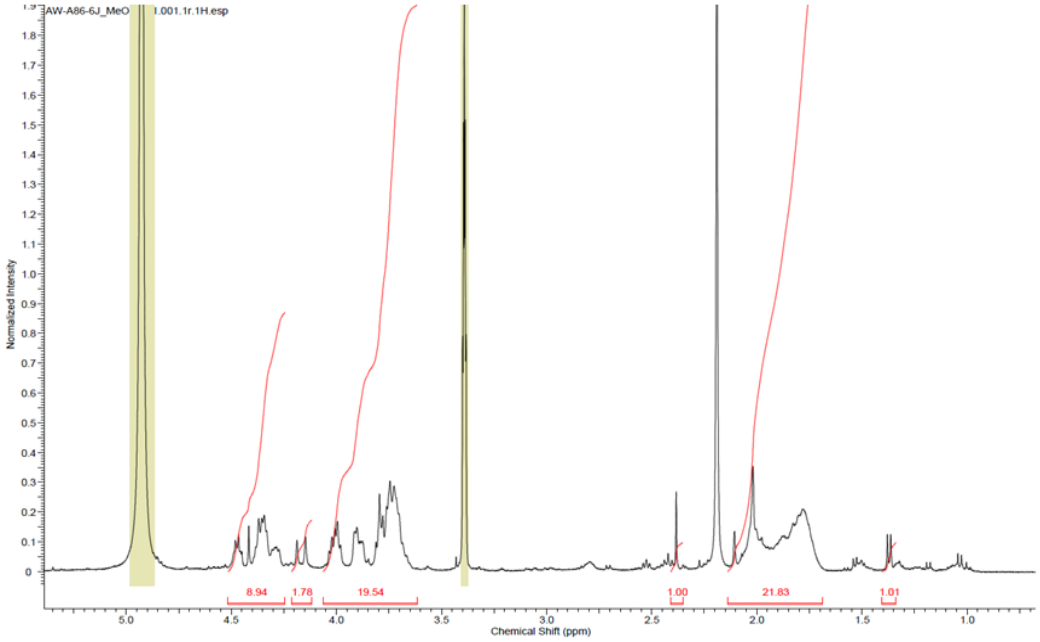
Aw- Circumdatin B



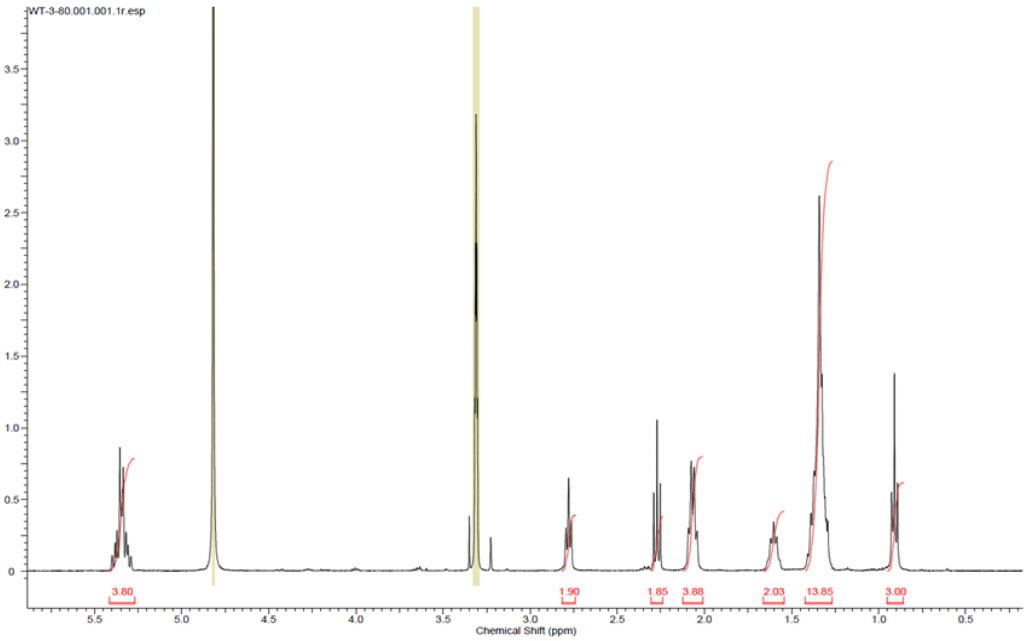
Aw- Circumdatin F



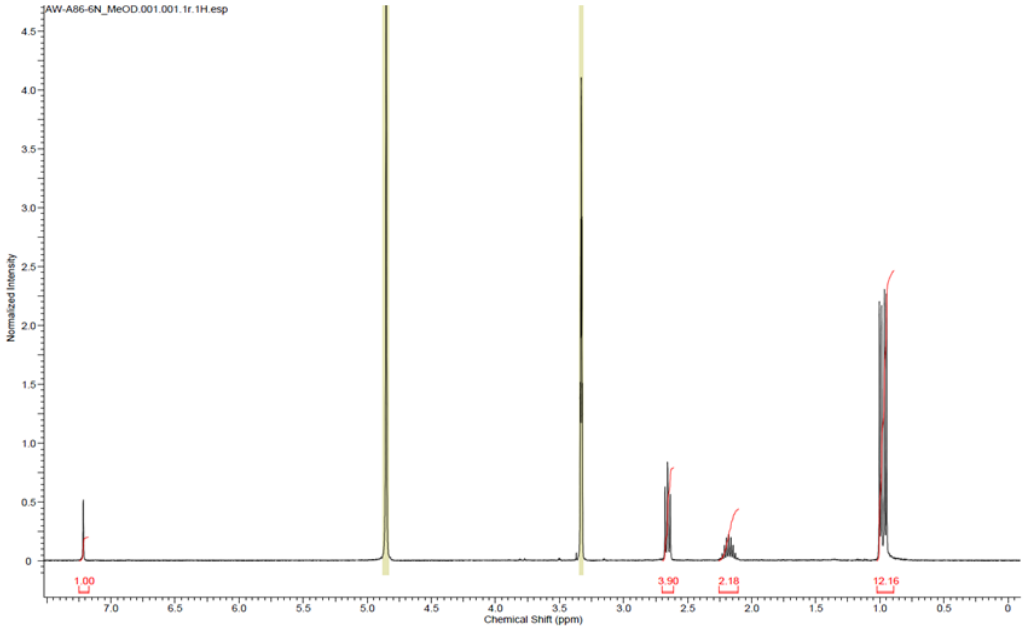
Aw- Desferrichrycin



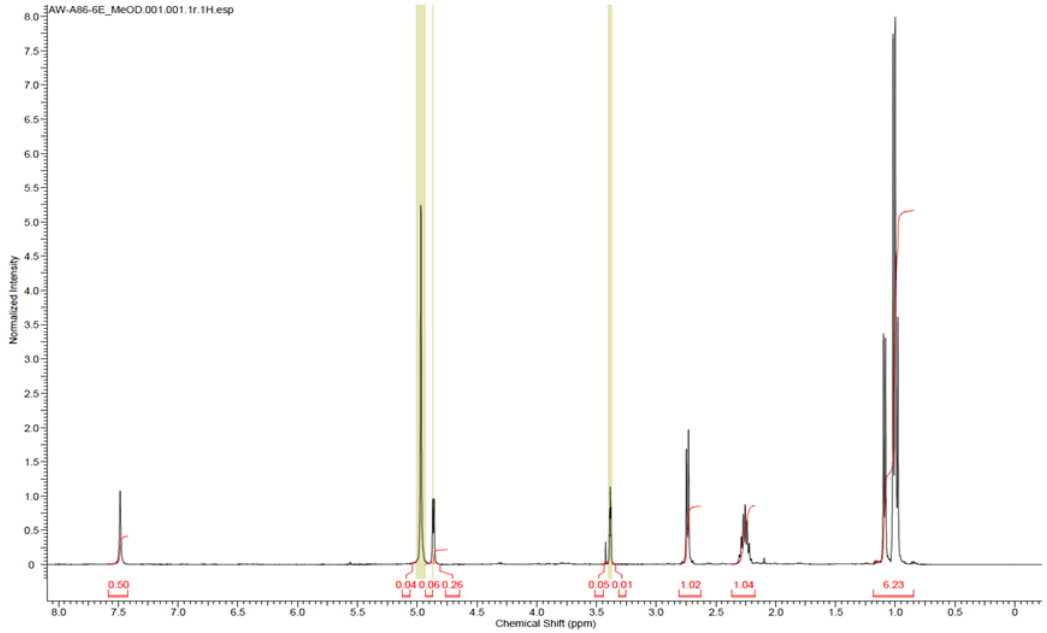
Aw- Linolenic acid



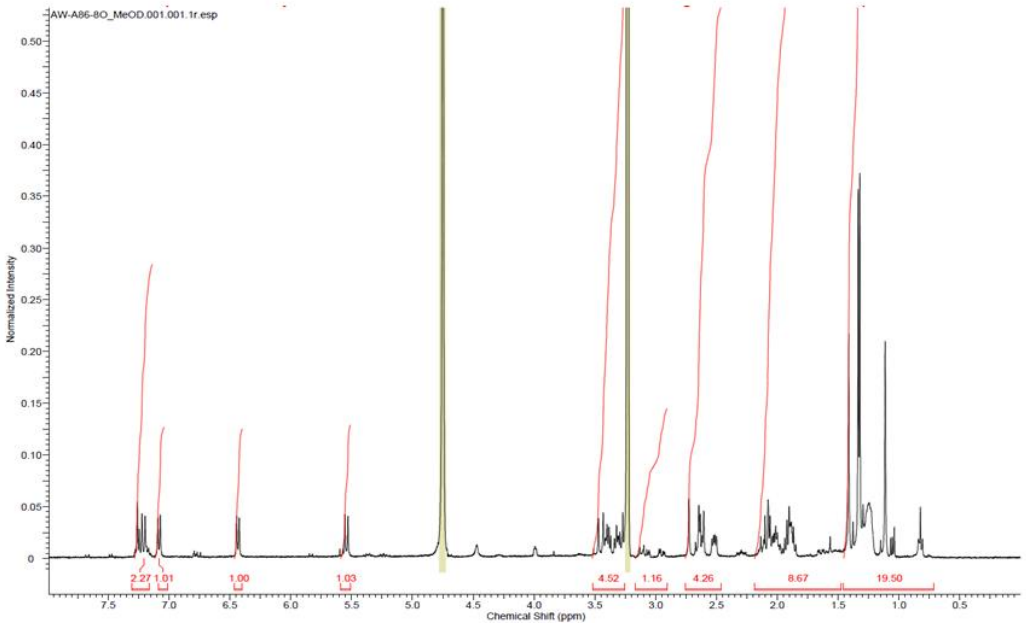
Aw- Neoaspergillic acid



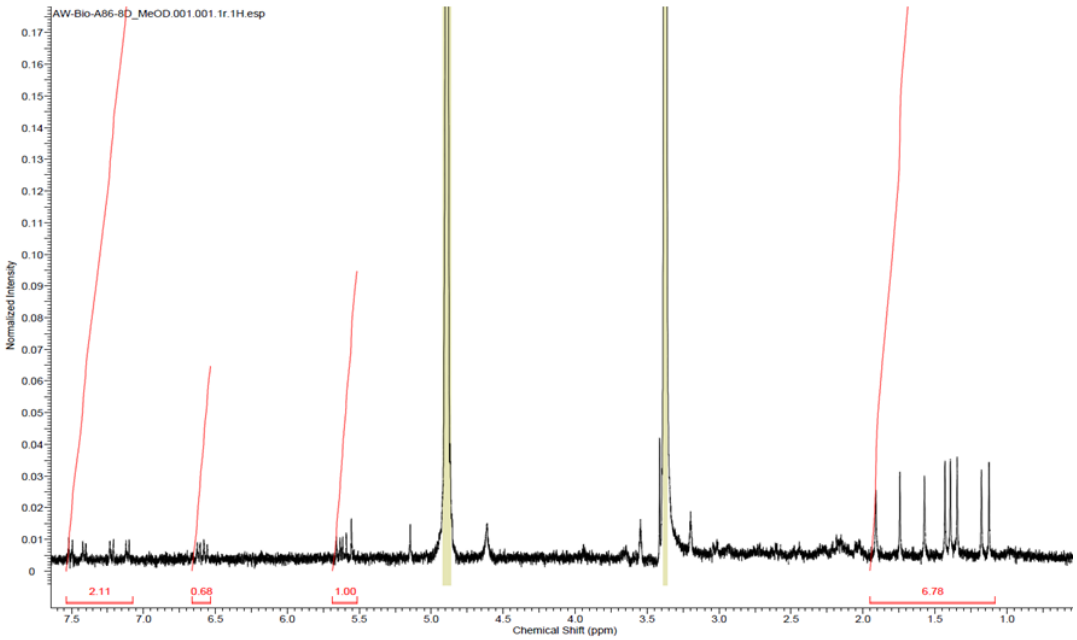
Aw- Neohydroxyaspergillic acid



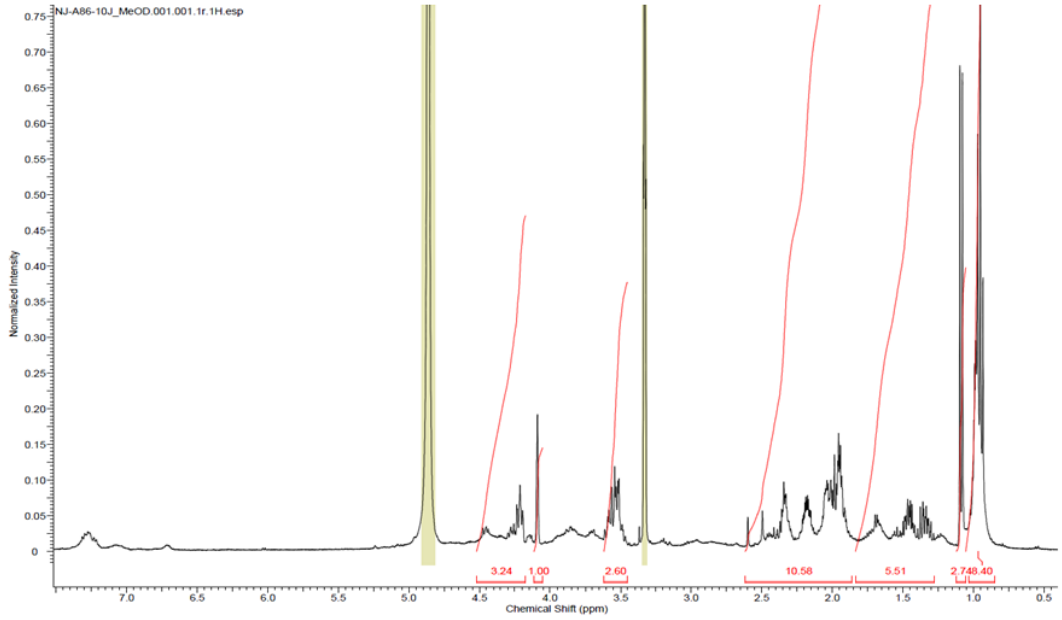
Aw- Notoamide B



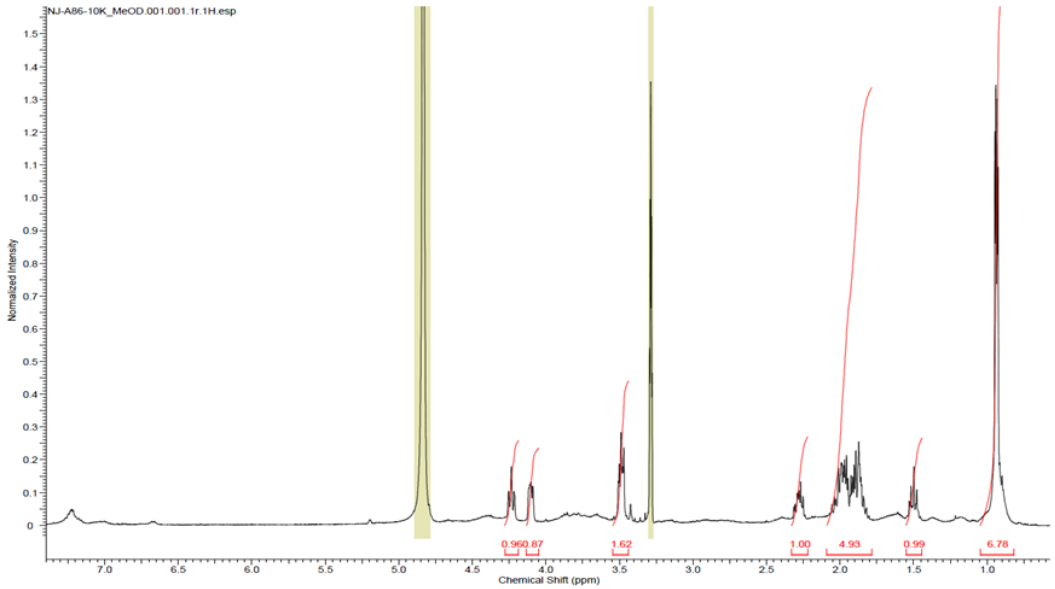
Aw- Stephacidin B



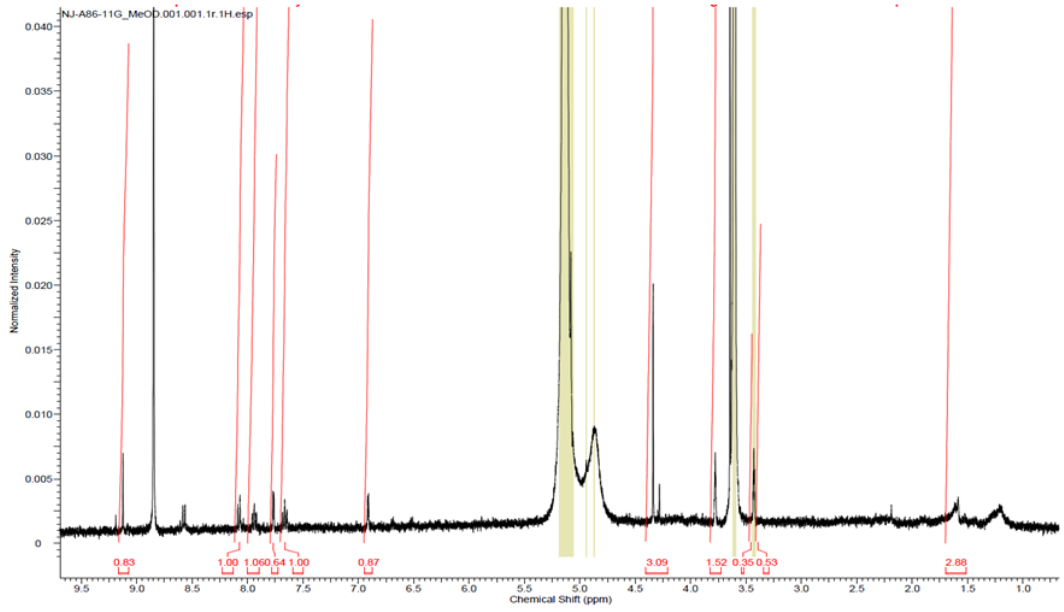
Nj – Cyclo(isoleucyl-propyl)



