THE USE OF MICROBIAL GENOME MINING FOR IN SILICO DISCOVERY OF NOVEL SECONDARY METABOLITE GENE CLUSTERS

BY

JAFAR ALAWNEH

ADISSERTATION SUBMITTED TO

RUTGERS-SCHOOL OF HEALTH PROFESSIONS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOMEDICAL INFORMATICS

DEPARTMENT OF HEALTH INFORMATICS

SCHOOL OF HEALTH PROFESSIONS

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

APRIL 2019

RUTGERS
School of Health Professions

FINAL DISSERTATION DEFENSE APPROVAL FORM

THE USE OF MICROBIAL GENOME MINING FOR THE IN SILICO DISCOVERY OF NOVEL ANTICANCER AGENTS AND NOVEL ANTIMICROBIAL COMPOUNDS

BY

JAFAR ALAWNEH

DISSERTATION COMMITTEE:

FREDERICK COFFMAN, PhD

SHANKAR SRINIVASAN, PhD

AMRO HANORA, PhD

APPROVED BY THE DISSERTATION COMMITTEE:

_____ **DATE:** _____

_____ **DATE:** _____

_____ **DATE:** _____

_____ **DATE:** _____

TABLE OF CONTENTS

ABSTRACT

Secondary metabolites (SMs) are small organic molecules that have various biological functions and produced by bacteria, fungi, archaea, and plants. Because of their diverse structures, different SMs have been shown to have antibacterial, antifungal, antiviral, and anticancer activities. Nonribosomal peptides (NRPs) and polyketides (PKs) are two diverse classes of SMs that are produced by nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS), respectively. One class of SMs, the epothilones, were discovered in the soil bacteria *S. cellulosum* and some epothilones have been shown to have antitumor activities similar to the taxanes. There is significant interest in expanding the available pool of structurally unique epothilones and other SMs as therapeutic candidates, however the distribution and structural variations in the microbial genomic landscape is currently poorly understood.

In this study, genome mining was used to find Epothilone-similar gene clusters (ESGCs) and other SMs gene clusters that potentially encode Epothilone-similar compounds and novel SMs. The sequences of genes (epo A-F) forming the *S. cellulosum* So ce90 Epothilone gene cluster were initially used to find epo A-F homologs (EAFHs) in other bacteria. These homologs were used to find gene clusters (ESGCs), and these newly discovered gene clusters were subsequently used to screen bacterial species and strains to find currently unidentified ESGCs and hybrid PKS-NRPS gene clusters that potentially encode novel SMs. The gene clusters identified in this study can be divided into three groups: 1) ESGCs highly similar to the Epothilone gene cluster and likely produce epothilone variants; 2) Gene clusters highly similar to those that encode genes which

produce other secondary metabolites; 3) Gene clusters that showed relatively low similarities with secondary metabolite gene clusters. Many of these gene clusters are reported for the first time in this study. Further, a number of EAFHs identified in this study were used for in silico design of ESGCs, which resulted in new gene clusters that could produce novel Epothilone-similar compounds with predictable molecular structures. These results suggest that directed manipulation of modular EGSC components is a viable approach to producing a large number of new secondary metabolites for testing against pathogens and cancer cells.

ACKNOWLEDGMENT

This dissertation would not have been a reality without the help of three groups: our GOD Almighty, my thesis committee members, and my family.

First and foremost, I would like to thank our God Almighty for the blessings he has bestowed upon me and for giving me good health, patience, and the strength to finish this dissertation

I am indebted to my thesis advisor Dr. Fredrick Coffman for the continuous support of my PhD study, for his guidance, motivation, and patience.

I would like to express my sincere and profound gratitude to Dr. Shankar Srinivasan for his sustained support, motivation and persistent inspiration throughout my PhD study

Also, I would like to express my heartfelt gratitude to Dr. Amro Hanora for his inspired guidance, encouragement, precious assistance, and scholarly knowledge.

Last but not least, my deepest gratitude goes to my family members and especially to my beloved wife Hajar Qasem, for their continuous support, patience, love, and sacrifices they have made.

DEDICATION

This dissertation is lovingly dedicated to:

My father, Mohammad Qasseim Alawneh. Although he is no longer alive, he continues to inspire me to be the best version of myself every day.

My mother, Aisha Qasem, for her endless love and encouragement.

My beloved wife, Hajar Qasem who has been a constant source of support and encouragement during the challenges of life and PhD study

My lovely children, Mohammad, Rama, Yareen, Leen , Rana, and Hiba who have always loved me unconditionally

My beloved family and to all my friends

LIST OF FIGURE

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Microorganisms play a central role in the development of several biotechnological processes that result in the production of nutritional products and supplements that fulfill the increasing needs of the rapidly-growing world population. Indeed, the microbial industrial processes result in less hazardous pollutants than chemical methods. Therefore, microorganisms can alternatively be used in the production of supplements and nutritional products without creating environmental problems caused by chemical processes. Among the products that can be produced by microorganisms are the secondary metabolites. Secondary metabolites (SMs) are organic compounds with variable and sophisticated structures that are biosynthesized by some microbial strains [1]. Microorganisms might use these metabolites as a mean of defense against other competing microorganisms, as a mean of chemical communication, or to facilitate their interaction with their hosts [2,3]. SMs can enhance the symbiotic relationship between microorganisms and their host (e.g. nematodes, insects, plants, and animals) and can also be used as effectors against ecological competitors [1]. Although we are too far from realizing the reason behind the production of SMs by microorganisms, these compounds are of great importance for human health and nutrition [4]. Moreover, these compounds can be used as food additives, biofuels, antibiotics, anti-cancer compounds, and agrochemicals [1].

1. Medically Relevant SMs

1.1 Antibacterial drugs

1.1.1 β-lactams

β-lactam antibiotics have been used for more than 50 years [5]. However, they are still the most commonly prescribed antibiotics [6]. Some of the β-lactam antibiotics, (e.g cephalosporins and penicillins) are still in use because some bacteria are still sensitive to these natural antibiotics. Furthermore, the development of semisynthetic derivatives that are effective against resistant bacteria extends the lifespan of β-lactam antibiotics [7]. Fungal genera that produce cephalosporin and penicillin include *Aspergillus,* and *Cephalosporium*. Bacterial species that belong to *Lysobacter, Nocardia*, *Flavobacterium*, and *Streptomyces* can also produce cephalosporin and penicillin [8].

All members of β-lactam antibiotics have a four-membered cyclic amide (β-lactam). With only one exception among the classes of β-lactams, the β-lactam ring is usually fused with a five- or six-membered ring. The classification of β-lactams is based on the nature of these rings and/or the presence of some atoms connected to these rings. There are six classes of β-lactams: 1) penams (five-membered ring with sulfur atom, 2) clavams (five-membered ring containing an oxygen or sulfur), 3) carbapenems (five-membered ring containing double bond but no atoms 4) cephems (six-membered unsaturated ring with a sulfur atom, 5) oxacephems (six-membered unsaturated ring with an oxygen atom 6) monobactams (contain only the β-lactam ring connected to other functional groups) [9].

1.1.2 Vancomycin

Vancomycin is a glycopeptide antibiotic produced by *Streptococcus orientalis* [10]. Vancomycin is usually prescribed to treat severe infections caused by Gram-positive bacteria and prescribed to combat staphylococcal strains resistant to methicillin or to patients with allergy to cephalosporins and penicillins [10,11]. This antibiotic inhibits the growth of bacteria by interfering with peptidoglycan biosynthesis and therefore the cell wall construction [12]. Despite its effectiveness in treating serious infections and its inhibitory effect against methicillin-resistant staphylococcal strains, vancomycin can cause various side effects. Adverse effects such as ototoxicity, phlebitis, nephrotoxicity, hypotension, and tachycardia [13], and other complications associated with peripheral intravenous infusion of vancomycin have been reported [14]. Due to the development of vancomycin resistance by some bacteria, the vancomycin semi-synthetic analogs dalbavancin, oritavancin, and telavancin were developed [15]. The vancomycin gene cluster has recently been identified in *Amycolatopsis orientalis* subsp. *orientalis* Strain KCTC 9412 [16].

1.1.3 Erythromycin

Erythromycin is a macrolide antibiotic that was first isolated in 1952 [17]. Erythromycin exerts its antibacterial effect by inhibiting bacterial protein synthesis [18]. A new semi-synthetic derivative of the macrolide erythromycin, EP-013420, also known as EP-420, is under clinical development [19]. Some of the unique pharmacokinetics features of EP-013420 are its long half-life (17-18 h) and increased systemic exposure which

makes it appropriate for once-daily dosing [20]. The polyketide synthetase's gene cluster of erythromycin was identified in *Saccharopolyspora erythraea* [21].

1.1.4 Teicoplanin

Teicoplanin is another glycopeptide antibiotic produced by an NRPS of *Actinoplanes teichomyceticus* and used against vancomycin-resistant bacteria. This antibiotic inhibits the growth of bacteria by blocking the biosynthesis of the cell wall [12]. Teicoplanin refers to a complex of structurally related natural compounds produced by *Actinoplanes teichomyceticus nov. sp*. (ATCC 31121) [22]. This antibiotic is a glycoprotein and it is chemically and microbiologically similar to vancomycin [23]. It has been used as the first choice antibiotic to treat enterococcal endocarditis and became an important choice to treat fatal infections caused by methicillin-resistant *Staphylococcus aureus* [24].

1.1.5 Tetracyclines

Tetracyclines are a group of bacteriostatic compounds that exert their action by inhibiting bacterial protein synthesis. The discovery of these antibacterial compounds was in the 1940s from *Streptomyces* species [25]. Due to the increased bacterial resistance, semi-synthetic analogs of tetracyclines have been developed. Tigecycline is one of these semi-synthetic analogs, and it was approved by the FDA in 2005 to treat infection caused by drug-resistant gram-positive bacteria [26]. The tetracycline units are assembled by a Type II PKS [27].

1.1.6 Capreomycin

Capreomycin refers to a mixture of four antibiotics with similar structure. This antibiotic is produced by *Saccharothrix mutabilis* and used along with other drugs to combat multidrug-resistant tuberculosis. Capreomycin is used to treat bacteria that are resistant to first and second lines of streptomycin [28]. This antibiotic has the ability to kill replicating and non-replicating *Mycobacterium tuberculosis* [29], and interfere with the bacterial ribosomes [30]. The capreomycin biosynthetic gene cluster was identified in *S. mutabilis subsp. Capreolus* and showed a close relationship with the tuberactinomycin family of NRPS' antibiotics that are effective against multidrug-resistant tuberculosis [28]

1.1.7 Bleomycin

The bleomycins refer to a family of glycosylated peptides that have antitumor activity [31]. The mechanism of action of bleomycin involves chelating ferrous iron which subsequently results in the formation of an iron-bleomycin complex. This complex has the ability to remove the hydrogen of C4' of DNA which in turn causes single or double DNA breaks [32].

1.1.8 Quinoxalines

The quinoxalines family includes many bisintercalator natural products. The bisintercalators are molecules that interact with DNA by placing their two planar chromophores between DNA bases and occupying the minor groove of DNA by their cyclic depsipeptide backbone. Because of their structural differences, the quinoxalines can have different biological activities and sequence selectivity. For instance, thiocoraline

(a compound produced by a marine actinomycete strain L-13-ACM2-092) binds to supercoiled DNA and consequently inhibiting DNA polymerase α [33], Whereas echinomycin is a potent inhibitor of DNA transcription catalyzed by RNA polymerase [34]. Quinoxalines are medically important because of their anti-tumor activity and a potential anti-HIV activity [28].

1.2 Antifungal compounds

1.2.1 Polyene antibiotics

Polyene antibiotics are PKSs produced by some *Streptomyces* species and exert their effect by disturbing the permeability of fungi [35]. Examples of Polyene antibiotics is Partricin that was first isolated from *S. aureofaciens* [36]. Tow semi-synthetic derivatives were developed from Partricin A: SPA-S-753 and SPK-843/SPAS-843. These derivatives have similar efficacy to that of amphotericin B, however, they have shown lower toxicity and longer serum half-life [37,38].

1.2.2 Cispentacin

Cispentacin is a water-soluble antibiotic that was first isolated from the culture broth of *Bacillus cereus* L450-B2. This antibiotic showed weak antifungal in vitro; however it showed strong protection against *Candida albicans* in murine models [39]. Icofungipen (PLD-118) is a synthetic derivative of cispentacin that showed activity against *C. albicans.* This compound exerts it action by inhibiting the isoleucyl-tRNA synthetase and thus interfering with the protein biosynthesis [40].

1.3 Antiparasitic Compounds

1.3.1 Fumagillin

Fumagillin is a secondary metabolite which was first isolated from *Aspergillus fumigates*. This compound has been thoroughly investigated as a potential treatment for microsporidiosis [41], amebiasis [42], and as an angiogenesis inhibitor [43]. This compound has been used for the treatment of Microsporidiosis of honey bees caused by *Nosema apis* [44]. The biosynthetic gene cluster of Fumagillin containing PKS, terpene cyclization, and tailoring modification genes, has been identified [45].

1.3.2 Paromomycin

Paromomycin is an antibiotic that was first isolated from *S. rimosus* var. *paromomycinus.* This antibiotic has been used to treat various parasitic infections [46]. Paromomycin is a drug of choice in the treatment of cutaneous and visceral leishmaniasis. The *in-vitro* use of paromomycin showed that it can selectively target ribosomes of *Leishmania* but barely affect ribosomes of mammalian cells [47].

1.4 Antiviral Drugs

1.4.1 FK506

FK506 is a macrolide compound that was first isolated from *S. tsukubaenis* 9993 and showed an immunosuppressive activity [48]. Later, an *in-vitro* experiment showed that FK506 can interfere with HIV-1 replication in T cells via an unknown mechanism [49]. It has been also reported that FK506 has a strong effect on Orthopoxvirus replication in BSC-40 cells [50].

1.4.2 RP 71955

RP 71955 is a secondary metabolite isolated from a broth culture of a *Streptomyces* strain SP9440. This compound showed an *in-vitro* inhibitory effect against the cytopathogenic effect and the reverse transcriptase of HIV-1 [51].

1.5 Drugs acting on the immune system

1.5.1 Rapamycin (Sirolimus)

Rapamycin is an antifungal antibiotic that was first isolated in 1975 from *S. hygroscopicus* [52]. This antibiotic was initially developed to be used as an antifungal drug, however, due to its immunosuppressive and anti-proliferative activities, the use of this drug as an antifungal was discontinued [53]. Today, this compound is used as an immunosuppressant in transplantation [54] and the treatment of autoimmune diseases [55].

1.5.2 Cyclosporin A

Cyclosporin A was first isolated along with Cyclosporin B from *Tolypocladium inflatum*. Cyclosporin A is a nonpolar cyclic peptide which is composed of 11 amino acids [56]. This NRP initially showed antifungal activity but it was found that it has immunosuppressive and anti-inflammatory activities that made it a good choice for the treatment of autoimmune diseases and the prevention of graft rejection [57]. The immunosuppressive and anti-inflammatory activities of Cyclosporin A come from its ability to suppress the proliferation of T-cells [58].

1.6 Microbial SMs with Cytotoxic activity

1.6.1 Enediynes

Enediynes are a group of antibiotics that have an antitumor effect. The most significant compound that belongs to this group is calicheamicin $\gamma 1^{I}$. The calicheamicins were first discovered in *Micromonospora echinospora* spp. *calichensis* in 1987. Calicheamicin $\gamma 1^{I}$ and its closely related compound dynemicin A represent the precursors of enediynes. The mode of action of enediynes involves activation and a unique rearrangement that leads to their interaction with DNA and subsequently cleavage of its double strand [59]. Calicheamicin has been used as a Calicheamicin Conjugated Monoclonal Antibody and showed strong apoptosis-inducing activity in pediatric acute lymphoblastic leukemia cells [60].

1.6.2 Actinomycins

Actinomycins (actinomycin A and actinomycin B) were first isolated from soil *Streptomyces* in 1940 [61]. Twelve years after, another actinomycin (actinomycin C) was isolated and this antibiotic also showed antitumor activity [62]. Actinomycin D was isolated later and it was approved as a treatment for malignant tumors. However, this compound has limited use because of its high toxicity and the need for careful manipulation before giving it to cancer patients [59]. Five derivatives of Actinomycin D were isolated from *S. iakyrus* strain DSM 41873 in 2006 [63]. Actinomycin D was also isolated from a marine strain (*Streptomyces* sp. MS449) and showed an anti-tuberculosis activity [64]. Actinomycin D was used as pulsed and 5-day chemotherapy to treat patients with low-risk gestational trophoblastic neoplasia. The overall complete remission for the 34

patients who received pulsed and 5-day chemotherapy was 62% and 73%, respectively [65]. Actinomycin D was also used as a second-line treatment for patients with gestational trophoblastic neoplasia (GTN) who received methotrexate as a first-line treatment. All Patients with methotrexate-resistant GTN who received Actinomycin D as second-line and third-line treatment had an overall survival rate of 100% [66].

1.6.3 Anthracyclines

The most valuable anthracyclines are daunorubicin and doxorubicin. Daunorubicin was firstly isolated from *S. peucetius*. Daunorubicin has anti-cancer activity and was approved by FDA in the1960s for the treatment of lymphoblastic or myeloblastic lymphoma. Doxorubicin is structurally similar to daunorubicin and was used to treat soft tissue sarcomas, aggressive lymphomas, and solid tumors in children. However, these compounds can cause high toxicity for the heart that may develop into congestive heart failure. Moreover, these compounds can be used for a limited time and cancer cells can resist their activity [59]. Therefore, thousands of anthracyclines derivatives have been synthesized to obtain safer forms of these antitumor compounds [59].

1.6.4 Rapamycin

Rapamycin was originally isolated from the soil bacterium *S. hygroscopicus* and was intended to be used as antifungal compound [67]. Due to its immunosuppressant effects, this compound was neglected as an antifungal compound. However, in 1984, rapamycin was found to have an anti-tumor activity [68] and its base structure was used for the synthesis of new compounds with several pharmacological properties especially anticancer activities [59].

1.6.5 Geldanamycin

The first geldanamycin was isolated in 1970 from *S. hygroscopicus* var *geldanus*. Geldanamycin was initially isolated to be used as an antibacterial compound, but later its antitumor properties were recognized. The early studies showed that this compound inhibits the tyrosine-specific kinase vSrc that plays a role in cell proliferation and growth regulation [69]. However, other studies showed that geldanamycin can indirectly cause cell death by interacting with the heat shock protein 90 (HSP90) [70] and specifically to the ATP binding site of HSP90 [71].

1.6.6 Histone deacetylase inhibitors

As their names imply histone deacetylase inhibitors (HDACI) are a group of secondary metabolites that hinder the transcriptional regulation of histone deacetylases. The first HDACI was isolated in 1976 from strains of *S. hygroscopicus* and exhibited antifungal properties [72]. Romidepsin (FK228) is another HDACI that was isolated in 1994 from the soil Gram-negative bacterium *Chromobacterium violaceum* [73]. Romidepsin (FK228) was approved by the FDA in 2009 to treat relapsed cutaneous cell lymphoma and in 2011 to treat refractory peripheral T-cell lymphoma [28]. The gene cluster encoding Romidepsin (FK228) in *C. violaceum* was identified in 2007. This HDACI is biosynthesized by an NRPS-PKS-NRPS hybrid in addition to other proteins [74].

1.6.7 Bryostatins

Bryostatins is a group of 20 macrocyclic lactones isolated from the Brown bryozoans *Bugula neritina*. These secondary metabolites are believed to be biosynthesized by the bacterial symbiont (*Candidatus Endobugula sertula*) of the

bryozoan *Bugula neritina*. Bryostatin has been tried in more than 30 phase I and II cancer clinical trials, alone or combined with other anticancer compounds, however, bryostatin has not been efficient enough to be used in phase III clinical trials [75].

1.6.8 Epothilones

The epothilones are novel secondary metabolites that were first discovered while exploring secondary metabolites produced by the soil myxobacterium *Sorangium cellulosum* [76]. However, the antitumor properties of epothilones were recognized in 1995 and their mode of action was found to be similar to paclitaxel [77]. The effectiveness of epothilones against paclitaxel-resistant cells made them attractive for further modification [78].

The epothilones were initially isolated from *Sorangium cellulosum* strain So ce90 in 1996 [76]. These metabolites showed antifungal activity and a cytotoxic activity similar to that of taxol [77]. Epothilones are 16-membered macrolides with a side-chain bound to a thiazole ring [78]. Although epothilones-A and B have a very similar structure, epothilones B differs from epothilones A in containing a methyl group at C12. This structural difference affects the efficiency of these epothilones, as shown by the lower activity of Epothilone A [79]. Therefore, only Epothilone B has been used for developing new semi-syhthetic analogs of epothilones. At first, this Epothilone showed promising results against tumor cell lines; however, its efficiency was not high in tumor xenografts [80]. The inefficiency of Epothilone B in rodent's xenografts was attributed to the high rate of Epothilone B conversion by plasma esterase in these models [81]. However, the esterase

activity in humans turned out to be lower than that of rodents and this encouraged a drug developer, Novartis, to synthesize a semisynthetic version of Epothilone B [78].

During the biosynthesis of epothilones A and B, *S. cellulosum* also synthesizes small amounts of intermediate compounds named epothilones C and D. These precursors differ from epothilones A and B in lacking the C12-C13 epoxide. Epothilone D differs from Epothilone C in having a methyl group at C12. The *in-vitro* experiments showed that epothilones C and D are more potent than epothilones A and B. Similar to Epothilone B, Epothilone D showed sensitivity to rodent's esterases and therefore Kosan Biosciences developed two semisynthetic analogs from Epothilone D. However, these developed analogs did not reach the stage of clinical trials [78].

The relatively simple structure of epothilones makes its chemical synthesis possible and aids in developing new semi- synthetic analogs. Specifically, their 16-membered macrolide structure is helpful to perform a huge number of synthetic methods and this property was exploited to perform several structure and activity relationship (SAR) studies which resulted in the production of various semi- synthetic analogs [76]. Current analogs of epothilones are derived from epothilone-B (ixabepilone and sagopilone), or derived from epothilone-D (KOS-862, KOS-1584, and KOS -1591) [78].

The pioneers of developing analogs of epothilones at the end of the 20th century were Bristol-Myers Squibb (BMS) Company which focused on semi- synthetic approaches instead of finding natural forms of these compounds. This company focused on the production of epothilones analogs that can withstand plasma esterases. Therefore, they exchanged the lactone group of Epothilone B with a lactam and obtained the second

generation semi- synthetic analog (BMS-247550) which was given the generic name ixabepilone [82]. This analog showed improved resistance to plasma esterases and also improved pharmacokinetic properties in rodent models. For instance, the half-life of this analog was over forty times more than Epothilone B [83]. Sagopilone is another Epothilone analog that was that was fully synthesized based on Epothilone B structure [84].

The analog KOS-862 was used in a phase II study of patients with metastatic androgen-independent prostate cancer [85]. However, the investigators of this study did not recommend further studies and clinical trials using this compound due to its inefficiency and its high toxicity. Therefore, other second-generation semi-synthetic analogs were developed and they were named KOS-1584, and KOS -1591 [86].

1.2 Genome mining for the discovery of novel secondary metabolites

The advancement in next-generation sequencing (NGS) strategies has provided insight into microbial genetic variation and led to the discovery of more microbial genes encoding secondary metabolites. Before the era of NGS, searching for microbial secondary metabolites was mainly performed using traditional approaches that directly detect the production of metabolites in microbial growth media. These approaches include the bioactivity-directed and compound-directed high-throughput approaches [87], and both approaches may only lead to re-identification of known compounds. For instance, streptomycin and streptothricin can be frequently isolated from 1% and 10% of actinomycetes, respectively [88]. The frequent re-isolation of secondary metabolites from related microbial species led to the belief that microbes can no longer be a rich source of new secondary metabolites such as antibacterials, anticancer compounds, and others. In

the era of NGS, and the availability of bioinformatics tools, the inspection of microbial genomes to decipher their contents of secondary metabolites has become easier than ever. For instance, the genomes of *S. avermitilis* [89] and *S. coelicolor* [90] were thoroughly investigated and found to encode previously unidentified natural products. These findings drew the attention of scientists to the presence of undiscovered genes in other microorganisms that would also encode bioactive compounds. Since that time, the concept of genome mining has been adopted. Genome mining is defined by Paterson et al as "systematic collection, analysis and interpretation of mounds of data ('big data') with the goal of discovering significant new traits, relationships, patterns or trends" [91]. In terms of discovery of secondary metabolites, genome mining refers to deriving information gathered from inspection of microbial genomes and using this information to predict which products, such as bioactive compounds or natural products, can different genes encode. It has been found that even well-studied genomes of microorganisms may contain previously undiscovered gene clusters that encode bioactive compounds. Therefore, it is necessary to re-inspect microbial genomes to discover more gene clusters that encode new bioactive compounds. There are many methods that are used for mining of microbial genomes: classical, phylogeny-based genome mining, resistance based and target based mining, regulator based mining and comparative genome mining [92].

1.2.1 Classical methods of genome mining

Classical genome mining involves searching for the genes encoding enzymes that are expected to participate in the biosynthesis of secondary metabolites. These methods rely on the assumption that many biosynthetic machineries of secondary metabolites are conserved. A closer look at the amino acid sequences of the core biosynthetic enzymes of

biosynthetic machineries that are involved in the biosynthesis of PKs, NRPs, and post-translationally modified peptides (RiPPs) shows this conservation. This conservation can be exploited for the *in silico* mining of specific genes or gene clusters involved in the biosynthesis of many secondary metabolites [92]. For instance, the use of a sequence or multiple sequences of a core enzyme can be used as a seed to find its/their homolog(s) and consequently the potential gene cluster encoding these enzymes.

1.2.2 Phylogeny-based genome mining

The principle of phylogeny-based method is based on the possibility of co-evolution of a biosynthetic gene along with its gene cluster. In this regard, a sequence of a core biosynthetic gene along with known and unknown sequences is used to build a phylogenetic tree to infer their evolutionary relationships and consequently the evolutionary relationships of gene clusters containing these genes. The clustering or divergence of the gene(s) of interest with known genes might indicate how closely related is/are the gene(s) of interest to other known genes. Subsequently, this evolutionary relationship can help us infer how novel or similar are the gene clusters containing the gene of interest to known gene clusters [93].

1.2.3 Resistance and target based genome mining

Resistance based and target based mining methods are relatively new. These methods rely on that fact that the microorganism producing antibiotics or other secondary products that affect other organisms must have resistance mechanisms for their own protection. It is now known that biosynthetic gene clusters (BGCs) do not only contain genes that form the biosynthetic machineries but also other genes that encode regulatory

proteins, transporter proteins and resistance proteins [92,94]. Resistance-based and target-based mining methods find BGCs in microbial genomes by locating the resistance genes within the genome and therefore the potential gene cluster that biosynthesizes the secondary metabolite of interest. Thaker et al used this principle to identify gene clusters that encode novel antibiotics. The potential gene clusters were compared to previously known gene clusters and their relatedness helped in the identification of bacterial strains that are potentially capable of producing novel antibiotics without the need for full genome sequencing of these strains [95]. [1]

1.2.4 Regulator-based genome mining

Regulators play an important role in the activation of gene clusters. The activation of many fungal gene clusters is controlled by cluster-specific regulators. The highly conserved binding sites of many regulators of BGCs can be exploited to find related but novel BGCs. Recently, a software named Cluster Assignment by Islands of Sites (CASSIS) implemented a promoter-based method for prediction of BGCs in eukaryotic genomes. CASSIS relies on the presence of common regulatory patterns in BGCs regulators to find undiscovered and potentially novel BGCs [96].

1.2.5 Comparative genome mining

In comparative genome mining methods, the goal is to search for groups of genes that make up the gene clusters of SMs [92]. One of the most valuable tools that are based on the comparative genome methods is antiSMASH. This software identifies the potential BGCs in the query sequences and then compares those BGCs with a large number of microbial BGCs to find their degree of novelty [97].

1.3 Discovery of microbial SMs by genome mining

The discovery of novel SMs by mining microbial genomes has become an important approach. With the availability of thousands of microbial genomes that are freely open to all, it is now possible to explore those genomes for gene clusters and link them to the SMs they encode. Moreover, the rapid development of high-throughput sequencing methods and the availability of many genome mining techniques made it easier than before to discover novel microbial SMs.

Genome mining of *S. xinghaiensis* NRRL B-24674T led to the discovery of many gene clusters including the one that encodes a compound that has anti-complement activity and the gene clusters that encode xiamycin analogs. The updated genome sequence of this strain was analyzed using antiSMASH to find potential secondary metabolites BGCs. This analysis showed that this strain genome contains 24 secondary metabolites BGCs not found in the phylogenetically close species to *S. xinghaiensis* NRRL B-24674T which indicates that this species is a potential source of novel secondary metabolites. In addition to the BGC that encodes an anticomplement compound, the BGC that encodes xiamycin analogs, and the BGCs that encode carotenoid, desferrioxamine B, ectoine, and neomycin, eighteen novel secondary metabolites BGCs were also identified. The antiSMASH analysis was confirmed by the gene knock out experiments of the BGC that encode the anticomplement, and HPLC separation that purified nine xiamycin analogs [98].

Genome mining of e *S. marokkonensis* M10 a close strain to *S. marokkonensis* Ap1T that is known for the production of the antifungal polyene macrolide known as

pentaenes, revealed plenty of BGCs in this strain. The entire genomic DNA sequence of this strain was analyzed by antiSMASH that identified 29 BGCs. The most interesting secondary metabolites BGC was the Type I PKS (*pks1*) that showed high similarity with candicidin/FR008 gene cluster that encodes an antifungal polyene macrolide. To correlate this *in silico* prediction of a *pks1* gene cluster with antifungal activity, crude extracts of *S. marokkonensis* M10 were fractioned and tested against *Candida albicans*. The fractions that showed growth inhibition were separated by Sephadex LH-20 chromatography and this resulted in the identification of tow polyene family compounds PF1 and PF2 [99].

Genome mining of the rare actinomycete *Streptosporangium* sp. CGMCC 4.7309 revealed that the genome of this species encodes several SMs. Analysis by antiSMASH of its gnome revealed the presence of more than 20 BGCs. The focus of this study was on one of the BGC that contains type II PKSs. This gene cluster contains 33 open reading frames and shows high similarity to *hex* gene cluster of *S. roseum* NI 9100T. The product of *hex* gene cluster of *S. roseum* NI 9100T is colorful SM that resembles other pentangular polyphenols. To study the product of *hex* gene cluster of *Streptosporangium* sp. CGMCC 4.7309, a mutant (CIM-H001) with a disrupted *hex* gene cluster was constructed. Therefore, this gene cluster product would not produce the expected colorful products. The growth medium of *Streptosporangium* sp. CGMCC 4.7309 contained red compounds that were absent in the CIM-H001mutant which indicates that the *hex* gene cluster is the gene cluster that produces these compounds. Analysis by HPLC and fractionation by chromatography processes resulted in the isolation of hexaricins A, B, and C [100].

Searching for SMs gene clusters in *Streptomyces* sp. IB2014/011-12 led to the discovery of novel gene clusters including a hybrid NRPS-trans-AT-PKS. Analyzed by antiSMASH, the genome of *Streptomyces* sp. IB2014/011-12 predicted to contain 29 BGCs. The culture broth of this species was analyzed by dereplication to identify SMs. By using the dereplication method the known compounds are neglected and only unknown compounds can be purified and analyzed. The LC-MS data of the extracts of an organism are compared to data banks and only compounds that are not seen in the data banks are subjected to further analysis. The LC-MS data of the extracts of *Streptomyces* sp. IB2014/011-12 identified several polycyclic tetramate macrolactams (PTMs). One of these PTMs was identified as alteramide A and several isomers were not distinguished. These compounds were identified by HPLC and mass spectrometry. Based on the structure of the isolated compounds, the synthesis of these compounds was thought to be a function of PKS and an NRPS and it is the NRPS in this cluster that adds the glycine to the structure. There was only one gene cluster that was though to add glycine. This cluster encodes a hybrid NRPS-trans-AT-PKS. Gene deletion experiments performed on this gene cluster confirmed that it is responsible for the production of alpiniamide A and its derivatives [101].

The genomes of Subsection V cyanobacterial genomes in addition to the genomes of *H. welwitschii* UH strain IC-52-3 and *W. intricata* UH strain HT-29-1were mined for SMs gene clusters. A wide range of SMs gene clusters was identified including hapalosin, and microcystin. Moreover, the post-translationally modified peptides (RiPPs) gene clusters were also identified including those encoding microviridin, cyanobactin,

and a bacteriocin. The SMs gene clusters were identified by antiSMASH and BLASTp [102].

Rhizopodin biosynthetic gene cluster was discovered in *Stigmatella aurantiaca* Sg a15 by combining genome mining and statistical tools. This study used the 454 sequencing method to obtain the draft genome of *S. aurantiaca* Sg a15. This method resulted in the synthesis of 7.8 kb contigs, and this length of the DNA sequence hinders the discovery of BGCs with sizes between 30 and 80 kb. However, closer inspection of these contigs may show some domains of PKSs and NRPSs. Many contigs contained domains that together did not show similarity to any known gene clusters. Two of these contigs showed similarity to trans-AT PKS/NRPS hybrids and they were thought to belong to the same gene cluster. To identify the product that this potential gene cluster can produce, tow *S. aurantiaca* Sg a15 mutants were constructed by deleting the KS domains that correspond to KS domains of these contigs. By comparing the SMs profiles of these mutants with their wild type and the use of statistical tools to detect minor variations in the mutants' metabolic profiles, it was found that this gene cluster is responsible for producing Rhizopodin [103]. Rhizopodin is a potent actin-binding anticancer compound [104]. In order to identify the whole gene cluster the gaps between the two contigs were filled by PCR and this gene cluster was found to be PKS/NRPS hybrid gene cluster containing 13 genes.

The rhizosphere bacterium *Pseudomonas sp*. SH-C52 was mined for antimicrobials and this led to the identification of twelve BGCs. Six of these gene clusters were NRPSs gene clusters. The complete genome of this species was analyzed with antiSMASH and NP.searcher. The 16S rRNA phylogenetic analysis placed

Pseudomonas sp. SH-C52 within *P. fluorescens* clade, and within the subgroup I of this clade. To identify BGCs that encode antibiotics and other SMs compounds the genome of *Pseudomonas sp*. SH-C52 was screened with BLAST in order to find ORFs of common compounds produced by members of *P. fluorescens* clade. The in silico genome mining of *Pseudomonas sp*. SH-C52 identified six NRPS gene clusters out of the twelve identified.

1.4 Statement of the problem

Natural products including SMs have become a very important source of compounds that have diverse applications including medical and pharmaceutical applications. It is currently estimated that more than 50% of the compounds used for medicinal purposes are originated from natural products [105]. Therefore, mining genomes of microbes using different bioinformatics tools can pave the way for the discovery of novel SMs with different beneficial applications. With the development of high-throughput sequencing methods, there is a wealth of microbial genomes data that can be mined through *in silico* approaches to discover gene clusters that potentially encode novel SMs.

The World Health Organization (WHO) estimated that cancer is one of the world's leading causes of death in 2015 [106]. Treatment of many types and different stages of cancer may result in the development of drug resistance and subsequently to death. We are in need for a wide range of novel SMs with anticancer activities to be used as an arsenal to fight cancer drug resistance and to be used with other cancer therapeutic means, or in anticancer drug cocktails.

The WHO also warned that there is a need for developing new antibiotics to face the emerging antibiotic-resistant bacteria. The WHO also stated in its recent report that among the antibiotics that are in the clinical pipeline or being clinically developed, only a few of them can be considered novel. Therefore, we are in need of more novel antibiotics to treat serious infections caused by antibiotic-resistant bacteria [106].

Emerging viral infections (e.g H5N1 influenza, severe acute respiratory syndrome (SARS), West Nile fever) may appear due to changes in people's behavior such as the expansion of their living areas and the ease of global travel [107]. The low number of antivirals available to treat life-threatening and emerging viral infections necessitates the presence of more antivirals.

Several publications have cautioned that fungi can infect billions of people around the globe annually [108,109]. Although most of the fungal infections are not life-threatening, however, under certain conditions they can be as fatal as tuberculosis and malaria [108]. The development of antifungals is advancing slowly [110]; therefore, the control of life-threatening fungal infections requires the availability of a broad range of novel antifungals.

The *In silico* mining of microbial genomes is a valuable approach with which we can focus on the gene clusters that can potentially encode novel anticancer and antimicrobial compounds that have better pharmaceutical properties and fewer side effects. In this study, the microbial genome mining approach was used to identify microbial gene clusters that potentially encode novel anticancer agents and antimicrobials. Initially, the gene sequences of *Sorangium cellulosum* So ce90's genes

comprising *Sorangium cellulosum* So ce90's Epothilone gene cluster were used to identify their homologs in bacteria and subsequently to find epothilone-similar gene clusters (ESGCs). Additionally, the entire genomic DNA sequences of bacteria that contained ESGCs were analyzed by antiSMASH to find hybrid PKS/NRPS hybrid gene clusters. Using this approach, not only ESGCs but also new gene clusters similar to those encoding cytotoxic compounds and antimicrobials were identified.

1.5 The need for this study

Noncommunicable diseases (NCDs), also named as chronic diseases, are the diseases that last for a long period of time (http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases). The leading types of NCDs are cardiovascular diseases, cancers, and diabetes. According to the World Health Organization (WHO) estimation in 2016, the number of deaths in millions (percentage) of deaths of all global NCDs that resulted from cardiovascular diseases, cancers, and respiratory diseases were 17.9 (44%), 9.0 (22%), and 3.8 (9%), respectively (https://www.who.int/gho/ncd/mortality_morbidity/en/). According to the WHO's estimation in 2015, the world's first or second leading cause of death among people under 70 years of age in 91 of 172 countries was cancer. Cancer is also the third or fourth leading cause of death in another 22 countries [111].

Chemotherapy drugs are currently used to treat many types and different stages of cancers. However, the use of anticancer drugs may result in the development of chemotherapeutic resistance and consequently can lead to death. The development of resistance to chemotherapeutic drugs is attributed to macroscopic and microscopic factors

[112]. One of the main macroscopic factors that affect the activity of drugs is their pharmacokinetics. Pharmacokinetics is a term used to describe the reaction of the body to administered drugs. Pharmacokinetics involves four basic steps Absorption, Distribution, Metabolism, and Excretion (ADME) [113]. The absorption of drugs is affected by the presence of a multidrug drug resistance protein known as permeability glycoprotein (P-gp). This protein is present on the entire length of the gastrointestinal tract [114]. The oral administration of some anticancer drugs may result in the overexpression of P-gp which consequently results in the reduction of bioavailability of these drugs [115]. The absorption of anticancer drugs is also affected by diet. For instance, grape juice has been found to inhibit the function of the intestinal CYP3A4 that act as a barrier to small foreign organic molecules (xenobiotics). The inhibition of CYP3A4 may boost the absorption of drugs including anticancer drugs with low bioavailability [116]. Metabolism of anticancer drugs involves the action of groups of enzymes known as cytochrome P450 (CYP450) and glutathione S–transferase (GSTs). CYP450 can activate [117] or inactivate [118] some drugs. Cytochrome P450 enzymes activate anticancer drugs by adding polar or reactive groups [112]. However, overexpression of CYP450 may lead to accelerated inactivation of the anticancer drug, and hence resistance. GSTs overexpression may also lead to resistance to anticancer drugs because of their action as drug inactivators [119].

Resistance to anticancer drugs could also be attributed to microscopic factors such as drug residency. Drug residency can be affected by the expression of certain proteins. Multidrug resistance protein 1 (MDR1) plays a role in metabolites secretion. MDR1 has a positive effect on the expression of CYP3A4 [120]. The overexpression of CYP3A4 can

lead to deactivation of some anticancer drugs [112]. Another protein that controls trafficking metabolites, including anticancer drugs, is the mitoxantrone resistance protein (MXR). MRPs are multidrug resistance-associated proteins and their overexpression was found to be associated with multidrug-resistance of human lung cancer cell line [121]. MXR can play the same role of MDR1 and MRPs proteins in controlling metabolites trafficking and causing resistance to anticancer drugs [122].

Other microscopic factors can also impact the resistance of anticancer drugs. The microenviroment of oxygen, glucose concentration, and pH of cancer cells play a role in developing resistance to anticancer drugs. Most anticancer drugs exert their effect on cancer cells by activating apoptosis pathway that is induced by the presence of free radicals [123]. Therefore, resistance to anticancer drugs can occur under low oxygen concentration [123]. Glucose plays a role in enhancing the growth of tumors [124], and also in developing resistance to anticancer drugs [125].

The need for a wide range of anticancer drugs is becoming necessary as these drugs can be used as anticancer drug cocktails, or used with other therapeutic means such as radiotherapy, biotherapy, targeted anticancer drugs, antimetastatic drugs, etc. The combination of cytotoxic anticancer drugs with other treatment strategies can provide good solutions for cancer treatments [126]. Due to multiple genetic and molecular changes that occur in cancer cells, the use of single anticancer drugs would not be effective for the treatment of cancers [126]. Several studies showed that treatment of cancers by using cocktails of drugs rather than single doses would be more effective in controlling cancer progression [127-129].

More novel antibiotics are also needed to combat emerging antibiotic-resistant bacteria. The World Health Organization (WHO) has recently warned that development of new antibiotics that are able to counteract the growing threat of antimicrobial resistance is seriously insufficient. According to a recent report of WHO, only eight out of 51 new compounds being clinically developed to combat antibiotic-resistant pathogens can be considered novel. This report also mentioned that current antibiotics in the clinical pipeline are not novel antibiotics but modifications of known antibiotic types, and this boundaries the number of antibiotics that can treat serious infections caused by antibiotic-resistant pathogens. There is also a need for more oral antibiotics that can be used for treating infections outside health care facilities [106].

There is also a need for more antivirals to treat deadly and emerging viral infections. Factors contributing to emerging viral infections (e.g H5N1 influenza, severe acute respiratory syndrome (SARS), West Nile fever) include changes in people's behavior such as the expansion of their living areas and the ease of global travel [107].

There are a low number of antivirals available for treating viral infections due to the high cost of developing these drugs. Most of the antivirals that have been developed were those used against HIV and hepatitis C virus (HCV) infections due to their medical significance and their continuous resistance to antivirals [107]. Although vaccination can protect against viral infections, however, more new antivirals are needed to cope with the developed antiviral resistance and emerging viral infections.

Several publications have warned that fungal infection can really impact the health of billions of people around the world [108,109]. Although most of the fungal

infections are not serious infections, however, they can be as fatal as malaria and tuberculosis [108]. The treatment and control of serious fungal infections require the availability of a broad range of antifungal drugs, which has not achieved yet. Indeed, the development of antifungal drugs that can treat fatal fungal infection has advanced slowly [110], and only 3 classes of antifungal drugs have been developed since the beginning of the 20th century [130]. An example of the slow advancement of developing antifungal drugs is amphotericin B. Since its development in the late 1950s, amphotericin B has been used as a drug of choice for the treatment of the fatal fungal infection caused by Cryptococcal meningitis, and some other emerging fungal infections in immunocompromised patients. Taking together the relative paucity of antifungal drugs and the emergence of certain species of *Candida*, such as *C. glabrata* and *C. auris* [131], necessitate the need for more novel antifungal drugs.

Other SMs with medical applications are also required. Immunosuppressant drugs are used to prevent the body's immune response from rejecting organ transplants. Several immunosuppressant drugs of microbial origin, such as cyclosporine A, tacrolimus, sirolimus, gliotoxin, and mofetil have been used in organ transplantation [132]. Most of the patients who take immunosuppressant drugs can have some complications, such as hypertension, diabetes, malignancy, etc. and these complications might increase mortality and morbidity among those patients [133]. Therefore, there should be a wider range of immunosuppressant drugs with fewer side effects.

The use of microbial enzyme inhibitors is growing as they can potentially be used in medicine. Amylase inhibitors are compounds that interfere with starch hydrolysis that yields dextrins and sugars. Therefore, these inhibitors can be used to control obesity and

to lower blood glucose levels in type-2 diabetic patients [134]. The clavulanic acid is used to block β-lactamase produced by some antibiotic-resistant bacteria and therefore, it can help tackle antibiotic resistance [135].

Mining of microbial genomes to discover novel SMs is of great importance. The main limitation of exploring microbial genetic and metabolic diversity is the uncultivability of most microorganisms. According to the American Academy of Microbiology (AAM) report entitled "The Microbial World: Foundation of the Biosphere", millions of microorganisms are still undiscovered. AAM also stated in this report "it is conservatively estimated that less than 1 % of bacterial species and less than 5 % of fungal species are currently known" [136]. Therefore, the exploration of more microbial genomes, including genomes of unculturable microbes, might provide more SMs with novel biological activities. Moreover, SMs can be exploited in combinatorial biosynthesis to produce novel bioactive compounds. Combinatorial biosynthesis can be defined as "the application of genetic engineering to modify biosynthetic pathways to natural products in order to produce new and altered structures using nature's biosynthetic machinery" [137]. A notable example of using combinatorial biosynthesis to produce new structures is the scale-up production of Epothilone D from Epothilone B [138]. SMs also can be used as pharmacophores (molecular structures in drugs that determine their biological or pharmacological interaction with their targets) which then can be modified to synthesize new analogs with similar or enhanced activity [139].

SMs as a part of natural products are important sources of leads that can be used for the production of drugs via combinatorial chemistry. In combinatorial chemistry, thousands of compounds are produced in mixtures without knowledge of the compounds'

structures. The biological activity of the whole mixture is tested for specific activities and only when the desired biological activity is found then the compounds that show this activity are identified [140]. The use of scaffolds of natural origin in combinatorial chemistry can result in the production of library members with promising bioavailabilities and bioactivities [141].

CHAPTER II

LITERATURE REVIEW

2. Nonribosomal peptides (NRPs) and polyketides (PKs)

Nonribosomal peptides (NRPs) and polyketides (PKs) are two classes of SMs produced by various organisms. The diverse structural and biological activities of NRPs and PKs are attributed to their mode of synthesis. For example, the antibiotics penicillin, erythromycin, vancomycin [142], Monensin A, 6-methylsalicyclic acid [143], the antifungal compounds amphotericin B [143], Myxothiazols [144] and Echinocandins [145], the anticancer substances romidepsin [146] , bleomycin, and epothilone[142], and the immunosuppressive agent cyclosporine A [142] are biosyntesized by non-ribosomal peptide synthetases (NRPSs) or polyketide synthases (PKSs).

NRPs are biosynthesized by biosynthetic mega enzymes called non-ribosomal peptide synthetases (NRPSs) [147]. NRPSs are composed of modules within which there are different domains, and each domain is responsible for performing a certain task during the biosynthesis of NRPs [142]. NRPs are chiefly produced by microorganisms with a length range from 2 to 48 residues. The building blocks of NRPs involve D-amino acids, N-methylated amino acids and other non-proteinogenic residues [142]. These building blocks can be linked together to form linear, branched or cyclic structures. The biosynthesis of the final products of NRPs may include some modifications such as heterocyclizaton, acylation, or glycosylation [142].

PKs are biosynthesized by another group of modular mega-enzymes called polyketide synthases (PKSs) [147]. PKs are successively biosynthesized by a process known as decarboxylative condensations, which resemble the biosynthesis of fatty acids. PKSs are categorized into three types. PKSs Type I biosynthesize considerably reduced polyketides that range from polyethers, macrolides, polyenes, and hybrids of PKs/NRP [142]. PKSs Type I might contain one module or several modules. Each module might contain several domains, and each domain has a different function during the elongation of products [142]. Bacterial PKSs Type I contain non-iterative modules that are used for one cycle of polyketide chain elongation and modification. On the other hand, fungal PKSs only contain one module, and each of these modules is iteratively used during the biosynthesis of PKs [142]. Type II PKS are multifunctional enzymes used by bacteria to biosynthesize multi-aromatic compounds. Examples of Type II PKSs are those involved in the biosynthesis of tetracenomycin and actinorhodin [148].

The third type of PKSs was first discovered by two groups [149,150] and found to be involved in the biosynthesis of bacterial small aromatic products. The sequence, mechanism of catalysis and primary structure of Type III PKSs differentiate them from Type I and Type II PKSs [151]. Unlike Type I and II, the substrates of Type III PKSs are free CoA thioesters and their enzymatic reaction does not involve the use of 4'-phosphopantetheine residues on acyl carrier proteins [148].

2.1 Historical View

With the discovery of the ribosomal protein synthesis at the beginning of the 1960s, scientists start thinking about how *Bacillus* species can produce cyclic peptides

from D-amino acids. Mach et al [152] showed that the biosynthesis of tyrocidine by *Bacillus* species was not affected by the use of ribosome inhibitors. Based on this result those authors assumed that the biosynthesis of this natural product may involve the use of different mechanisms of protein synthesis.

Further studies on the biosynthesis of ediene [153] and gramicidin S [154,155] in *Bacillus brevis* and the studies on the biosynthesis of polymixin B in *B. polymyxa* [156] indicated that blockage of protein synthesis using RNases or ribosome inhibitors did not affect the production of these peptides. Therefore, the conclusion that was drawn from these experiments is that the biosynthesis of these compounds must involve ribosome-independent mechanisms. It turned out later that these non-ribosomal proteins and other natural products are biosynthesized by NRPSs and PKSs.

2.2 Natural Functions of SMs

The production of SMs in the natural environment of microorganisms might significantly improve their survival and adaptability. For instance, the SMs producers can scavenge more nutrients than non-SM producers and they also have a better ability to eliminate their competitors [157]. Many organisms might biosynthesize their SMs as a result of their morphological and developmental fluctuations [158]. In SM-producing bacteria, SMs are produced in the stationary phase of their growth where exhaustion of nutrients can result in changing the bacterial regulatory events which in turn affects the morphology of bacteria and enhances production of SMs [159].

2.3 Natural Functions of SMs in Fungi

It is obvious that the biosynthesis of SMs can be advantageous and provide their producing microorganisms with better adaptation to their environments [157]. However, the reasons behind the production of these peptides by only some microorganisms remain unclear. In fungi, SMs are produced as antibiotics to help them survive better than other competing microorganisms in their environment. For instance, some soil fungi produce antibiotics when there is high competition from other inhabitants [160].

SMs might also be produced by fungi during their asexual reproduction. Fungi might produce pigmented SMs to absorb light energy that causes destruction to their spores' genetic material [160]. For instance, *Alternaria alternate* produces melanin for better survival and protection from UV light. The mutants of this fungus with disrupted melanin gene have smaller conidia and show more sensitivity to UV light compared to their wild type counterparts [161].

Siderophores are other SMs that have been thoroughly investigated in fungi [162]. Siderophores are NRPs that play a role in controlling the interaction between fungi and their hosts. These compounds are essential for iron acquisition, virulence, protection from oxidative stress and iron-induced toxicity, and influence sexual and asexual development [163].

Some studies showed the importance of siderophores in the pathogenicity of *Aspergillus fumigatus* in animal models [164,165]. Another compound with similar functional characteristics similar to siderophores has been identified in *A. fumigatus.* [166]. Activation

of this compound's pathway resulted in augmentation of the pathogenicity of *A. fumigatus*.

Siderophores may play a role in maintaining the mutualistic mycorrhizal relationships of some fungi with their host plants. For instance, some fungal siderophores facilitate the solubilization of iron oxides, and thus augment the iron uptake by the host plant [167]. A study by Haselwandter et al [168] showed that at low iron concentration, certain strains of *Ceratobasidium* and *Rhizoctonia* produce basidiochrome, a siderophore that enhances the mutualistic mycorrhizal relationships between these fungi and their host plants. Another compound named Epichloënin A, encoded by a non-ribosomal peptide synthetase gene (sidN) of the filamentous fungus *Epichloë festucae*, is also essential for maintaining the mutualistic mycorrhizal relationship and the internal growth of *E. festucae* in perennial ryegrass (*Lolium perrene*). Experimental knocking down of sidN abolished production of epichloënin A both in vitro and in *L. perrene*, and this consequently affected the sporulation and the polarized hyphal growth of ΔsidN mutants of *E. festucae.* Moreover, the disruption of *sidN* deformed the hyphal ultrastructure of *E. festucae* and its normal localization and distribution within *L. perrene* [169].

Some SMs play a role in the pathogenicity of some fungi by changing their cell wall structure which in turn affects their hydrophobicity. For instance, the plant pathogen *Fusarium graminearum* that causes Fusarium head blight can produce an NRP (FgNRPS4) encoded by NRPS4 gene and affect the shape of its cells. It has been found that Knocking down of NRPS4 homologs in other fungi affected their ability to repel water. Moreover, deletion of this gene in these fungi resulted in significant morphological changes which indicate that this product has a significant role in maintaining the structure

of *F. graminearum's* cell wall [170]. Another NRP, an orthologue of FgNRPS4, produced by the corn pathogen *Cochliobolus heterostrophus* and has a similar structure to FgNRPS4 also has the same effect on hydrophobicity [171]. *Alternaria brassicicola*, a necrotrophic pathogen that causes black spot disease produces an NRP that is entirely involved in conidia formation. The deletion of AbNPS2 gene that produces this NRP adversely affected the hydrophobicity and spore cell wall morphology of *A. brassicicola*. Abnps2 mutants of *A. brassicicola* showed swollen cell wall and more lipid bodies compared to their wild type counterparts. Moreover, the *in-vivo* and *in-vitro* sporulation rates of those mutants were lower. These results indicate the importance of AbNPS2 in the development and pathogenicity of *A. brassicicola* [172].

2.3 Natural Functions of SMs in Bacteria

The reasons behind the production of SMs in bacteria are still not well understood. However, some SMs may play a vital role in the survival of some bacteria by suppressing the growth of other competitors in their surrounding environment. For instance, *Agrobacterium radiobacter* K84 can produce the antibiotic agrocin 84 in the rhizospheres of plants. This antibiotic inhibits the growth of tumorigenic strains of *Agrobacterium tumefaciens*. Therefore, *A. radiobacter* K84 is being used to control crown gall disease around the world [172].

Some SMs regulate the growth and the mode of interaction of bacteria with their hosts. González et al showed that *Pseudomonas aeruginosa* harbor genes encoding putative multi-modular-NRPS (MM-NRPS). Mutation in MM-NRPS genes affects the synthesis of bacterial cyclodipeptides (CDPs) that function in a similar way to auxin.

Therefore, minimized amounts of CPDs lead to a decrease in the quorum-sensing-LasR dependent signaling and therefore to repression of primary root growth of the host plant [173].

Bacterial SMs may play a role in inter-kingdom signaling as a mean of communication. *Ralstonia solanacearum,* a soil-borne pathogen that causes harmful fading of various plants, can biosynthesize ralsolamycin that facilitates an inter-kingdom signaling and communication with other fungi. Ralsolamycin, is a PKS-NRP hybrid that has a role in facilitating the penetration of fungal tissues by the bacterium *R. solanacearum.* This compound also induces tens of fungal species to produce chlamydospores, the important structure that helps fungi survive in unfavorable conditions [174]. In *R. solanacearum*, ralsolamycin production is encoded by *rmy* genes that are controlled by PhcBSR quorum sensing system. Mutation in the sensing system gene *phcB* has a significant negative impact on the expression of genes involved in ralsolamycin production and therefore the colonization of *R. solanacearum* in fungi and the production of the chlamydospores [175].

Bacterial SMs may function as mineral scavengers. For instance, *P. Putida* and *P. fluorescens* suppress plant pathogens in the rhizosphere of plants by excreting high amounts of siderophores that in turn sequestering most of the available iron available for those pathogens [176].

Lipopeptide biosurfactants (LPs) is another category of SMs that is naturally produced by bacteria. LPs is a big group of biosurfactants that can be subdivided into several structurally diverse groups. LPs structure takes the form of short oligopeptides

attached to a lipid tail. Bacterial LPs can be synthesized by NRPSs or PKSs [177]. LPs provide their producing bacteria with many benefits that may include inhibition of other microorganisms, attachment of bacterial cells in biofilms, and motility [178].