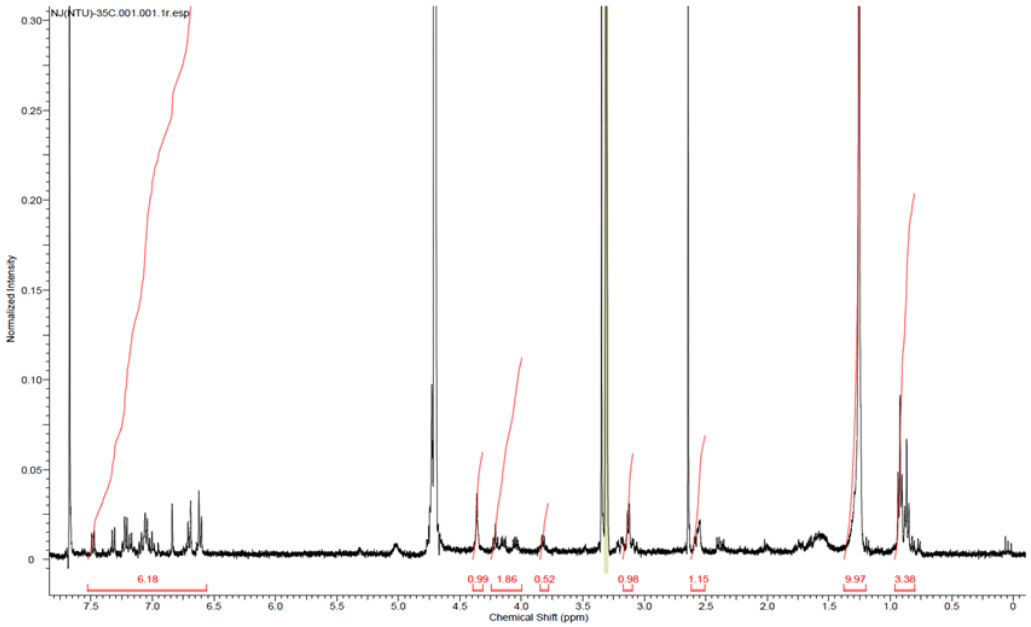
Nj – Cyclo(leucyl-propyl)



Nj –Flazine methyl ester

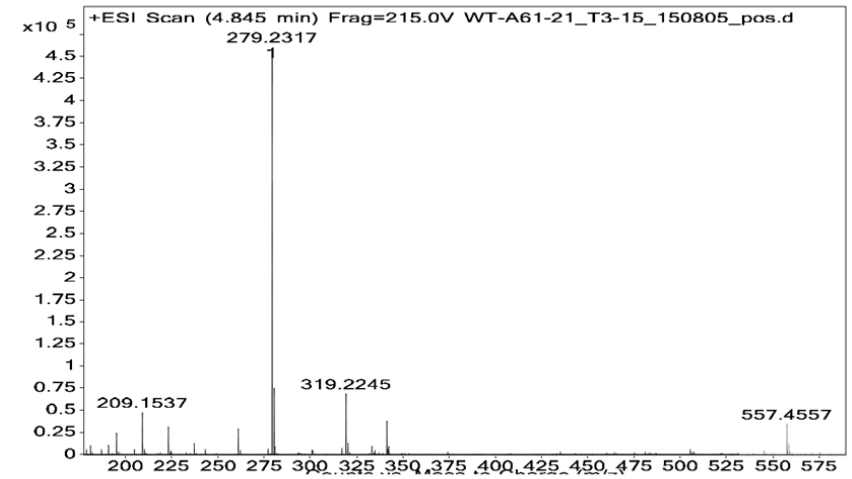


Nj –unidentified peptide

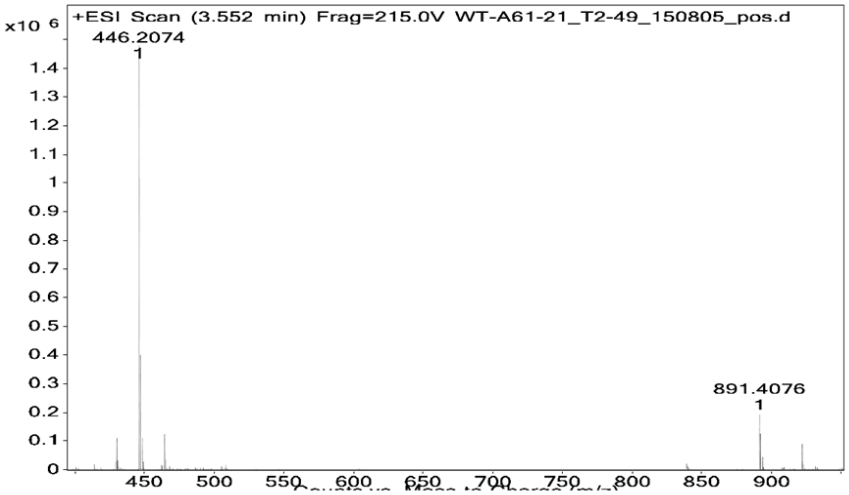


Appendix XI: MS data

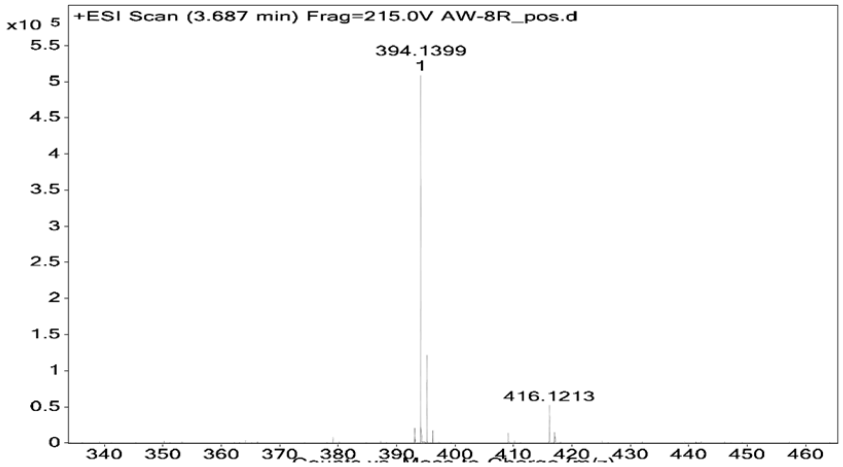
Aw- 13-hydroxyoctadeca-9,11-dienoic acid