2.4 Characteristics of NRPSs and PKSs

NRPSs and PKSs share similar structural and biosynthetic properties [179]. Both are mega-enzymes composed of modules, and each module contains different domains that carry out partial reactions to produce the final SM product [180]. Moreover, both types exploit carrier domains: NRPSs use PCPs and PKSs use ACPs as carrier proteins of the substrates, and 4′- phosphopantetheine cofactor for the pre-activation of the substrates that are incorporated in the biosynthesis of peptides. The amino acids activation of NRPSs yields amino acyl-S-PCPs and the activation of carboxylic acids of PKSs yields acyl-S-ACPs. NRPSs and PKSs also use the same strategy to elongate their chains. NRPSs exploit the C domain to form the C–N bond that catalyzes the nucleophilic attack of the peptidyl-S-PCP of the upstream module with a nitrogen nucleophile of the amino group of the downstream amino acyl-S-PCP. PKSs elongate their chains and form the C–C bond between their subunits using their KS domains that catalyze the nucleophilic attack of the upstream acyl-S-ACP with a carbon nucleophile of the downstream acyl-S-ACP [142]. Hybrid NRPS/PKS are biosynthesized by hybrid NRPSs/PKSs. Their production involves intermediates bound to NRPSs that are elongated further by PKSs or *vice versa* [181].

2.4.1 Characteristics of Bacterial and Fungal PKSs

The classification of PKSs relies on the structural arrangement of their domains. Type I PKSs are usually found as a single large polypeptide chain (module), containing several functional domains that catalyze the reactions necessary to the biosynthesis of PKs. Type I PKSs are usually sub classified into two main groups: Iterative and non-iterative [99]. The iterative Type I PKSs are mainly found in fungi whereas non-iterative are mainly found in prokaryotes [182]. The fungal iterative type I PKSs usually contain a single-module consisting of one copy of functional domains. The domains of iterative type I PKSs can repetitively perform several catalytic cycles to produce PKs. Prokaryotic non-iterative type I PKSs contain several modules and each module have several functional domains required for extending and tailoring the PKs. Type II PKSs are mainly found in bacteria and produced as separate modules, each with several domains that are iteratively involved in catalyzing the extension and tailoring the PKs [183]. The PKSs that biosynthesize actinorhodin and tetracenomycin are examples of bacterial Type II PKSs [148]. Type III PKSs can be found in bacteria and fungi and other organisms [184]. Unlike other PKS types, type III PKSs typically contains one domain (KS) that performs all reactions required for PKs biosynthesis [185]. Although type III PKSs are structurally and mechanistically simpler than Type I and type II PKSs, this type of PKSs can utilize a wide range of substrates. Moreover, their structural simplicity can be exploited for bioengineering and generation of PKs libraries [186].

2.4.2 Characteristics of Bacterial and Fungal NRPSs

The major characteristics that distinguish bacterial NRPSs from fungal NRPSs are their variable abilities to select and incorporate different amino acids during biosynthesis of NRPs. The Norine database was used to characterize the differences between bacterial and fungal NRPSs. Based on the analysis of >1000 peptides that represent >10,000 monomer positions and >500 monomer types, more than 34% of the peptides in this database was of fungal origin [187,188]. The most common monomer found in this database was aminoisobutyric acid (Aib) that belongs to fungal peptaibols. Other monomers of fungal NRPs were found to be valinol, leucinol, isovaline, and phenylalaninol [189]. One of the modifications that NRPSs employ is cyclization of NRPs. Cyclization is the step that precedes the release of the NRPs. Bacterial and fungal NRPSs use different mechanisms to perform cyclization of NRPs. Cyclization of bacterial NRPs is performed using thioesterase domains or termination (TE) domain, whereas fungal NRPSs use a specialized condensation domain (C_T) [190]. The fungal C_T domain and the bacterial termination (TE) domain are not only structurally different, but they also have a totally different catalytic mechanism. Thus far, all fungal NRPSs that catalyze peptide cyclization contain C_T domain at the end of their synthetic machinery [162].

2.5 Biosynthesis Strategies of NRPs and PKs

NRPSs have been identified in the three-domain system (bacteria, eukarya, and archaea). They are predominant in bacteria, less common in eukarya, and uncommon in archaea. Among the bacterial phyla, NRPSs predominate in *Actinobacteria*, *Cyanobacteria*, *Firmicutes*, and *Proteobacteria*. It has been noted that the abundance of

NRPS's clusters is directly proportional to the genome size of producing microorganisms [191].

NRPs and PKS are biosynthesized by NRPSs and PKSs enzyme complexes, respectively [192]. NRPs and PKS type I share the same biosynthetic mechanism. However, PKS type II and PKS type III use a different biosynthetic mechanism from that of NRPs [193]. Those mega enzyme complexes are composed of modules that form biosynthetic machineries [194]. A module is a segment of the NRPS enzyme complexes that adds one amino acid into NRPs [195]. Each module of NRPSs is composed of three main domains that catalyze steps of NRPs biosynthesis [196]. Domains of NRPSs modules contain conserved sequence motifs that aids in their *in silico* identification by bioinformatics tools [197]. The genes encoding NRPSs and their tailoring enzymes come together in gene clusters that may reach 200 Kb in size [191]. Modules of NRPSs are composed of three core domains: 1) adenylation (A), peptidyl carrier protein (PCP) domain, and condensation (C) domain. NRPSs may also have additional domains responsible for modification of the produced peptides [198]. Adenylation (A) domains works as gatekeepers as they select the amino acids substrates that will be incorporated in the final product. This domains is also responsible for activation of the selected amino acid substrates and converting them into aminoacyl adenylates, a step that consumes one molecule of ATP [199]. The selectivity of (A) domains is attributed to the presence of a stretch of ten amino acids that are considered as codons of NRPSs [200]. The identification of these codons can help in the accurate prediction of the substrates that would be selected by (A) domains [201]. The peptidyl carrier protein (PCP) domain, also called thiolation-(T)-domain, is responsible for the movement of activated amino acids between catalytic centers of NRPSs [202,203].

These catalytic centers contain the condensation(C) domain which is responsible for the formation of the peptide bond and elongation of peptides [196,204].

PKSs have three domains analogous to those of NRPSs. The acyl transferase (AT) domain, analogous to A domain of NRPSs, is the selector of substrates that are incorporated into the produced peptide. This domain transfer (CoA) thioesters of activated malonate derivatives into acyl-carrier protein (ACP) domain. The ACP domain, analogous to PCP of NRPSs, enables activated substrates to move to Ketosynthase (KS) domain, analogous to C-domain of NRPSs, to form the polyketide bonds and initiate polyketide sequence synthesis [205].

2.5.1 Priming

NRPs biosynthesis starts with the priming step which involves enzymatic modification of the carrier proteins (PCP) domains. PCP domains are first produced in their inactive (apo-form), but they require conversion into their active (holo) forms to initiate the biosynthesis of the NRPs [206]. The conversion of the PCP domains involves covalent bridging of a serine side chain of the PCP domains and Coenzyme-A-derived phosphopantetheine (P-pant) group [207]. This conversion of the PCP domains which is needed only once to keep them catalytically active is catalyzed by an enzyme belonging to the phosphopantetheinyl transferase (PPTase) enzyme family [206]. There are more than 20 homologs of PPTases that can be separated into two groups: the first group that converts the ACPs of fatty acid synthases and PKSs into their active form and the other group that converts the PCPs of NRPSs into their active form [193].

2.5.2 Initiation

The initiation step of NRPSs and PKSs biosynthesis starts with the selection of their substrates by A domain and AT domain, respectively. A domain is considered as a gatekeeper as it is responsible for activation and loading of substrates with high specificity. The high specificity of A domain was confirmed by a study that used the crystal structure of gramicidin S synthetase 1 (GrsA) as a guide and compared its specificity to that of 160 A domains. This study revealed that the presence of a variable region within A domain is what determines its specificity [208]. Similarly, another study revealed that specificity of the (AT) domains of PKSs is attributed to the presence of a short segment in this domain [209].

NRPSs and PKSs produce their natural products in a step-by-step fashion starting from their N-terminal loading units. The starting domains of NRPSs gramicidin S synthetase complex (GrsA) can be used as a model to explain the initiation step of NRPSs. GrsA starts the NRPS production by activating L-phenylalanine and converting it into AMP-ester at the A domain. The next step involves the formation of a covalent bond between the activated L-phenylalanine-AMP-ester and the thiol group of the P-pant hanging arm on the PCP domain. The epimerization (E) domain converts L-phenylalanine into D-phenylalanine while loaded on the PCP domain, and this form will function as a donor for the growing sequence on the next module [193].

The PKSs initiation follows the same manner of NRPSs. An example that illustrates how PKSs initiate the synthesis of their products is the DEBS1module of the PKS 6-deoxyerythronolide B synthase. The AT domain of DEBS1pick out propionyl-

CoA which already has an activated propionyl group and transfer this group to the P-pant of the neighboring ACP domain. The propionyl group will then serve as a donor for the next module [210].

2.5.3 Elongation

Elongation of sequences produced by NRPSs and PKSs require the presence of a minimum set of domains. Elongation by NRPSs requires the presence of PCP as an upstream domain followed by a downstream module composed of A, C, and PCP domains. Similarly, elongation by PKSs requires the presence of an ACP as an upstream domain followed by a module composed of KS, AT, and ACP domains. Although the catalytic mechanism of C domains in NRPSs has not yet been known, however, the possibility of forming a covalent acyl–C domain intermediate cannot be excluded [197]. Moreover, mutagenesis studies showed that the conserved double histidine (HH) motif is essential for the catalytic mechanism of C domains [196]. In PKSs, the upstream ACP donates the growing chain to the conserved cysteine of the KS domain and thus forming a thioester bond between this domain and the growing chain. The KS domain decarboxylates methyl- or malonyl-*S*-ACP and thus forms a carbanion that moves the growing sequence toward the downstream ACP domain. The growing chain on the downstream ACP now becomes the upstream donor for the next module in the PKS [193].

2.5.4 Termination

The movement of the produced sequences of NRPs and PKs on the modular assembly line terminates at a specialized thioesterase (TE) domain of the last module which catalyzes the release of the final product. The presence of a conserved serine

residue in TE domains that act as a nucleophile facilitates the acceptance of the product from the last carrier domain. The products are released from TE domains via two mechanisms: water hydrolysis or cyclization. Some NRPs has their TE domains repositioned in upstream sites where they can hydrolyze (e.g surfactin) or cyclize (e.g DEBS) short forms of products [193]. Some PKSs and NRPSs may have, in addition to the carboxy-terminal TE domain, a separate TE domain that is believed to release stalled mis-acylated monomers or acyl sequences from carrier protein domains [211].

2.6 Other Biosynthesis Strategies of NRPs

2.6.1 Heterocyclization

One of the structural features of NRPs is the presence of heterocyclic rings. These structures are made by heterocyclization (Cy) domain that catalyze heterocyclization of the functional side chains of the amino acids cysteine, serine and threonine, and the backbone of peptide sequence [194]. Cy domain was first identified in the bacitracin synthetase operon of *Bacillus licheniformis* ATCC 10716 [212] . Patel et al. studied in detail the function of Cy domain in catalyzing reactions of the epothilone, pyochelin, and vibriobactin systems [213].

2.6.2 Epimerization and Amino Acid Racemization

Another structural feature of NRPSs is the presence of D-Amino acids in some of their peptide sequences. Two studies showed that the integration of D-Amino acids is determined by the specificity of the A domains. One of these studies showed that the A domain of cyclosporine synthetase is responsible for the incorporation of D-Alanine in cyclosporine A [214]. The other study showed that the A domain of hybrid NRP/PKS hybrid

producing the cyanobacterial toxin microcystin is responsible for the incorporation of D-Glu in this toxin [215]. NRPSs also exploit one of their additional domains for amino acid racemization, i.e, changing L-amino acids to D-amino acids. Racemerization is catalyzed by the E domains that can be found within the initiation or the elongation modules [194]. Racemerization of amino acids may provide the producing organism with products (e.g., antibiotics) that are more resistant to degradation by proteases that selectively target L-amino acids [216].

2.6.3 Methylations

The fungal NRPs are often characterized by the presence of methylated nitrogen atoms. The incorporation of methyl groups in NRPs is the function of the N-methyl transferase (N-Mt) domain or methyltransferase domain (C-Mt). In NRPSs, the N-Mt domains are inserted into the C-terminal subdomain of their A domains. These domains are composed of ~420 residues and show some sequence homology with S-adenosyl-L-methionine (SAM)-dependent methyltransferases [217]. The N-methylation of amino acids of NRPs strengthens their backbones and prevents their premature proteolytic degradation [194].

2.6.4 Oxidation

Products of NRPSs and PKSs can be oxidized before their release. The oxidation reactions of NRPSs are performed by oxidation (O) domains. These domains are composed of ~200 amino acids and share homology with NAD-binding proteins [194].

2.6.5 Reduction

The C-terminal carboxyl group of NRPs and PKs chains may contain reduced units. The reduced structures are carried out by a catalytic domain which alternatively uses NADPH and NADH to release the intermediate product bound to the last PCP domain. An example of reductive domains is the C-terminal domain of NRPS involved in myxochelin production (MxcG). This domain alternatively uses NADPH and NADH to release the PCP-bound thioester as an aldehyde and then reduces it to produce an alcohol structure that can be found in myxochelin A [218].

2.7 Regulation of NRPS Genes

NRPSs and PKSs are so diverse to study their mode of regulation. However, some information from the well-studied regulatory mechanisms of microbial NRPSs or PKSs can shed light on the regulation of these mega enzymes. DegU is a global two-component regulator that works at the transcriptional level and can positively regulate the expression of *bac* gene cluster which encodes bacilysin of *Bacillus amyloliquefaciens* FZB42. DegU controls the expression of bacilysin by interacting with the σ^A-dependent promoter of the *bac* operon. The interaction between DegU occurs at three positions within the *bac* operon. On the other hand, the regulator ScoC can negatively regulate the biosynthesis of bacilysin in *B. amyloliquefaciens* FZB4 [219]. Binding of ScoC occurs at two ScoC boxes in the *bac* promoter: 1) SoC box 1 which is enclosed by positions −50 and −42 and 2) SoC box 2 which is enclosed by positions −12 and −4 [220]. Comp/ComA is another two-component regulator that controls the expression of *srfA* gene cluster that encodes surfactin. The initiation of transcription of *srfA* genes starts with the activation of ComP

by ComX. This step result in autophosphorylation of ComP and in turn transfer of phosphate group to ComA. The phosphorylated ComA initializes the transcription of *srfA* by recognizing a specific sequence (ComA boxes) upstream of the *srfA* promoter [221-223]. In *B. subtilis*, the H_2O_2 stress-responsive regulator (PerR) positively regulates *srfA.* The transcription of *srfA* is initiated by direct binding of PerR to two PerR boxes of its upstream promoter [224]. One of the negative regulators of bacterial biosurfactins is CodY. This protein represses the expression of *srfA* when the concentrations of amino acids Ile, Leu, and Val reach high levels. Similarly, other negative regulators of surfactin biosynthesis (AbrB and Spx) can repress the transcription of *srfA* genes [222]. The biosynthesis of surfactin in *B. subtilis* requires the function of Sfp enzyme that posttranslationally phosphorylates a serine residue of PCP domains of surfactin synthetase. This modification is necessary to form docking sites for loading amino acids and peptide bond formation [222].

The biosynthesis of the antitumor doxorubicin produced by *Streptomyces peucetius* is regulated by a set of transcription regulators that includes *dnrI*, *dnrN*, *dnrO*, and *doxR*. The activation of genes involved in the transcription of doxorubicin starts with the activation of *dnrN* by binding of a regulator protein produced by *dnrO* gene to the *dnrN*/*dnrO* promoter region. Production of DnrN enhances the transcription of *dnrI* gene which in turn starts the transcription of genes involved in doxorubicin biosynthesis. Suppression of doxorubicin production is indirectly controlled by the product of *doxR* gene that belongs to the transcription regulators family IclR [225].

2.8 Structure of NRPs

NRPSs are characterized by their distinct modular structures. The function of each module is to incorporate a specific monomer in the final nonribosomal peptide product. The structural characteristics and the order of modules specifically determine which monomers are to be activated and included in the final product, and also specify and control the chemical reactions that occur during the sequential production of the final products. The modules of NRPSs also determine the length and identity of the final product at the end of the production line [226]. The NRPSs composed of three universal domains necessary for the synthesis of nonribosomal peptides: 1) The A domain is responsible for selection and activation of amino acids, 2) The PCP domain moves the activated amino acids between catalytic domains, 3) The C domain is responsible for condensation of amino acids. The TE domain that is located in the distal module is responsible for releasing the final product [226].

2.8.1 Adenylation (A) domain

The function of A domain is the selection and activation of amino acids prior to the begging of the biosynthesis on the assembly line of NRPSs. These domains use ATP to catalyze the activation of the carboxylate group of the selected amino acids and form an aminoacyl-AMP intermediate that is subsequently attached to the 4'-Ppant of the adjoining PCP domain [28]. The A domain is considered as “a gate keeper” because of its specificity in choosing the amino acids that will be incorporated into the nonribosomal peptides [227]. Moreover, these domains are responsible for the huge diversity of nonribosomal peptides because they do not only select from the typical proteinogenic

amino acids, but also from hundreds of other precursors. The crystal structures of A domain that activates L-Phe for the biosynthesis of gramicidin S [228] and A domain DhbE that activates aryl acid 2,3-dihydroxybenzoate for the biosynthesis of bacillibactin [201], revealed the presence of specific residues in these domains that are essential in substrate recognition.

2.8.2 Thiolation (T) domain

The T domain, also described as PCP domain, comes into play during NRPs synthesis after the formation of aminoacyl-AMP by A domains. Both A domains and T domains work together to form aminoacylthioester intermediate. The aminoacylthioester intermediate is formed by attacking the carboxyl group of aminoacyl-AMP intermediate by the terminal thiol of the 4'-Ppant attached to the T domains [28]. The overall folds of T domains resemble the topology of the acyl carrier proteins (ACPs) of polyketide and fatty acid synthase [229]. The T domain also resembles ACPs in its post-translational modification with a 4'-Ppant group. The addition of this group by 4'-Ppant transferase on the centrally located and conserved seryl residue converts the T domain to its holo form [230]. The T domain not only interacts with A domain during NRPs biosynthesis but ~~they~~ also plays an essential role in peptide bond formation and modification and release of the finalized NRPs from the biosynthetic line [28]. The T domain can take two forms: 1) the apo form 2) and the holo form (when loaded with phosphopantetheine group). Each of these forms can have three conformations (A state, A/H state, and H state). The apo and holo forms share a similar A/H state but differ in their A state and H-states. These states are essential for NRPSs because they determine the position of phosphopantetheine group and determine the way the T domains interact with other domains in the NRPSs

biosynthesis line [231]. The interactions of the T domain with other domains of NRPSs have been determined by NMR studies. Several studies showed that the formation of PCP-TE [232] and PCP-C didomains [233] and C-A-PCP-TE tetra-domain [234] occurs while the T domain is in its A/H state.

2.8.3 Condensation (C) domain

The function of the C domain during the biosynthesis of nonribosomal peptides is to catalyze the formation of peptides bonds between two amino acids. The C domain forms a bond between the amine group of a PCP-bound phosphopantetheinyl amino acid and the peptidyl thioester found on the upstream PCP and this will consequently move the growing peptide to the downstream PCP domain [231]. The C domains are homologs of the acyltransferases family. The members of this family share a conserved HHXXXDG motif which is believed to be essential for the catalysis of peptide bond formation [231]. The first structure of a C domain was revealed by X-ray crystallography of the NRPS VibH which catalyzes the formation of peptide bond formation during the biosynthesis of the *Vibrio cholera*'s vibriobactin [235].

2.8.4 Thioesterase (TE) domain

TE domain is usually present at the C-terminus of NRPSs and its function is to catalyze the release of nonribosoaml peptides at the end of the biosynthesis line by either hydrolysis of the phosphopantetheinyl thioester or by intramolecular cyclization [231]. TE domains belong to a larger family of α,β-hydrolases that are characterized by having a catalytic triad containing a nucleophilic serine residue that cleaves the phosphopantetheinyl thioester.

2.9 Production of novel NRPS through reconstruction and engineering

Manipulations of NRPSs are essential to obtain new forms of non-ribosomal peptides that have new characteristics compared to their typical forms. Manipulations of domains of NRPSs can alter their specificity to substrates. Fusaricidins are antimicrobials produced by a NRPS of *Paenibacillus polymyxa* DBB1709 that have antimicrobial effect against Gram-positive bacteria and fungi [236]. Fusaricidins structures differ from each other because the NRPS that produces them has an A domain (FusA-A) in the third module that can recruit more than one type of amino acids. This domain can bring to the synthase machinery L-Phe, L-Ile, L-Tyr, L-Val, or L-allo-Ile, and thus leads to the biosynthesis of different analogs of fusaricidins [237]. However, the analog that contains L-Phe (LI-F07) has the most potent antimicrobial effect. Han et al. [237] changed the binding pocket of the FusA-A3 by manipulating a specific sequence that selectively recruits L-Phe. In this study, six mutants were created and only one strain was able to produce more LI-F07 than the wild type.

Polymyxin derivatives were also developed in order to decrease their toxicity and to increase their antimicrobial effect. Kim et al [238] genetically engineered polymyxin synthetases in a two-step method in which the A domains of this NRPS were substituted. Polymyxin A synthetase is encoded by pmxABCDE gene cluster of *Paenibacillus polymyxa* E681. The substitution of A domains involved replacing the A domains of Polymyxin A synthetase by A domains from other species. To produce Polymyxin E synthetase, the L-leucine-specific A-domain obtained from *P. polymyxa* ATCC21830 was used to substitute the L-threonine-specific A-domain sequence of pmxA gene. The production of Polymyxin B synthetase was achieved by substitution of the D-leucine-

specific A-domain sequence with the D-phenylalanine-specific sequence of *P. polymyxa* F4. The genetically manipulated pmxA gene containing the modified A domains were fused to pmxBCDE and expressed in *Bacillus subtilis* BSK4dA and resulted in new recombinant strains that can produce Polymyxins E and B and can be potentially manipulated further to produce novel Polymyxins.

Recently, genetic engineering of NRPs peptides was used in order for obtaining new properties of these natural products. Neilsen et al [239] exploited the possibility of recombining PKS-NRPS hybrids from different species to produce SMs with new functionalities. This group was able to combine modules of PKS-NRPS hybrids (CcsA) involved in the biosynthesis of cytochalasin E of *Aspergillu clavatus*, with those of the related hybrid (Syn2) from *Magnaporthe oryzae* and then they successfully expressed the new chimeric PKS-NRPS in *A. nidulans*. Using this method it was possible to produce novel compounds.

Cell-free protein synthesis (CFPS) is another strategy used recently to reconstruct the bio synthesis of NRPSs straight from DNA [240]. This strategy has some advantages over the conventional cell-based strategy: first) it is easier to control the production conditions, second) Atypical precursors can be added to the reaction and this, in turn, can result in the production of novel NRPs. This group used CFPS method to express the first two modules, GrsA and GrsB, of the NRPS that are involved in the biosynthesis of gramicidin S. GrsA and GrsB modules can produce a cyclized natural shunt product named Phe-L-Pro diketopiperazine (DKP). Using this method, it was possible to express GrsA and GrsB separately, which in turn work in concert to produce higher concentrations of Phe-L-Pro DKP than cell-based expression. The cyclic DKPs are

produced by bacteria, fungi, plants [241] and humans [242] and have shown various biological activities.

CHAPTER III

METHODS

3.1 Genes of *S. sorangium's* Epothilone Gene Cluster

The amino acid sequences of individual genes of *Sorangium cellulosum* So ce90's Epothilone gene cluster were obtained from GenBank sequence database (www.ncbi.nlm.nih.gov/genbank/), using the accession number AF210843.1. This gene cluster is a PKS-NRPS hybrid cluster that contains 22 open reading frames (ORFs) which span 68,750 base pairs of *S. cellulosum* So ce90's genome. Within the 22 ORFs there are five polyketide synthase (PKS) genes (e*poA* and *epoB–epoE*), one nonribosomal peptide synthetase (NRPS) gene (e*poP*), and one gene (e*poF*) that codes for cytochrome P450 oxygenase. The ORFs 3 and 14 code for putative antibiotic transport proteins. Each of e*poA*, e*poB*, and e*poE* form one PKS module within this cluster, whereas e*poD* and e*poC* form 2 and 4 PKS modules, respectively **(Fig. 1).**

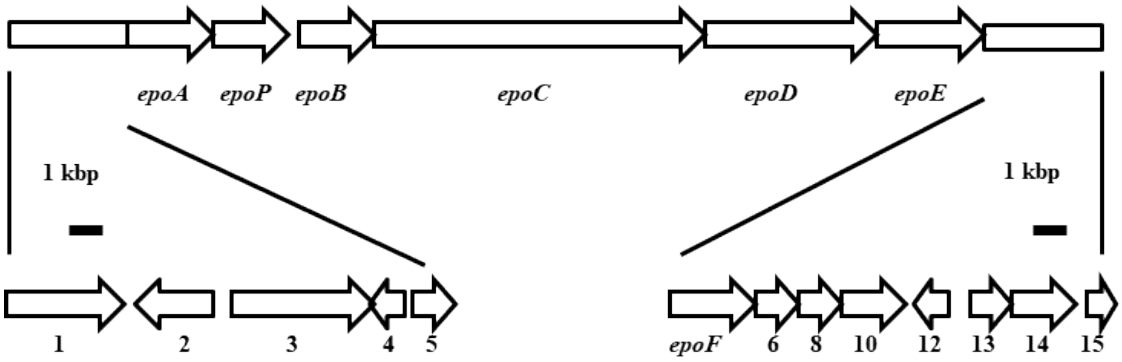


Figure 1. Epothilone gene cluster of *S. cellulosum* So ce90's showing the arrangement of 22 ORFs (Epos A-F and 15 other ORFs) [243].

3.2 Finding *EpoA-F* Homologs (the EAFHs)

The amino sequences of *epo A-F* of the *S. cellulosum* So ce90's gene cluster were used to find their homologs in other bacteria. The Basic Local Alignment Search Tool (BLASTp) was used for comparing protein sequences of *epo A-F* against other bacterial sequences. The identity cutoff of >40% was used to increase the level of certainty that the EAFHs and their counterparts of *S. cellulosum* So ce90's gene cluster share a functional similarity. All parameters of BLAST algorithm were set to their default.

3.3 Locating ~~*EpoA-F* (~~the EAFHs~~)~~ in Bacterial Genomes

The NCBI's Sequence Viewer 3.28.0.1 (https://www.ncbi.nlm.nih.gov/projects/sviewer/) was used to find the EAFHs in bacterial genomes and by using their accession numbers obtained from BLASTp search. Only the EAFHs with > 40% identity were used for this purpose. When multiple BLASTp hits were obtained for the same homolog, only the accession number of the hit with the highest query coverage value was considered **(Table 1).** The length of sequences of the EAFHs and the boundaries of bacterial DNA sequences that contained three or more of the EAFHs and expected to contain ESGCs were determined and recorded for further analyses.

3.4 Screening of Bacterial DNA Sequences for ESGCs

The bacterial DNA sequences found to contain three or more of the EAFHs were subjected to initial antiSMASH analysis to identify any ESGCs, even when these genes are separated from each other by long sequences. The antiSMASH analysis was only carried out when the complete genome sequences of bacteria were available.

3.5 Identification of ESGCs in Bacterial Whole Genomes

To find more ESGCs and other PKS, NRPS and hybrid NRPS-PKS gene clusters, the initial antiSMASH analysis of any bacterial DNA sequence that had been initially found to contain ESGCs was repeated by using the whole genome sequence of those bacteria. Furthermore, to increase the chance of finding more ESGC, the whole genomes of bacteria that are related to those found to have ESGCs were also analyzed by AntiSMASH. The gene cluster 4751261-4894265 was analyzed by PRISM (http://magarveylab.ca/prism/) the computational resource that identifies biosynthetic gene clusters [244].

3.6 *In silico* Design of ESGCs

In silico designed ESGCs were generated using the EAFHs identified by antiSMASH in ~~these~~ some gene clusters. The boundaries of the the EAFHs were determined by antiSMASH and then cut by FasParser software [245] from ESGCs. Then, the DNA sequences of the EAFHs were used to replace their *epo A-F* counterparts of the *S. cellulosum* So ce90's gene cluster. The new gene clusters were reanalyzed by antiSMASH to identify their domains and to predict the scaffolds of the products they might encode.

CHAPTER IV

RESULTS

4.1 Distribution of the EAFHs in bacterial genomes

Ninety-three bacterial genera were found to carry at least one of the EAFHs. However, 19 of these genera had three or more the EAFHs of *S. cellulosum* So ce90: *Byssovorax* (4), *Calothrix* (3), *Candidatus* (4), *Chondromyces* (6), *Corallococcus* (4), *Cystobacter* (6), *Melittangium* (4), *Micromonospora* (4), *Moorea* (3), *Myxococcus* (4), *Nannocystis* (3), *Nocardia* (3), *Nostoc* (4), *Nostocales* (3), *Polyangium* (3), *Pseudoalteromonas* (3), *Stigmatella* (5), *Streptomyces* (3), and *Tolypothrix* (3) (**Table 1**). The other 74 genera with one or two ECSGs are shown in **Supplementary Table 2**.

Table 1**.** Bacterial genera and their species or strains found to carry the EAFHs.

Genus	EAFHs	Protein encoded by the EAFHs (Species or strain)	Identity	Accession
Byssovorax				
	B	polyketide synthase (Byssovorax cruenta)	43%	AOG74796.1
	C	polyketide synthase (Byssovorax cruenta)	52%	AOG74797.1
	D	polyketide synthase (Byssovorax cruenta)	51%	AOG74797.1
	E	polyketide synthase (Byssovorax cruenta)	45%	AOG74797.1
Calothrix				
	P	non-ribosomal peptide synthetase (Calothrix sp.	44%	WP_0194

		PCC 7103)		96853.1
	P	non-ribosomal peptide synthetase (Calothrix parasitica)	44%	WP_0966 54065.1
	P	non-ribosomal peptide synthetase (Calothrix parasitica)	50%	WP_0966 57254.1
	A	short-chain dehydrogenase (Calothrix brevissima)	50%	WP_0966 48942.1
	E	type I polyketide synthase (Calothrix sp. PCC 7103)	40%	WP_0194 89581.1
Candidatus				
	A	hypothetical protein OT06_34100 (Candidatus Thiomargarita nelsonii)	50%	KHD0764 9.1
	B	hypothetical protein ETSY1_00130 (Candidatus Entotheonella sp. TSY1)	43%	ETX0340 2.1
	E	hypothetical protein ETSY1_25090 (Candidatus Entotheonella sp. TSY1)	41%	ETW9682 4.1
	P	hypothetical protein (Candidatus Scalindua sp. husup-a2)	47%	WP_0968 93538.1
	P	hypothetical protein ETSY1_00125 (Candidatus Entotheonella sp. TSY1)	47%	ETX0340 1.1
Chondromyces				
	A	type I polyketide synthase (Chondromyces crocatus)	51%	WP_0504 32595.1
	A	type I polyketide synthase	52%	WP_0504

		(Chondromyces crocatus)		33113.1
	A	hypothetical protein (Chondromyces crocatus)	52%	WP_082362832.1
	A	polyketide synthase (Chondromyces crocatus)	52%	CAQ18829.1
	B	polyketide synthase (Chondromyces crocatus)	55%	CAJ46689.1
	B	type I polyketide synthase (Chondromyces crocatus)	46%	WP_050432597.1
	B	polyketide synthase (Chondromyces crocatus)	46%	AKT40700.1
	B	type I polyketide synthase (Chondromyces crocatus)	52%	WP_050432507.1
	B	uncharacterized protein CMC5_048570 (Chondromyces crocatus)	46%	AKT40701.1
	B	type I polyketide synthase (Chondromyces crocatus)	43%	WP_050432982.1
	B	Malonyl CoA-acyl carrier protein transacylase (Chondromyces apiculatus DSM 436)	44%	EYF00131.1
	B	hypothetical protein (Chondromyces apiculatus)	44%	WP_081865839.1
	B	hypothetical protein (Chondromyces apiculatus)	41%	WP_081864958.1
	B	type I polyketide synthase (Chondromyces crocatus)	49%	WP_050433117.1
	B	Malonyl CoA-acyl carrier protein transacylase (Chondromyces apiculatus	41%	EYF05471.1

		DSM 436)		
	D	polyketide synthase (Chondromyces crocatus)	45%	CAQ18829.1
	D	hypothetical protein (Chondromyces crocatus)	45%	WP_082362832.1
	E	type I polyketide synthase (Chondromyces crocatus)	53%	WP_050432507.1
	E	uncharacterized protein CMC5_048570 (Chondromyces crocatus)	47%	AKT40701.1
	E	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Chondromyces apiculatus)	46%	WP_052375665.1
	E	type I polyketide synthase (Chondromyces apiculatus)	44%	WP_052375398.1
	E	type I polyketide synthase (Chondromyces crocatus)	44%	WP_050433113.1
	E	type I polyketide synthase (Chondromyces crocatus)	43%	WP_050433115.1
	E	hypothetical protein (Chondromyces crocatus)	41%	WP_082362807.1
	E	type I polyketide synthase (Chondromyces crocatus)	42%	WP_050433117.1
	F	cytochrome P450 (Chondromyces crocatus)	50%	WP_050434271.1
	F	cytochrome P450 (Chondromyces crocatus)	48%	WP_050434274.1
	F	cytochrome P450 (Chondromyces apiculatus)	46%	WP_044248321.1

	P	non-ribosomal peptide synthetase (Chondromyces apiculatus)	54%	WP_044244473.1
	P	non-ribosomal peptide synthetase (Chondromyces crocatus)	48%	WP_050433116.1
	P	nonribosomal polypetide synthetase (Chondromyces crocatus)	48%	CAQ18831.1
Corallococcus				
	A	polyketide synthase (Corallococcus coralloides)	53%	AQW44888.1
	B	polyketide synthase (Corallococcus coralloides)	46%	AQW44891.1
	B	polyketide synthase (Corallococcus coralloides)	45%	AQW44888.1
	B	polyketide synthase (Corallococcus coralloides)	45%	AQW44892.1
	D	polyketide synthase (Corallococcus coralloides)	46%	AQW44888.1
	E	polyketide synthase (Corallococcus coralloides)	44%	AQW44889.1
Cystobacter				
	A	polyketide synthase (Cystobacter fuscus)	52%	WP_095988602.1
	B	polyketide synthase (Cystobacter fuscus)	47%	WP_095988603.1
	B	polyketide synthase (Cystobacter fuscus)	46%	WP_095988602.1
	B	CtaD (Cystobacter fuscus)	43%	AAW033

				27.1
	D	polyketide synthase (Cystobacter fuscus)	46%	WP_095988602.1
	D	MmxC (Cystobacter fuscus)	47%	ABA29781.1
	D	hypothetical protein (Cystobacter ferrugineus)	49%	WP_071900713.1
	D	type I polyketide synthase (Cystobacter fuscus)	48%	WP_002624106.1
	D	hypothetical protein (Cystobacter fuscus)	48%	WP_095988851.1
	E	polyketide synthase (Cystobacter fuscus)	44%	WP_095988601.1
	F	cytochrome P450 (Cystobacter fuscus)	48%	WP_002631959.1
	F	cytochrome P450 (Cystobacter ferrugineus)	46%	WP_071903224.1
	F	cytochrome P450 (Cystobacter fuscus)	46%	WP_095987172.1
	F	cytochrome P450 (Cystobacter fuscus)	46%	WP_095990522.1
	F	cytochrome P450 (Cystobacter ferrugineus)	46%	WP_071896186.1
	F	cytochrome P450 (Cystobacter ferrugineus)	46%	WP_071899248.1
	F	cytochrome P450 (Cystobacter fuscus)	45%	WP_002627453.1
	F	cytochrome P450 (Cystobacter fuscus)	45%	WP_095986811.1
	F	cytochrome P450	45%	WP_0026

		(Cystobacter fuscus)		28875.1
	F	cytochrome P450 (Cystobacter ferrugineus)	46%	WP_071904357.1
	F	cytochrome P450 (Cystobacter fuscus)	39%	WP_095983905.1
	F	putative cytochrome P450 hydroxylase (Cystobacter fuscus DSM 2262)	38%	EPX57901.1
	F	cytochrome P450 (Cystobacter fuscus)	38%	WP_081713978.1
	P	CtaD (Cystobacter fuscus)	43%	AAW03327.1
Melittangium				
	B	polyketide synthase (Melittangium boletus)	46%	WP_095980636.1
	B	MelD protein (Melittangium lichenicola)	43%	CAD89775.1
	B	non-ribosomal peptide synthetase (Melittangium boletus)	43%	WP_095982063.1
	B	polyketide synthase (Melittangium boletus)	43%	WP_095980635.1
	D	hypothetical protein (Melittangium boletus)	48%	WP_095975708.1
	D	type I polyketide synthase (Melittangium boletus DSM 14713)	48%	ATB26815.1
	F	cytochrome P450 (Melittangium boletus)	47%	WP_095981990.1
	F	cytochrome P450	45%	WP_0959

		(Melittangium boletus)		76970.1
	F	cytochrome P450 (Melittangium boletus)	41%	WP_095978808.1
	P	MelD protein (Melittangium lichenicola)	45%	CAD89775.1
	P	non-ribosomal peptide synthetase (Melittangium boletus)	44%	WP_095982063.1
Micromonospora				
	A	beta-ketoacyl synthase (Micromonospora echinospora)	52%	WP_088984494.1
	C	Acyl transferase domain-containing protein (Micromonospora humi)	42%	SCG63841.1
	C	Acyl transferase domain-containing protein (Micromonospora humi)	41%	SCG79465.1
	C	Acyl transferase domain-containing protein (Micromonospora echinospora)	42%	SCF29238.1
	C	KR domain-containing protein (Micromonospora echinospora)	68%	WP_088985802.1
	C	Acyl transferase domain-containing protein (Micromonospora coxensis)	42%	SCG75708.1
	C	Acyl transferase domain-containing protein (Micromonospora nigra)	42%	SCL14128.1
	C	type I polyketide synthase	42%	WP_0738

		(Micromonospora sp. TSRI0369)		27812.1
	C	type I polyketide synthase (Micromonospora carbonacea)	43%	WP_083303087.1
	C	type I polyketide synthase (Micromonospora citrea)	42%	WP_091106414.1
	C	type I polyketide synthase (Micromonospora sp. CNB394)	42%	WP_018789130.1
	C	type I polyketide synthase (Micromonospora carbonacea)	40%	WP_043962713.1
	E	beta-ketoacyl synthase (Micromonospora echinospora)	44%	WP_088984494.1
	E	hypothetical protein (Micromonospora pallida)	43%	WP_091640838.1
	E	hypothetical protein (Micromonospora echinospora)	43%	WP_088984502.1
	P	hypothetical protein (Micromonospora nigra)	44%	WP_091075556.1
	P	non-ribosomal peptide synthetase (Micromonospora viridifaciens)	44%	WP_089006614.1
	P	hypothetical protein (Micromonospora halophytica)	44%	WP_091296371.1
Moorea				
	A	polyketide synthase	49%	WP_0833

		(Moorea producens)		05302.1
	A	hypothetical protein BJP34_22190 (Moorea producens PAL-8-15-08-1)	49%	AOX04464.1
	A	polyketide synthase (Moorea producens)	50%	WP_071104822.1
	A	short-chain dehydrogenase (Moorea producens)	50%	WP_070394216.1
	A	short-chain dehydrogenase (Moorea producens)	50%	WP_083373684.1
	A	short-chain dehydrogenase (Moorea producens JHB)	50%	AOY84655.1
	A	hypothetical protein (Moorea producens)	47%	WP_070395080.1
	E	polyketide synthase (Moorea producens)	41%	WP_083305302.1
	E	hypothetical protein BJP34_22190 (Moorea producens PAL-8-15-08-1)	41%	AOX04464.1
	E	beta-ketoacyl synthase (Moorea producens PAL-8-15-08-1)	41%	AOX01784.1
	E	hypothetical protein (Moorea producens)	40%	WP_071106601.1
	E	hypothetical protein BJP34_29070 (Moorea producens PAL-8-15-08-1)	39%	AOX02953.1
	E	polyketide synthase (Moorea producens)	39%	WP_083305401.1
	E	hypothetical protein BI334_26040 (Moorea	39%	OLT69482.1

		producens PAL)		
	E	polyketide synthase (Moorea producens)	39%	WP_008191793.1
	E	hypothetical protein (Moorea producens)	39%	WP_070393504.1
	E	polyketide synthase (Moorea producens)	40%	WP_070395515.1
	E	polyketide synthase (Moorea producens)	40%	WP_083305402.1
	E	hypothetical protein BJP34_29085 (Moorea producens PAL-8-15-08-1)	40%	AOX04584.1
	E	polyketide synthase (Moorea producens)	40%	WP_008191799.1
	E	hypothetical protein BI334_26055 (Moorea producens PAL)	40%	OLT69483.1
	P	non-ribosomal peptide synthetase (Moorea producens)	44%	WP_070393503.1
	P	non-ribosomal peptide synthetase (Moorea producens)	46%	WP_070391439.1
	P	non-ribosomal peptide synthetase (Moorea bouillonii)	45%	WP_075898318.1
	P	non-ribosomal peptide synthetase (Moorea producens)	45%	WP_070392180.1
Myxococcus				
	B	polyketide synthase	46%	WP_0115

		(Myxococcus xanthus)		54523.1
	B	polyketide synthase (Myxococcus virescens)	46%	WP_090494067.1
	B	type I polyketide synthase (Myxococcus hansupus)	47%	WP_021780891.1
	B	polyketide synthase (Myxococcus stipitatus)	45%	WP_015349881.1
	B	polyketide synthase (Myxococcus stipitatus)	44%	WP_015349880.1
	B	polyketide synthase (Myxococcus virescens)	44%	WP_090494064.1
	B	polyketide synthase (Myxococcus xanthus)	44%	WP_011554522.1
	B	type I polyketide synthase (Myxococcus hansupus)	43%	WP_021780892.1
	B	polyketide synthase (Myxococcus stipitatus)	44%	WP_015349883.1
	B	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Myxococcus stipitatus)	42%	WP_015350515.1
	D	polyketide synthase (Myxococcus stipitatus)	46%	WP_015349880.1
	D	type I polyketide synthase (Myxococcus fulvus)	46%	WP_052770859.1
	D	type I polyketide synthase (Myxococcus hansupus)	46%	WP_021780892.1
	D	polyketide synthase (Myxococcus virescens)	46%	WP_090494064.1
	D	polyketide synthase	46%	WP_0115

		(Myxococcus xanthus)		54522.1
	D	hypothetical protein (Myxococcus fulvus)	48%	WP_074958389.1
	E	polyketide synthase (Myxococcus stipitatus)	44%	WP_015349884.1
	E	polyketide synthase (Myxococcus stipitatus)	41%	WP_015349879.1
	P	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Myxococcus stipitatus)	44%	WP_015350515.1
Nannocystis				
	B	polyketide synthase (Nannocystis sp. MB1016)	53%	ALD82522.1
	B	polyketide synthase (Nannocystis sp. MB1016)	55%	ALD82524.1
	F	cytochrome P450 (Nannocystis exedens)	42%	WP_096327420.1
	F	cytochrome P450 (Nannocystis pusilla)	42%	ATG32080.1
	F	Cytochrome P450 (Nannocystis exedens)	37%	SFE94421.1
	F	cytochrome P450 (Nannocystis exedens)	37%	WP_096331714.1
	P	nonribosomal peptide synthetase (Nannocystis pusilla)	46%	ATG32076.1
	P	non-ribosomal peptide synthetase (Nannocystis exedens)	46%	WP_096327416.1

	P	epothilone synthetase B (Nannocystis exedens)	46%	SFE85729.1
Nocardia				
	B	type I polyketide synthase (Nocardia exalbida)	43%	WP_040871896.1
	B	type I polyketide synthase (Nocardia pneumoniae)	44%	WP_040785698.1
	B	type I polyketide synthase (Nocardia abscessus)	43%	WP_043691701.1
	B	type I polyketide synthase (Pseudonocardia spinosispora)	43%	WP_028933378.1
	B	type I polyketide synthase (Nocardia sp. NRRL WC-3656)	42%	WP_030517776.1
	C	type I polyketide synthase (Pseudonocardia spinosispora)	41%	
	D	polyketide synthase (Nocardia terpenica)	42%	WP_098698800.1
	D	polyketide synthase (Nocardia sp. CNY236)	42%	WP_028478665.1
Nostoc				
	A	short-chain dehydrogenase (Nostocales cyanobacterium HT-58-2)	50%	WP_087538336.1
	A	type I polyketide synthase (Nostoc punctiforme)	50%	WP_012408705.1
	A	type I polyketide synthase (Nostoc punctiforme)	47%	WP_012409763.1

	A	hypothetical protein (Nostocales cyanobacterium HT-58-2)	50%	WP_087541151.1
	A	hypothetical protein BZZ01_15445 (Nostocales cyanobacterium HT-58-2)	50%	ARV63050.1
	A	MULTISPECIES: short-chain dehydrogenase (Nostocales)	49%	WP_096584887.1
	B	type I polyketide synthase (Nostoc punctiforme)	40%	WP_012412906.1
	B	type I polyketide synthase (Nostoc punctiforme)	42%	WP_012408706.1
	E	type I polyketide synthase (Nostoc punctiforme)	42%	WP_012409763.1
	E	polyketide synthase (Nostocales cyanobacterium HT-58-2)	41%	WP_087539588.1
	P	hypothetical protein (Nostoc sp. NIES-4103)	47%	WP_096555087.1
	P	non-ribosomal peptide synthetase (Nostoc linckia)	45%	WP_099073120.1
	P	non-ribosomal peptide synthetase (Nostoc linckia)	45%	WP_099071360.1
	P	MULTISPECIES: non-ribosomal peptide synthetase (Nostocales)	45%	WP_052335078.1
	P	amino acid adenylation protein (Nostoc linckia z13)	45%	PHK20210.1
	P	amino acid adenylation protein (Nostoc linckia z1)	45%	PHJ57115.1

	P	non-ribosomal peptide synthetase (Nostoc sp. 'Peltigera membranacea cyanobiont' 213)	45%	WP_094331875.1
	P	non-ribosomal peptide synthetase (Nostoc carneum)	45%	WP_096726971.1
	P	non-ribosomal peptide synthetase (Nostoc sp. PCC 7120)	41%	WP_010996801.1
Nostocales				
	A	short-chain dehydrogenase (Nostocales cyanobacterium HT-58-2)	50%	WP_087538336.1
	A	hypothetical protein (Nostocales cyanobacterium HT-58-2)	50%	WP_087541151.1
	A	hypothetical protein BZZ01_15445 (Nostocales cyanobacterium HT-58-2)	50%	ARV63050.1
	A	MULTISPECIES: short-chain dehydrogenase (Nostocales)	49%	WP_096584887.1
	E	polyketide synthase (Nostocales cyanobacterium HT-58-2)	41%	WP_087539588.1
	P	MULTISPECIES: non-ribosomal peptide synthetase (Nostocales)	45%	WP_052335078.1
Polyangium				
	A	polyketide synthase (Polyangium spumosum)	53%	ANI24099.1

	B	polyketide synthase (Polyangium spumosum)	60%	ANI24099.1
	D	polyketide synthase (Polyangium spumosum)	45%	ANI24099.1
Pseudoalteromonas				
	B	hypothetical protein (Pseudoalteromonas luteoviolacea)	39%	WP_063364358.1
	B	hypothetical protein (Pseudoalteromonas sp. HM-SA03)	39%	WP_095727193.1
	E	hypothetical protein (Pseudoalteromonas luteoviolacea)	39%	WP_063364354.1
	P	non-ribosomal peptide synthetase (Pseudoalteromonas luteoviolacea)	42%	WP_063379439.1
	P	non-ribosomal peptide synthetase (Pseudoalteromonas sp. HM-SA03)	42%	WP_095727196.1
Stigmatella				
	A	StiJ protein (Stigmatella aurantiaca Sg a15)	51%	CAD19093.1
	A	polyketide synthase (Stigmatella erecta)	52%	WP_093524456.1
	A	polyketide synthase (Stigmatella aurantiaca)	52%	WP_075006161.1
	A	StiC protein (Stigmatella aurantiaca Sg a15)	50%	CAD19087.1

	A	type I polyketide synthase (Stigmatella aurantiaca)	50%	WP_013374824.1
	B	polyketide synthase (Stigmatella aurantiaca)	46%	WP_075006160.1
	B	polyketide synthase (Stigmatella erecta)	46%	WP_093524457.1
	B	MxaD (Stigmatella aurantiaca)	46%	AAK57188.1
	B	MxaC (Stigmatella aurantiaca)	45%	AAK57187.1
	B	polyketide synthase (Stigmatella erecta)	45%	WP_093524456.1
	B	polyketide synthase (Stigmatella aurantiaca)	45%	WP_075006161.1
	B	polyketide synthase (Stigmatella aurantiaca)	44%	WP_013375988.1
	B	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Stigmatella aurantiaca)	41%	WP_013376004.1
	B	polyketide synthase (Stigmatella aurantiaca)	43%	WP_049805256.1
	B	Polyketide synthase AufC (Stigmatella aurantiaca DW4/3-1)	43%	ADO71771.1
	B	non-ribosomal peptide synthetase (Stigmatella aurantiaca)	42%	WP_075010837.1
	B	non-ribosomal peptide synthetase (Stigmatella erecta)	41%	WP_093515620.1

	B	polyketide synthase (Stigmatella aurantiaca)	41%	WP_0026 19013.1
	B	Acyl transferase domain-containing protein (Stigmatella aurantiaca)	44%	SEM0718 1.1
	B	hypothetical protein (Stigmatella erecta)	44%	WP_0935 22399.1
	D	polyketide synthase (Stigmatella aurantiaca)	46%	WP_0750 06161.1
	D	KR domain-containing protein (Stigmatella aurantiaca)	40%	WP_0133 76680.1
	D	polyketide synthase (Stigmatella erecta)	47%	WP_0935 24456.1
	D	MxaC (Stigmatella aurantiaca)	46%	AAK5718 7.1
	E	Acyl transferase domain-containing protein (Stigmatella aurantiaca)	45%	SEM0718 1.1
	E	hypothetical protein (Stigmatella erecta)	45%	WP_0935 22399.1
	E	StiF protein (Stigmatella aurantiaca Sg a15)	43%	CAD1909 0.1
	E	hypothetical protein (Stigmatella erecta)	41%	WP_0935 19600.1
	E	ype I polyketide synthase (Stigmatella aurantiaca)	83%	WP_0133 74824.1
	E	hypothetical protein (Stigmatella aurantiaca)	41%	WP_0750 07769.1
	F.	cytochrome P450 (Stigmatella aurantiaca)	48%	WP_0750 06780.1

	F.	cytochrome P450 (Stigmatella erecta)	48%	WP_093516259.1
	F.	cytochrome P450 (Stigmatella aurantiaca)	47%	WP_002611914.1
	F.	Cytochrome P450 (Stigmatella aurantiaca DW4/3-1)	47%	ADO73095.1
	F.	cytochrome P450 (Stigmatella aurantiaca)	48%	WP_075006780.1
	P	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Stigmatella aurantiaca)	44%	WP_013376004.1
	P	amino acid adenylation domain protein (Stigmatella aurantiaca DW4/3-1)	44%	EAU62130.1
	P	non-ribosomal peptide synthetase (Stigmatella aurantiaca)	43%	WP_075010837.1
	P	non-ribosomal peptide synthetase (Stigmatella erecta)	43%	WP_093515620.1
Streptomyces				
	A	polyketide synthase (Streptomyces melanosporofaciens)	52%	WP_093467521.1
	A	polyketide synthase (Streptomyces castelarensis)	51%	WP_086710540.1
	A	polyketide synthase (Streptomyces hygroscopicus)	51%	WP_078646104.1
	A	polyketide synthase	51%	ASQ9929

		(Streptomyces sp. 11-1-2)		5.1
	C	beta-ketoacyl synthase (Streptomyces hygroscopicus)	41%	AQW47576.1
	C	Beta-ketoacyl synthase (Streptomyces violaceusniger Tu 4113)	41%	AEM84951.1
	C	type I modular polyketide synthase (Streptomyces malaysiensis)	41%	ATL80750.1
	C	MerC (Streptomyces violaceusniger)	41%	ABJ97439.1
	C	hypothetical protein (Streptomyces katrae)	47%	WP_079432438.1
	C	Acyl transferase domain-containing protein (Streptomyces sp. 2314.4)	45%	SEE95671.1
	C	Acyl transferase domain-containing protein (Streptomyces sp. 2112.2)	45%	SEF10106.1
	C	Type I polyketide synthase (Streptomyces caniferus)	45%	CUW01176.1
	C	Acyl transferase domain-containing protein (Streptomyces sp. 2323.1)	45%	SOE09189.1
	C	type I polyketide synthase (Streptomyces sp. 2323.1)	45%	WP_096215242.1
	C	hypothetical protein (Streptomyces angustmyceticus)	44%	WP_086719259.1
	C	type I polyketide synthase 1 (Streptomyces sp.)	44%	APD71607.1

	C	Acyl transferase domain-containing protein (Streptomyces sp. yr375)	43%	SES19425.1
	C	polyketide synthase (Streptomyces sp. 211726)	44%	ARM20280.1
	C	hypothetical protein (Streptomyces sp. TSRI0107)	43%	WP_079186763.1
	C	Acyl transferase domain-containing protein (Streptomyces sp. SceaMP-e96)	45%	SCK51529.1
	C	hypothetical protein SAM40697_6427 (Streptomyces ambofaciens)	43%	ANB10380.1
	C	putative polyketide synthase B (Streptomyces ambofaciens ATCC 23877)	43%	CAJ88175.1
	C	modular polyketide synthase (Streptomyces neyagawaensis)	46%	BAW35653.1
	C	type I polyketide synthase (Streptomyces sp. NBRC 109706)	41%	WP_062212858.1
	C	polyketide synthase type I (Streptomyces aizunensis)	42%	AAX98184.1
	C	type I polyketide synthase (Streptomyces albus)	44%	WP_061405414.1
	C	beta-ketoacyl synthase (Streptomyces sampsonii)	44%	WP_067410113.1
	C	beta-ketoacyl synthase (Streptomyces sp. FR-008)	44%	WP_075987756.1

	C	FscD (Streptomyces sp. FR-008)	44%	AAQ82568.1
	C	beta-ketoacyl synthase (Streptomyces sp. ScaeMP-6W)	44%	WP_093844724.1
	C	type I polyketide synthase (Streptomyces sp. CNY228)	44%	WP_018894183.1
	C	type I polyketide synthase (Streptomyces sp. LaPpAH-202)	44%	WP_018470932.1
	C	beta-ketoacyl synthase (Streptomyces albidoflavus)	44%	WP_071338830.1
	C	beta-ketoacyl-ACP synthase II (Streptomyces albus)	44%	WP_015508322.1
	C	polyketide synthase (Streptomyces sp. NRRL 30748)	43%	ABC87511.1
	C	beta-ketoacyl synthase (Streptomyces albireticuli)	43%	WP_095584536.1
	C	beta-ketoacyl synthase (Streptomyces sp. CS227)	44%	WP_087777615.1
	C	Acyl transferase domain-containing protein (Streptomyces melanosporofaciens)	43%	SED50232.1
	C	beta-ketoacyl synthase (Streptomyces albireticuli)	42%	WP_087930866.1
	C	ScnS2 (Streptomyces albireticuli)	42%	ARZ72143.1
	C	type I polyketide synthase (Streptomyces aureus)	44%	WP_037627924.1

	C	type I polyketide synthase (Streptomyces sp. NBRC 109706)	42%	WP_062213694.1
	C	TamAI (Streptomyces sp. 307-9)	39%	ADC79637.1
	C	TrdAI (Streptomyces sp. SCSIO 1666)	39%	ADY38531.1
	C	hypothetical protein (Streptomyces sp. NBRC 109706)	42%	WP_078857235.1
	C	3-ketoacyl-ACP reductase (Streptomyces malaysiensis)	43%	WP_099013783.1
	C	3-ketoacyl-ACP reductase (Streptomyces antioxidans)	43%	WP_075200352.1
	C	polyketide synthase 12 (Streptomyces sp. DI166)	41%	SBT91292.1
	C	polyketide synthase (Streptomyces sp. SolWspMP-5a-2)	42%	WP_093829180.1
	C	3-ketoacyl-ACP reductase (Streptomyces melanosporofaciens)	43%	WP_093461884.1
	D	Herb (Streptomyces chromofuscus)	44%	AEZ64505.1
	D	Beta-ketoacyl synthase (Streptomyces sp. 769)	43%	AJC56296.1
	D	hypothetical protein (Streptomyces yunnanensis)	42%	WP_073450222.1
	D	polyketide synthase (Streptomyces sp. 769)	43%	WP_039633289.1