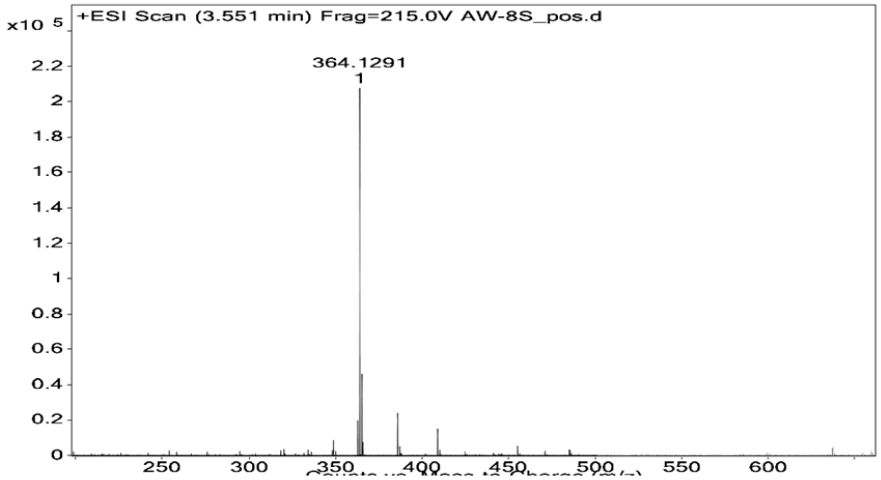
Aw- Avrainvillamide



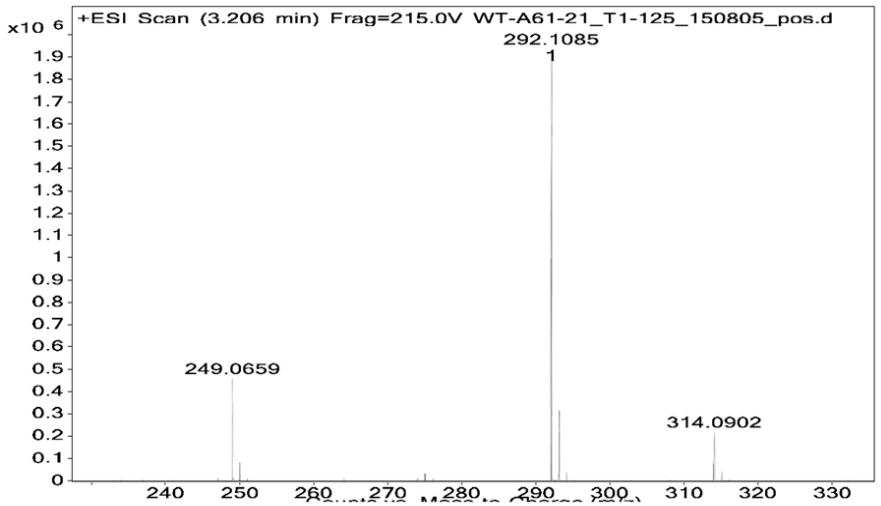
Aw- Circumdatin A



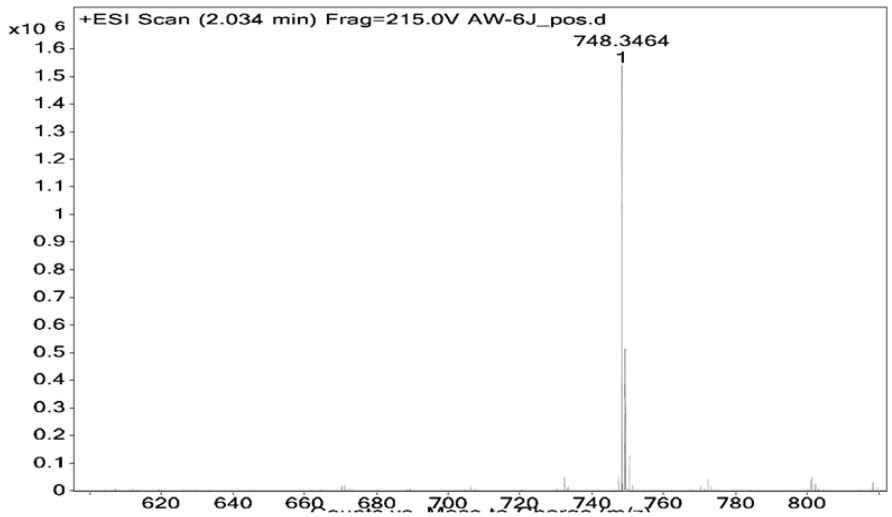
Aw- Circumdatin B



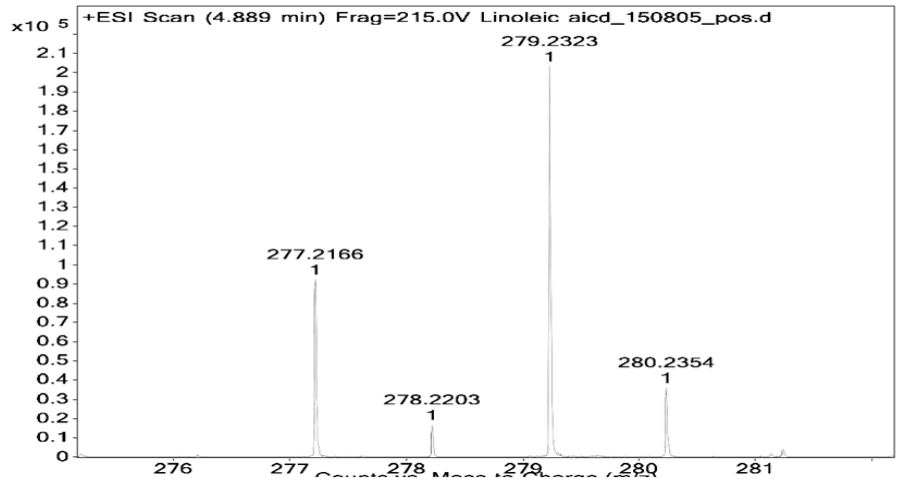
Aw- Circumdatin F



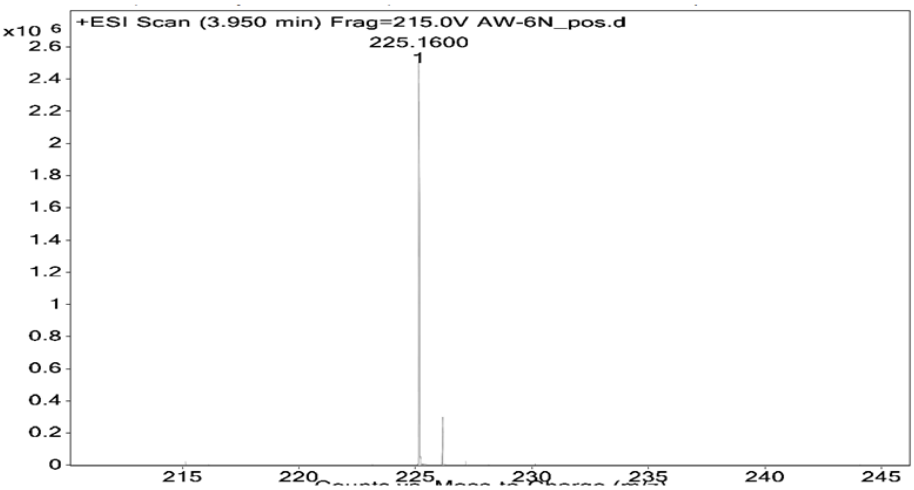
Aw- Desferrichrycin



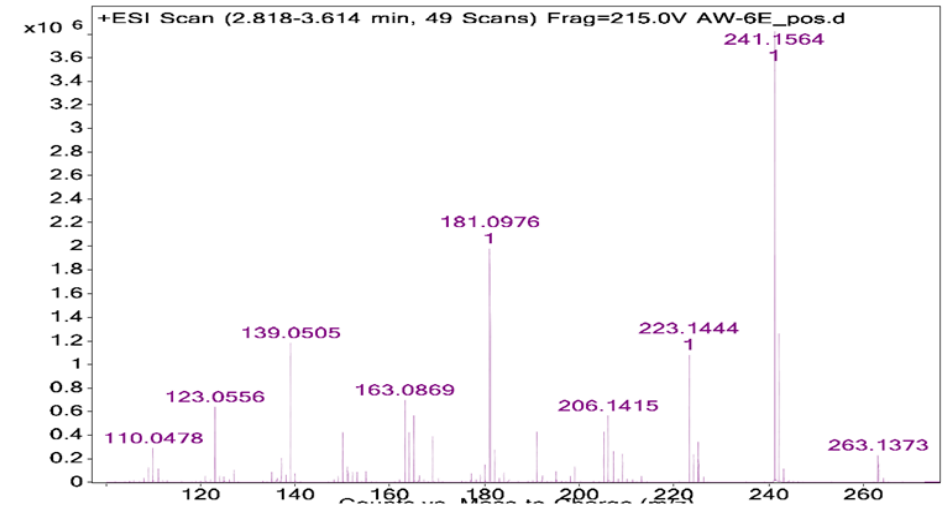
Aw- Linolenic acid



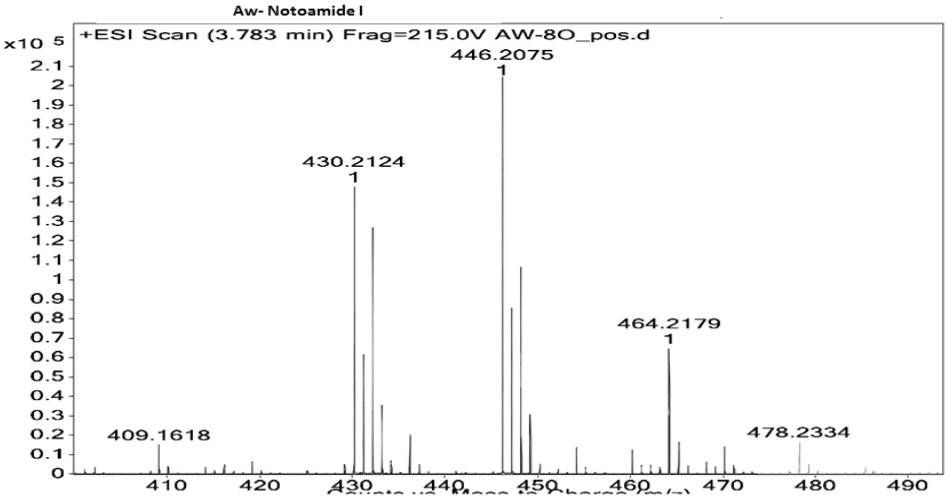
Aw- Neoaspergillic acid



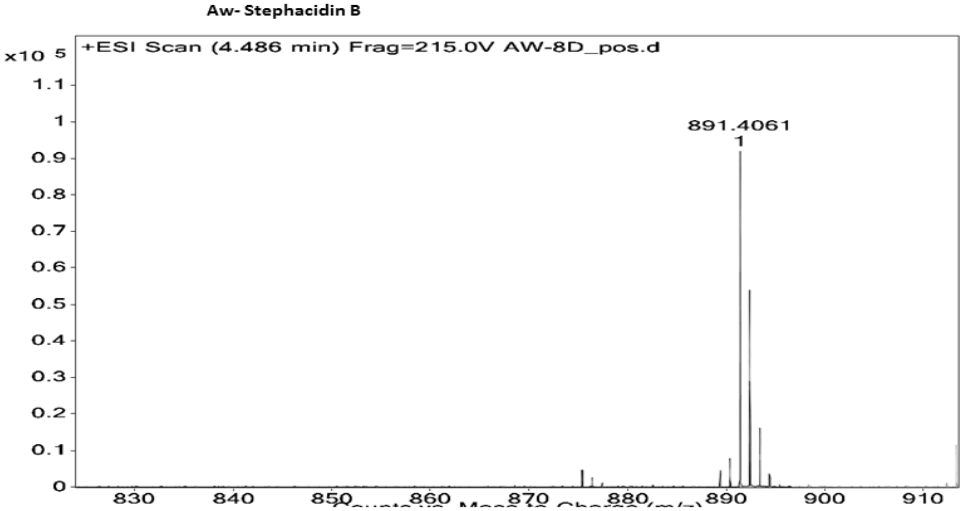
Aw- Neohydroxyaspergillic acid



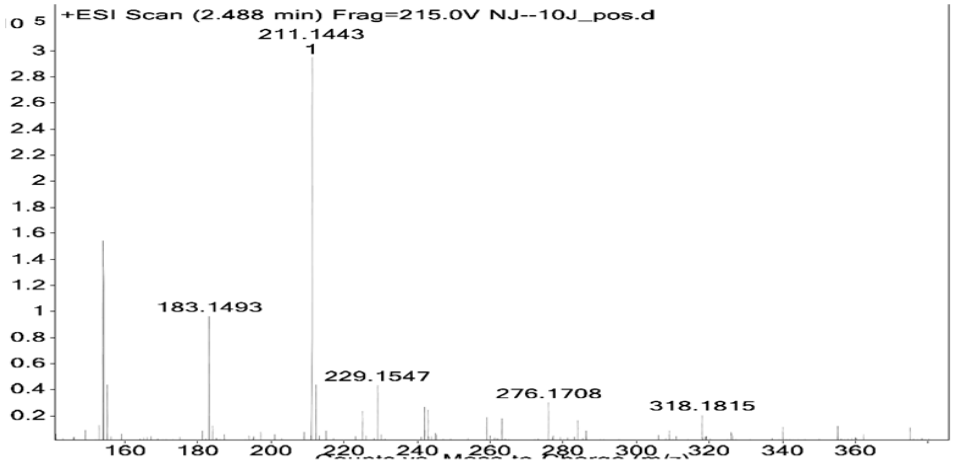
Aw- Notoamide I



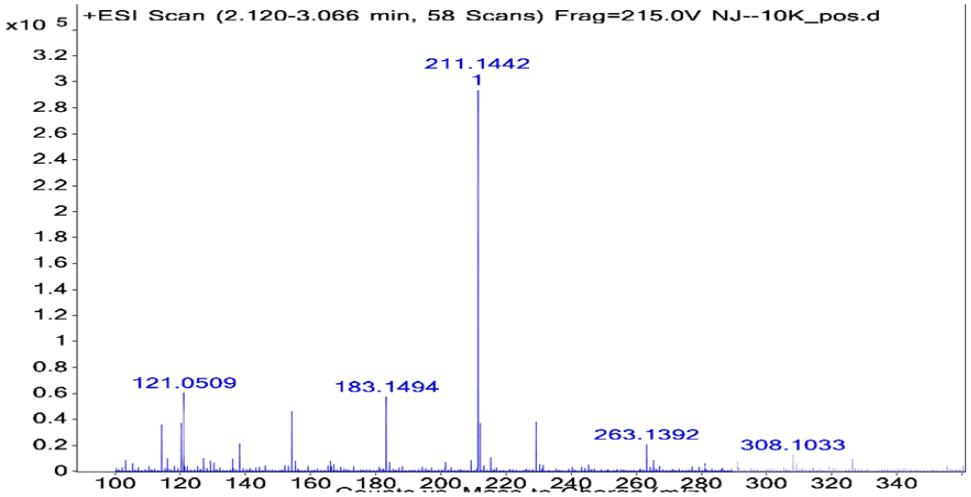
Aw- Stephacidin B



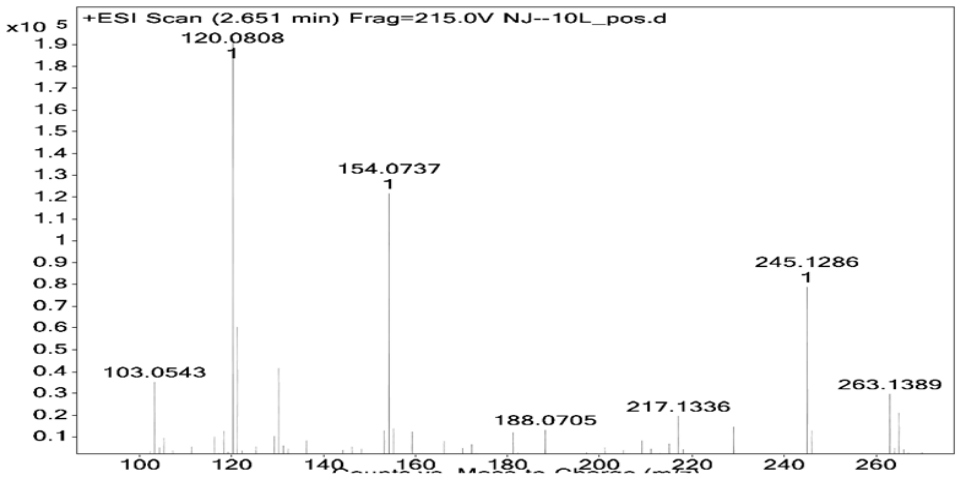
Nj – Cyclo(isoleucyl-propyl)



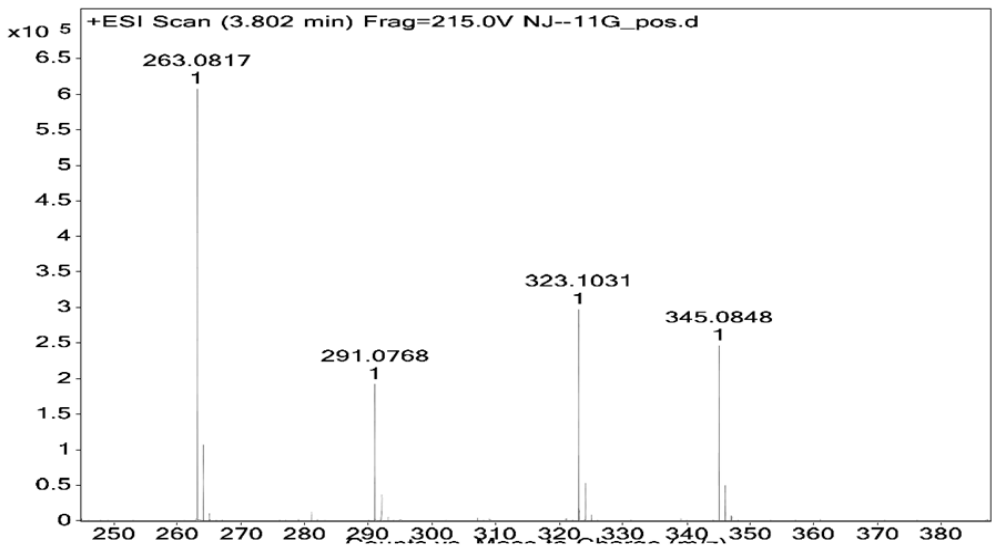
Nj – Cyclo(leucyl-propyl)



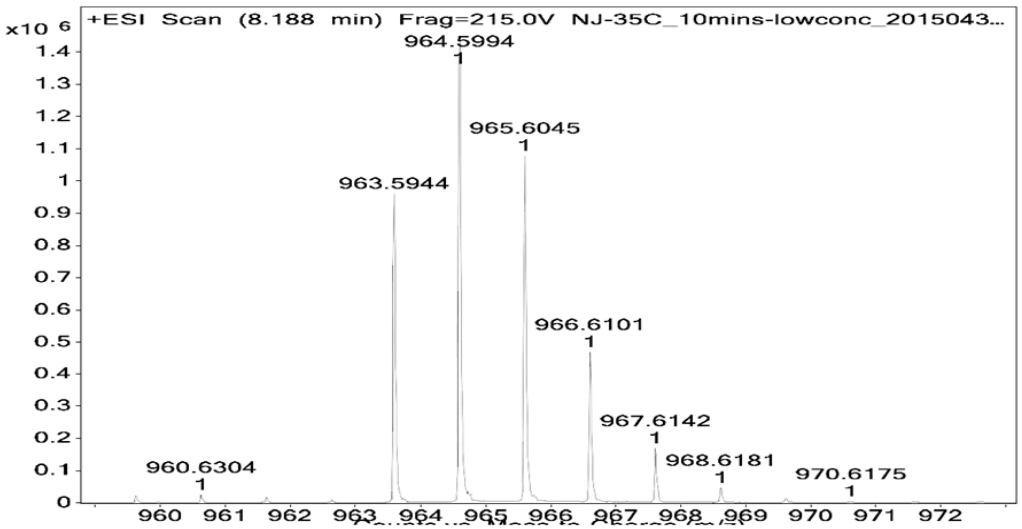
Nj – Cyclo(Phenylalanyl-prolyl)



Nj –Flazine methyl ester



Nj –unidentified peptide



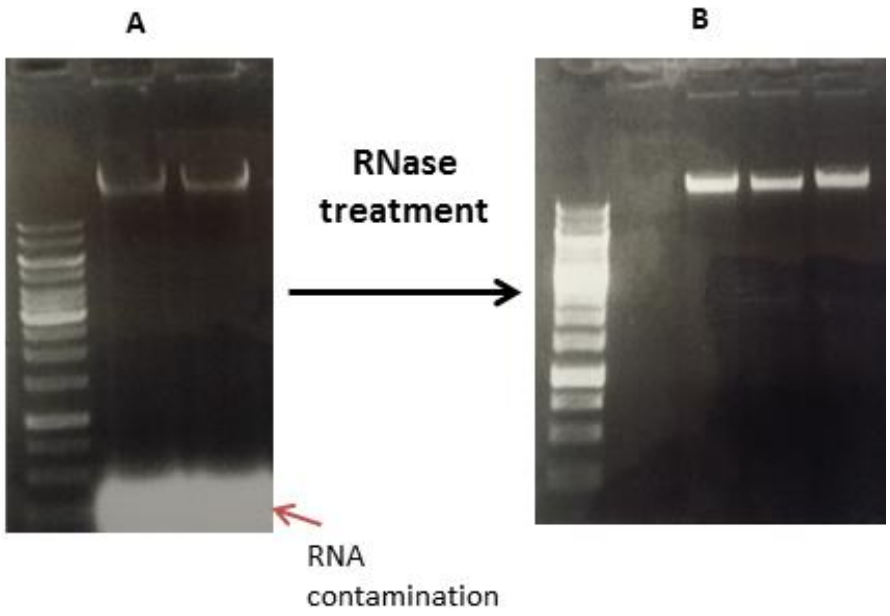
Appendix XII: Comparison of OTA biosynthesis cluster across *Aspergilli*

A.niger CBS 513.88	*A.carbonarius ITEM5010*	*A.westerdijkae CBS112803*
	Acar5010_173470 fungal zinc cluster transcription factors	Putative dehydrogense
	Acar5010_053516 Acid phosphatase	putative Sorbitol dehydrogenase
	Acar5010_007814 Nucleoside/nucleotide kinase (NK) superfamily	Protein Kinase superfamily
	Acar5010_517145 Tartrate dehydrogenase	putative transcription factor (zinc finger)
An15g07930 Putative oxidoreductase	Acar5010_517146 Oxidoreductase/phenylcoumaran benzylic ether reductase	orAwota Putative amino acid oxidoreductase
An15g07920 Polyketide synthase	Acar5010_173482 Polyketide synthase	pksAwota Polyketide synthase
An15g07910 Nonribosomal peptide synthetase	Acar5010_132610 Nonribosomal peptide Synthetase	nrpsAwota Nonribosomal peptide synthetase
An15g07900 Cytochrome p450	Acar5010_517149 Cytochrome p450	cpAwota Cytochrome p450
An15g07890 bZIP Transcription factor	Acar5010_007821 bZIP Transcription factor	tfAwota bZIP Transcription factor
An15g07880 Halogenase	Acar5010 209543 Halogenase	hlgAwota Halogenase
An15g07870 Putative oxidoreductase / dehydrogenase	Acar5010_007823 HP/putative aspartic peptidase/ pepsin-like aspartate proteases	hydAwota Abhydrolase_4 superfamily
An15g07860 Putative oxidoreductase	Acar5010_132614 Putative zinc cluster TF	Hypothetical Protein
An15g07850 Uncharacterized	Acar5010_209555 Aconitase family protein	Glutathione-dependent formaldehyde-activating enzyme domain containing protein
An15g07840 Uncharacterized	Acar5010_209556 Oxaloacetate hydrolase	Putative sugar transporter
An15g07830 Predicted hydrolase activity/metallo beta lactamase	Acar5010_209557 Hypothetical Protein	

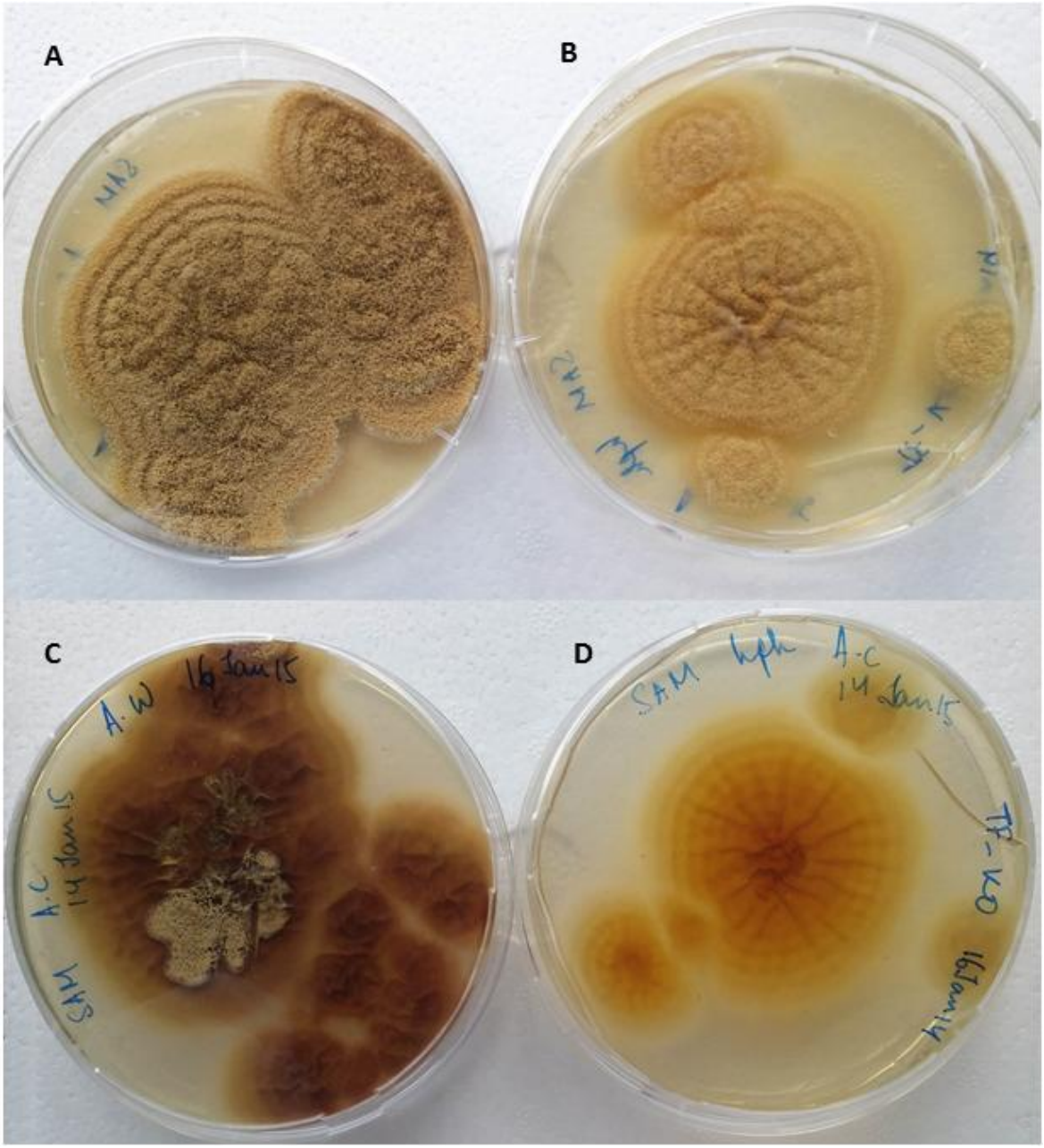
Appendix XIII: List of genes in the OTA biosynthesis cluster of *A. westerdijkiae*

Gene	Function	Nearest Homology
Dehydrogenase	Putative dehydrogenase	Oxidoreductase, short chain dehydrogenase/reductase family - *Aspergillus flavus* NRRL3357 [XP_002385119.1] - (71% identity)
Sorbitol dehydrogenase	Putative sorbitol dehydrogenase	Pc16g10990 - *Penicillium rubens* Wisconsin 54-1255 XP_002561405.1 (81% identity)
Protein kinase	Protein kinase superfamily	CMGC/SRPK protein kinase - *Blastomyces dermatitidis* ATCC 26199 [EQL29799.1] - (52% identity)
Transcription factor	Putative transcription factor (zinc cluster)	Hypothetical protein AN2786.2 - *Aspergillus nidulans* FGSC A4 [XP_660389.1] - (52% identity)
orAwota	Putative amino acid oxidoreductase	Hypothetical protein AN2786.2 - *Aspergillus nidulans* FGSC A4 [XP_660390.1] - (67% identity)
pksAwota	Polyketide synthase	Hypothetical protein - ANI_1_1836134 *Aspergillus niger* CBS 513.88 (An15g07920) [XP_001397313.2] - (62% identity) PKS LC35-12 - *Aspergillus ochraceus* [AAT92023.1] - (99% identity)
nrpsAwota	Nonribosomal peptide synthetase	Hypothetical protein - ANI_1_1832134 *Aspergillus niger* CBS 513.88 (An15g07910) [XP_001397312.2] - (54% identity)
cpAwota	Cytochrome p450	Cytochrome P450 - *Aspergillus niger* CBS 513.88 (An15g07900) [XP_001397311.1] - (63% identity)
tfAwota	bZIP Transcription factor	Un-named protein product - *Aspergillus niger* (An15g07890) [CAK42676.1] - (53% identity)
hlgAwota	Halogenase	radH flavin-dependent halogenase - *Aspergillus niger* CBS 513.88 (An15g07880) [XP_001397309.2] - (75% identity)
hydAwota	Abhydrolase_4 superfamily	Predicted protein - *Aspergillus terreus* NIH2624 [XP_001215315.1] - (57% identity)
Hypothetical protein	Hypothetical protein	Hypothetical protein - *Neosartorya fischeri* NRRL 181 [XP_001259653.1] - (40% identity)
Glutathione-dependent formaldehyde-activating	Glutathione-dependent formaldehyde-activating enzyme domain containing protein	DUF636 domain protein - *Aspergillus flavus* NRRL3357 [XP_002381437.1] - (67% identity)
Sugar transporter	Putative sugar transporter	Putative transporter - *Aspergillus oryzae* 3.042 [EIT79634.1] - (68% identity)

Appendix XIV: *A. westerdijkiae* genomic DNA gel picture and close-up view of of wild type vs Δ*tf*Awota colonies.

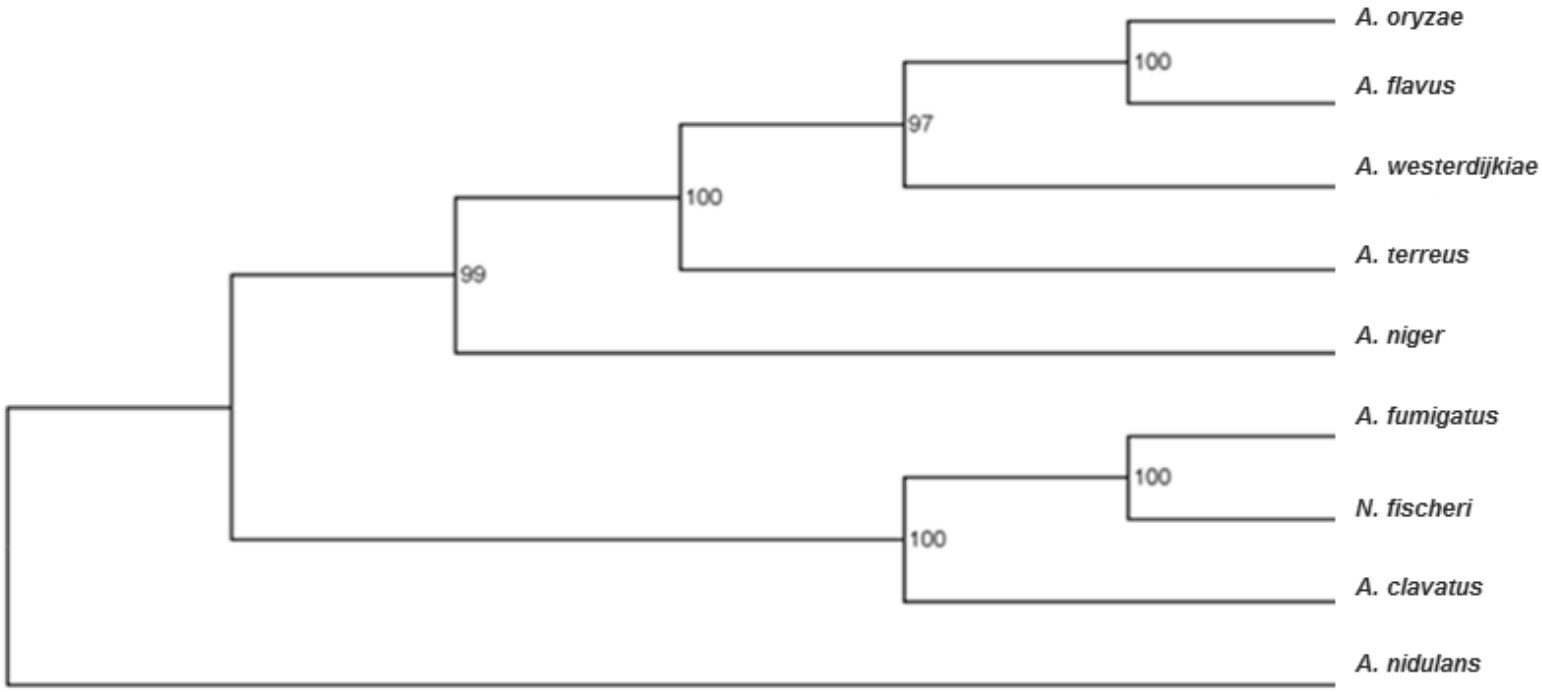


The genomic DNA isolated using the protocols described in Chapter 2 (Materials and further subjected to QC analysis by the sequencing company before it was accepted for sequencing.Method section) was subjected to multiple RNase treatment until all the RNA contaminant was successfully removed. The DNA isolated did not show a visible smear against a 1Kb ladder and the concentration was measured to 1000ng/µl. This DNA sample was



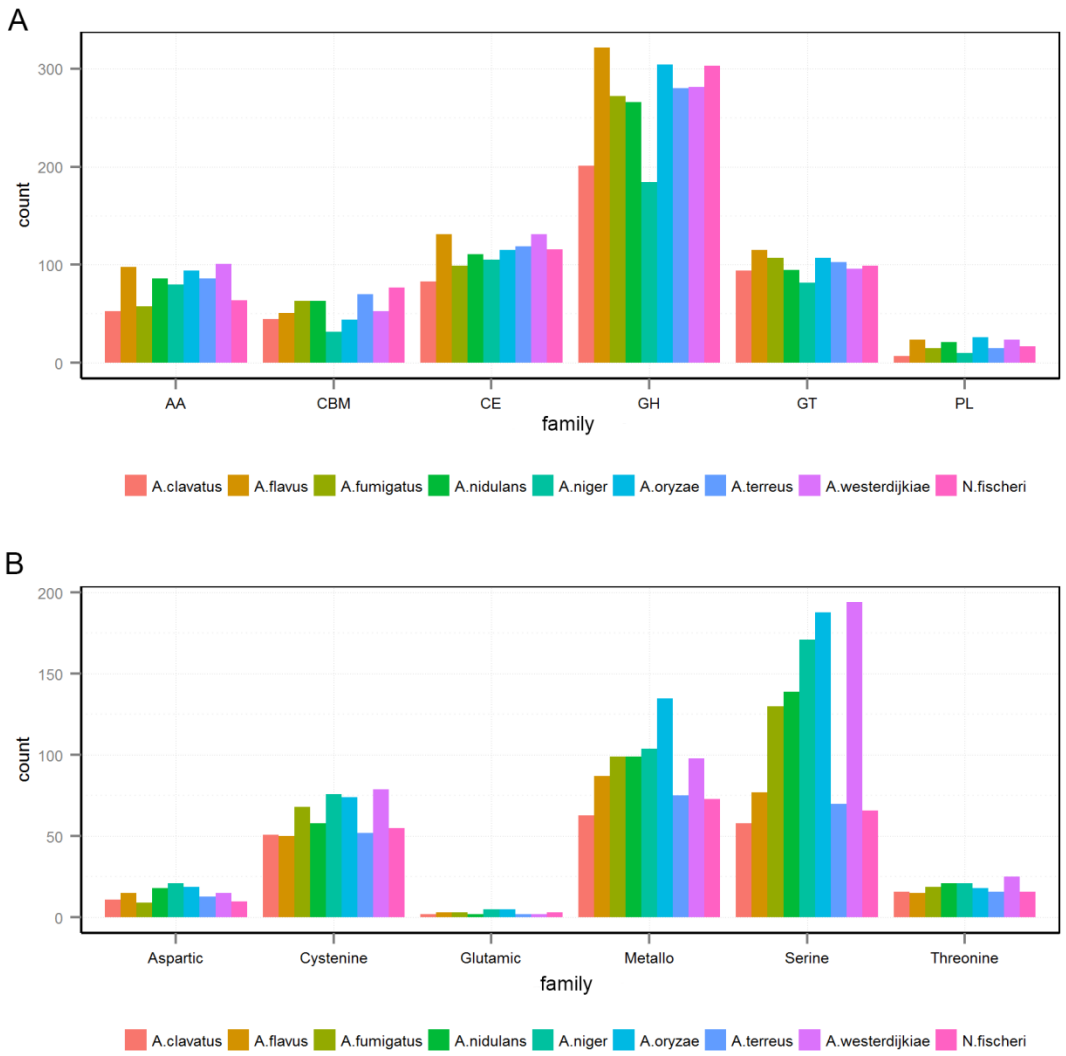
Close-up view of of wild type vs Δ*tf*Awota colonies: Δ*tf*Awota shows paler appearance and less spore production (panel B and D) when compared to the wild type (panel A and C).

Appendix XV: Phylogenetic relationship of *A. westerdijkiae* with other reported species of Aspergillus.



Phylogenetic relationship of *A. westerdijkiae*. The amino acid sequences of predicted proteins were used to build the phylogenetic tree and performing blast search. (A) A maximum likelihood phylogenomic tree was constructed on the basis of 561 concatenated orthologous single-copy protein-coding gene of nine different members of the genus *Aspergillus* (*A. oryzae, A. flavus, A. westerdijkiae, A. terreus, A. niger, A. fumigatus, N. fischeri, A. clavatus, A. nidulans*) using the Dayhoff model in the TREE-PUZZLE.

Appendix XVI: Comparison of enzyme families (CAZymes and proteases)



Comparison of enzyme families among multiple species (*A. oryzae*, *A. flavus*, *A. westerdijkiae*, *A. terreus*, *A. niger*, *A. fumigatus*, *N. fischeri*, *A. clavatus*, *A. nidulans*). (A) Bar-plot distribution of carbohydrate-active enzymes. The horizontal axis indicates the family name of CAZyme. Each bar indicates the count of the proteins matched to a given family, where different aspergilli species are shown by different colors. Enzyme families of all species were predicted using HMMER search against CAZy database and represented by their classes (GH:Glycoside Hydrolase, GT: GlycosylTransferase, PL: Polysaccharide Lyase, CE: Carbohydrate Esterase, AA: Auxiliary Activity, CBM: Carbohydrate-Binding Module). (B) Bar-plot distribution of proteases. The horizontal axis indicates the family name from MEROPS database. Each bar indicates the count of the proteins matched to a given family, and different aspergilli species are shown by different colors.

CAZyme list in *A. westerdijkiae*

[The numbers assigned to the putative CAZymes in *A. westerdijkiae* in this list; have been given for our reference. Since the genome sequence we submitted to NCBI is unannotated, we do not have the accession numbers of these proteins yet. The accession number of CAZy related proteins of other Aspergilli can be found in www.cazy.org.]

Protein	Signal peptide	Transmembrane	CAZyme family
awe00030	SP='NO'	TM="O"	AA1
awe00052	SP='NO'	TM="M"	GT22
awe00054	SP='NO'	TM="O"	AA2
awe00080	SP='YES'	TM="O"	GH10
awe00099	SP='NO'	TM="M"	GH47
awe00104	SP='NO'	TM="O"	CE4
awe00131	SP='NO'	TM="M"	GH71
awe00161	SP='YES'	TM="O"	PL1
awe00169	SP='YES'	TM="O"	CBM20 ,GH15
awe00170	SP='NO'	TM="M"	GT31
awe00174	SP='YES'	TM="O"	GH31
awe00177	SP='YES'	TM="O"	GH13
awe00186	SP='NO'	TM="M"	GT54
awe00215	SP='NO'	TM="M"	GT34
awe00217	SP='NO'	TM="M"	CE16
awe00274	SP='NO'	TM="M"	GT76
awe00376	SP='YES'	TM="O"	GH5
awe00412	SP='NO'	TM="O"	CE10
awe00420	SP='YES'	TM="O"	GH81
awe00496	SP='NO'	TM="O"	GH109
awe00500	SP='NO'	TM="M"	GH5
awe00501	SP='YES'	TM="O"	CE10
awe00558	SP='YES'	TM="O"	GH3
awe00578	SP='YES'	TM="O"	GH2
awe00588	SP='YES'	TM="O"	GH128
awe00595	SP='YES'	TM="O"	GH16
awe00598	SP='YES'	TM="O"	CE2
awe00659	SP='YES'	TM="O"	GH28
awe00712	SP='NO'	TM="O"	GH109
awe00743	SP='YES'	TM="O"	GH43
awe00749	SP='YES'	TM="O"	GH93
awe00780	SP='YES'	TM="O"	CE5
awe00793	SP='NO'	TM="O"	CE7
awe00799	SP='YES'	TM="O"	GH15
awe00806	SP='YES'	TM="O"	GH71
awe00815	SP='NO'	TM="M"	GH76
awe00832	SP='NO'	TM="O"	CE10
awe00858	SP='YES'	TM="O"	GH32
awe00877	SP='YES'	TM="O"	GH2

awe00879	SP='NO'	TM="M"	GT2
awe00882	SP='YES'	TM="O"	GH43
awe00883	SP='YES'	TM="O"	CE10
awe00885	SP='YES'	TM="O"	GH13
awe00894	SP='NO'	TM="M"	GT2
awe00914	SP='NO'	TM="O"	GH76
awe00941	SP='NO'	TM="O"	CBM21
awe00958	SP='NO'	TM="M"	GT54
awe00971	SP='YES'	TM="O"	AA7
awe00994	SP='NO'	TM="M"	GH43
awe00999	SP='NO'	TM="O"	GT32
awe01025	SP='NO'	TM="O"	CE10
awe01036	SP='NO'	TM="O"	CE10
awe01038	SP='YES'	TM="O"	CE10
awe01051	SP='NO'	TM="M"	GH18
awe01104	SP='YES'	TM="O"	GH43
awe01112	SP='NO'	TM="M"	GH127
awe01115	SP='YES'	TM="O"	GH2
awe01152	SP='YES'	TM="O"	GH18
awe01154	SP='YES'	TM="O"	GH88
awe01171	SP='YES'	TM="O"	CE1
awe01181	SP='YES'	TM="O"	CBM18 ,GH18
awe01186	SP='NO'	TM="O"	CBM50
awe01202	SP='YES'	TM="O"	AA7
awe01205	SP='YES'	TM="O"	AA9
awe01218	SP='YES'	TM="O"	GH16
awe01223	SP='YES'	TM="O"	AA5 ,CBM32
awe01228	SP='YES'	TM="O"	GH25
awe01230	SP='NO'	TM="O"	GH13
awe01234	SP='YES'	TM="O"	CE5
awe01235	SP='NO'	TM="O"	GH5
awe01236	SP='YES'	TM="O"	GH2
awe01272	SP='NO'	TM="O"	GH5
awe01278	SP='NO'	TM="M"	GT8
awe01288	SP='NO'	TM="M"	GT69
awe01298	SP='NO'	TM="M"	GH109
awe01303	SP='YES'	TM="O"	GH43
awe01305	SP='NO'	TM="O"	GT69
awe01354	SP='YES'	TM="O"	CBM63
awe01357	SP='NO'	TM="M"	GT2
awe01383	SP='NO'	TM="O"	GH109
awe01426	SP='NO'	TM="M"	GT62
awe01451	SP='NO'	TM="O"	CE10
awe01488	SP='YES'	TM="O"	CE5
awe01492	SP='NO'	TM="O"	GT1
awe01497	SP='YES'	TM="O"	GH18

awe01510	SP='YES'	TM="O"	AA7
awe01513	SP='YES'	TM="O"	AA7
awe01515	SP='NO'	TM="O"	GH13
awe01522	SP='NO'	TM="O"	GH2
awe01554	SP='YES'	TM="O"	CBM35 ,GH43
awe01568	SP='YES'	TM="O"	GH71
awe01583	SP='YES'	TM="O"	GH28
awe01589	SP='YES'	TM="O"	GH3
awe01659	SP='YES'	TM="O"	CBM50
awe01662	SP='NO'	TM="O"	CE10
awe01679	SP='YES'	TM="O"	GH2
awe01700	SP='YES'	TM="O"	CE3
awe01723	SP='YES'	TM="M"	GH78
awe01730	SP='NO'	TM="O"	GT1
awe01731	SP='NO'	TM="O"	CE10
awe01734	SP='YES'	TM="O"	GH79
awe01736	SP='NO'	TM="O"	CBM67 ,GH78
awe01743	SP='YES'	TM="O"	CE10
awe01756	SP='YES'	TM="M"	CBM13 ,GH27
awe01758	SP='YES'	TM="O"	AA3
awe01785	SP='YES'	TM="O"	CBM20 ,GH13
awe01787	SP='YES'	TM="O"	GH92
awe01794	SP='NO'	TM="O"	CE10 ,CE7
awe01853	SP='NO'	TM="O"	GT20
awe01899	SP='NO'	TM="O"	AA2
awe01902	SP='YES'	TM="O"	CBM50
awe01906	SP='YES'	TM="O"	AA5
awe01909	SP='NO'	TM="O"	GH31
awe01916	SP='YES'	TM="O"	CBM18 ,GH18
awe01933	SP='YES'	TM="O"	GH11
awe01949	SP='NO'	TM="O"	AA3
awe01957	SP='NO'	TM="M"	GT71
awe01968	SP='YES'	TM="O"	GH28
awe01970	SP='YES'	TM="O"	CBM1 ,GH5 ,GH7
awe02033	SP='NO'	TM="O"	CBM67 ,GH78
awe02034	SP='NO'	TM="O"	AA3
awe02041	SP='NO'	TM="O"	GH76
awe02051	SP='NO'	TM="O"	GH1
awe02053	SP='NO'	TM="O"	GH3
awe02065	SP='NO'	TM="M"	GH2
awe02066	SP='YES'	TM="O"	AA3
awe02085	SP='YES'	TM="O"	GH75
awe02112	SP='NO'	TM="O"	AA3
awe02114	SP='NO'	TM="O"	AA3
awe02116	SP='NO'	TM="O"	AA7
awe02135	SP='YES'	TM="O"	AA3