	D	polyketide synthase (Streptomyces sp. SPMA113)	44%	WP_069865677.1
	D	polyketide synthase (Streptomyces malaysiensis)	44%	WP_099013784.1
	D	polyketide synthase (Streptomyces graminofaciens)	45%	BAJ16467.1
	D	modular polyketide synthase (Streptomyces blastmyceticus)	44%	BAW35635.1
	D	polyketide synthase (Streptomyces autolyticus)	44%	WP_079257696.1
	D	hypothetical protein (Streptomyces hygroscopicus)	43%	WP_078638585.1
	D	hypothetical protein M271_46355 (Streptomyces rapamycinicus NRRL 5491)	44%	AGP60636.1
	D	polyketide synthase (Streptomyces sp. NRRL 30748)	42%	ABC87509.1
	D	Acyl transferase domain-containing protein (Streptomyces melanosporofaciens)	42%	SED50090.1
	D	Type I polyketide synthase (Streptomyces caniferus)	41%	CUW01175.1
	D	Acyl transferase domain-containing protein (Streptomyces sp. 2314.4)	41%	SEE95640.1
	D	type I polyketide synthase (Streptomyces	42%	WP_0497

		caatingaensis)		17537.1
	D	TamAII (Streptomyces sp. 307-9)	41%	ADC79638.1
	D	TamAI (Streptomyces sp. 307-9)	41%	ADC79637.1
	D	TrdAI (Streptomyces sp. SCSIO 1666)	41%	ADY38531.1
	D	hypothetical protein (Streptomyces malaysiensis)	43%	WP_099012513.1
	D	AlmHI (Streptomyces sp. A1(2016))	43%	ANC94966.1
	D	Acyl transferase domain-containing protein (Streptomyces sp. 2323.1)	41%	SOE09190.1
	D	MerA (Streptomyces violaceusniger)	42%	ABJ97437.1
	D	3-ketoacyl-ACP reductase (Streptomyces malaysiensis)	42%	WP_099012633.1
	D	hypothetical protein A6A29_38940 (Streptomyces sp. TSRI0281)	40%	OKI40766.1
	D	hypothetical protein (Streptomyces sp. MUSC 1)	43%	WP_071384984.1
	D	polyketide synthase subunit (Streptomyces bikiniensis)	43%	AAS79459.1
	D	type I polyketide synthase (Streptomyces violaceusniger)	42%	WP_014058439.1
	D	hypothetical protein (Streptomyces autolyticus)	43%	WP_079256297.1

	D	hypothetical protein M271_00360 (Streptomyces rapamycinicus NRRL 5491)	42%	AGP51711.1
	D	NlmA7 (Streptomyces nanchangensis)	42%	AAS46345.1
	D	polyketide synthase (Streptomyces sp. CB02120-2)	43%	PJN20247.1
	D	hypothetical protein (Streptomyces hygroscopicus)	43%	WP_078638741.1
	D	hypothetical protein CGL27_05440 (Streptomyces sp. 11-1-2)	43%	ASQ92665.1
	D	polyketide synthase (Streptomyces autolyticus)	42%	WP_087684118.1
	D	polyketide synthase (Streptomyces autolyticus)	42%	AQA15888.1
	D	polyketide synthase (Streptomyces sp. CB02056)	43%	OKH96871.1
	D	Type I PKS (Streptomyces hygroscopicus)	42%	CCF23202.1
	D	polyketide synthase subunit (Streptomyces parvulus)	41%	WP_079163899.1
	D	beta-ketoacyl synthase (Streptomyces tsukubensis NRRL18488)	44%	EIF94454.1
Tolypothrix				

	A	type I polyketide synthase (Tolypothrix bouteillei)	52%	WP_038071985.1
	A	type I polyketide synthase (Tolypothrix campylonemoides)	47%	WP_052490492.1
	A	hypothetical protein SD81_24665 (Tolypothrix campylonemoides VB511288)	47%	KIJ74914.1
	A	short-chain dehydrogenase (Tolypothrix campylonemoides)	50%	WP_071838261.1
	A	polyketide synthase (Tolypothrix sp. NIES-4075)	51%	WP_089125746.1
	A	short-chain dehydrogenase (Tolypothrix sp. NIES-4075)	50%	WP_089125744.1
	E	hypothetical protein SD81_24665 (Tolypothrix campylonemoides VB511288)	42%	KIJ74914.1
	E	type I polyketide synthase (Tolypothrix campylonemoides)	42%	WP_052490492.1
	P	non-ribosomal peptide synthetase (Tolypothrix sp. NIES-4075)	47%	WP_089130891.1

4.2 Initial antiSMASH Analysis of Genome Sequences Containing the EAFHs

To start screening for the presence of ESGCs within the genomes of bacterial species and strains, the ends of DNA sequences found to contain individual the EAFHs were determined using the NCBI's Sequence Viewer 3.28.0.1 (Methods). Two species of *Cystobacter*, *C. fuscus* and *C. ferrugineus*, were found to have the EAFHs. *C. fuscus* was found to carry the EAFHs that show similarity with *epoA*, *epoB*, *epoD*, *epoE*, *epoF* and *epoP*. However, *C. ferrugineus* was found to have only 2 the EAFHs that are similar to *epoD* and *epoF* **(Table 1)**. The Sequence viewer showed that the the EAFHs are located within the DNA sequence 5637787-7833173 of *C. fuscus* DSM 52655. The antiSMASH analysis of this sequence detected only one ESGC. The boundaries of the DNA sequences containing each of the the EAFHs and the boundaries of the DNA sequence of *C. fuscus* strain DSM 52655 encompassing three or more of the EAFHs and expected to contain ESGCs, and the number of ESGCs detected by antiSMASH are shown in **Table 2**.

Table 2. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *C. fuscus* DSM 52655.

EAFHs	**DNA Location of the EAFHs**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOA	7817547-7833173	5637787-7833173	Cluster 1/16%
EPOB	7817547-7833173		
EPOD	7817547-7833173		
EPOE	7811056-7817550		
EPOF	5637787-5639037		

Two species of *Chondromyce*, *C. crocatus* and *C. apiculatus*, were found to have the EAFHs. *C. crocatus* had the EAFHs that are similar to *epo A, epoB, epoD, epoE, epoF* and *epoP* , whereas *C. apiculatus* had the EAFHs that are similar to *epoB*, *epoE*, *epoF epoP* (**Table 1**). The the EAFHs in *C. apiculatus* genome couldn't be found using the accession numbers of the EAFHs shown in **Table 1,** probably because the accession numbers of *C. apiculatus*'s the EAFHs found by BLASTp did not belong to *C. apiculatus* DSM 436, the strain that its whole genome sequence is available. Therefore, *C. crocatus* strain Cm c5 was chosen to identify the the EAFHs in its genome. Inspection by NCBI's Sequence Viewer showed that the the EAFHs are located within the DNA sequence 6407444-9367812 of this strain's genome. The antiSMASH analysis of this DNA sequence showed the presence of three ESGCs (**Table 3**).

Table 3. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *C. crocatus* strain Cm.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOA	7499052-7505570	6407444-9367812	Cluster 1/38%
EPOP	7522530-7526765		Cluster 2/71%
EPOB	6574167-6579932		Cluster 3/46%
EPOD	7505577-7515803		
EPOE	6407444-6414040		
EPOF	9366562-9367812		

Two species of *Melittangium* were found to have the EAFHs. *M. boletus* had the EAFHs that are similar to *epoB, epoD, epoF* and *epoP* whereas *M. lichenicola* had only 2 the EAFHs that show similarity with *epoB* and *epoP* (**Table 1)**. Therefore, those genes were only identified in *M. boletus* DSM 14713 and found to be contained within the DNA sequence 2088765-9326762. This sequence was analyzed using antiSMASH which detected only one ESGC (**Table 4)**.

Table 4. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *M. boletus* DSM 14713.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOP	9316908-9326762	2088765-9326762	Cluster 1/29%
EPOB	7210099-7215579		
EPOD	296175-305216		
EPOF	2088765-2090023		

Ten species of *Micromonospora* were found to have the the EAFHs *epoA*, *epoC*, *epoE*, and *epoP*. The only species that contained all these the EAFHs was *M. echinospora*. Therefore, those genes were searched in *M. echinospora* strain DSM 43816 to determine the DNA sequence they are located in. The DNA sequence that contained *epoA*, *epoC*, *epoE*, *epoP* was analyzed by antiSMASH and this analysis showed the presence of only one ESGC (**Table 5)**.

Table 5. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *M. echinospora* strain DSM 43816.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOA	6826726-6833310	5556236-6859090	Cluster 1/16%
EPOC	5556236-5562139		
EPOE	6852563-6859090		

Five species of *Myxococcus* **(Table 1)** were found to have the EAFHs similar to *epoB*, *epoD*, *epoE*, and *epoP* **(Table 1)**. However, the DNA sequence that contains the the EAFHs was only determined in *M. stipitatus* DSM 14675. The antiSMASH analysis of the DNA sequence (5519785-6388370) containing the the EAFHs was analyzed by antiSMASH and found to have only one ESGC **(Table 6).**

Table 6. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *M. stipitatus* DSM 14675.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOB	5526213-5541800	5519785-6388370	Cluster 1/33%

EPOB	5526213-5541800		
EPOB	5552948-5558590		
EPOD	5526213-5541800		
EPOE	5558613-5565143		
EPOE	5519785-5526216		
EPOP	6378771-6388370		

Three species of *Nannocystis* were found to have the EAFHs similar to *epoB, epoF*, *epoP* **(Table 1)**. The results of BLASTp search showed that only *N. exedens* carries *epoF*, *epoP*. The search within the complete genome sequence of *N. exedens* strain DSM 71 utg0 showed that the the EAFHs that are similar to *epoF*, *epoP* are found within the DNA sequence 2764121-2781992. The AniSMASH analysis of this sequence revealed the presence of only one ESGC **(Table 7)**.

Table 7. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *N. exedens* strain DSM 71 utg0.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOF	2780763-2781992	2764121-2781992	Cluster 1/25%
EPOP	2764121-2768398		

The BLASTp search showed that two species of *Stigmatella, S. aurantiaca* and *S. erecta,* have the EAFHs that are similar to *epoA, epoB, epoD, epoE, epoF*, and *epoP* **(Table 1)**. To determine the boundaries of the DNA sequence where the the EAFHs are

located, the complete genome sequence of *S. aurantiaca* DW4/3-1 was inspected using the NCBI's Sequence Viewer 3.28.0.1. The the EAFHs were found within the DNA sequence 070776-6608426. The antiSMASH analysis of this sequence revealed the presence of one ESGC (**Table 8).**

Table 8. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *S. aurantiaca* DW4/3-10.

EAFHs	**Location of the EAFHs in Genome**	**Boundaries of the Sequence containing the EAFHs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* Epothilone Gene Cluster**
EPOA	2070776-2077351	2070776-6608426	Cluster 1/71%
EPOB	4771261-4777539		
EPOB	4849134-4859009		
EPOB	4777536-4783184		
EPOB	4783216-4788771		
EPOD	6431955-6442136		
EPOE	2070776-2077351		
EPOF	6607176-6608426		
EPOP	4849134-4859009		

Species and strains of *Calothrix*, *Moorea*, and *Nostoc* were found to have the EAFHs **(Table 1)**. However, the antiSMASH analysis of their DNA sequences in which the EAFHs were identified did not show the presence of ESGCs **(Supplementary Tables 3, 4, 5)**.

4.3 AntiSMASH Analysis of Bacterial Complete genomes

To widen the search of ESGCs and to find other novel secondary metabolites gene clusters, the complete genome sequences of the species and strains that had found to have ESGCs were analyzed using AntiSMASH. Only clusters containing genes with more than 10% similarity to those of *S. cellulosum's* Epothilone gene cluster or to other SMs gene clusters are shown in the tables. The antiSMASH analysis of the complete genome of *C. fuscus*-strain-DSM-52655 revealed that this strain encompasses 3 ESGCs in its genome. Moreover, this strain was rich in other hybrid T1PKS-NRPS gene clusters **(Table 9).** Other secondary gene clusters including terpenes, bacteriocin-Proteusin, NRPS and T3PKS were also identified in this strain **(Supplementary Table 6).** SMs and their biological activities, which are probably encoded by BGCs and ESGCs, are listed in **Table 18.**

Table 9. ESGCs and other T1PKS-NRPS gene clusters found in the genome of *C. fuscus*-strain-DSM-52655.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T1PKS-NRPS	6907986-6971837	Cystothiazole A	41
		Landepoxcin (A and B)	11
		Epoxomicin	11
		Clarexpoxcin	11
		Tubulysin	17

		Lorneic acid A	11
T1PKS-NRPS	7304520-7370079	Gephyronic acid	14
T1PKS-NRPS	7411411-7493967	Dkxanthene	76
		Ajudazol	35
		Chlorizidine	23
		Pyralomicin	17
		Stigmatellin	41
		Nostophycin	11
		Guadinomine	11
		Myxothiazol	11
T1PKS-NRPS	7786476-7871447	Phenalamide (A2)	58
		Myxalamid(s)	58
		Chondramide	16
		Pellasoren (A)	50
		Stigmatellin	58
		Nostophycin	16
		Epothilone	25
		Epothilone	25
		Epothilone	25
		Epothilone	25
T1PKS-NRPS	8185030-8310009	Phenalamide (A2)	41

		Ajudazol	50
		Epothilone	25
		Epothilone	25
		Epothilone	25
		Epothilone	25
		Myxalamid(s)	33
		Napsamycin	16
		Pacidamycin(s)	16
T1PKS-NRPS	9828130-9896725	Cystomanamides	80
Cf_Putative	10596190-10604249	Epothilone	12

Because of the relatedness among species and strains that belong to the same genus, it was expected to find ESGCs in other species or strains that are related to those that had been found to carry ESGCs. Therefore, the complete DNA sequence of *C. fuscus*-DSM-2262 was searched for the presence of ESGCs. The antiSMASH analysis of this bacterium's complete DNA sequence showed the presence of one ESGC with a 12% similarity to epothilone gene cluster's genes **(Table 10).** However, this strain has fewer secondary metabolites gene clusters than *C. fuscus*-strain-DSM-52655 **(supplementary Table 7)**.

Table 10. ESGCs and other T1PKS-NRPS gene clusters detected in the genome of *C. fuscus*-DSM-2262.

Cluster	**Location in**	**Gene clusters it shares**	**Percent of genes of known gene**

type	genome	similarity with	clusters that showed similarity to the identified BGC
Cf_Putative	805372-813432	Epothilone	12

The initial antiSMASH analysis of the DNA sequence of *C. crocatus* strain Cm that contained the EAFHs showed that this sequence contains three ESGCs. The analysis of the complete DNA sequence of this strain showed the presence of four ESGCs, in addition to ~~another~~ other hybrid T1PKS-NRPS gene clusters **(Table 11).** Additionally, this strain encompasses other secondary metabolites gene clusters including terpenes, thiopeptide-NRPS, and T3PKS **(supplementary Table 8).**

Table 11. ESGCs and other T1PKS-NRPS gene clusters identified in the genome of *C. crocatus* strain Cm.

Cluster type	Location in genome	Gene clusters it shares similarity with	Percent of genes of known gene clusters that showed similarity to the identified BGC
T1PKS-NRPS	3931595-4020018	Crocacin	100
		Stigmatellin	42
		Ajudazol	38
		Myxothiazol	14
		Epothilone	23
		Epothilone	19
		Epothilone	19

		Cystothiazole A	9
T1PKS-NRPS	6372147-6446585	Chondramide	100
		Nannocystin	53
		Phenalamide (A2)	38
		Pellasoren (A)	46
		Epothilone	38
		Epothilone	38
		Epothilone	38
		Epothilone	38
		Myxalamid	38
T1PKS	6542321-6614252	Myxothiazol	28
		Epothilone	71
		Epothilone	71
		Epothilone	71
		Epothilone	71
		Nostophycin	28
		Microsclerodermins (A and B)	28
		Ajudazol	64
		Nystatin-like Pseudonocardia polyene	28
T1PKS-NRPS	7151424-7435649	Chondrochloren (A and B)	100

		Thuggacin (A and B)	88
		Pellasoren (A)	38
		Kedarcidin	27
		JBIR-100	27
		Soraphen(A)	22
		Guadinomine	27
		Thuggacin (A1 and B2)	22
		Concanamycin A	27
		Nostopeptolide	22
T1PKS-NRPS	7479052-7589890	Ajudazol	100
		Epothilone	46
		Epothilone	46
		Epothilone	46
		Epothilone	38
		Cystothiazole A	38
		Nostophycin	15
		Stigmatellin	69
		Curacin	69

The initial antiSMASH analysis of the EAFH-containing DNA sequence of *M. boletus* DSM 14713 showed the presence of one ESGC. This analysis was repeated for the complete DNA sequence of this strain and showed the presence of two ESGCs, as well as two hybrid T1PKS-NRPS gene clusters **(Table 12).** Additional secondary

metabolites gene clusters were also detected in this strain which includes NRPS, terpene, T2PKS, Microviridin-Lantipeptide, and Butyrolactone-Lantipeptide-T1PKS-NRPS gene clusters **(Supplementary Table 9).**

Table 12. ESGCs and other T1PKS-NRPS gene clusters identified in the genome of *M. boletus* DSM 14713.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T2PKS	273912-332916	Ajudazol	46
		Microsclerodermins (A and B)	15
		Phenalamide (A2)	30
		Stigmatellin	53
		Epothilone	23
		Epothilone	23
		Epothilone	23
		Epothilone	23
		Myxalamid(s)	23
T1PKS-NRPS	1994615-2075125	Dkxanthene	70
		Chlorizidine A	23
		Myxothiazol	11

		Pyralomicin	17
		Pyrrolomycin	17
		Stigmatellin	27
Cf_Fatty_Acid-T1PKS-NRPS	7124563-7262502	Cystothiazole A	29
		Melithiazol	29
		Myxothiazol	23
		Epothilone	29
		Epothilone	29
		Epothilone	23
		Epothilone	29
		Ajudazol	23
		Crocacin	23
T1PKS-NRPS	9281795-9353915	Melithiazol	94
		Cystothiazole A	47
		Ajudazol	42
		Myxothiazol	31
		Leupyrrin	21
		Antalid	21
		Anatoxin	15
		Cryptophycin	10
		Stigmatellin	47

Myxococcus stipitatus DSM 14675 was initially found to contain one ESGC within the DNA sequence that contained the EAFHs **(Table 6).** The complete DNA sequence of this strain was searched for the presence of any other ESGCs. As shown in **(Table 13),** this analysis confirmed the presence of only one ESGC cluster, in addition to six hybrid T1PKS-NRPS gene clusters. Other ESGCs, which included T3PKS, terpene, Cf_Saccharide, and Cf-Putative, in addition to other gene clusters were also detected **(Supplementary Table 10).**

Table 13. ESGCs and other T1PKS-NRPS gene clusters detected in the genome of *M. stipitatus* DSM 14675.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T1PKS-NRPS	2487532-2677586	Puwainaphycins (F/G)	40
		Jamaicamide	20
		Bleomycin	30
		Nostopeptolide	20
		Microsclerodermins	30
		Nostophycin	20
		Crocacin	20
T1PKS-NRPS	3002152-3075411	Puwainaphycins (F/G)	50
		Nostopeptolide	30
		Lysobactin	20

		Microsclerodermins	20
		Ralsolamycin	20
		Nostocyclopeptide	20
		Oxazolomycin	20
		Bleomycin	20
		Chondrochloren (A and B)	20
T1PKS-NRPS	4497133-4599110	Nostopeptolide	25
		Cystothiazole A	25
		Myxothiazol	25
		Melithiazol	25
		Bleomycin	25
		Microsclerodermins (A and B)	37
		Tubulysin	25
		Jamaicamide	25
T1PKS-NRPS	5255821-5335729	Myxoprincomide	66
		Bacillomycin	66
		Iturin	66
		Amphotericin (B)	66
T1PKS-NRPS	5466423-5644506	Myxochelin	91
		Phenalamide (A2)	100
		Pellasoren (A)	50
		Paenibactin	33

		Bacillibactin	33
		Griseobactin	33
		Fuscachelin	33
		Myxalamid(s)	66
		Myxothiazol	16
T1PKS-NRPS	6343701-6431814	Cystothiazole A	47
		Melithiazol	47
		Myxothiazol	35
		Tubulysin	35
		Ajudazol	47
		Microsclerodermins	23
		Antalid	11
		Hectochlorin	17
		Cylindrospermopsin	17
Bacteriocin-T1PKS-NRPS	7721562-7824172	Dkxanthene	17
		Pyralomicin	17
		Chlorizidine A	17
		Leupyrrin	23
		Microsclerodermins (A and B)	17
		Cystothiazole A	17
		Ajudazol	23
		Myxothiazol	11

		Epothilone	17
		Epothilone	17

The whole DNA sequences of other *Myxococcus* species were searched for the presence of other ESGCs. Analysis of these sequences using antiSMASH showed the presence of one ESGC and five hybrid T1PKS-NRPS gene clusters for both *M. hansupus*-strain-mixupus and *M. xanthus*-DK-1622 species **(Tables 14 and 15).** The genomes of these strains contain other secondary metabolites gene clusters **(Supplementary Tables 11 and 12).**

Table 14. ESGCs and other T1PKS-NRPS gene clusters detected in the genome of *M. hansupus*-strain-mixupus.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T1PKS-NRPS	2771375-2902802	Phenalamide (A2)	58
		Myxalamid(s)	58
		Epothilone	33
		Epothilone	33
		Epothilone	33
		Epothilone	25
		Myxothiazol	33
		Crocacin	33
		Cystothiazole A	41

		Melithiazol	41
T1PKS-NRPS	3002152-3075411	Puwainaphycins (F/G)	50
		Nostopeptolide	30
		Lysobactin	20
		Microsclerodermins	20
		Ralsolamycin	20
		Nostocyclopeptide	20
		Oxazolomycin	20
		Bleomycin	20
		Chondrochloren (A and B)	20
T1PKS-NRPS	3129770-3216154	Dkxanthene	76
		Chlorizidine A	23
		Calcimycin	17
		Pyrrolomycin	17
		Marinopyrrole	17
		Curacin (A)	29
		Stigmatellin	29
		Nostophycin	11
		Guadinomine	11
T1PKS-NRPS	3446609-3512502	Myxochromide (B3)	100

T1PKS-NRPS	3924291-4013404	Myxoprincomide	100
		Bacillomycin	66
		Iturin	66
		Amphotericin	66
T1PKS-NRPS	4135121-4237928	Microsclerodermins	18
		Myxalamid(s)	18

Table 15. ESGCs and other T1PKS-NRPS gene clusters detected in the genome of *M. xanthus*-DK-1622.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with s**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T1PKS-NRPS	4200394-4368749	Myxochelin (A)	75
		Paenibactin	33
		Bacillibactin	33
		Griseobactin	33
		Fuscachelin	33
		Microsclerodermins	41
		Esmeraldin(s)	33
		Microsclerodermins (A and B)	33
		Caerulomycin A	25

T1PKS-NRPS	4481281-4564105	Myxoprincomide	100
		Bacillomycin	66
		Iturin	66
T1PKS-NRPS	4976865-5045896	Myxochromide (B3)	100
T1PKS-NRPS	5235098-5311992	Dkxanthene	76
		Cystothiazole A	11
		Microsclerodermins	17
		Chlorizidine A	17
		Stigmatellin	35
		Nostophycin	11
		Guadinomine	11
		Ajudazol	23
T1PKS-NRPS	5387416-5488927	Puwainaphycins (F/G)	50
		Microsclerodermins	20
		Nostopeptolide	30
		Ralsolamycin	20
T1PKS-NRPS	5580883-5673548	Phenalamide (A2)	58
		Myxalamid(s)	58
		Pellasoren (A)	50
		Chondrochloren (A and	33

	B)	
	Stigmatellin	58
	Epothilone	25
	Epothilone	25
	Epothilone	25
	Epothilone	25
	Antalid	16

The initial antiSMASH analysis of the DNA sequence of *N. exedens*-strain-DSM-71-utg0 that was found to contain the EAFHs revealed the presence of only one ESGC **(Tables 7).** The antiSMASH analysis of the whole DNA sequence of this strain showed the presence of one ESGC and another hybrid T1PKS-NRPS gene cluster **(Tables 16).** Other secondary metabolites gene clusters, which include phenazine, terpene, terpene-lantipeptide-T1PKS-NRPS, and an Unknown gene cluster are shown in **(Supplementary Table 13).**

Table 16. ESGCs and other T1PKS-NRPS gene clusters detected in the genome of *N. exedens*-strain-DSM-71-utg0.

Cluster type	**Location in genome**	**Gene clusters it shares similarity with s**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
Terpene-Lantipeptide-T1PKS-NRPS	1833443-1941377	Nostopeptolide	37
		Microsclerodermins	37

		Epothilone	25
		Kedarcidin	25
		Jagaricin	25
		Tallysomycin(s)	25
		Bleomycin	25
		Sporolide	25
T1PKS-NRPS	2733381-2794934	Coumermycin (A1)	42
		Clorobiocin	42
		phenalinolactone (A-D) pyrrole	28
		Chlorothricin MSAS/OSAS	28

The initial gene cluster analysis using antiSMAH showed that *S. aurantiaca* DW4/3-1 had only one ESGC within the DNA sequence 2070776-6608426 **(Table 8).** The antiSMASH analysis of the complete DNA sequence of this species revealed the presence of one ESGC and two hybrid T1PKS-NRPS gene clusters **(Table 17)**. Additionally, various secondary metabolites gene clusters which include Terpene, T3PKS, T1PKS-Cf_Saccharide, NRPS, Bacteriocin-T1PKS-NRPS, Bacteriocin-Lantipeptide-T1PKS-NRPS, T1PKS, T1PKS-Ectoine, Microviridin, and fatty acid were also found in this strain **(Supplementary Table 14).**

Table 17: ESGCs and other T1PKS-NRPS gene clusters detected in the genome *S. aurantiaca* DW4/3-1.

Cluster type	**Location in**	**Homologous known**	**Percent of genes of known gene clusters that**

	genome	gene clusters	showed similarity to the identified BGC
T1PKS-NRPS	6854-83186	Nostopeptolide	25
T1PKS - Cf_Saccharide-NRPS	4751261-4894265	Myxothiazol	100
		Cystothiazole A	100
		Melithiazol	85
		Epothilone	71
		Epothilone	71
		Epothilone	71
		Epothilone	71
		Myxalamid	100
		Stigmatellin	100
T1PKS-NRPS	5092487-5154581	Cystothiazole A	11
		Myxothiazol	11
		Tubulysin	11
		Microsclerodermins	17
		Jamaicamide	11
		Melithiazol	11

4.4 Phylogenetic relationship among the species and strains that have ESGCs

To study the evolutionary relationships among bacterial species found to harbor ESGCs, and to understand the horizontal transfer of ESGCs, BGCs, and their genes, a phylogenetic analysis was conducted on their 16s rRNA gene sequences **(Figure 2).**

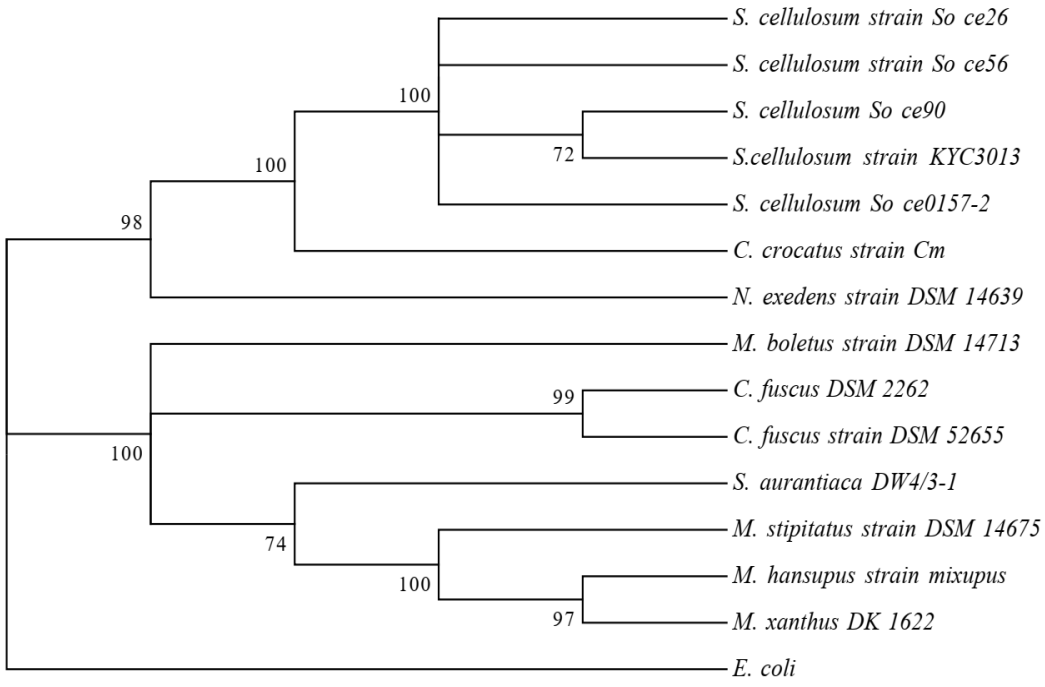


Figure 2. Evolutionary relationships among the species and strains that have ESGCs. The phylogenetic tree was constructed by MEGA 7.0 software using the Maximum Likelihood method based on the Tamura-Nei model [246]. 10,000 bootstrap replicates were used to infer tree topology and the branches with more than 50% bootstrap support are shown. The analysis involved 15 nucleotide sequences of the16s rRNA gene, and the *E. coli* sequence was used to root the tree.