awe02182	SP='YES'	TM="O"	CE10
awe02243	SP='YES'	TM="O"	GH12
awe02314	SP='NO'	TM="O"	GT8 ,GT90
awe02337	SP='YES'	TM="O"	GH3
awe02390	SP='NO'	TM="M"	AA4
awe02391	SP='NO'	TM="M"	GH47
awe02403	SP='NO'	TM="O"	GT21
awe02429	SP='YES'	TM="O"	PL3
awe02435	SP='YES'	TM="O"	CE10
awe02461	SP='YES'	TM="O"	CE1
awe02472	SP='NO'	TM="O"	AA7
awe02474	SP='NO'	TM="O"	AA3
awe02476	SP='YES'	TM="O"	CBM43 ,GH72
awe02525	SP='NO'	TM="M"	GT69
awe02537	SP='NO'	TM="O"	CE1
awe02542	SP='YES'	TM="O"	AA7
awe02545	SP='NO'	TM="M"	CE10 ,PL22
awe02548	SP='NO'	TM="O"	AA7
awe02565	SP='NO'	TM="O"	GH3
awe02592	SP='YES'	TM="O"	GH16
awe02623	SP='NO'	TM="O"	AA7
awe02707	SP='YES'	TM="O"	GH5
awe02709	SP='YES'	TM="O"	CE1
awe02718	SP='NO'	TM="O"	CE1
awe02733	SP='YES'	TM="O"	GH24
awe02744	SP='NO'	TM="O"	AA7
awe02753	SP='YES'	TM="O"	CE1
awe02758	SP='NO'	TM="O"	CE12
awe02768	SP='NO'	TM="O"	CE10
awe02775	SP='NO'	TM="M"	GT15
awe02805	SP='YES'	TM="O"	GH92
awe02806	SP='YES'	TM="O"	GH76
awe02809	SP='YES'	TM="O"	CBM48 ,GH13 ,GH28
awe02826	SP='YES'	TM="O"	PL1
awe02829	SP='YES'	TM="M"	PL1
awe02852	SP='NO'	TM="O"	CE10
awe02865	SP='NO'	TM="O"	GT34
awe02883	SP='YES'	TM="O"	GH1
awe02887	SP='NO'	TM="O"	GH109
awe02901	SP='YES'	TM="O"	GH65
awe02904	SP='YES'	TM="O"	CBM1 ,GH6
awe02922	SP='NO'	TM="O"	AA7
awe02927	SP='YES'	TM="O"	GH11
awe02935	SP='NO'	TM="O"	GH105
awe02973	SP='NO'	TM="O"	AA3
awe02991	SP='NO'	TM="O"	GT20

awe02996	SP='YES'	TM="O"	CE5
awe03015	SP='YES'	TM="O"	CE2
awe03026	SP='NO'	TM="O"	CBM22
awe03040	SP='YES'	TM="O"	GH55
awe03060	SP='YES'	TM="O"	GH27
awe03069	SP='NO'	TM="O"	GH43
awe03078	SP='YES'	TM="O"	GH79
awe03086	SP='NO'	TM="O"	CE10
awe03096	SP='NO'	TM="O"	GT41
awe03100	SP='NO'	TM="M"	GT66
awe03113	SP='NO'	TM="M"	GT39
awe03130	SP='NO'	TM="O"	CE9
awe03147	SP='YES'	TM="O"	CE10
awe03148	SP='NO'	TM="O"	AA3
awe03229	SP='NO'	TM="O"	GH3
awe03245	SP='YES'	TM="O"	CE10
awe03256	SP='YES'	TM="O"	CE4
awe03257	SP='NO'	TM="O"	GH3
awe03266	SP='NO'	TM="M"	GH18
awe03280	SP='YES'	TM="O"	AA3
awe03287	SP='NO'	TM="M"	GT32
awe03326	SP='YES'	TM="O"	AA9
awe03331	SP='YES'	TM="O"	GH62
awe03337	SP='YES'	TM="O"	GH28
awe03365	SP='NO'	TM="O"	AA7
awe03406	SP='NO'	TM="M"	GT39
awe03470	SP='YES'	TM="O"	GH18
awe03484	SP='YES'	TM="O"	PL3
awe03500	SP='NO'	TM="O"	AA3
awe03518	SP='NO'	TM="M"	GT22
awe03522	SP='YES'	TM="O"	CE10
awe03537	SP='YES'	TM="O"	GH132
awe03581	SP='NO'	TM="O"	CE10
awe03601	SP='YES'	TM="O"	GH20
awe03615	SP='NO'	TM="O"	GT35
awe03650	SP='YES'	TM="O"	AA9
awe03654	SP='NO'	TM="M"	GT2
awe03789	SP='NO'	TM="M"	GT71
awe03797	SP='NO'	TM="O"	GH17
awe03828	SP='NO'	TM="O"	AA3
awe03851	SP='YES'	TM="O"	GH92
awe03928	SP='NO'	TM="O"	CE10
awe03937	SP='NO'	TM="O"	GH51
awe03955	SP='YES'	TM="O"	GH6
awe03976	SP='YES'	TM="O"	AA9 ,CBM1
awe04019	SP='YES'	TM="O"	GH5

awe04092	SP='YES'	TM="O"	AA3
awe04109	SP='YES'	TM="O"	GH74 ,GH93
awe04120	SP='YES'	TM="O"	PL1
awe04152	SP='NO'	TM="O"	CBM67 ,GH78
awe04155	SP='YES'	TM="O"	CE10
awe04165	SP='YES'	TM="O"	GH43
awe04170	SP='YES'	TM="O"	GH42
awe04171	SP='YES'	TM="O"	AA9 ,CBM1
awe04173	SP='YES'	TM="O"	AA9 ,CBM1 ,GH10
awe04175	SP='YES'	TM="O"	GH43
awe04179	SP='NO'	TM="O"	GT1
awe04180	SP='NO'	TM="O"	GH71
awe04219	SP='YES'	TM="O"	CBM1 ,GH5
awe04220	SP='NO'	TM="O"	CE1
awe04228	SP='YES'	TM="O"	GH47
awe04258	SP='NO'	TM="O"	GH1
awe04274	SP='NO'	TM="O"	GH13
awe04283	SP='NO'	TM="O"	CE4
awe04291	SP='NO'	TM="O"	AA7
awe04308	SP='NO'	TM="O"	AA3
awe04310	SP='NO'	TM="O"	GH3
awe04315	SP='YES'	TM="O"	GH55
awe04364	SP='NO'	TM="M"	CE10
awe04379	SP='NO'	TM="M"	GT69
awe04380	SP='YES'	TM="M"	CBM37 ,GH35
awe04388	SP='NO'	TM="O"	AA7
awe04407	SP='YES'	TM="O"	AA7
awe04414	SP='YES'	TM="O"	CE10
awe04442	SP='YES'	TM="O"	AA7
awe04448	SP='YES'	TM="O"	GH43
awe04463	SP='YES'	TM="O"	GH3
awe04465	SP='YES'	TM="O"	AA7
awe04466	SP='NO'	TM="O"	AA7
awe04488	SP='NO'	TM="O"	CE10
awe04517	SP='NO'	TM="M"	GH74
awe04520	SP='YES'	TM="O"	GH105
awe04522	SP='YES'	TM="O"	CBM20 ,CE10 ,GH15
awe04527	SP='YES'	TM="O"	GH28
awe04529	SP='NO'	TM="O"	GT8
awe04563	SP='NO'	TM="M"	CBM24 ,GH71
awe04590	SP='YES'	TM="O"	CE10
awe04597	SP='YES'	TM="O"	AA7
awe04602	SP='YES'	TM="O"	AA3
awe04605	SP='NO'	TM="O"	CE15
awe04611	SP='YES'	TM="M"	PL4
awe04638	SP='NO'	TM="M"	GT33

awe04639	SP='YES'	TM="O"	AA7
awe04653	SP='NO'	TM="O"	CE10
awe04654	SP='YES'	TM="O"	CE10
awe04657	SP='YES'	TM="O"	GH92
awe04664	SP='NO'	TM="O"	CE10
awe04725	SP='NO'	TM="O"	AA2
awe04740	SP='YES'	TM="O"	CBM1 ,GH10
awe04744	SP='NO'	TM="O"	CE10
awe04746	SP='NO'	TM="O"	AA6
awe04806	SP='NO'	TM="O"	GT20
awe04841	SP='NO'	TM="M"	GT48
awe04847	SP='NO'	TM="O"	GH17
awe04857	SP='YES'	TM="M"	GH47
awe04861	SP='YES'	TM="O"	CBM43 ,GH72
awe04884	SP='NO'	TM="O"	AA7
awe04886	SP='NO'	TM="O"	CE10
awe04900	SP='NO'	TM="O"	GH37
awe04909	SP='NO'	TM="O"	GT28
awe04965	SP='NO'	TM="M"	GT22
awe04966	SP='YES'	TM="M"	CBM43 ,GH72
awe04987	SP='YES'	TM="O"	PL22
awe04989	SP='NO'	TM="O"	GH31
awe05056	SP='NO'	TM="M"	GT69
awe05091	SP='YES'	TM="O"	GH3
awe05096	SP='YES'	TM="O"	PL1
awe05098	SP='NO'	TM="O"	PL1
awe05109	SP='NO'	TM="O"	GH2
awe05125	SP='YES'	TM="O"	GH3
awe05142	SP='YES'	TM="O"	GH5
awe05144	SP='NO'	TM="M"	AA8
awe05147	SP='NO'	TM="O"	GT1
awe05159	SP='YES'	TM="O"	PL4
awe05162	SP='YES'	TM="O"	GH5
awe05164	SP='NO'	TM="M"	GT90
awe05185	SP='YES'	TM="M"	CE12
awe05187	SP='YES'	TM="O"	GH28
awe05231	SP='NO'	TM="M"	GT57
awe05235	SP='NO'	TM="O"	CE1
awe05244	SP='YES'	TM="O"	GH78
awe05247	SP='YES'	TM="O"	GH125
awe05267	SP='NO'	TM="O"	GH18
awe05289	SP='YES'	TM="O"	AA7
awe05309	SP='YES'	TM="O"	CBM24 ,GH71
awe05323	SP='NO'	TM="O"	GH109
awe05330	SP='YES'	TM="O"	AA7
awe05333	SP='NO'	TM="O"	GT26

awe05352	SP='NO'	TM="M"	GT62
awe05399	SP='NO'	TM="O"	CE10
awe05401	SP='YES'	TM="O"	GH51
awe05403	SP='NO'	TM="M"	GH131
awe05434	SP='NO'	TM="M"	GT34
awe05439	SP='NO'	TM="O"	AA3
awe05453	SP='YES'	TM="O"	GH43
awe05460	SP='NO'	TM="M"	GT90
awe05467	SP='YES'	TM="O"	CE5
awe05483	SP='YES'	TM="M"	CE3
awe05497	SP='YES'	TM="O"	GH95
awe05515	SP='NO'	TM="O"	CBM46 ,GH5
awe05567	SP='YES'	TM="O"	CE5
awe05581	SP='NO'	TM="M"	CBM67
awe05592	SP='NO'	TM="O"	AA4
awe05609	SP='NO'	TM="O"	AA8
awe05653	SP='NO'	TM="O"	CE10
awe05658	SP='NO'	TM="O"	GT3
awe05660	SP='NO'	TM="O"	AA7
awe05663	SP='NO'	TM="M"	GH16
awe05681	SP='YES'	TM="O"	GH32
awe05683	SP='YES'	TM="O"	GH71
awe05691	SP='YES'	TM="O"	GH12
awe05704	SP='YES'	TM="O"	AA8
awe05707	SP='YES'	TM="M"	GH72
awe05738	SP='NO'	TM="O"	AA7
awe05741	SP='YES'	TM="O"	CE16
awe05753	SP='YES'	TM="O"	CBM44 ,PL1
awe05761	SP='NO'	TM="O"	AA7
awe05767	SP='NO'	TM="O"	GH109
awe05772	SP='YES'	TM="O"	GH20
awe05790	SP='NO'	TM="O"	GH3
awe05799	SP='NO'	TM="O"	GH3
awe05810	SP='YES'	TM="O"	GH3
awe05833	SP='YES'	TM="O"	AA7
awe05853	SP='NO'	TM="O"	GH71
awe05854	SP='NO'	TM="O"	GH18
awe05862	SP='NO'	TM="O"	CE10
awe05864	SP='YES'	TM="O"	GH18
awe05903	SP='NO'	TM="O"	CE1
awe05950	SP='NO'	TM="O"	GT20
awe05979	SP='YES'	TM="M"	GH5
awe05997	SP='NO'	TM="O"	GT1
awe06009	SP='YES'	TM="O"	CE8
awe06027	SP='NO'	TM="O"	GH2
awe06038	SP='YES'	TM="O"	GH78

awe06039	SP='NO'	TM="M"	GT50
awe06041	SP='YES'	TM="O"	CE10
awe06051	SP='YES'	TM="O"	GH7
awe06068	SP='YES'	TM="O"	PL11
awe06085	SP='YES'	TM="O"	GH35
awe06089	SP='YES'	TM="O"	AA7
awe06108	SP='YES'	TM="O"	GH75
awe06113	SP='NO'	TM="M"	PL4
awe06119	SP='YES'	TM="O"	AA1
awe06140	SP='NO'	TM="O"	GH109
awe06159	SP='NO'	TM="O"	CE10
awe06161	SP='YES'	TM="O"	AA7
awe06182	SP='NO'	TM="O"	CE10
awe06194	SP='YES'	TM="O"	AA3
awe06235	SP='NO'	TM="O"	GH47
awe06264	SP='YES'	TM="O"	GH53
awe06283	SP='NO'	TM="O"	CBM67 ,GH78
awe06284	SP='NO'	TM="M"	GT4
awe06300	SP='YES'	TM="O"	GH78
awe06301	SP='NO'	TM="O"	GH31
awe06343	SP='NO'	TM="O"	AA3
awe06361	SP='NO'	TM="O"	CE10
awe06364	SP='YES'	TM="O"	GH3
awe06368	SP='YES'	TM="O"	AA7
awe06397	SP='YES'	TM="O"	GH115
awe06398	SP='YES'	TM="O"	PL4
awe06420	SP='NO'	TM="M"	CE10
awe06456	SP='NO'	TM="O"	CBM18 ,CE4
awe06457	SP='NO'	TM="M"	CE10
awe06486	SP='NO'	TM="O"	GH109
awe06509	SP='NO'	TM="O"	GT34
awe06515	SP='NO'	TM="M"	GT71
awe06552	SP='YES'	TM="O"	AA7
awe06555	SP='YES'	TM="O"	GH26
awe06575	SP='NO'	TM="O"	AA3
awe06647	SP='YES'	TM="O"	CE1
awe06653	SP='NO'	TM="O"	GH63
awe06660	SP='YES'	TM="O"	AA7
awe06669	SP='NO'	TM="O"	AA4
awe06737	SP='NO'	TM="O"	GT1
awe06770	SP='YES'	TM="M"	GH13 ,GT5
awe06776	SP='NO'	TM="M"	GT59
awe06798	SP='YES'	TM="O"	AA9
awe06824	SP='YES'	TM="O"	AA7
awe06833	SP='YES'	TM="O"	AA7
awe06835	SP='NO'	TM="M"	GH43

awe06843	SP='YES'	TM="O"	CE8
awe06891	SP='YES'	TM="O"	GH63
awe06916	SP='YES'	TM="O"	GH75
awe06948	SP='YES'	TM="O"	GH31
awe06951	SP='NO'	TM="O"	AA7
awe06970	SP='NO'	TM="M"	GT32
awe06974	SP='YES'	TM="O"	GH35
awe06986	SP='NO'	TM="O"	AA3
awe06989	SP='NO'	TM="M"	GT4
awe07000	SP='YES'	TM="O"	GH3
awe07007	SP='NO'	TM="O"	GH29
awe07036	SP='YES'	TM="O"	AA9 ,CBM19
awe07058	SP='NO'	TM="O"	AA3
awe07071	SP='YES'	TM="O"	CBM42 ,GH54
awe07072	SP='NO'	TM="O"	GH3
awe07077	SP='YES'	TM="O"	AA3
awe07079	SP='YES'	TM="O"	GH72
awe07089	SP='NO'	TM="O"	CE10
awe07094	SP='NO'	TM="O"	GH109
awe07116	SP='NO'	TM="M"	CE14
awe07166	SP='NO'	TM="M"	GT71
awe07189	SP='NO'	TM="O"	CBM50
awe07206	SP='NO'	TM="O"	GT4
awe07258	SP='YES'	TM="O"	GH106
awe07267	SP='YES'	TM="O"	CE10
awe07278	SP='YES'	TM="O"	GH16
awe07279	SP='YES'	TM="O"	GH43
awe07288	SP='YES'	TM="O"	GH67
awe07294	SP='NO'	TM="O"	GH88
awe07309	SP='NO'	TM="O"	GH31
awe07321	SP='NO'	TM="O"	CE1
awe07399	SP='NO'	TM="M"	GT21
awe07415	SP='YES'	TM="M"	GH115
awe07434	SP='NO'	TM="O"	CE1
awe07450	SP='NO'	TM="O"	GH132
awe07485	SP='NO'	TM="O"	GH5
awe07502	SP='YES'	TM="O"	AA7
awe07508	SP='NO'	TM="O"	CE16
awe07530	SP='NO'	TM="O"	GH43
awe07558	SP='YES'	TM="O"	AA7
awe07566	SP='NO'	TM="O"	GT1
awe07585	SP='YES'	TM="O"	PL3
awe07586	SP='YES'	TM="M"	GH28
awe07634	SP='NO'	TM="O"	GT2
awe07657	SP='NO'	TM="O"	CBM48
awe07666	SP='NO'	TM="O"	GH127

awe07721	SP='NO'	TM="O"	AA3
awe07726	SP='NO'	TM="O"	CE10
awe07741	SP='NO'	TM="O"	CE10
awe07768	SP='YES'	TM="O"	GH28
awe07770	SP='YES'	TM="O"	GH92
awe07775	SP='YES'	TM="O"	CE12
awe07787	SP='YES'	TM="O"	CE16
awe07808	SP='NO'	TM="M"	GH1 ,GH95
awe07867	SP='YES'	TM="O"	GH72
awe07893	SP='YES'	TM="O"	GH32
awe07908	SP='YES'	TM="O"	CE5
awe07919	SP='YES'	TM="O"	CE1
awe07920	SP='NO'	TM="O"	GH2
awe07925	SP='NO'	TM="O"	CE10
awe07930	SP='YES'	TM="O"	GH12
awe07931	SP='YES'	TM="O"	GH17
awe07933	SP='YES'	TM="O"	AA3
awe07944	SP='YES'	TM="O"	AA9
awe07962	SP='NO'	TM="M"	CE10
awe07981	SP='YES'	TM="O"	CE10
awe08051	SP='NO'	TM="O"	CE1
awe08056	SP='YES'	TM="O"	CE12
awe08079	SP='YES'	TM="O"	GH43
awe08081	SP='NO'	TM="O"	PL9
awe08085	SP='NO'	TM="O"	GT90
awe08092	SP='YES'	TM="M"	GT31
awe08101	SP='YES'	TM="O"	GH43
awe08115	SP='NO'	TM="M"	GT32
awe08141	SP='NO'	TM="O"	GH109
awe08149	SP='YES'	TM="O"	CBM6 ,GH51
awe08154	SP='NO'	TM="O"	AA7
awe08175	SP='NO'	TM="O"	AA3
awe08193	SP='NO'	TM="M"	GT90
awe08203	SP='NO'	TM="O"	CE4
awe08205	SP='YES'	TM="O"	PL1
awe08259	SP='NO'	TM="M"	GT15
awe08276	SP='NO'	TM="M"	GT58
awe08317	SP='NO'	TM="M"	CE4
awe08353	SP='NO'	TM="O"	CBM50
awe08395	SP='NO'	TM="O"	GH93
awe08405	SP='NO'	TM="O"	CE10
awe08463	SP='NO'	TM="O"	GH3
awe08466	SP='YES'	TM="O"	AA7
awe08494	SP='YES'	TM="M"	GH16
awe08497	SP='NO'	TM="O"	GT4
awe08498	SP='YES'	TM="O"	GH95

awe08500	SP='NO'	TM="M"	CE14
awe08506	SP='YES'	TM="M"	GH43
awe08508	SP='NO'	TM="M"	GT32
awe08516	SP='YES'	TM="O"	CE10
awe08518	SP='NO'	TM="O"	GH5
awe08533	SP='YES'	TM="O"	GH43
awe08563	SP='YES'	TM="O"	GH28
awe08590	SP='NO'	TM="O"	GH38
awe08600	SP='NO'	TM="O"	CE1
awe08624	SP='NO'	TM="M"	GH114
awe08634	SP='YES'	TM="O"	AA3
awe08635	SP='NO'	TM="M"	GT4
awe08636	SP='NO'	TM="M"	CE1
awe08645	SP='NO'	TM="M"	GT57
awe08655	SP='YES'	TM="O"	CE8
awe08659	SP='NO'	TM="O"	GH32
awe08675	SP='YES'	TM="M"	CE1
awe08684	SP='NO'	TM="O"	GH55
awe08690	SP='NO'	TM="O"	AA3
awe08694	SP='NO'	TM="O"	CBM50 ,GH18
awe08697	SP='YES'	TM="O"	AA7
awe08699	SP='NO'	TM="O"	GH13
awe08704	SP='NO'	TM="M"	CE5 ,GT25
awe08724	SP='YES'	TM="O"	GT32
awe08730	SP='NO'	TM="M"	GH43
awe08744	SP='NO'	TM="O"	CE10
awe08796	SP='YES'	TM="O"	CBM50
awe08814	SP='NO'	TM="M"	GT2
awe08821	SP='NO'	TM="O"	CBM50 ,GH18
awe08831	SP='NO'	TM="O"	GT1
awe08880	SP='YES'	TM="O"	GT24
awe08926	SP='YES'	TM="O"	CBM67 ,GH35
awe08934	SP='YES'	TM="O"	GH53
awe08935	SP='YES'	TM="O"	CBM18 ,GH16
awe08946	SP='YES'	TM="O"	GH76
awe08961	SP='NO'	TM="O"	GH13
awe09025	SP='NO'	TM="O"	GT20
awe09028	SP='YES'	TM="O"	GH71
awe09043	SP='NO'	TM="O"	GT20
awe09061	SP='YES'	TM="O"	GH5
awe09062	SP='NO'	TM="O"	CE16
awe09103	SP='YES'	TM="O"	GH7
awe09143	SP='NO'	TM="O"	GH109
awe09152	SP='NO'	TM="M"	GT39
awe09216	SP='YES'	TM="O"	GH28
awe09232	SP='NO'	TM="O"	AA7

awe09235	SP='YES'	TM="O"	CE12
awe09251	SP='YES'	TM="O"	PL4
awe09254	SP='YES'	TM="O"	AA3
awe09292	SP='YES'	TM="O"	AA9
awe09295	SP='NO'	TM="O"	CE10
awe09301	SP='NO'	TM="O"	CE1
awe09316	SP='NO'	TM="O"	CE10
awe09340	SP='NO'	TM="O"	CE10
awe09395	SP='NO'	TM="O"	GT62
awe09483	SP='YES'	TM="M"	GT15
awe09494	SP='YES'	TM="O"	AA9
awe09557	SP='YES'	TM="M"	GH18 ,GH31
awe09615	SP='YES'	TM="M"	GT22
awe09684	SP='NO'	TM="O"	CE1
awe09697	SP='NO'	TM="M"	GH17
awe09744	SP='NO'	TM="O"	CE10
awe09752	SP='YES'	TM="M"	GH28
awe09757	SP='NO'	TM="O"	GH18
awe09854	SP='YES'	TM="O"	GH105
awe09856	SP='YES'	TM="O"	GH51
awe09859	SP='YES'	TM="O"	GH23
awe09865	SP='YES'	TM="O"	GH79
awe09882	SP='NO'	TM="M"	GT2
awe09891	SP='NO'	TM="O"	CE10
awe09898	SP='YES'	TM="O"	AA9 ,CBM1
awe09921	SP='NO'	TM="O"	AA6
awe09924	SP='NO'	TM="O"	GH18
awe09947	SP='NO'	TM="M"	GH3
awe09957	SP='YES'	TM="O"	PL1
awe09974	SP='YES'	TM="O"	GH3
awe09983	SP='NO'	TM="O"	GH36
awe09988	SP='YES'	TM="O"	GH28
awe09992	SP='YES'	TM="M"	GH76
awe09995	SP='NO'	TM="M"	CE10
awe10002	SP='NO'	TM="O"	CE10
awe10025	SP='NO'	TM="O"	CE1
awe10043	SP='NO'	TM="O"	CE10
awe10061	SP='NO'	TM="O"	GH18
awe10080	SP='NO'	TM="M"	GT2
awe10091	SP='NO'	TM="O"	CBM50
awe10103	SP='NO'	TM="O"	CBM18 ,GH18
awe10174	SP='YES'	TM="O"	GH16
awe10183	SP='NO'	TM="O"	CE1
awe10185	SP='NO'	TM="M"	GT90
awe10191	SP='NO'	TM="O"	CBM67 ,GH78
awe10204	SP='NO'	TM="O"	GH31

awe10227	SP='NO'	TM="O"	GH18
awe10238	SP='NO'	TM="O"	GH88
awe10239	SP='NO'	TM="O"	GH28
awe10245	SP='NO'	TM="O"	AA6
awe10264	SP='YES'	TM="O"	CE10
awe10273	SP='NO'	TM="O"	CE1
awe10324	SP='YES'	TM="O"	GH76
awe10333	SP='NO'	TM="O"	CE10
awe10339	SP='YES'	TM="O"	PL1
awe10355	SP='YES'	TM="M"	AA7
awe10357	SP='NO'	TM="O"	CE10
awe10361	SP='YES'	TM="M"	GH24
awe10372	SP='YES'	TM="M"	GH76
awe10383	SP='YES'	TM="O"	GH106
awe10385	SP='YES'	TM="O"	CE16
awe10392	SP='NO'	TM="M"	CE16
awe10393	SP='YES'	TM="O"	CE16
awe10408	SP='NO'	TM="O"	GT32
awe10419	SP='NO'	TM="O"	GH3
awe10420	SP='NO'	TM="O"	AA7
awe10421	SP='YES'	TM="O"	GH76
awe10424	SP='NO'	TM="O"	AA7 ,CE1
awe10427	SP='YES'	TM="O"	PL1
awe10428	SP='YES'	TM="O"	GH3
awe10429	SP='NO'	TM="O"	CBM50
awe10451	SP='NO'	TM="M"	GT32
awe10461	SP='NO'	TM="O"	GT71
awe10487	SP='NO'	TM="M"	CE4
awe10527	SP='NO'	TM="M"	GT2
awe10534	SP='YES'	TM="O"	CE4
awe10590	SP='NO'	TM="O"	GH109
awe10627	SP='YES'	TM="O"	CE16
awe10673	SP='NO'	TM="M"	GT2
awe10728	SP='YES'	TM="O"	CE10
awe10730	SP='NO'	TM="O"	CBM67 ,GH78
awe10736	SP='YES'	TM="O"	GH78
awe10746	SP='YES'	TM="O"	PL4
awe10748	SP='NO'	TM="O"	GH36
awe10778	SP='YES'	TM="O"	GH20
awe10789	SP='YES'	TM="O"	GH28
awe10805	SP='YES'	TM="O"	GH17
awe10825	SP='NO'	TM="O"	GH5

Number of putative genes per CAZy family for nine Aspergilli genomes

Glycoside Hydrolases (GHs)	GH1	GH2	GH3	GH5	GH6	GH7	GH10	GH11	GH12	GH13	GH15	GH16	GH17	GH18	GH20	GH23	GH24	GH25	GH26	GH27	GH28	GH29	GH30	GH31	GH32	GH33	GH35	GH36	GH37	GH38	GH39	GH42	GH43	GH45	GH46	GH47	GH51	GH53	GH54
Asperdillus clavatus	3	3	12	9	2	4	2	3	3	16	6	12	5	14	2	0	0	3	0	2	3	0	2	5	1	1	2	2	1	1	0	0	13	0	1	5	2	0	1
Aspergillus flavus	3	8	24	15	1	3	4	4	5	15	3	12	4	17	3	0	2	1	1	3	21	0	0	9	4	1	7	2	1	1	0	0	20	0	0	5	4	2	1
Aspergillus fumigatus	5	6	18	13	1	4	4	3	4	15	5	14	5	18	2	1	0	3	0	5	13	0	1	6	4	1	5	2	1	1	0	0	18	1	0	5	2	1	1
Aspergillus nidulans	3	10	20	16	2	3	3	2	1	12	2	13	5	18	2	0	1	3	3	3	10	0	0	10	2	1	3	3	1	1	1	0	17	0	0	7	3	1	1
Aspergillus niger	3	3	14	5	1	0	1	2	1	11	1	8	3	10	1	0	0	0	1	4	20	1	1	4	1	1	4	1	1	0	0	0	8	0	0	4	2	1	0
Aspergillus oryzae	3	7	23	12	1	3	4	4	4	15	3	12	5	18	3	0	2	1	1	3	21	0	0	10	4	1	5	2	1	1	0	0	19	0	0	5	3	1	1
Aspergillus terreus	3	10	21	18	2	4	4	2	6	13	2	8	4	21	2	0	2	2	0	4	8	2	1	11	6	2	4	3	1	1	1	0	20	0	0	6	3	1	1
Aspergillus westerdijkiae	4	10	22	16	2	3	3	2	3	10	3	8	5	18	3	1	2	1	1	2	15	1	0	8	4	0	4	2	1	1	0	1	19	0	0	5	4	2	1
N. fischeri	5	6	19	16	2	5	4	4	5	15	5	14	5	22	2	1	0	3	0	5	13	0	1	8	4	1	5	2	1	1	0	0	20	1	0	6	2	1	1

Glycoside Hydrolases (GHs)	GH55	GH62	GH63	GH65	GH67	GH71	GH72	GH74	GH75	GH76	GH78	GH79	GH81	GH88	GH89	GH92	GH93	GH95	GH105	GH106	GH109	GH114	GH115	GH125	GH127	GH128	GH131	GH132	Total
Asperdillus clavatus	4	2	2	1	1	3	6	2	4	8	0	0	1	0	1	3	1	1	2	0	9	1	1	1	2	1	1	2	201
Aspergillus flavus	3	2	2	1	1	8	8	1	4	11	12	5	1	3	1	6	3	2	4	2	22	1	3	1	1	2	3	2	321
Aspergillus fumigatus	6	2	2	1	1	8	7	3	4	8	5	0	1	2	1	4	3	2	3	0	11	1	1	1	1	2	2	2	272
Aspergillus nidulans	6	2	2	1	1	5	5	3	2	7	9	1	1	2	0	5	2	3	4	0	12	2	1	1	1	2	1	2	266
Aspergillus niger	2	0	2	0	1	4	4	1	2	10	7	3	1	1	0	4	1	2	2	2	11	1	0	1	1	1	0	2	184
Aspergillus oryzae	3	2	2	1	1	8	7	1	4	11	9	5	1	3	1	6	3	3	4	1	19	1	4	1	0	2	1	2	304
Aspergillus terreus	4	3	2	1	2	7	6	2	2	10	4	3	1	0	0	6	4	3	2	0	7	1	2	1	3	2	1	2	280
Aspergillus westerdijkiae	3	1	2	1	1	9	6	2	3	9	11	3	1	3	0	5	3	3	3	2	13	1	2	1	2	1	1	2	281
N. fischeri	11	2	2	1	1	9	7	4	4	9	7	0	1	2	1	5	3	2	3	1	13	1	1	1	1	2	2	2	303

GlycosylTransferases (GTs)	GT1	GT2	GT3	GT4	GT5	GT8	GT15	GT20	GT21	GT22	GT24	GT25	GT26	GT28	GT31	GT32	GT33	GT34	GT35	GT39	GT41	GT45	GT48	GT50	GT51	GT54	GT57	GT58	GT59	GT61	GT62	GT66	GT69	GT71	GT76	GT78	GT90	GT91	Total
Asperdillus clavatus	3	13	1	6	3	4	3	7	2	4	1	1	0	1	3	8	1	4	1	3	0	0	1	1	1	1	2	1	1	0	3	1	3	4	1	0	5	0	94
Aspergillus flavus	8	13	1	11	3	5	3	6	3	4	1	6	0	0	5	10	1	2	1	3	0	0	1	1	0	2	2	1	1	0	3	1	3	5	1	0	8	0	115
Aspergillus fumigatus	5	13	1	8	3	4	3	7	2	4	1	2	0	2	4	8	1	3	1	3	1	0	1	1	0	1	2	1	1	1	3	1	3	3	2	1	9	1	107
Aspergillus nidulans	7	10	1	6	2	5	3	4	2	4	1	4	0	1	5	8	1	3	1	3	1	0	1	1	0	3	2	1	1	0	3	1	3	2	1	0	4	0	95
Aspergillus niger	9	7	0	7	4	3	1	1	2	3	1	3	0	0	3	9	1	2	0	1	0	0	0	1	0	2	2	1	1	0	1	1	4	4	2	0	6	0	82
Aspergillus oryzae	8	10	1	10	3	4	3	6	3	4	1	5	0	0	5	8	1	2	1	3	0	1	1	1	0	0	2	1	1	0	3	1	4	5	2	0	7	0	107
Aspergillus terreus	7	10	1	7	3	2	3	5	2	4	1	4	1	0	7	10	1	3	1	3	0	0	1	1	0	2	2	1	1	0	3	1	4	4	1	0	6	1	103
Aspergillus westerdijkiae	8	10	1	5	1	3	3	6	2	4	1	1	1	1	2	8	1	4	1	3	1	0	2	2	1	1	0	3	1	5	5	1	0	6	0	96
N. fischeri	5	13	1	8	3	5	3	7	2	4	1	1	0	1	2	7	1	4	1	3	1	0	1	1	0	1	2	1	1	0	3	1	3	2	1	0	9	0	99

Carbohydrate Esterases (CEs)	CE1	CE2	CE3	CE4	CE5	CE7	CE8	CE9	CE10	CE12	CE14	CE15	CE16	Total
Asperdillus clavatus	20	1	1	6	3	1	2	1	39	2	3	0	4	83
Aspergillus flavus	22	1	5	4	5	3	5	1	71	5	5	0	4	131
Aspergillus fumigatus	21	1	1	7	5	1	4	2	44	4	3	1	5	99
Aspergillus nidulans	21	2	7	9	4	0	3	1	54	3	2	0	5	111
Aspergillus niger	9	0	1	8	5	2	3	1	69	3	1	0	3	105
Aspergillus oryzae	22	1	3	4	5	3	4	1	59	5	4	0	4	115
Aspergillus terreus	19	3	3	5	8	1	1	1	65	3	3	1	6	119
Aspergillus westerdijkiae	23	2	2	8	8	2	3	1	65	5	2	1	9	131
N. fischeri	22	1	2	7	7	2	4	1	56	4	3	1	6	116

Polysaccharide Lyases (PLs)	PL1	PL3	PL4	PL5	PL7	PL9	PL11	PL17	PL20	PL22	Total
Asperdillus clavatus	2	1	2	0	0	0	0	0	2	0	7
Aspergillus flavus	12	3	3	0	1	1	0	0	2	2	24
Aspergillus fumigatus	6	3	3	1	0	1	0	0	1	0	15
Aspergillus nidulans	8	5	4	0	0	1	1	0	2	0	21
Aspergillus niger	4	0	2	0	0	0	0	0	1	3	10
Aspergillus oryzae	12	3	5	0	1	1	0	0	2	2	26
Aspergillus terreus	7	3	3	0	0	1	0	1	0	0	15
Aspergillus westerdijkiae	11	3	6	0	0	1	1	0	0	2	24
N. fischeri	6	3	3	1	1	1	0	0	2	0	17

Auxiliary Activities (AAs)	AA1	AA2	AA3	AA4	AA5	AA6	AA7	AA8	AA9	Total
Asperdillus clavatus	2	3	13	2	1	2	20	2	8	53
Aspergillus flavus	2	3	32	1	2	2	45	3	8	98
Aspergillus fumigatus	1	3	11	5	1	2	25	2	8	58
Aspergillus nidulans	2	3	25	5	0	1	38	1	11	86
Aspergillus niger	3	1	29	4	0	1	37	2	3	80
Aspergillus oryzae	2	3	28	1	2	2	45	2	9	94
Aspergillus terreus	5	3	22	3	1	1	34	4	13	86
Aspergillus westerdijkiae	2	3	29	3	2	3	44	3	12	101
N. fischeri	2	3	14	3	2	1	29	2	8	64

Carbohydrate-Binding Modules (CBMs	CBM	CBM	CBM	CBM	CBM	CBM	CBM1	CBM1	CBM1	CBM1	CBM2	CBM2	CBM2	CBM2	CBM3	CBM3	CBM3	CBM3	CBM4	CBM4	CBM4	CBM4	CBM4	CBM5	CBM5	CBM6	CBM6	CBM6	CBM6	Total
Asperdillus clavatus	15	0	1	1	0	1	1	1	5	1	6	1	0	0	0	1	0	0	1	3	0	0	1	5	0	1	0	0	0	45
Aspergillus flavus	5	0	0	0	0	3	3	0	4	1	1	1	0	3	2	2	0	0	1	5	0	1	2	9	0	1	1	0	6	51
Aspergillus fumigatus	17	0	0	0	1	1	1	0	6	1	5	1	0	4	1	1	0	1	1	4	0	1	1	8	0	1	1	2	4	63
Aspergillus nidulans	6	0	0	2	0	0	1	0	12	0	4	1	1	1	0	1	1	0	1	3	0	1	1	18	0	1	1	1	6	63
Aspergillus niger	1	0	0	0	0	1	1	0	4	0	1	1	0	1	0	1	1	0	0	3	0	0	0	11	1	0	0	1	4	32
Aspergillus oryzae	3	0	0	0	0	3	2	0	4	1	1	1	0	2	2	1	0	0	1	5	0	1	1	8	0	1	1	1	5	44
Aspergillus terreus	15	1	0	1	0	2	1	0	10	1	4	1	2	2	2	2	0	0	1	3	0	1	1	14	0	1	0	2	3	70
Aspergillus westerdijkiae	8	0	0	1	0	1	0	0	5	1	3	1	1	2	1	1	1	0	1	3	1	1	2	10	0	1	0	0	8	53
N. fischeri	22	0	0	0	0	1	2	0	7	0	6	1	1	5	1	1	2	2	1	4	0	1	1	12	0	1	1	1	4	77