4.5 In Silico Design of ESGCs

To design *in silico* ESGCs, the the EAFHs identified by antiSMASH in some bacterial ESGCs gene clusters were used. The the EAFHs boundaries were determined by antiSMASH and then excised and replaced their counterparts in the Epothilone gene cluster of *S. cellulosum* So ce90. The new gene clusters were reanalyzed by antiSMASH

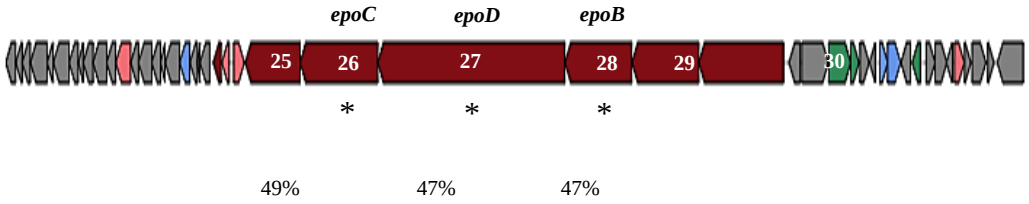
to see the effect of gene exchange on their domain types and the predicted scaffolds they would produce.

The the EAFHs 26, 27 and 28 of the T1PKS-NRPS gene clusters (7786476-7871447) of *C. fuscus*-strain-DSM-5265 showed similarity with *epoC, epoD, epoB* respectively. Therefore, *epoC, epoD*, *epoB* of Epothilone gene cluster of *S. cellulosum* So ce90 were replaced with the DNA sequences of these genes. The new gene cluster, the size of its genes, its domain composition, and the scaffold structure it is expected to produce are shown in **(Figure 3).**

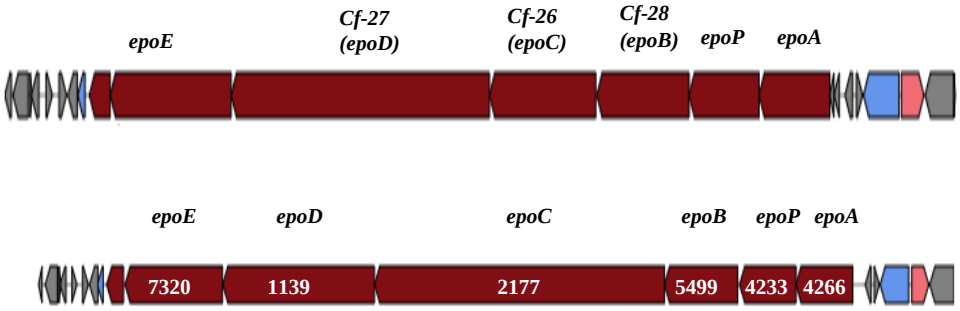
The the EAFHs 18, 16, 17 of the T1PKS-NRPS gene cluster (3931595-4020018) of *C. crocatus* strain Cm were used to substitute (*epoA, epoB), epoC,* and *epoD,* of Epothilone gene cluster of *S. cellulosum* So ce90, respectively, to generate a new ESGC. The new gene cluster, the size of its genes, its domain composition, and the scaffold structure it is expected to produce are shown in **(Figure 4).**

To generate a new ESGC from the T1PKS-NRPS gene cluster (6372147-6446585) of *C. crocatus* strain Cm, the DNA sequences of the EAFH 17 replaced (*epoA, epoC,* and *epoE),* whereas the EAFH 18 replaced *epoB* and *epoD* of Epothilone gene cluster of *S. cellulosum* So ce90 (**Figure 5**).

A.



B.

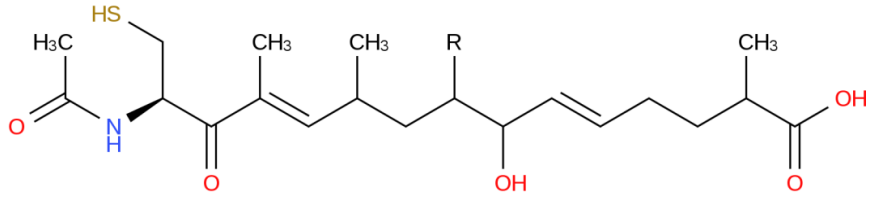


C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
epoA	KS-AT-ER-ACP	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
Cf-28 (epoB)	KS-AT-DH-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
Cf-26 (epoC)	KS-AT-DH-ER-KR-ACP	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-KR-ACP-KS-AT-DH-ER-KR-ACP
Cf-27 (epoD)	KS-AT-DH-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP
epoE	KS-AT-DH-ER-KR-ACP-TE	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

Gene cluster (7786476-7871447)



Epothilone Gene cluster

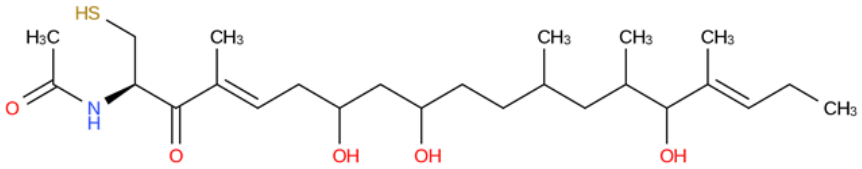
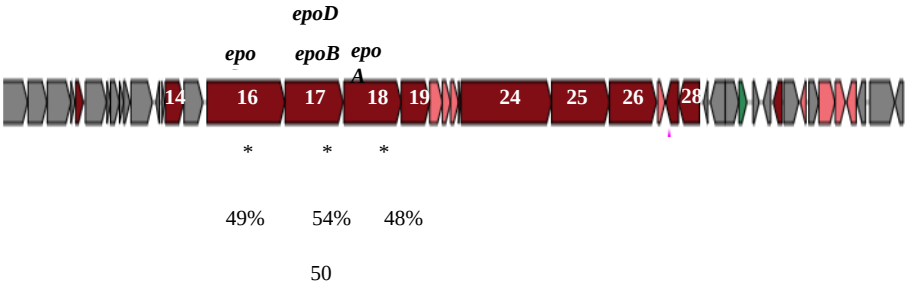


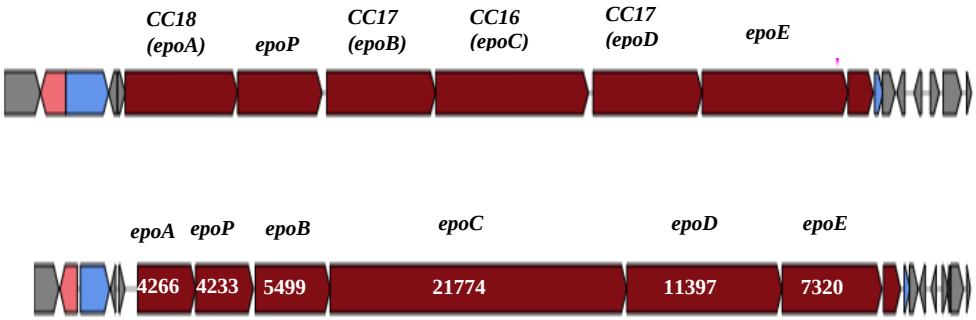
Figure 3. In silico design of ESGC using genes from the gene cluster (7786476-7871447) of *C. fuscus*-strain-DSM-5265. **(A)** Gene cluster 7786476-7871447(84954bp). The genes marked with asterisks are the genes that showed similarity to those of Epothilone gene cluster of *S. cellulosum* So ce90. The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 28, 26, 27, and *epoB*, *epoC*, *epoD* of *S. cellulosum* So ce90, respectively. **(B)** The new gene cluster (57513 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 26, 27, and 28 that replaced *epoC*, *epoD*, and *epoB*, of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the Gene cluster (7786476-7871447) (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). The antiSMASH software predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.

Core genes | Additional biosynthetic genes | Transport-related genes | Regulatory genes | Other genes

A.



B.



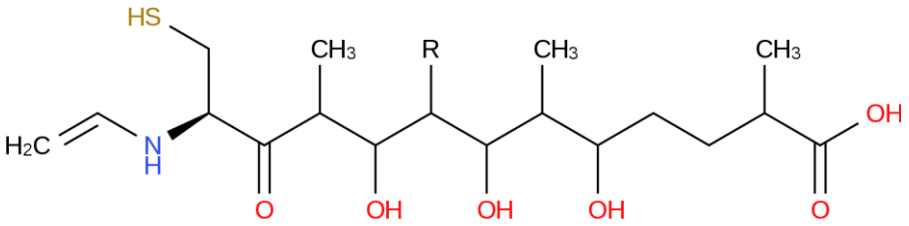
C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
CC18 (epoA)	KS-AT-ACP-DH-KR-ACP	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
CC17 (epoB)	KS-AT-OMT-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
CC16 (epoC)	CAL-ACP-KS-AT-OMT-KR-ACP	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-KR-ACP-KS-AT-DH-ER-KR-ACP
CC17 (epoD)	KS-AT-OMT-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP

epoE	KS-AT-DH-ER-KR-ACP-TE	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

Gene cluster (3931595-4020018)



Epothilone Gene cluster

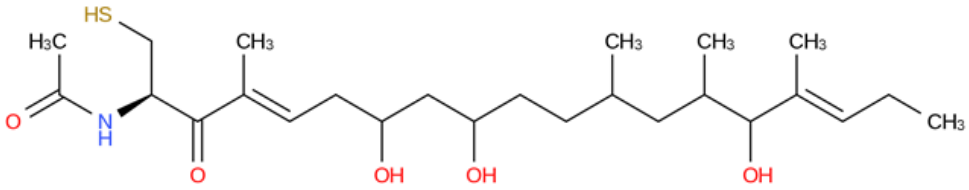
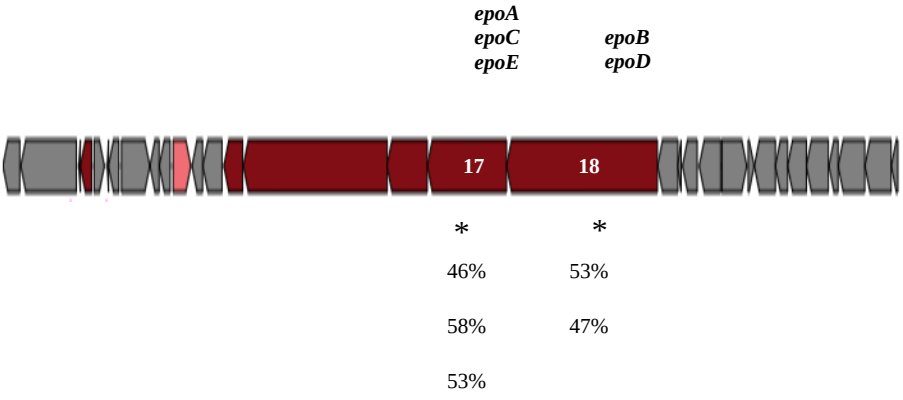


Figure 4. In silico design of ESGC using genes from the gene cluster (3931595-4020018) of C. crocatus strain Cm. **(A)** Gene cluster 3931595-4020018 (88424bp). The genes marked with asterisks are the genes that showed similarity to those of Epothilone gene cluster of *S. cellulosum* So ce90 (shown above the gene cluster). The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 18, 16, 17, and (*epoA, epoB), epoC, epoD,* of *S. cellulosum* So ce90, respectively. **(B)** The new gene cluster (48503 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 18, 16, 17 that replaced *(epoA, epoB), epoC, epoD,* of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the Gene cluster (7786476-7871447) (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). AntiSMASH predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.

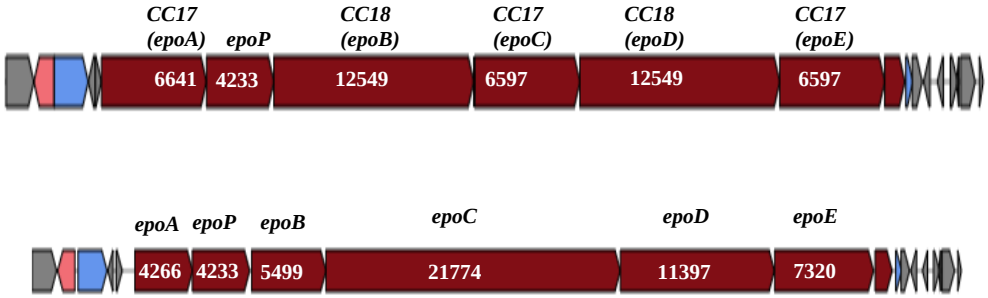
Core genes Additional biosynthetic Transport-related Regulatory genes Other genes

genes genes

A.



B.



C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
CC17 *(epoA)*	KS-AT-DH-ER-KR-ACP	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
CC18 (epoB)	ACP-KS-AT-AT-DH-KR-ACP-KS-AT-DH-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
CC17 (epoC)	KS-AT-DH-ER-KR-ACP	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-KR-ACP-KS-AT-DH-ER-KR-ACP
CC18 (epoD)	ACP-KS-AT-AT-DH-KR-ACP-KS-AT-DH-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP
CC17 (epoE)	KS-AT-DH-ER-KR-ACP	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

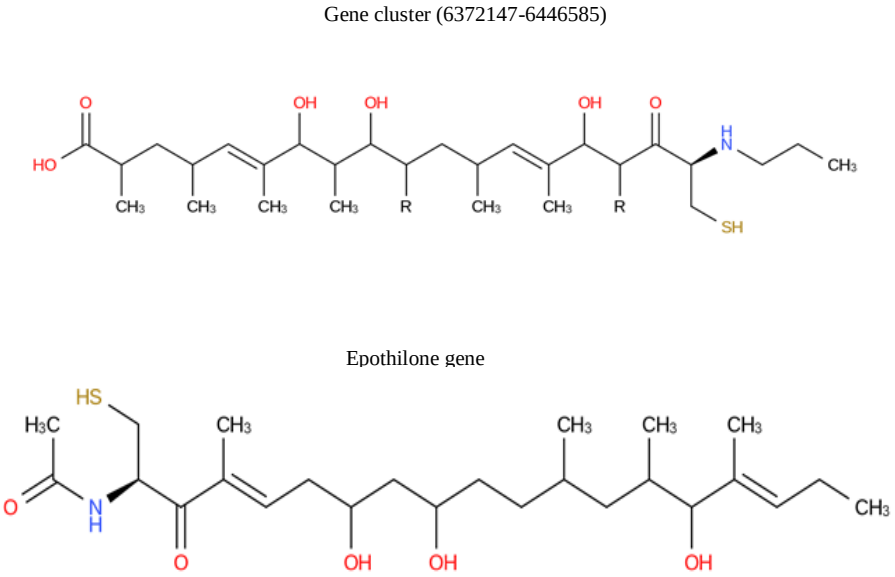


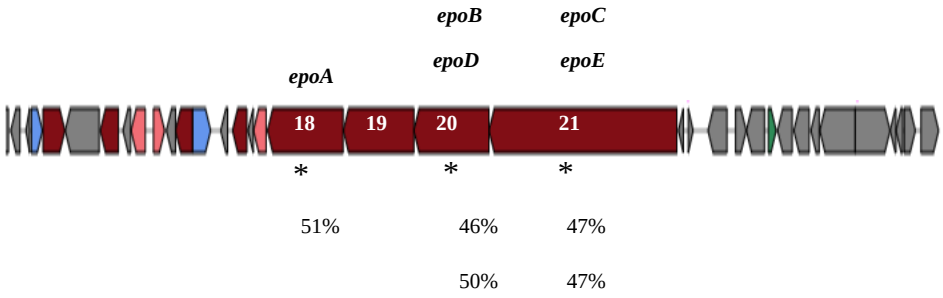
Figure 5. In silico design of ESGC using genes from the gene cluster (6372147-6446585) of C. crocatus strain Cm c5. **(A)** Gene cluster 6372147-6446585 (74439 bp). The genes marked with asterisks are the genes that showed similarity to Epothilone gene cluster of *S. cellulosum* So ce90 (shown above the gene cluster). Above the gene cluster is the percent of genes of this cluster that showed similarity to Epothilone gene cluster of *S. cellulosum* So ce90. The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 17, 18 and (*epoA, epoC, epoE), (epoB, epoD),* of *S. cellulosum* So ce90, respectively. **(B)** The new gene cluster (61446 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 17, 18 that replaced (*epoA, epoC, epoE), (epoB, epoD),* of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the new gene cluster (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). AntiSMASH predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.

Another ESGC derived from the T1PKS-NRPS gene cluster (6542321-6614252) of *C. crocatus* strain Cm c5 was designed by replacing (epoA), (epoB, epoD), (epoC, epoE) of Epothilone gene cluster of *S. cellulosum* So ce90 with the genes 18, 20, 21, respectively **(Figure 6).**

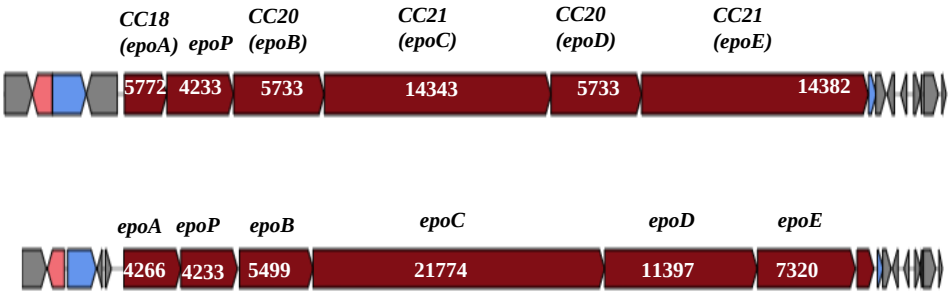
The gene cluster **(**273912-332916) of *M. boletus* DSM 14713 contained the EAFHs 20, 21 that showed similarity to (*epoA)* and *(epoC, epoD),* respectively of the Epothilone gene cluster of *S. cellulosum* So ce90. Therefore, generation of a new ESGC involved substitution of (*epoA)* and *(epoC, epoD)* with the the EAFHs 20 and 21, respectively **(Figure 7).**

Substitution of (*epoC),(epoD), (epoB),* of Epothilone gene cluster of *S. cellulosum* So ce90 with the the EAFHs 18, 19, 20 of the T1PKS-NRPS gene cluster (5580883-5673548) of *M. xanthus*-DK-1622 resulted in a new ESGC as shown in **(Figure 8).**

A.



B.



C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
CC18 (epoA)	KS-AT-DH-KR-ACP	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
CC20 (epoB)	KS-AT-DH-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
CC21 (epoC)	CAL-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-ER-KR-ACP	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-KR-ACP-KS-AT-DH-ER-KR-ACP

CC20 (epoD)	KS-AT-DH-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP
CC21 (epoE)	CAL-ACP-KS-AT-DH-KR-ACP-KS-AT-DH-ER-KR-ACP	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

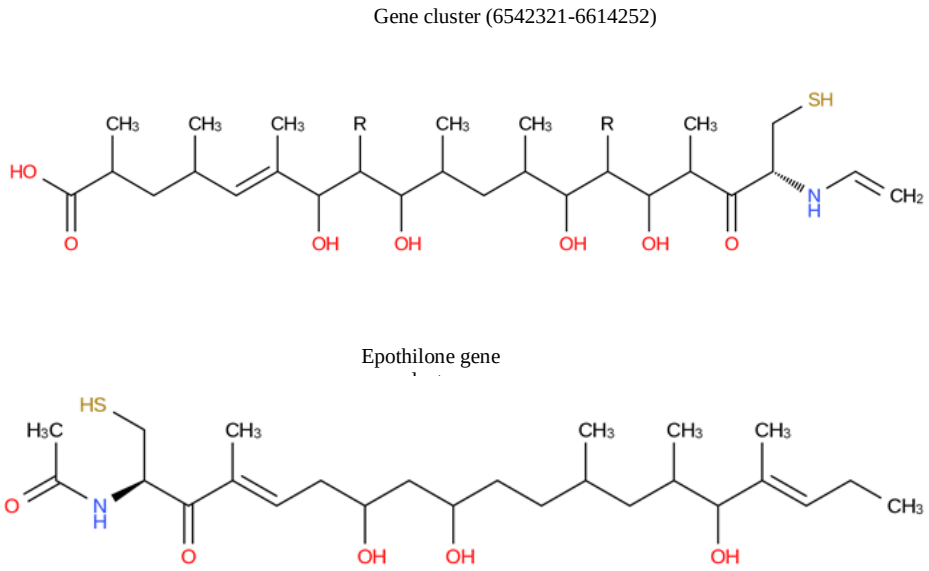
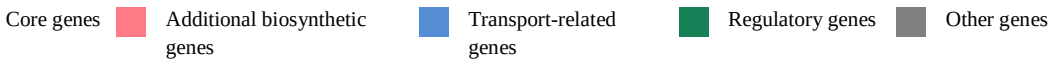
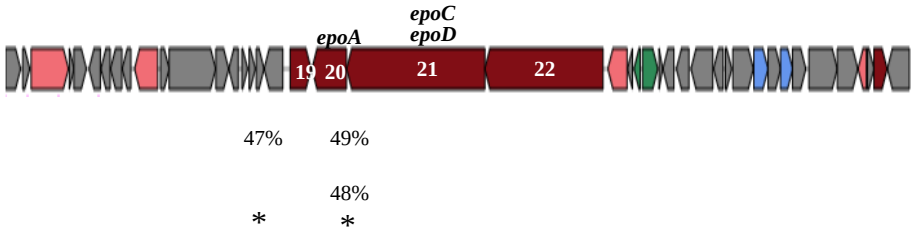


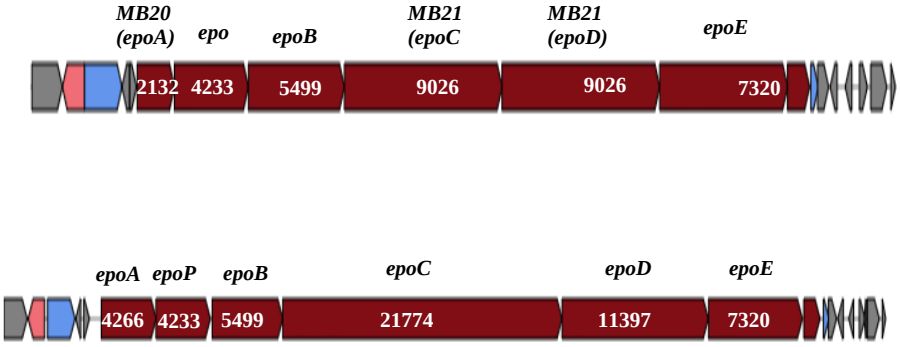
Figure 6. In silico design of ESGC using genes from the gene cluster (6542321-6614252) of C. crocatus strain Cm c5. **(A)** Gene cluster 6542321-6614252 (71932 bp). The genes marked with asterisks are the genes that showed similarity to those of Epothilone gene cluster of *S. cellulosum* So ce90 (shown above the gene cluster). The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 18, 19, 20 and (*epoC),(epoD), (epoB),* of *S. cellulosum* So ce90, respectively. **(B)** The new gene cluster (62481 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 18, 20, 21 that replaced (*epoA), (epoB, epoD), (epoC, epoE),* of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the new gene cluster (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). The antiSMASH software predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.



A.



B.



C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
MB20 (epoA)	KS	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
epoB	KS-AT-DH-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
MB21	KS-AT-KR-ACP-KS-DHt-DH-KR-	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-

(epoC)	ACP		KR-ACP-KS-AT-DH-ER-KR-ACP
MB21 (epoD)	KS-AT-KR-ACP-KS-DH-DH-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP
epoE	KS-AT-DH-ER-KR-ACP-TE	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

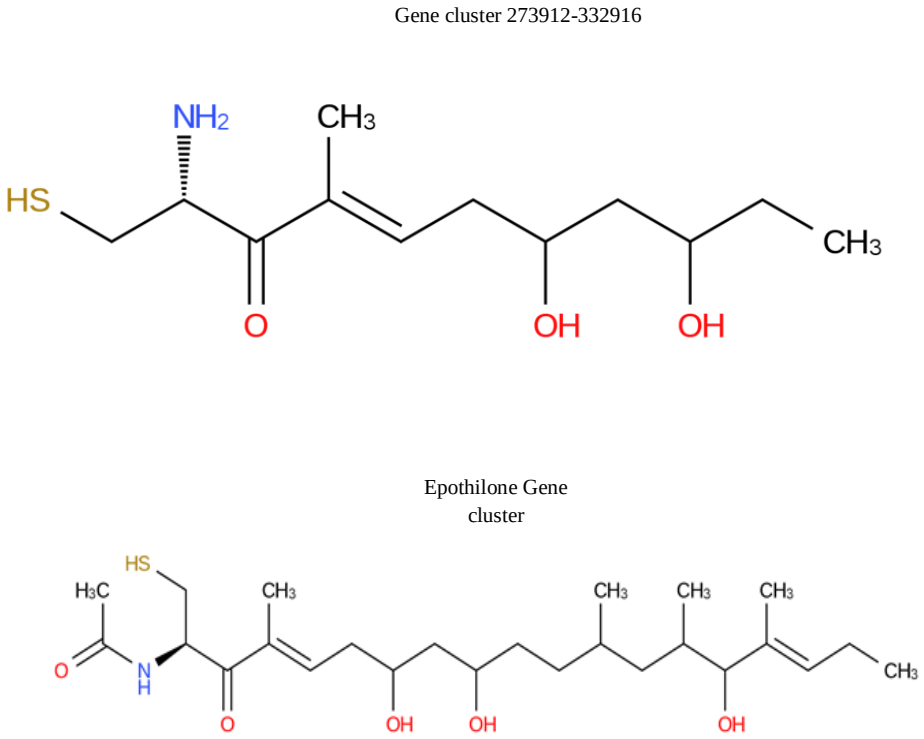
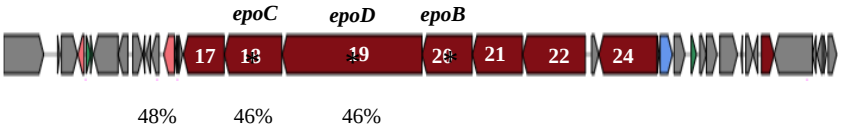


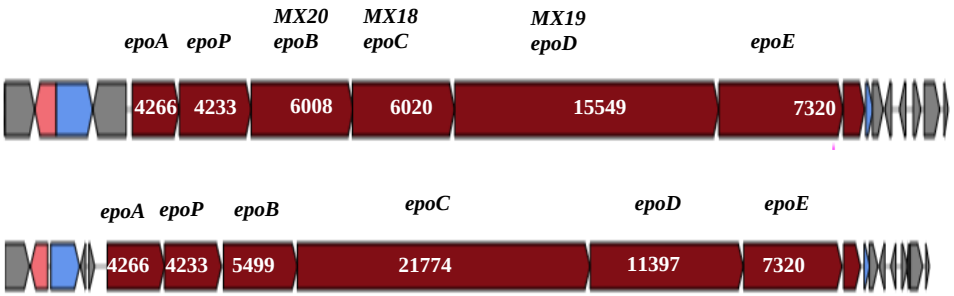
Figure 7. In silico design of ESGC using genes from the gene cluster (273912-332916) of M. boletus DSM 14713. **(A)** Gene cluster 273912-332916 (59005bp). The genes marked with asterisks are the genes that showed similarity to those of Epothilone gene cluster of *S. cellulosum* So ce90 (shown above the gene cluster). The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 20, 21 and (*epoA),(epoC, epoD),* respectively of *S. cellulosum* So ce90. **(B)** The new gene cluster (49533 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 20, 21 that replaced (*epoA),(epoC, epoD),* of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the Gene cluster (273912-332916) (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). The antiSMASH software

predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.