Number of putative gene per Plant Polysaccharide Degradation (PDD) related CAZy family for nine Aspergilli genomes

Glycoside Hydrolases (GHs)	GH1	GH2	GH3	GH5	GH6	GH7	GH10	GH11	GH12	GH13	GH15	GH26	GH27	GH28	GH29	GH31	GH32	GH35	GH36	GH43	GH45	GH51	GH53	GH54	GH62	GH67	GH74	GH78	GH88	GH93	GH95	GH105	GH115	Total	Percentage
Asperdillus clavatus	3	3	12	9	2	4	2	3	3	16	6	0	2	3	0	5	1	2	2	13	0	2	0	1	2	1	2	0	0	1	1	2	1	104	51.7%
Aspergillus flavus	3	8	24	15	1	3	4	4	5	15	3	1	3	21	0	9	4	7	2	20	0	4	2	1	2	1	1	12	3	3	2	4	3	190	59.2%
Aspergillus fumigatus	5	6	18	13	1	4	4	3	4	15	5	0	5	13	0	6	4	5	2	18	1	2	1	1	2	1	3	5	2	3	2	3	1	158	58.1%
Aspergillus nidulans	3	10	20	16	2	3	3	2	1	12	2	3	3	10	0	10	2	3	3	17	0	3	1	1	2	1	3	9	2	2	3	4	1	157	59.0%
Aspergillus niger	3	3	14	5	1	0	1	2	1	11	1	1	4	20	1	4	1	4	1	8	0	2	1	0	0	1	1	7	1	1	2	2	0	104	56.5%
Aspergillus oryzae	3	7	23	12	1	3	4	4	4	15	3	1	3	21	0	10	4	5	2	19	0	3	1	1	2	1	1	9	3	3	3	4	4	179	58.9%
Aspergillus terreus	3	10	21	18	2	4	4	2	6	13	2	0	4	8	2	11	6	4	3	20	0	3	1	1	3	2	2	4	0	4	3	2	2	170	60.7%
Aspergillus westerdijkiae	4	10	22	16	2	3	3	2	3	10	3	1	2	15	1	8	4	4	2	19	0	4	2	1	1	1	2	11	3	3	3	3	2	170	60.5%
N. fischeri	5	6	19	16	2	5	4	4	5	15	5	0	5	13	0	8	4	5	2	20	1	2	1	1	2	1	4	7	2	3	2	3	1	173	57.1%

Carbohydrate Esterases (CEs)	CE1	CE8	CE12	CE15	Total	Percentage
Asperdillus clavatus	20	2	2	0	24	####
Aspergillus flavus	22	5	5	0	32	####
Aspergillus fumigatus	21	4	4	1	30	####
Aspergillus nidulans	21	3	3	0	27	####
Aspergillus niger	9	3	3	0	15	####
Aspergillus oryzae	22	4	5	0	31	####
Aspergillus terreus	19	1	3	1	24	####
Aspergillus westerdijkiae	23	3	5	1	32	####
N. fischeri	22	4	4	1	31	####

Polysaccharide Lyases (PLs)	PL1	PL3	PL4	PL9	PL11	Total	Percentage
Asperdillus clavatus	2	1	2	0	0	5	71.4%
Aspergillus flavus	12	3	3	1	0	19	79.2%
Aspergillus fumigatus	6	3	3	1	0	13	86.7%
Aspergillus nidulans	8	5	4	1	1	19	90.5%
Aspergillus niger	4	0	2	0	0	6	60.0%
Aspergillus oryzae	12	3	5	1	0	21	80.8%
Aspergillus terreus	7	3	3	1	0	14	93.3%
Aspergillus westerdijkiae	11	3	6	1	1	22	91.7%
N. fischeri	6	3	3	1	0	13	76.5%

Appendix XVII: Accession number list for *A. westerdijkiae* genome

#	Accession	Name	Length
1	LKBE01000001	scaffold1	1959216
2	LKBE01000002	scaffold2	1444596
3	LKBE01000003	scaffold3	993355
4	LKBE01000004	scaffold4	1646026
5	LKBE01000005	scaffold5	279310
6	LKBE01000006	scaffold6	9192
7	LKBE01000007	scaffold7	736687
8	LKBE01000008	scaffold8	1383725
9	LKBE01000009	scaffold9	2817998
10	LKBE01000010	scaffold10	3128
11	LKBE01000011	scaffold11	3221
12	LKBE01000012	scaffold12	1469756
13	LKBE01000013	scaffold13	3171
14	LKBE01000014	scaffold14	3084807
15	LKBE01000015	scaffold15	17828
16	LKBE01000016	scaffold17	1603627
17	LKBE01000017	scaffold18	59885
18	LKBE01000018	scaffold19	9755
19	LKBE01000019	scaffold20	3201
20	LKBE01000020	scaffold21	9771
21	LKBE01000021	scaffold23	257841
22	LKBE01000022	scaffold24	9488
23	LKBE01000023	scaffold25	1430477
24	LKBE01000024	scaffold26	12112
25	LKBE01000025	scaffold27	3359036
26	LKBE01000026	scaffold31	3238
27	LKBE01000027	scaffold33	3140
28	LKBE01000028	scaffold34	822893
29	LKBE01000029	scaffold35	2604827
30	LKBE01000030	scaffold36	13010
31	LKBE01000031	scaffold38	3156
32	LKBE01000032	scaffold40	10181
33	LKBE01000033	scaffold41	493414
34	LKBE01000034	scaffold42	3194
35	LKBE01000035	scaffold43	2212477
36	LKBE01000036	scaffold44	6163
37	LKBE01000037	scaffold45	1089895
38	LKBE01000038	scaffold46	1025085
39	LKBE01000039	scaffold47	9582
40	LKBE01000040	scaffold48	9281
41	LKBE01000041	scaffold49	3252
42	LKBE01000042	scaffold52	18628
43	LKBE01000043	scaffold53	3258
44	LKBE01000044	scaffold56	3185
45	LKBE01000045	scaffold57	9274
46	LKBE01000046	scaffold58	9172
47	LKBE01000047	scaffold60	3180
48	LKBE01000048	scaffold62	669811
49	LKBE01000049	scaffold63	654656
50	LKBE01000050	scaffold64	3194
51	LKBE01000051	scaffold65	3287
52	LKBE01000052	scaffold66	3264
53	LKBE01000053	scaffold70	12146
54	LKBE01000054	scaffold71	796945
55	LKBE01000055	scaffold72	15500
56	LKBE01000056	scaffold73	3168
57	LKBE01000057	scaffold74	11929
58	LKBE01000058	scaffold75	929178
59	LKBE01000059	scaffold77	3302
60	LKBE01000060	scaffold79	3228
61	LKBE01000061	scaffold80	3163
62	LKBE01000062	scaffold82	3199
63	LKBE01000063	scaffold83	9184
64	LKBE01000064	scaffold87	9253
65	LKBE01000065	scaffold88	9255
66	LKBE01000066	scaffold89	348230
67	LKBE01000067	scaffold90	3220
68	LKBE01000068	scaffold91	9511
69	LKBE01000069	scaffold92	3201
70	LKBE01000070	scaffold94	9256
71	LKBE01000071	scaffold95	3341

72	LKBE01000072	scaffold97	136940
73	LKBE01000073	scaffold98	3258
74	LKBE01000074	scaffold99	3294
75	LKBE01000075	scaffold100	3208
76	LKBE01000076	scaffold101	9220
77	LKBE01000077	scaffold102	3318
78	LKBE01000078	scaffold106	9333
79	LKBE01000079	scaffold108	3271
80	LKBE01000080	scaffold109	9532
81	LKBE01000081	scaffold111	8866
82	LKBE01000082	scaffold114	3967
83	LKBE01000083	scaffold115	3220
84	LKBE01000084	scaffold117	3162
85	LKBE01000085	scaffold118	3279
86	LKBE01000086	scaffold120	3305
87	LKBE01000087	scaffold121	3288
88	LKBE01000088	scaffold122	3202
89	LKBE01000089	scaffold127	3689
90	LKBE01000090	scaffold129	9181
91	LKBE01000091	scaffold130	6198
92	LKBE01000092	scaffold133	9291
93	LKBE01000093	scaffold135	9289
94	LKBE01000094	scaffold139	284447
95	LKBE01000095	scaffold140	3352
96	LKBE01000096	scaffold143	3397
97	LKBE01000097	scaffold145	3375
98	LKBE01000098	scaffold146	3407
99	LKBE01000099	scaffold150	108867
100	LKBE01000100	scaffold151	379913
101	LKBE01000101	scaffold153	3612
102	LKBE01000102	scaffold154	4455
103	LKBE01000103	scaffold157	6520
104	LKBE01000104	scaffold164	2499
105	LKBE01000105	C6100	1006
106	LKBE01000106	C6110	1011
107	LKBE01000107	C6112	1014
108	LKBE01000108	C6154	1042
109	LKBE01000109	C6162	1045
110	LKBE01000110	C6168	1048
111	LKBE01000111	C6194	1060
112	LKBE01000112	C6208	1065
113	LKBE01000113	C6254	1114
114	LKBE01000114	C6272	1135
115	LKBE01000115	C6276	1154
116	LKBE01000116	C6280	1160
117	LKBE01000117	C6310	1187
118	LKBE01000118	C6344	1224
119	LKBE01000119	C6352	1226
120	LKBE01000120	C6358	1232
121	LKBE01000121	C6360	1232
122	LKBE01000122	C6394	1296
123	LKBE01000123	C6448	1364
124	LKBE01000124	C6540	1493
125	LKBE01000125	C6548	1504
126	LKBE01000126	C6572	1554
127	LKBE01000127	C6592	1579
128	LKBE01000128	C6594	1589
129	LKBE01000129	C6600	1594
130	LKBE01000130	C6604	1596
131	LKBE01000131	C6606	1597
132	LKBE01000132	C6632	1643
133	LKBE01000133	C6640	1654
134	LKBE01000134	C6652	1668
135	LKBE01000135	C6682	1730
136	LKBE01000136	C6690	1742
137	LKBE01000137	C6702	1766
138	LKBE01000138	C6712	1778
139	LKBE01000139	C6714	1784
140	LKBE01000140	C6728	1798
141	LKBE01000141	C6764	1853
142	LKBE01000142	C6806	1937
143	LKBE01000143	C6816	1950
144	LKBE01000144	C6820	1951
145	LKBE01000145	C6826	1963
146	LKBE01000146	C6868	2026

147	LKBE01000147	C6874	2032
148	LKBE01000148	C6876	2033
149	LKBE01000149	C6880	2047
150	LKBE01000150	C6892	2064
151	LKBE01000151	C6902	2082
152	LKBE01000152	C6920	2140
153	LKBE01000153	C6934	2154
154	LKBE01000154	C6944	2176
155	LKBE01000155	C6964	2216
156	LKBE01000156	C6976	2245
157	LKBE01000157	C7004	2318
158	LKBE01000158	C7020	2345
159	LKBE01000159	C7022	2361
160	LKBE01000160	C7042	2450
161	LKBE01000161	C7054	2494
162	LKBE01000162	C7062	2513
163	LKBE01000163	C7088	2563
164	LKBE01000164	C7106	2606
165	LKBE01000165	C7142	2661
166	LKBE01000166	C7230	2876
167	LKBE01000167	C7280	2991
168	LKBE01000168	C7284	2999
169	LKBE01000169	C7286	3013
170	LKBE01000170	C7320	3094
171	LKBE01000171	C7356	3202
172	LKBE01000172	C7374	3231
173	LKBE01000173	C7394	3283
174	LKBE01000174	C7412	3340
175	LKBE01000175	C7424	3361
176	LKBE01000176	C7430	3384
177	LKBE01000177	C7506	3602
178	LKBE01000178	C7508	3612
179	LKBE01000179	C7520	3643
180	LKBE01000180	C7522	3644
181	LKBE01000181	C7550	3682
182	LKBE01000182	C7564	3714
183	LKBE01000183	C7578	3755
184	LKBE01000184	C7586	3766
185	LKBE01000185	C7612	3840
186	LKBE01000186	C7614	3841
187	LKBE01000187	C7630	3893
188	LKBE01000188	C7652	3954
189	LKBE01000189	C7666	4005
190	LKBE01000190	C7682	4043
191	LKBE01000191	C7712	4158
192	LKBE01000192	C7726	4215
193	LKBE01000193	C7742	4259
194	LKBE01000194	C7768	4317
195	LKBE01000195	C7772	4334
196	LKBE01000196	C7792	4420
197	LKBE01000197	C7796	4425
198	LKBE01000198	C7804	4471
199	LKBE01000199	C7818	4528
200	LKBE01000200	C7842	4610
201	LKBE01000201	C7856	4681
202	LKBE01000202	C7874	4762
203	LKBE01000203	C7914	4869
204	LKBE01000204	C7916	4874
205	LKBE01000205	C7918	4879
206	LKBE01000206	C7936	4946
207	LKBE01000207	C7962	5069
208	LKBE01000208	C7964	5074
209	LKBE01000209	C7966	5082
210	LKBE01000210	C7970	5105
211	LKBE01000211	C8006	5264
212	LKBE01000212	C8014	5295
213	LKBE01000213	C8038	5360
214	LKBE01000214	C8056	5479
215	LKBE01000215	C8080	5551
216	LKBE01000216	C8092	5656
217	LKBE01000217	C8096	5668
218	LKBE01000218	C8134	5817
219	LKBE01000219	C8148	5892
220	LKBE01000220	C8178	6079
221	LKBE01000221	C8208	6293

222	LKBE01000222	C8268	6546
223	LKBE01000223	C8278	6620
224	LKBE01000224	C8284	6651
225	LKBE01000225	C8290	6693
226	LKBE01000226	C8292	6701
227	LKBE01000227	C8300	6741
228	LKBE01000228	C8348	6948
229	LKBE01000229	C8416	7368
230	LKBE01000230	C8444	7507
231	LKBE01000231	C8452	7555
232	LKBE01000232	C8472	7721
233	LKBE01000233	C8506	7892
234	LKBE01000234	C9122	12565
235	LKBE01000235	C9142	12731
236	LKBE01000236	C9300	14259
237	LKBE01000237	C9338	14629
238	LKBE01000238	C9448	16008
239	LKBE01000239	C10106	27120

Appendix XVIII: Accession number list for *N. jinanensis* genome

#	Accession	Name	Length
1	LNDA01000001	C3070	1148
2	LNDA01000002	C3078	1199
3	LNDA01000003	C3086	1209
4	LNDA01000004	C3100	1260
5	LNDA01000005	C3106	1299
6	LNDA01000006	C3120	1436
7	LNDA01000007	C3146	1569
8	LNDA01000008	C3174	1674
9	LNDA01000009	C3180	1712
10	LNDA01000010	C3186	1754
11	LNDA01000011	C3192	1802
12	LNDA01000012	C3196	1823
13	LNDA01000013	C3256	2369
14	LNDA01000014	C3352	3089
15	LNDA01000015	C3374	3191
16	LNDA01000016	C3378	3241
17	LNDA01000017	C3388	3325
18	LNDA01000018	C3390	3336
19	LNDA01000019	C3398	3400
20	LNDA01000020	C3400	3420
21	LNDA01000021	C3442	3880
22	LNDA01000022	C3476	4170
23	LNDA01000023	C3496	4577
24	LNDA01000024	C3578	5592
25	LNDA01000025	C3688	7039
26	LNDA01000026	C3712	7358
27	LNDA01000027	C3714	7394
28	LNDA01000028	C3734	7816
29	LNDA01000029	C3856	10372
30	LNDA01000030	C3876	10879
31	LNDA01000031	C3880	11062
32	LNDA01000032	C3888	11177
33	LNDA01000033	C3946	12708
34	LNDA01000034	C3952	12865
35	LNDA01000035	C3968	13089
36	LNDA01000036	C3990	13871
37	LNDA01000037	C4008	14253
38	LNDA01000038	C4090	17347
39	LNDA01000039	C4132	18596
40	LNDA01000040	C4154	19808
41	LNDA01000041	C4260	28167
42	LNDA01000042	C4332	35748
43	LNDA01000043	C4354	39401
44	LNDA01000044	SCAFFOLD1	9416
45	LNDA01000045	SCAFFOLD2	64315
46	LNDA01000046	SCAFFOLD3	306047

47	LNDA01000047	SCAFFOLD5	359073
48	LNDA01000048	SCAFFOLD6	94549
49	LNDA01000049	SCAFFOLD7	40565
50	LNDA01000050	SCAFFOLD8	85280
51	LNDA01000051	SCAFFOLD9	15480
52	LNDA01000052	SCAFFOLD10	30835
53	LNDA01000053	SCAFFOLD11	44562
54	LNDA01000054	SCAFFOLD12	89445
55	LNDA01000055	SCAFFOLD13	36802
56	LNDA01000056	SCAFFOLD14	135563
57	LNDA01000057	SCAFFOLD15	314023
58	LNDA01000058	SCAFFOLD16	3061
59	LNDA01000059	SCAFFOLD17	66838
60	LNDA01000060	SCAFFOLD18	2146
61	LNDA01000061	SCAFFOLD19	39005
62	LNDA01000062	SCAFFOLD20	96342
63	LNDA01000063	SCAFFOLD21	80100
64	LNDA01000064	SCAFFOLD22	87614
65	LNDA01000065	SCAFFOLD23	2532
66	LNDA01000066	SCAFFOLD24	45505
67	LNDA01000067	SCAFFOLD26	35795
68	LNDA01000068	SCAFFOLD27	34447
69	LNDA01000069	SCAFFOLD28	58417
70	LNDA01000070	SCAFFOLD29	62271
71	LNDA01000071	SCAFFOLD31	69380
72	LNDA01000072	SCAFFOLD32	114069
73	LNDA01000073	SCAFFOLD35	51069
74	LNDA01000074	SCAFFOLD37	93845
75	LNDA01000075	SCAFFOLD38	230940
76	LNDA01000076	SCAFFOLD41	354037
77	LNDA01000077	SCAFFOLD42	65142
78	LNDA01000078	SCAFFOLD53	217375
79	LNDA01000079	SCAFFOLD59	3666
80	LNDA01000080	SCAFFOLD64	85899
81	LNDA01000081	SCAFFOLD69	15427
82	LNDA01000082	SCAFFOLD74	73008
83	LNDA01000083	SCAFFOLD77	150961
84	LNDA01000084	SCAFFOLD90	165664
85	LNDA01000085	SCAFFOLD96	25119
86	LNDA01000086	SCAFFOLD101	11938
87	LNDA01000087	SCAFFOLD102	9797
88	LNDA01000088	SCAFFOLD103	17290
89	LNDA01000089	SCAFFOLD104	38491
90	LNDA01000090	SCAFFOLD105	26109
91	LNDA01000091	SCAFFOLD106	36684
92	LNDA01000092	SCAFFOLD107	24192
93	LNDA01000093	SCAFFOLD108	16116
94	LNDA01000094	SCAFFOLD109	10222
95	LNDA01000095	SCAFFOLD110	20543
96	LNDA01000096	SCAFFOLD111	46081