A.



B.



C.

New gene cluster		Epothilone gene cluster	
Gene	Domains	Gene	Domains
epoA	AT-ER-ACP	*epoA*	KS-AT-ER-ACP
epoP	C-A-ACP	*epoP*	C-A-ACP
MX20 *(epoB)*	KS-AT-DH-KR-ACP	*epoB*	KS-AT-DH-KR-ACP
MX18	KS-AT-DH-ER-KR-ACP	*epoC*	KS-AT-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-ER-

(epoC)			KR-ACP-KS-AT-DH-ER-KR-ACP
MX19 (epoD)	KS-AT-DH-KR-ACP-KS-AT-KR-ACP-KS-AT-DH-KR-ACP	*epoD*	KS-AT-KR-ACP-KS-AT-DH-CMT-KR-ACP
epoE	KS-AT-DH-ER-KR-ACP-TE	*epoE*	KS-AT-DH-ER-KR-ACP-TE

D.

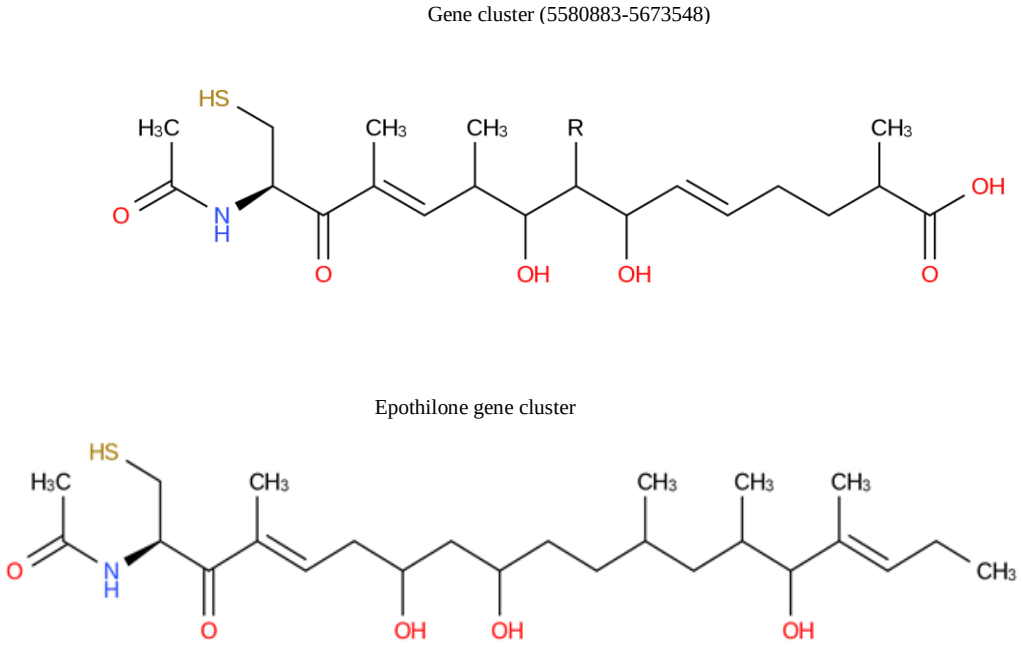


Figure 8. In silico design of ESGC using genes from the gene cluster (5580883-5673548) of M. xanthus-DK-1622. **(A)** Gene cluster 5580883-5673548 (92666). The genes marked with asterisks are the genes that showed similarity to those of Epothilone gene cluster of *S. cellulosum* So ce90 (shown above the gene cluster). The numbers in the boxes are the numbers of genes of this gene cluster. Below the gene cluster is the percent similarity between the genes 18, 19, 20 and (*epoC),(epoD), (epoB),* of Epothilone gene cluster of *S. cellulosum* So ce90, respectively. **(B)** The new gene cluster (55599 bp, top) and Epothilone gene cluster of *S. cellulosum* So ce90 (68750 bp, bottom). Above the gene cluster are the genes 20, 21 that replaced (*epoA),(epoC, epoD),* of *S. cellulosum* So ce90, respectively. The numbers in the boxes represent gene sizes (bp). **(C)** Domains found in each gene (module) of gene clusters. **(D)** Predicted core scaffold structures of the new gene cluster (top) and Epothilone gene cluster of *S. cellulosum* So ce90 (bottom). The

antiSMASH software predicts scaffold structures based on assumed PKS/NRPS collinearity without taking tailoring reactions into account.

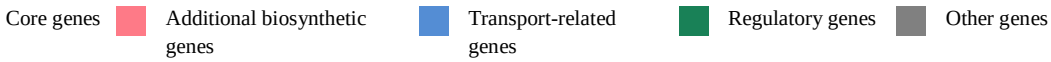


Table 18. Biological activities of SMs that are probably encoded by ESGCs.

Compound	Biological Activity	Reference
Cystothiazole A	Antifungal/novel inhibitor of mitochondrial oxidation	[247]
Landepoxcin (A and B)	Potent epoxyketone proteasome inhibitors	[248]
Epoxomicin	Proteasome inhibitor/ Potent *Plasmodium falciparum* Gametocytocidal Activity	[249]
Clarexpoxcin	Potent epoxyketone proteasome inhibitors	[248]
Tubulysin	Cytotoxic peptide/antimitotic activity	[250]
Gephyronic acid	A cytostatic polyketide/ selective eukaryotic protein synthesis inhibitor	[251]
Dkxanthene	Developmental sporulation	[252]
Ajudazol	Antifungal	[253]
Chlorizidine	Cytotoxic substance	[254]
Pyralomicin	Antibiotic	[255]
Stigmatellin	Antifungal/ Inhibitory of the mitochondrial bc_1 complex	[256]
Nostophycin	A compound with cytotoxic activity	[257]
Microsclerodermins (A and B)	Antifungal	[258]

Guadinomine	Antibiotic/inhibitor of the type III secretion system (TTSS) of Gram-negative bacteria	[259]
Thuggacin (A1 and B2)	Antibiotics active against Mycobacterium tuberculosis	[260]
Concanamycin A	Antibiotic	[261]
Nostopeptolide	A compound that controls cellular differentiation	[262]
Epothilone	Antitumor	[243]
Nostophycin	Weak Cytotoxic activity	[257]
Curacin	A potent cancer cell toxin	[263]
Phenalamide (A2)	inhibit respiratory complex I/inhibit HIV-1 replication	[264,265]
Pellasoren (A)	Antitumor	[266]
Myxothiazol	Inhibitory of the mitochondrial bc_1 complex	[267]
Chondrochloren (A and B)	Antibiotic	[147]
Kedarcidin	Antitumor antibiotic	[268]
JBIR-100	Antibiotic	[268]
Soraphen(A)	Inhibitor of HIV-1 virus	[269]
Melithiazol	Antifungal	[270]
Paenibactin	A catecholate siderophore	[271]
Leupyrrin	Antifungal	[272]
Antalid	No activity reported	[273]
Anatoxin	A neurotoxin	[274]
Puwainaphycins (F/G)	Necrosis inducers via cell membrane	[275]

	permeabilization	
Jamaicamide	A neurotoxin	[276]
Lysobactin	Antibiotic	[277]
Ralsolamycin	Inducer of chlamydospore formation in fungi	[278]
oxazolomycin	Antibiotic	[279]
Bacillomycin	Antifungal	[280]
Iturin	Antifungal	[281]
Amphotericin B	Antifungal/ treatment of leishmaniasis	[282]
Myxochelin (A)	Antitumor agent	[283]
Griseobactin	A catechol-peptide siderophore	[284]
Fuscachelin	A siderophore	[285]
Mirubactin	A siderophore	[286]
Heterobactins	siderophores	[287]
Nataxazole	Antitumor agent	[288]
A33853	Antibiotic	[289]
Cetoniacytone A	Antitumor agent	[290]
Hectochlorin	Antifungal/cytotoxic activity	[291]
Cylindrospermopsin	Cyanobacterial toxin	[292]
Bleomycin	Antitumor	[293]
Sporolide	HIV-1 reverse transcriptas	[294]
Lomofungi	Antimicrobial agent	[295]
Pyocyanine	Antibiotic	[296]
Endophenazines	Antibiotics	[297]

Esmeraldin	Antitumor	[298]
Caerulomycin A	Immunosuppressive	[299]
Kedarcidin	Antitumor	[300]
Jagaricin	Antifungal	[301]
Tallysomycin	Antitumor	[302]
Coumermycin	Antibiotic	[303]
Clorobiocin	Antibiotic	[304]
Phenalinolactone (A-D)	Antibiotics	[305]
Chlorothricin MSAS/OSAS	Antibiotics	[306]
Cystomanamides	Unknown	[307]

CHAPTER V

DISCUSSION

Hybrid PKS-NRPS gene clusters encode a wide variety of complex molecules that possess several biological activities. Compounds such as aspyridone A (cytotoxic), cytochalasin E (anti-angiogenic), equisetin F (inhibitor of HIV-1 integrase) [308], and Epothilone (antitumor) [243] are just some examples of the huge number of compounds encoded by hybrid PKS-NRPS gene clusters. Since the phylogenetically similar strains share similar gene clusters [309] it was anticipated that ESGCs will be detected in those species and strain that are more closely related to *S. cellulosum* So ce90's. Therefore, to find gene clusters that potentially encode novel ESGCs and other biologically-active secondary metabolites, the genes forming the hybrid PKS-NRPS Epothilone gene cluster of *S. cellulosum* So ce90's were used to search for BGCs within bacterial genome sequences available in the gene bank. Amino acid sequences of individual genes related to this gene cluster were blasted against NCBI's non-redundant protein database. The the EAFHs were found in 93 bacterial genera. Nineteen genera were found to have 3 or more the EAFHs in one or more of their species or strains (**Table 1).** This study focused species or strains of *Myxococcales' genera Cystobacter, Chondromyces, Melittangium,*

Myxococcus, *Nannocystis*, and *Stigmatella* that had been found to have 3 or more the EAFHs and also closely related to *S. cellulosum*. By doing so, there would be a higher chance of *in silico* detection and characterization of ESGCs in those strains' genomes. Screening of the DNA sequences that contained the the EAFHs is helpful but not enough to find all ESGCs within bacterial genomes. Initial analysis of the bacterial DNA segments that contained the the EAFHs relied on the determination of the boundaries of this segment before analyzing it with antiSMASH or other bioinformatics tools. Screening analysis of the DNA sequence of *C. fuscus*-strain-DSM-52655 that contained the the EAFHs by antiSMASH showed the presence of only 1 ESGCs **(Table 2).** However, the antiSMASH analysis of its complete genome showed the presence of 3 ESGCs **(Table 2).** This applies also to other species and strains. Although the screening analysis did not detect all biosynthetic gene clusters in question, however, this method would be the starting point for further analyses.

The gene clusters identified in this study can be grouped into three categories: 1) ESGCs that showed high similarity with the Epothilone gene cluster, and those would potentially encode Epothilone-similar compounds 2) Gene clusters that showed high similarity to those that encode known SMs. Those SMs were isolated from bacteria but their gene clusters have not been identified yet. 3) Gene clusters that showed relatively low similarities to more than one secondary metabolite gene clusters. In accordance with what previously expected [310], these gene clusters are likely to encode novel SMs. It is worth mentioning that the genes of most of the identified gene clusters share similarity with those forming gene clusters that encode cytotoxic or antimicrobial compounds.

Therefore, the possibility that these gene clusters encode novel anticancer compounds cannot be excluded. However, wet lab experiments are required.

The genome of *C. fuscus*-strain-DSM-52655 was found to have many hybrid PKS-NRPS gene clusters. The T1PKS-NRPS gene clusters (7786476-7871447) and (8185030-8310009) showed similarity with the Epothilone gene cluster were. However, they also showed similarity with other gene clusters encoding cytotoxic compounds. The gene cluster (7786476-7871447) showed its highest similarity with gene clusters that encode Phenalamide, Myxalamide, Pellasoren, and Stigmatellin. The gene cluster (8185030-8310009) showed its highest similarity with gene clusters that encode Phenalamide, Ajudazol, and Myxalamide. None of these compounds have been isolated from this strain nor have their gene clusters been identified. However, methylmyxalamide D was isolated from, and its gene cluster was characterized in *C. fuscus* strain AJ-13278 [311,312]. The Phenalamide gene cluster has been characterized in another myxobacterium (*M. stipitatus* DSM 14675) [313]. Ajudazol, Stigmatellin, and Pellasoren gene clusters were respectively identified in *C. crocatus* Cm c5 [314], and *S. aurantiaca* Sg a15 [315]. The relatively low similarities of these gene clusters with the Epothilone and other gene clusters that encode cytotoxic compounds may indicate their potential to encode novel SMs. The T1PKS-NRPS gene cluster (6907986-6971837) showed the highest similarity to Cystothiazole A, where 41% of its genes show similarity to Cystothiazole gene cluster, a compound that has antifungal activity [247]. Although the Cystothiazole gene cluster has never been characterized in *C. fuscus*-strain-DSM-52655, however, Cystothiazole A has been isolated from, and its gene cluster has been characterized in the closely related strain *C. fuscus* strain AJ-13278 [316]. The low similarity of this gene cluster may indicate its

potential to encode a Cystothiazole-like compound or novel SMs. The T1PKS-NRPS gene cluster (7304520-7370079) showed very low similarity to the Gephyronic acid gene cluster. Gephyronic acid has only been isolated from another *Cystobacter* species, *C. violaceus* strain Cb vi76. The T1PKS-NRPS gene cluster (7411411-7493967) showed a 76% similarity with the Dkxanthene gene cluster. Dkxanthene is an SM that is required for developmental sporulation of other myxobacteria. However, the Dkxanthene gene cluster has not been identified in this strain but in another myxobacterium, *M. xanthus* DK1050 [252]. Therefore, there is a possibility that this gene cluster encodes Dkxanthene-like compounds. The T1PKS-NRPS gene cluster (9828130-9896725) showed a high similarity (80%) with the Cystomanamides gene cluster. Cystomanamides gene cluster was identified in *C. fuscus* MCy9118 but not in this strain [307]. This gene cluster is probably encoding a Cystomanamides-like compound. The gene cluster (10596190-10604249) only showed low similarity to the Epothilone gene cluster, indicating its potential to encode a novel SM.

The aniSMASH analysis of *C. crocatus* strain Cm showed the presence of 4 T1PKS-NRPS gene clusters, a T1PKS gene cluster, and other gene clusters **(Table 11)** and (**Supplementary Table 8)**. Although the T1PKS-NRPS gene clusters (3931595-4020018) and (6372147-6446585) showed similarity with the Epothilone gene cluster, both gene clusters showed a 100% similarity with Crocacin and Chondramide gene clusters, and both were previously identified in this strain [317,318]. The T1PKS gene cluster 6542321-6614252 showed the highest percent similarity with the Epothilone gene cluster among all ESGCs detected by antiSMASH. This gene cluster has the potential to encode a novel Epothilone-like compound, not only because it encompasses high percentage of

genes that share similarity with those of Epothilone gene cluster, but also because it is bordered by other biosynthetic genes that impact the tailoring of the final product encoded by this gene cluster. The T1PKS-NRPS gene cluster 7151424-7435649 contains genes that showed 88 % and 38 % similarity with Thuggacin and Pellasoren Gene clusters' genes, respectively. However, 100% of its genes showed similarity with the genes of the Chondrochloren gene cluster that has been previously identified in this strain [319]. The genes of T1PKS-NRPS gene cluster (7479052-7589890) showed 46 % and 100% similarity with Epothilone and Ajudazol gene clusters' genes, respectively. Ajudazol has been previously isolated from this strain [253], and its gene cluster has been identified in this strain.

M. boletus DSM 14713 genome contained 3 T1PKS-NRPS gene clusters, a T2PKS (**Table 12)**, and other BGCs **(Supplementary Table 9).** The genes of the T2PKS gene cluster (273912-332916) showed 53%, 46%, 30%, and 23% similarity with Stigmatellin, Ajudazol, Phenalamide, and Epothilone gene clusters, respectively. However, none of these gene clusters have been previously characterized in this strain. All these gene clusters encode compounds that have cytotoxic activities (**Table 18).** The relatively low smiliraity of this gene clusters with those gene clusters may indicate its potential to encodes a novel SM. The genes of the T1PKS-NRPS gene cluster (1994615-2075125) showed a 70% similarity with the genes of the Dkxanthene gene cluster. This gene cluster has never described in this strain. However, its high similarity to the Dkxanthene gene cluster identified in other myxobacterial species may indicate the possibility that it can encode a Dkxanthene-like compound. The gene cluster (7124563-7262502) showed close percent similarities (23-29%) to the Epothilone gene cluster and

other BGCs that encode cytotoxic compounds. This low similarity of this gene cluster to those gene clusters may indicate its potential to encode a novel SM. The high similarity of the T1PKS-NRPS gene cluster (9281795-9353915) to the Melithiazol gene cluster (94%) strain indicates the potential ability of this gene cluster to encode Melithiazol. The Melithiazol gene cluster has never been identified in this strain. Moreover, the production of Melithiazol by this strain has never been reported. However, the Melithiazol gene cluster was identified in another species (*M. lichenicola* Me l46).

The genome of *M. stipitatus* DSM 14675 contained the highest number of T1PKS-NRPS gene clusters among all myxobacterial species examined in this study. The T1PKS-NRPS gene clusters (2487532-2677586) and (3002152-3075411) share their highest similarities with Puwainaphycins gene cluster and to a lesser extent with other gene clusters encoding cytotoxic compounds (**Table 13)**. Puwainaphycins cytotoxic compounds are Cyanobacterial cyclic lipopeptides produced by cyanobacteria. The degree of similarity of these gene clusters with known gene clusters may indicate their potential to encode novel SMs. The highest similarity of the T1PKS-NRPS gene cluster (4497133-4599110) was with the gene cluster that encodes Microsclerodermins (37%) and to a lesser extent with other gene clusters encoding compounds with cytotoxic activities. This gene cluster has never been identified in *Mellitangium* species; however, it was identified in *Sorangium* and *Jahnella* species. The T1PKS-NRPS gene cluster 5255821-5335729 showed the same percent of similarity (66%) with gene clusters encoding the antifungals Myxoprincomide, Bacillomycin, Iturin, and Amphotericin (B). None of these compounds have been isolated from this strain. However, Myxoprincomide was isolated from its closely related species *M. xanthus* DK1622 [320].

Therefore, this gene cluster has the potential to encode a novel SM. The T1PKS-NRPS gene cluster (5466423-5644506) showed a 100% similarity with the Phenalamide gene cluster, the gene cluster that has been previously identified in this strain [313]. *M. stipitatus* DSM 14675 also contained a Bacteriocin-T1PKS-NRPS (7721562-7824172) that showed low similarity with the Epothilone gene cluster and other gene clusters that encode antifungals and cytotoxic compounds. The degree of similarity of this gene clusters with known gene clusters may indicate it potential to encode a novel SM.

M. hansupus-strain-mixupus genome contained 6 T1PKS-NRPS gene clusters. Two of these gene clusters (3446609-3512502) and (3924291-4013404) had a 100% similarity with the gene clusters that encode Myxochromide and Myxoprincomide, respectively. These gene clusters have never been described in this strain. However, since the genome of its closely related species, *M. xanthus* DK1622, has these gene clusters [320], it most likely that these gene clusters are encoding Myxochromide and Myxoprincomide. The T1PKS-NRPS genes of the gene cluster (2771375-2902802) share similarities with their counterparts of Phenalamide, Myxalamide, Myxothiazol, Crocacin, Cystothiazole A, Melithiazol, and Epothilone gene clusters. These percentages of similarities of this gene cluster with the gene clusters that encode these cytotoxic compounds may indicate its potential to encode a novel SM. The T1PKS-NRPS genes of the gene cluster (3129770-3216154) showed a high similarity (76%) with the BGC of Dkxanthene that has been identified in *M. xanthus*-DK-1050. This gene cluster has never been identified in this strain. Therefore, it is expected that this gene cluster is encoding a Dkxanthene-like compound. The T1PKS-NRPS genes of the gene cluster (4135121-4237928) shared a

low similarity with Microsclerodermins and Myxalamide, indicating its potential to produce a novel SM.

The antiSMASH analysis showed that *M. xanthus*-DK-1622 genome has 6 T1PKS-NRPS gene clusters. Two of these gene clusters (4481281-4564105) and (4976865-5045896) were adjacent to each other and showed a 100% similarity with Myxoprincomide and Myxochromide gene clusters, respectively. These gene clusters have been previously identified in this strain [320]. The T1PKS-NRPS gene cluster (4200394-4368749) showed a 75% similarity with the Myxochelin gene cluster. Although Myxochelin was isolated from this strain, the gene cluster encoding this compound has not been identified yet. Therefore, this gene cluster has the potential for encoding a Myxochelin-similar compound. However, confirmation of this possibility needs wet lab experiments. The T1PKS-NRPS gene cluster (5235098-5311992) showed a 76 % similarity with the Dkxanthene gene cluster. Dkxanthene has been isolated from this strain. However, the Dkxanthene gene cluster has not been previously identified in this strain but it was identified in *M. xanthus*-DK-1050, the very closely related strain to *M. xanthus*-DK-1622. Therefore, this gene cluster is likely the gene cluster that is responsible for the biosynthesis of a Dkxanthene-similar compound. The T1PKS-NRPS gene cluster (5580883-5673548) shared relatively low similarities with many gene clusters that encode cytotoxic compounds and antimicrobials such as Myxalamid, Pellasoren, Chondrochloren, Stigmatellin, and Epothilone. Therefore, wet lab experiments are needed to check the novelty of the SM produced by this gene cluster.

The antiSMASH software identified a Terpene-Lantipeptide-T1PKS-NRPS gene cluster (1833443-1941377) and a T1PKS-NRPS gene cluster (2733381-2794934) in *N.*

exedens-strain-DSM-71-utg0 genome. The gene cluster (1833443-1941377) share close percent similarities (25-37%) with the gene clusters that encode Nostopeptolide, Microsclerodermins, Epothilone, Kedarcidin, Jagaricin, Tallysomycin, and Bleomycin. None of these compounds have been isolated from this strain. Moreover, none of the BGCs encoding these compounds have been identified in this strain's genome. The low similarities of this gene cluster to the BGCs that encode cytotoxic compounds indicate its potential to biosynthesize a novel SM. However, to identify the compounds that are encoded by this gene cluster, wet lab experiments are needed.

The gene cluster (2733381-2794934) shared low similarities with the gene clusters that encode antibiotics. Therefore, this gene cluster has the potential to encode a novel SM. However, wet experiments are required to check the novelty of the SM encoded by this gene cluster.

S. aurantiaca DW4/3-1 genome has 2 T1PKS-NRPS gene clusters and a T1PKS - Cf_Saccharide-NRPS gene cluster (gene cluster containing T1PKS, NRPS and separated by a putative saccharide cluster. This gene cluster (4751261-4894265), shared a 100%, 100% , 85%, and 71% similarity with gene clusters encodingMyxothiazol, Cystothiazole A, Melithiazol, and Epothilon, respectively. The big size of this gene cluster and its high percent similarities with more than one gene cluster may indicate the possibility that there is more than one gene cluster within this genomic segment. Interestingly, a reanalysis of this gene cluster by PRISM indicated the presence of two gene clusters. The T1PKS-NRPS gene cluster (6854-83186) shared a similarity with the Nostopeptolide gene cluster, a compound that controls cellular differentiation of cyanobacteria [262]. The T1PKS-NRPS gene cluster (5092487-5154581) shared a low similarity (11-17%) with

gene clusters that encode cytotoxic compounds. The T1PKS -Cf_Saccharide-NRPS gene cluster (4751261-4894265) showed high similarities with gene clusters of Myxothiazol, Cystothiazole, Melithiazol, Epothilone, Stigmatellin, and Myxalamid. To connect all these gene clusters to compounds they encode, and to check the novelty of the compound it can produce, wet lab experiments are needed.

Some of the EAFHs found in many ESGCs were used for *in silico* design of some ESGCs. The re-analysis of these new ESGCs by antiSMASH resulted in the generation of various scaffolds that were similar to the Epothilone gene cluster's scaffold. The prediction of a scaffold by antiSMASH is based on the assumption that the PKS/NRPS is collinear. Therefore, the real structure and novelty of the compounds encoded by the *in silico*-designed ESGCs should be assessed by heterologous expression experiments and other wet lab experiments.

It is worth mentioning that the identification and checking the novelty of the compounds encoded by the gene clusters that were identified in this study need wet lab experiments. This includes the gene clusters that show any degree of similarity but have not been described before or assessed for the products they can produce.

CHAPTER VI

LIMITATIONS AND FUTURE SCOPES

The use of microbial genome mining in silico discovery of gene clusters in bacteria proved to be valuable. Using this approach it was possible to identify several gene clusters that have never been linked to any of known secondary metabolites. Most of these gene clusters contain genes that are similar to those found in gene clusters that encode cytotoxic and antimicrobial compounds. Therefore, these gene clusters have the potential to encode compounds that have promising bioactivities. The uncertainty about what each of the newly discovered secondary metabolite gene clusters encodes and the assessment of novelty of these gene clusters are some limitations of this study. Therefore, wet lab experiments are needed to assess the novelty of the gene clusters discovered in this study and to link them to specific secondary metabolites. Heterologous expression of these gene clusters [321] allows for the synthesis of the products they encode and also allows for the identification and assessment of the biological activities of these products. Also, knock down and gene deletion experiments [98,170,322] allows for

comparison of mutants lacking these gene clusters with their wild type and therefore isolation and assessment of the products of the gene clusters.

Ninety-three bacterial genera were found to have the EAFHs in their genomes. These the EAFHs were used for *in silico* design of some ESGCs. Therefore, the huge number of the identified the EAFHs can form the basis for constructing a large library of ESGCs that can be used in another study to perform heterologous expression experiments.

Supplementary

Supplementary Table 1. Amino acid sequences of epoA-F genes of *S. cellulosum* So ce90 epothilone's gene cluster.