97	LNDA01000097	SCAFFOLD112	64327
98	LNDA01000098	SCAFFOLD113	23994
99	LNDA01000099	SCAFFOLD114	18787
100	LNDA01000100	SCAFFOLD115	58039
101	LNDA01000101	SCAFFOLD116	21641
102	LNDA01000102	SCAFFOLD117	34508
103	LNDA01000103	SCAFFOLD118	31594
104	LNDA01000104	SCAFFOLD119	53406
105	LNDA01000105	SCAFFOLD120	31087
106	LNDA01000106	SCAFFOLD121	70709
107	LNDA01000107	SCAFFOLD122	57636

Appendix XIX: *A. westerdijkiae*- Details of genome sequencing and SOAPdenovo assembly

1. Filtered data after sequencing

Library	Original		Filtered	
	Total Reads	Total Bases	Total Reads	Total Bases
ASPERGILLUS-WESTERDIJKIAE_3kb	53,204,834	5,373,688,234	26,465,748	2,443,781,167
ASPERGILLUS-WESTERDIJKIAE_10kb	59,171,878	5,976,359,678	28,848,060	2,639,218,532

Sample	TotalBases	ReadCount	GC(%)	Q20(%)	Q30(%)
ASPERGILLUS-WESTERDIJKIAE_3kb	2,443,781,167	26,465,748	48.89	98.89	93.26
ASPERGILLUS-WESTERDIJKIAE_10kb	2,639,218,532	28,848,060	49.17	98.98	93.71

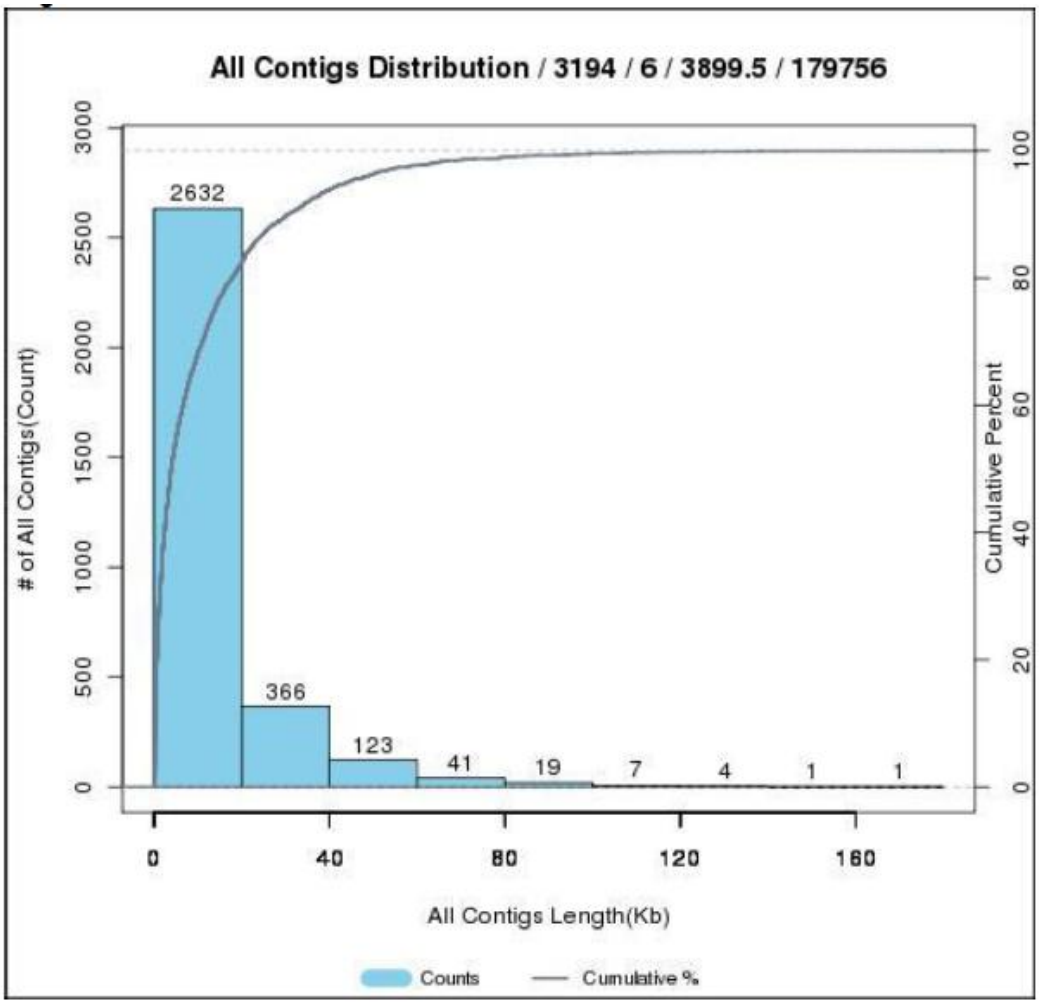
- Total Bases : The total number of bases in reads.
- Reads Count : The number of reads identified.
- GC(%) : The percentage of bases on a DNA that are either guanine or cytosine.
- Q20(%) : The percentage of bases called that have a quality score of 20 or above.
- Q30(%) : The percentage of bases called that have a quality score of 30 or above.

2. SOAPdenovo assembly summary (length>1000bp)

Number of Scaffolds	Scaffolds sum	N50	Longest Scaffold	Shortest Scaffold	Average length
322	36,523,223	1,603,627	3,359,036	1,003	113,426

- Number of Scaffolds : The number of Scaffolds identified.
- Scaffolds sum : The total number of bases in the Scaffolds.
- N50 : An N50 means that half of all bases reside in Scaffolds o this size or longer.
- Longest Scaffold : The sequence size of the longest Scaffold.
- Shortest Scaffold : The sequence size of the shortest Scaffold.
- Average length : The average Scaffold size.

3. Length distribution



4. Scaffold statistics

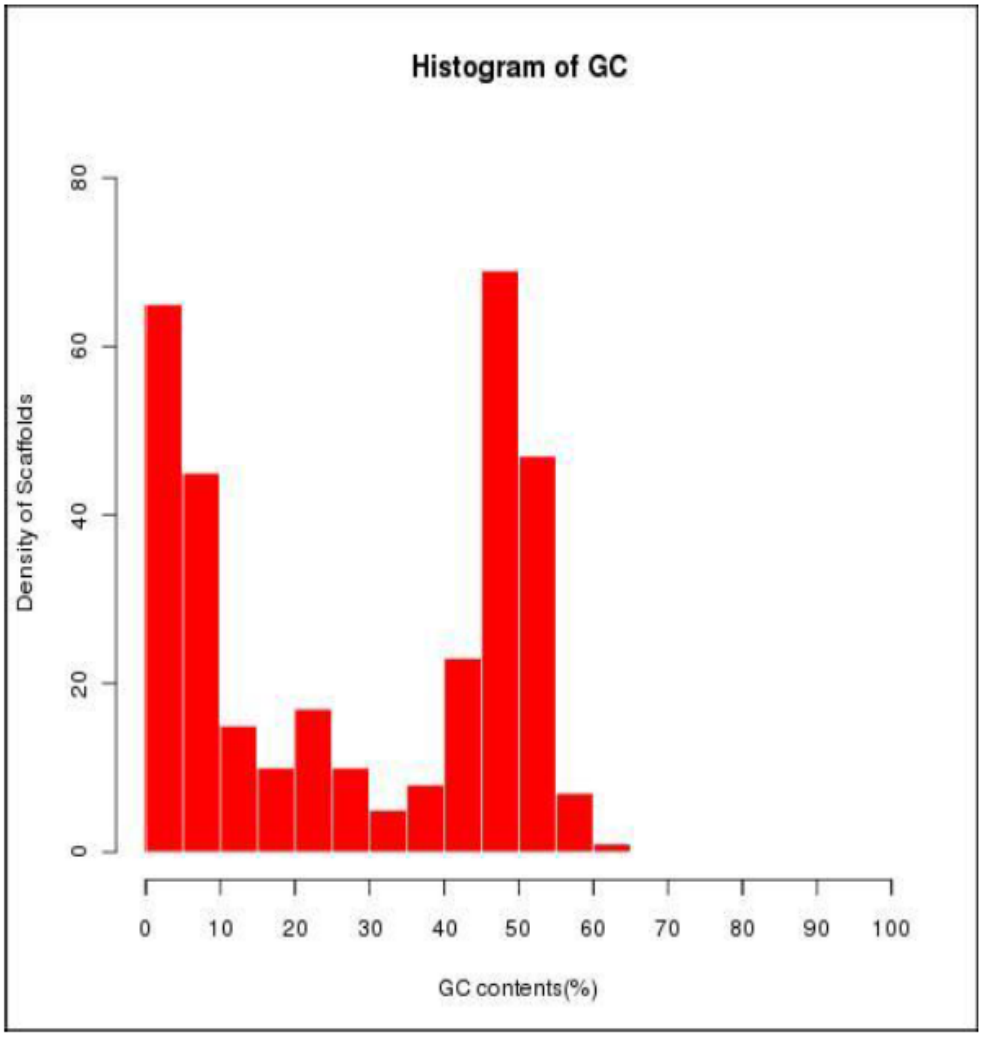
	Num of Scaffolds	Length	Avg.length	Sum
N10	2	3,084,807	3,221,921	6,443,843
N20	3	2,817,998	3,087,280	9,261,841
N30	4	2,604,827	2,966,667	11,866,668
N40	6	1,959,216	2,673,060	16,038,361
N50	8	1,603,627	2,411,001	19,288,014
N60	10	1,444,596	2,220,236	22,202,366
N70	13	1,089,895	2,008,189	26,106,463
N80	17	822,893	1,757,469	29,876,974
N90	22	493,414	1,510,385	33,228,487
N100	322	1,003	113,426	36,523,223

- Num of sequences : the number of sequences in the Scaffold statistics (N10~N100)
- Length : the length of sequence in the Scaffold statistics (N10~N100)
- Avg. length : the average length in the Scaffold statistics (N10~N100)
- Sum : the sum of the length in the Scaffold statistics (N10~N100)

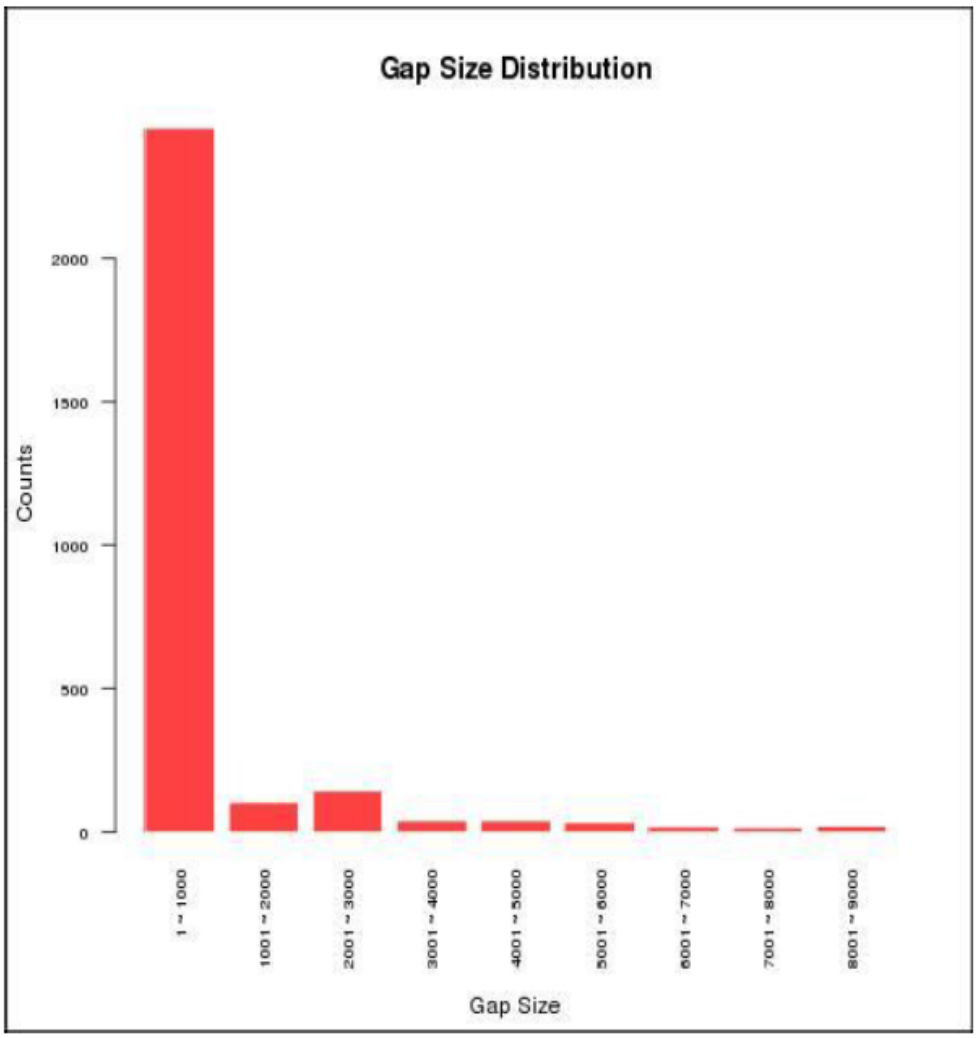
5. GC content

Num of A	Num of T	Num of G	Num of C	Num of N	GC contents
8,604,682	8,580,358	8,712,252	8,718,520	1,907,411	47.73%

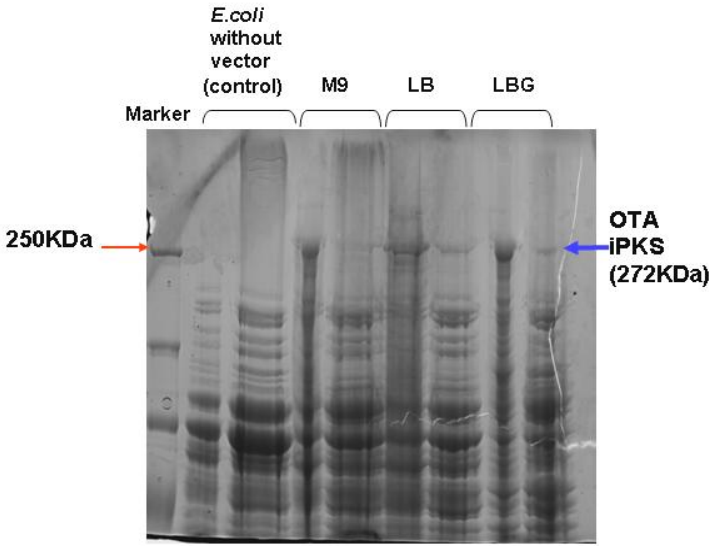
- Num of A : the total number of adenine(A)
- Num of T : the total number of thymine(T)
- Num of G : the total number of guanine(G)
- Num of C : the total number of cytosine(C)
- Num of N : the total number of ambiguous
- GC contents : the percentage of guanine-cytosine base pairs



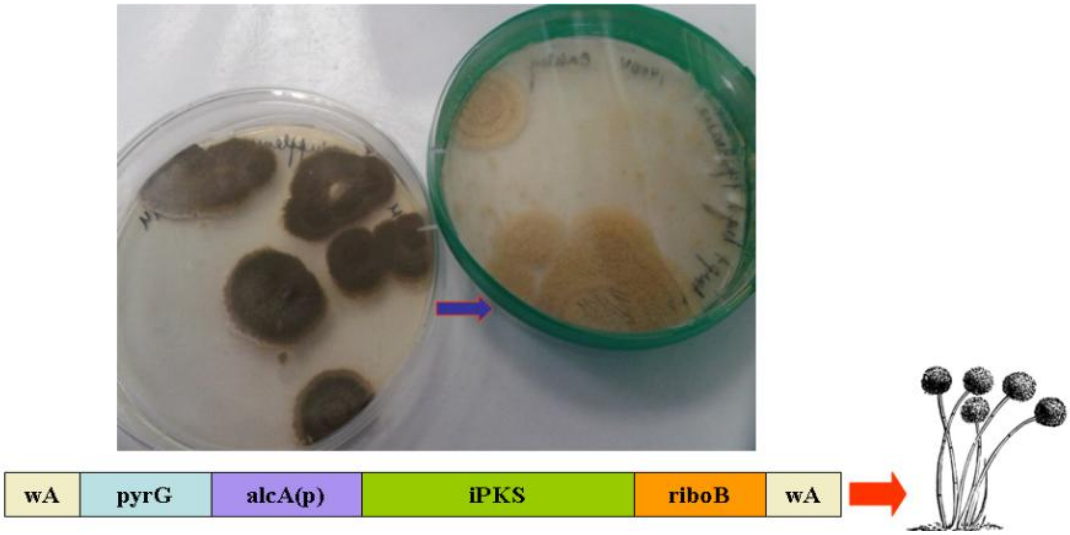
6. Gap size distribution



Appendix XX: Heterologous expression of pksAwota in various expression systems.



Protein gel picture showing the expression pksAwota (OTA iPKS) in *E. coli*. Three different media were used to optimize expression-M9, LB and LB with 10%glycerol (LBG). For each medium we have shown the profile of protein in the pellet and supernatant respectively.



Experiment design of heterologous expression of pksAwota in *A. nidulans* strain LO7020. Colony appearance of *A. nidulans* LO7020 strain before and after transformation with pksAwota (iPKS) gene. Since the wA locus is disturbed, pigmentation is lost in the transformed strain.

** The methods for these trials have been explained in Chapter 3 in details.*