Gene	Length	Sequence
EPOA	1421	>AAF26919.1 polyketide synthase (Sorangium cellulosum) MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGS RDTVGRVPAERWDAAAWFDPDPDAPGKTPV TRASFLSDVACFDASFFGISPREALRMDPAHRLLLEVCW EALENAAIAPSALVGTETGVFIGIGPSEYEA ALPQATASAEIDAHGGLGTMPSVGAGRISYALGLRGPCV AVDTAYSSSLVAVHLACQSLRSGECSTALAG GVSLMLSPSTLVWLSKTRALARDGRCKAFSAEADGFGR GEGCAVVVLKRLSGARADGDRILAVIRGSAIN HDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVE AHGTGTTLGDPIEIQALNAVYGLGRDVATPLL IGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQ ALNPRISWGDLRLTVTRARTPWPDWNTPRRA GVSSFGMSGTNAHVVLEEAPAATCTPPAPERPAELLVLS ARTASALDAQAARLRDHLETYPSQCLGDVAF SLATTRSAMEHRLAVAATSREGLRAALDAAAQGQTSPG AVRSIADSSRGKLAFLFTGQGAQTLGMGRGLY DVWSAFREAFDLCVRLFNQELDRPLREVMWAEPASVDA ALLDQTAFTQPALFTFEYALAALWRSWGVEPE LVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQAL PAGGAMVSIEAPEADVAAAVAPHAASVSIAAV NAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHA FHSPLMAPMLEAFGRVAESVSYRRPSIVLVSNL SGKACTDEVSSPGYWVRHAREVVRFADGVKALHAAGA GTFVEVGPKSTLLGLVPACMPDARPALLASSRA GRDEPATVLEALGGLWAVGGLVSWAGLFPSGGRRVPLP TYPWQRERYWIDTKADDAARGDRRAPGAGHDE VEEGGAVRGGDRRSARLDHPPPESGRREKVEAAGDRPF RLEIDEPGVLDHLVLRVTERRAPGLGEVEIAV

		DAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGR IVAVGEGVNGLVVGQPVIALSAGAFATHVTTS AALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP GERVLIHAATGGVGLAAVQWAQHVGAEVHATA GTPEKRAYLESLGVRYVSDSRSDRFVADVRAWTGGEGV DVVLNSLSGELIDKSFNLLRSHGRFVELGKRD CYADNQLGLRPFLRNLSFSLVDLRGMMLERPARVRALL EELLGLIAAGVFTPPPIATLPIARVADAFRSM AQAQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLD RLASAAPAARAAALEAFLRTQVSQVLRTPEI KVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLS TSPNIALLAQNLLDALATALSLERVAAENLR AGVQNDFVSSGADQDWEIIAL
EPOP	1410	>AAF26925.1 nonribosomal peptide synthetase (Sorangium cellulosum) MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLA RISEHKSTILTMLRQRLPAESIVPAPAERHA PFPLTDIQESYWLGRTGAFTVPSGIHAYREYDCTDLDVP RLSRAFRKVVARHDMLRAHTLPDMMQVIEPK VDADIEIIDLRGLDRSTREARLVSLRDAMSHRIYDTERPP LYHVVAVRLDERQTRLVLSIDLINVDLGSL SIIFKDWLSFYEDPETSLPVLELSYRDYVLALESRKKSEA HQRSMDYWKRRIAELPPPPTLPMKADPSTL KEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAF SEVIGRWSASPRFTLNITLFNRLPVHPRVND ITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHC DVSGIEVQREAARVLGIQRGALFPVVLTSAL NQQVVGVTSLQRLGTPVYTSTQTPQLLLDHQLYEHDGD LVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLT EEPWGEQVRCSLPPAQLEARASANATNALLSEHTLHGLF AARVEQLPMQLAVVSARKTLTYEELSRRSRR LGARLREQGARPNTLVAVVMEKGWEQVVAVLAVLESG AAYVPIDADLPAERIHYLLDHGEVKLVLTQPWL DGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLA YVIYTSGSTGLPKGVMIDHRGAVNTILDINE RFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA SKLRDPAHWAELIEREKVTVWNSVPALMRM LVEHFEGRPDSLARSLRLSLLSGDWIPVGLPGELQAIRPG VSVISLGGATEASIWSIGYPVRNVDLSWAS IPYGRPLRNQTFHVLDEALEPRPVWVPGQLYIGGVGLAL GYWRDEEKTRKSFLVHPETGERLYKTGDLGR YLPDGNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNV RDAVIVPVGNDAANKLLLAYVVPEGTRRRAA EQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGL

		RRDLDGKPVVDLTGQDPREAGLDVYARRRSVR TFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAGSTY PVQTYAYVKSGRIEGVDEGFYYYHPFEHR LLKLSDHGIERGAHVRQNFDVFDEAAFNLLFVGRIDAIES LYGSSSREFCLLEAGYMAQLLMEQAPSCNI GVCPVGQFNFEQVRPVLDLRHSDVYVHGMLGGRVDPR QFQVCTLGQDSSPRRATTRGAPPGREQHFADML RDFLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRER KDTSSPRHSGHTAPRDALEEILVAVVREVLG LEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITEL FQYPNLGSLASGLRRDSRDLDQRPNMQDRV EVRRKGRRRS
EPOB	1832	>AAF26920.1 polyketide synthase (Sorangium cellulosum) MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAV QRFSEQELAASGVDPALVLDPSYVRAGSVLED VDRFDAAFFGISPREAELMDPQHRIFMECAWEALENAG YDPTAYEGSIGVYAGANMSSYLTSNLHEHPAM MRWPGWFQTLIGNDKDYLATHVSYRLNLRGPSISVQTA CSTSLVAVHLACMSLLDRECDMALAGGITVRI PHRAGYVYAEGGIFSPDGHCRAFDAKANGTIMGNGCGV VLLKPLDRALSDGDPVRAVILGSATNNDGARK IGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTL LGDAIETAALRRVFGRDASARRSCAIGSVK TGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSID FASSPFYVNTSLKDWNTGSTPRRAGVSSF GIGGTNAHVVLEEAPAAKLPAAAPARSAELFVVSAKSA AALDAAAARLRDHLQAHQGISLGDVAFSLATT RSPMEHRLAMAAPSREALREGLDAAARGQTPPGAVRGR CSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEP VFHAALSACDRAIQAEAGWSLLAELAADEGSSQLERIDV VQPVLFALAVAFAALWRSWGVAPDVVIGHSM GEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAV TELSLAEAEAALRGYEDRVSVAVSNSPRSTV LSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPL REDLLAALGGLRPGAAAVPMRSTVTGAMVAG PELGANYWMNNLRQPVRFAEVVQAQLQGGHGLFVEMS PHPILTTSVEEMRRAAQRAGAAVGSLRRGQDER PAMLEALGTLWAQGYPVPWGRLFPAGGRRVPLPTYPW QRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQ TLSTQTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAA YLEMAISSGAEALGDGPLQITDVVLAEALAFA GDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVH ARGALLRVERTEVPAGLTLSAVRARLQASIPA AATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP

		DAAGSAAEYRLHPALLDACFQIVGSLFARSGE ATPWVPVELGSLRLLQRPSGELWCHARVVNHGHQTPDR QGADFWVVDSSGAVVAEVCGLVAQRLPGGVRR REEDDWFLELEWEPAAVGTAKVNAGRWLLLGGGGGLG AALRAMLEAGGHAVVHAAENNTSAAGVRALLAK AFDGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSA DVSPDALDPALVRGCDSVLWTVQALAGMGFRD APRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEH ADLRCARVDLDPARPEGELAALLAELLADDAEA EVALRGGERCVARIVRRQPETRPRGRIESCVPTDVTIRAD STYLVTGGLGGLGLSVAGWLAERGAGHLVL VGRSGAASVEQRAAVAALEARGARVTVAKADVADRAQ LERILREVTTSGMPLRGVVHAAGILDDGLLMQQ TPARFRKVMAPKVQGALHLHALTREAPLSFFVLYASGV GLLGSPGQGNYAAANTFLDALAHHRRAQGLPA LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEG LSALARLLESGRAQVGVMPVNPRLWVELYPAAA SSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARS ALLEPLLRAQISQVLRLPEGKIEVDAPLTSLG MNSLMGLELRNRIEAMLGITVPATLLWTYPTVAALSGH LAREACEAAPVESPHTTADSAVEIEEMSQDDL TQLIAAKFKALT
EPOC	7257	>AAF26921.1 polyketide synthase (Sorangium cellulosum) MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEP IAIVGIGCRFPGGADAPEAFWELLDAERDA VQPLDMRWALVGVAPVEAVPHWAGLLTEPIDCFDAAFF GISPREARSLDPQHRLLLEVAWEGLEDAGIPP RSIDGSRTGVFVGAFTADYARTVARLPREERDAYSATGN MLSIAAGRLSYTLGLQGPCLTVDTACSSSLV AIHLACRSLRAGESDLALAGGVSALLSPDMMEAAARTQ ALSPDGRCRTFDASANGFVRGEGCGLVVLKRL SDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETV LREALRSAHVEAGAVDYVETHGTGTSLGDPI EVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVA GLIKAALSLTHERIPRNLNFRTLNPRIRLEGS ALALATEPVPWPRTDRPRFAGVSSFGMSGTNAHVVLEE APAVELWPAAPERSAELLVLSGKSEGALDAQA ARLREHLDMHPELGLGDVAFSLATTRSAMSHRLAVAVT SREGLLAALSAVAQGQTPAGAARCIASSSRGK LAFLFTGQGAQTPGMGRGLCAAWPAFREAFDRCVALFD RELDRPLREVMWAEAGSAESLLLDQTAFTQPA LFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGV FSLEDGVRLVAARGRLMQGLSAGGAMVSLGAP EAEVAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAI

		AAGFAARGARTKRLHVSHAFHSPLMEPMLEEF GRVAASVTYRRPSVSLVSNLSGKVVTDELSAPGYWVRH VREAVRFADGVKALHEAGAGTFVEVGPKPTLL GLLPACLPEAEPTLLASLRAGREEAAGVLEALGRLWAA GGSVSWPGVFPTAGRRVPLPTYPWQRQRYWIE APAEGLGATAADALAQWFYRVDWPEMPRSSVDSRRAR SGGWLVLADRGGVGEAAAAALSSQGCSCAVLHA PAEASAVAEQVTQALGGRNDWQGVLYLWGLDAVVEA GASAEEVAKVTHLAAAPVLALIQALGTGPRSPRL WIVTRGACTVGGEPDAAPCQAALWGMGRVAALEHPGS WGGLVDLDPEESPTEVEALVAELLSPDAEDQLA FRQGRRRAARLVAAPPEGNAAPVSLSAEGSYLVTGGLG ALGLLVARWLVERGAGHLVLISRHGLPDREEW GRDQPPEVRARIAAIEALEAQGARVTVAAVDVADAEGM AALLAAVEPPLRGVVHAAGLLDDGLLAHQDAG RLARVLRPKVEGAWVLHTLTREQPLDLFVLFSSASGVFG SIGQGSYAAGNAFLDALADLRRTQGLAALSI AWGLWAEGGMGSQAQRREHEASGIWAMPTSRALAAM EWLLGTRATQRVVIQMDWAHAGAAPRDASRGRFW DRLVTATKEASSSAVPAVERWRNASVVETRSALYELVR GVVAGVMGFTDQGTLDVRRGFAEQGLDSLMAV EIRKRLQGELGMPLSATLAFDHPTVERLVEYLLSQALEL QDRTDVRSVRLPATEDPIAIVGAACRFPGGV EDLESYWQLLTEGVVVSTEVPADRWNGADGRVPGSGE AQRQTYVPRGGFLREVETFDAAFFHISPREAMS LDPQQRLLLEVSWEAIERAGQDPSALRESPTGVFVGAGP NEYAERVQELADEAAGLYSGTGNMLSVAAGR LSFFLGLHGPTLAVDTACSSSLVALHLGCQSLRRGECDQ ALVGGVNMLLSPKTFALLSRMHALSPGGRCK TFSADADGYARAEGCAVVVLKRLSDAQRDRDPILAVIR GTAINHDGPSSGLTVPSGPAQEALLRQALAHA GVVPADVDFVECHGTGTALGDPIEVRALSDVYGQARPA DRPLILGAAKANLGHMEPAAGLAGLLKAVLAL GQEQIPAQPELGELNPLLPWEALPVAVARAAVPWPRTD RPRFAGVSSFGMSGTNAHVVLEEAPAVELWPA APERSAELLVLSGKSEGALDAQAARLREHLDMHPELGL GDVAFSLATTRSAMNHRLAVAVTSREGLLAAL SAVAQGQTPPGAARCIASSSRGKLAFLFTGQGAQTPGM GRGLCAAWPAFREAFDRCVALFDRELDRPLRE VMWAEPGSAESLLLDQTAFTQPALFTVEYALTALWRSW GVEPELVAGHSAGELVAACVAGVFSLEDGVRL VAARGRLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAS VSIAAVNGPEQVVIAGVEQAVQAIAAGFAARGA RTKRLHVSHASHSPLMEPMLEEFGRVAASVTYRRPSVSL VSNLSGKVVADELSAPGYWVRHVREAVRFAD GVKALHEAGAGTFVEVGPKPTLLGLLPACLPEAEPTLLA

		SLRAGREEAAGVLEALGRLWAAGGSVSWPGV FPTAGRRVPLPTYPWQRQRYWPDIEPDSRRHAAADPTQ GWFYRVDWPEIPRSLQKSEEASRGSWLVLADK GGVGEAVAAALSTRGLPCVVLHAPAETSATAELVTEAA GGRSDWQVVLYLWGLDAVVGAEASIDEIGDAT RRATAPVLGLARFLSTVSCSPRLWVVTRGACIVGDEPAI APCQAALWGMGRVAALEHPGAWGGLVDLDPR ASPPQASPIDGEMLVTELLSQETEDQLAFRHGRRHAARL VAAPPQGQAAPVSLSAEASYLVTGGLGGLGL IVAQWLVELGARHLVLTSRRGLPDRQAWCEQQPPEIRA RIAAVEALEARGARVTVAAVDVADVEPMTALV SSVEPPLRGVVHAAGVSVMRPLAETDETLLESVLRPKVA GSWLLHRLLHGRPLDLFVLFSSGAAVWGSHS QGAYAAANAFLDGLAHLRRSQSLPALSVAWGLWAEGG MADAEAHARLSDIGVLPMSTSAALSALQRLVET GAAQRTVTRMDWARFAPVYTARGRRNLLSALVAGRDII APSPPAAATRNWRGLSVAEARVALHEIVHGAV ARVLGFLDPSALDPGMGFNEQGLDSLMAVEIRNLLQAE LDVRLSTTLAFDHPTVQRLVEHLLVDVLKLED RSDTQHVRSLASDEPIAIVGAACRFPGGVEDLESYWQLL AEGVVVSAEVPADRWDAADWYDPDPEIPGRT YVTKGAFLRDLQRLDATFFRISPREAMSLDPQQRLLLEV SWEALESAGIAPDTLRDSPTGVFVGAGPNEY YTQRLRGFTDGAAGLYGGTGNMLSVTAGRLSFFLGLHG PTLAMDTACSSSLVALHLACQSLRLGECDQAL VGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGY ARGEGCAVVVLKRLRDAQRAGDSILALIRGSA VNHDGPSSGLTVPNGPAQQALLRQALSQAGVSPVDVDF VECHGTGTALGDPIEVQALSEVYGPGRSGDRP LVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPAQP ELGELNPHLPWNTLPVAVPRKAVPWGRGARPR RAGVSAFGLSGTNVHVVLEEAPEVEPAPAAPARPVELVV LSAKSAAALDAAAARLSAHLSAHPELSLGDV AFSLATTRSPMEHRLAIATTSREALRGALDAAAQQKTPQ GAVRGKAVSSRGKLAFLFTGQGAQMPGMGRG LYETWPAFREAFDRCVALFDREIDQPLREVMWAAPGLA QAARLDQTAYAQPALFALEYALAALWRSWGVE PHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQ ALPAGGAMVAIAASEAEVAASVAPHAATVSIA AVNGPDAVVIAGAEVQVLALGATFAARGIRTKRLAVSH AFHSPLMDPMLEDFQRVAATIAYRAPDRPVVS NVTGHVAGPEIATPEYWVRHVRSAVRFGDGAKALHAA GAATFVEVGPKPVLLGLLPACLGEADAVLVPSL RADRSECEVVLAALGAWYAWGGALDWKGVFPDGARR VALPMYPWQRERHWMDLTPRSAAPAGIAGRWPLA GVGLCMPGAVLHHVLSIGPRHQPFLGDHLVFGKVVVPG

		AFHVAVILSIAAERWPERAIELTGVEFLKAIA MEPDQEVELHAVLTPEAAGDGYLFELATLAAPETERRW TTHARGRVQPTDGAPGALPRLEVLEDRAIQPL DFAGFLDRLSAVRIGWGPLWRWLQDGRVGDEASLATL VPTYPNAHDVAPLHPILLDNGFAVSLLSTRSEP EDDGTPPLPFAVERVRWWRAPVGRVRCGGVPRSQAFG VSSFVLVDETGEVVAEVEGFVCRRAPREVFLRQ ESGASTAALYRLDWPEAPLPDAPAERIEESWVVVAAPGS EMAAALATRLNRCVLAEPKGLEAALAGVSPA GVICLWEAGAHEEAPAAAQRVATEGLSVVQALRDRAV RLWWVTMGAVAVEAGERVQVATAPVWGLGRTVM QERPELSCTLVDLEPEADAARSADVLLRELGRADDETQV AFRSGKRRVARLVKATTPEGLLVPDAESYRL EAGQKGTLDQLRLAPAQRRAPGPGEVEIKVTASGLNFRT VLAVLGMYPGDAGPMGGDCAGVATAVGQGVR HVAVGDAVMTLGTLHRFVTVDARLVVRQPAGLTPAQA ATVPVAFLTAWLALHDLGNLRRGERVLIHAAAG GVGMAAVQIARWIGAEVFATASPSKWAAVQAMGVPRT HIASSRTLEFAETFRQVTGGRGVDVVLNALAGE FVDASLSLLSTGGRFLEMGKTDIRDRAAVAAAHPGVRY RVFDILELAPDRTREILERVVEGFAAGHLRAL PVHAFAITKAEAAFRFMAQARHQGKVVLLPAPSAAPLA PTGTVLLTGGLGALGLHVARWLAQQGVPHMVL TGRRGLDTPGAAKAVAEIEALGARVTIAASDVADRNAL EAVLQAIPAEWPLQGVIHAAGALDDGVLDEQT TDRFSRVLAPKVTGAWNLHELTAGNDLAFFVLFSSMSG LLGSAGQSNYAAANTFLDALAAHRRAEGLAAQ SLAWGPWSDGGMAAGLSAALQARLARHGMGALSPAQ GTALLGQALARPETQLGAMSLDVRAASQASGAAV PPVWRALVRAEARHTAAGAQGALAARLGALPEARRAD EVRKVVQAEIARVLSWSAASAVPVDRPLSDLGL DSLTAVELRNVLGQRVGATLPATLAFDHPTVDALTRWL LDKVLAVAEPSVSSAKSSPQVALDEPIAIIGI GCRFPGGVADPESFWRLLEEGSDAVVEVPHERWDIDAF YDPDPDVRGKMTTRFGGFLSDIDRFDPAFFGI SPREATTMDPQQRLLLETSWEAFERAGILPERLMGSDTG VFVGLFYQEYAALAGGIEAFDGYLGTGTTAS VASGRISYVLGLKGPSLTVDTACSSSLVAVHLACQALRR GECSVALAGGVALMLTPATFVEFSRLRGLAP DGRCKSFSAAADGVGWSEGCAMLLLKPLRDAQRDGDPI LAVIRGTAVNQDGRSNGLTAPNGSSQQEVIRR ALEQAGLAPADVSYVECHGTGTTLGDPIEVQALGAVLA QGRPSDRPLVIGSVKSNIGHTQAAAGVAGVIK VALALERGLIPRSLHFDAPNPHIPWSELAVQVAAKPVEW TRNGVPRRAGVSSFGVSGTNAHVVLEEAPAA AFAPAAARSAELFVLSAKSAAALDAQAARLSAHVVAHP

		ELGLGDLAFSLATTRSPMTYRLAVAATSREAL SAALDTAAQGQAPPAAARGHASTGSAPKVVFVFPGQGS QWLGMGQKLLSEEPVFRDALSACDRAIQAEAG WSLLAELAADETTSQLGRIDVVQPALFAIEVALSALWRS WGVEPDAVVGHSMGEVAAAHVAGALSLEDAV AIICRRSLLLRRISGQGEMAVVELSLAEAEAALLGYEDRL SVAVSNSPRSTVLAGEPAALAEVLAILAAK GVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATV SMRSTVTSTIMAGPELVASYWADNVRQPVRF AEAVQSLMEDGHGLFVEMSPHPILTTSVEEIRRATKREG VAVGSLRRGQDERLSMLEALGALWVHGQAVG WERLFSAGGAGLRRVPLPTYPWQRERYWVDAPTGGAA GGSRFAHAGSHPLLGEMQTLSTQRSTRVWETTL DLKRLPWLGDHRVQGAVVFPGAAYLEMALSSGAEALG DGPLQVSDVVLAEALAFADDTPAAVQVMATEER PGRLQFHVASRVPGHGGAAFRSHARGVLRQIERAEVPA RLDLAALRARLQASAPAAATYAALAEMGLEYG PAFQGLVELWRGEGEALGRVRLPEAAGSPAACRLHPAL LDACFHVSSAFADRGEATPWVPVEIGSLRWFQ RPSGELWCHARSVSHGKPTPDRRSTDFWVVDSTGAIVA EISGLVAQRLAGGVRRREEDDWFMEPAWEPTA VPGSEVMAGRWLLIGSGGGLGAALHSALTEAGHSVVHA TGRGTSAAGLQALLTASFDGQAPTSVVHLGSL DERGVLDADAPFDADALEESLVRGCDSVLWTVQAVAG AGFRDPPRLWLVTRGAQAIGAGDVSVAQAPLLG LGRVIALEHAELRCARIDLDPARRDGEVDELLAELLADD AEEEVAFRGGERRVARLVRRLPETDCREKIE PAEGRPFRLEIDGSGVLDDLVLRATERRPPGPGEVEIAVE AAGLNFLDVMRAMGIYPGPGDGPVALGAEC SGRIVAMGEGVESLRIGQDVVAVAPFSFGTHVTIDARML APRPAALTAAQAAALPVAFMTAWYGLVHLGR LRAGERVLIHSATGGTGLAAVQIARHLGAEIFATAGTPE KRAWLREQGIAHVMDSRSLDFAEQVLAATKG EGVDVVLNSLSGAAIDASLSTLVPDGRFIELGKTDIYADR SLGLAHFRKSLSYSAVDLAGLAVRRPERVA ALLAEVVDLLARGALQPLPVEIFPLSRAADAFRKMAQA QHLGKLVLALEDPDVRIRVPGESGVAIRADGA YLVTGGLGGLGLSVAGWLAEQGAGHLVLVGRSGAVSA EQQTAVAALEAHGARVTVARADVADRAQMERIL REVTASGMPLRGVVHAAGILDDGLLMQQTPARFRAVM APKVRGALHLHALTREAPLSFFVLYASGAGLLG SPGQGNYAAANTFLDALAHHRRAQGLPALSIDWGLFAD VGLAAGQQNRGARLVTRGTRSLTPDEGLWALE RLLDGDRTQAGVMPFDVRQWVEFYPAAASSRRLSRLM TARRVASGRLAGDRDLLERLATAEAGARAGMLQ EVVRAQVSQVLRLSEGKLDVDAPLTSLGMDSLMGLELR

		NRIEAVLGITMPATLLWTYPTVAALSAHLASH VVSTGDGESARPPDTGSVAPTTHEVASLDEDGLFALIDES LARAGKR
EPOD	3798	>AAF26922.1 polyketide synthase (Sorangium cellulosum) MTDREGQLLERLREVTLALRKTLNERDTLELEKTEPIAIV GIGCRFPGGAGTPEAFWELLDDGRDAIRPL EERWALVGVDPGDDVPRWAGLLTEAIDGFDAAFFGIAP REARSLDPQHRLLLEVAWEGFEDAGIPPRSLV GSRTGVFVGVCATEYLHAAVAHQPREERDAYSTTGNML SIAAGRLSYTLGLQGPCLTVDTACSSSLVAIH LACRSLRARESDLALAGGVNMLLSPDTMRALARTQALS PNGRCQTFDASANGFVRGEGCGLIVLKRLSDA RRDGDRIWALIRGSAINQDGRSTGLTAPNVLAQGALLRE ALRNAGVEAEAIGYIETHGAATSLGDPIEIE ALRAVVGPARADGARCVLGAVKTNLGHLEGAAGVAGL IKATLSLHHERIPRNLNFRTLNPRIRIEGTALA LATEPVPWPRTGRTRFAGVSSFGMSGTNAHVVLEEAPA VEPEAAAPERAAELFVLSAKSAAALDAQAARL RDHLEKHVELGLGDVAFSLATTRSAMEHRLAVAASSRE ALRGALSAAAQGHTPPGAVRGRASGGSAPKVV FVFPGQGSQWVGMGRKLMAEEPVFRAALEGCDRAIEAE AGWSLLGELSADEAASQLGRIDVVQPVLFAME VALSALWRSWGVEPEAVVGHSMGEVAAAHVAGALSLE DAVAIICRRSRLLRRISGQGEMALVELSLEEAE AALRGHEGRLSVAVSNSPRSTVLAGEPAALSEVLAALTA KGVFWRQVKVDVASHSPQVDPLREELIAALG AIRPRAAAVPMRSTVTGGVIAGPELGASYWADNLRQPV RFAAAAQALLEGGPALFIEMSPHPILVPPLDE IQTAAEQGGAAVGSLRRGQDERATLLEALGTLWASGYP VSWARLFPAGGRRVPLPTYPWQHERCWIEVEP DARRLAAADPTKDWFYRTDWPEVPRAAPKSETAHGSW LLLADRGGVGEAVAAALSTRGLSCTVLHASADA STVAEQVSEAASRRNDWQGVLYLWGLDAVVDAGASAD EVSEATRRATAPVLGLVRFLSAAPHPPRFWVVT RGACTVGGEPEASLCQAALWGLARVAALEHPAAWGGL VDLDPQKSPTEIEPLVAELLSPDAEDQLAFRSG RRHAARLVAAPPEGDVAPISLSAEGSYLVTGGLGGLGLL VARWLVERGARHLVLTSRHGLPERQASGGEQ PPEARARIAAVEGLEAQGARVTVAAVDVAEADPMTALL AAIEPPLRGVVHAAGVFPVRHLAETDEALLES VLRPKVAGSWLLHRLLRDRPLDLFVLFSSGAAVWGGKG QGAYAAANAFLDGLAHHRRAHSLPALSLAWGL WAEGGMVDAKAHARLSDIGVLPMATGPALSALERLVN

		TSAVQRSVTRMDWARFAPVYAARGRRNLLSALV AEDERAASPPVPTANRIWRGLSVAESRSALYELVRGIVA RVLGFSDPGALDVGRGFAEQGLDSLMALEIR NRLQRELGERLSATLAFDHPTVERLVAHLLTDVLKLEDR SDTRHIRSVAADDDIAIVGAACRFPGGDEGL ETYWRHLAEGMVVSTEVPADRWRAADWYDPDPEVPGR TYVAKGAFLRDVRSLDAAFFAISPREAMSLDPQ QRLLLEVSWEAIERAGQDPMALRESATGVFVGMIGSEH AERVQGLDDDAALLYGTTGNLLSVAAGRLSFF LGLHGPTMTVDTACSSSLVALHLACQSLRLGECDQALA GGSSVLLSPRSFVAASRMRLLSPDGRCKTFSA AADGFARAEGCAVVVLKRLRDAQRDRDPILAVVRSTAI NHDGPSSGLTVPSGPAQQALLRQALAQAGVAP AEVDFVECHGTGTALGDPIEVQALGAVYGRGRPAERPL WLGAVKANLGHLEAAAGLAGVLKVLLALEHEQ IPAQPELDELNPHIPWAELPVAVVRRAVPWPRGARPRRA GVSAFGLSGTNAHVVLEEAPAVEPVAAAPER AAELFVLSAKSAAALDAQAARLRDHLEKHVELGLGDVA FSLATTRSAMEHRLAVAASSREALRGALSAAA QGHTPPGAVRGRASGGSAPKVVFVFPGQGSQWVGMGR KLMAEEPVFRAALEGCDRAIEAEAGWSLLGELS ADEAASQLGRIDVVQPVLFAMEVALSALWRSWGVEPEA VVGHSMGEVAAAHVAGALSLEDAVAIICRRSR LLRRISGQGEMALVELSLEEAEAALRGHEGRLSVAVSNS PRSTVLAGEPAALSEVLAALTAKGVFWRQVK VDVASHSPQVDPLREELIAALGAIRPRAAAVPMRSTVTG GVIAGPELGASYWADNLRQPVRFAAAAQALL EGGPALFIEMSPHPILVPPLDEIQTAAEQGGAAVGSLRRG QDERATLLEALGTLWASGYPVSWARLFPAG GRRVPLPTYPWQHERYWIEDSVHGSKPSLRLRQLRNGA TDHPLLGAPLLVSARPGAHLWEQALSDERLSY LSEHRVHGEAVLPSAAYVEMALAAGVDLYGTATLVLEQ LALERALAVPSEGGRIVQVALSEEGPGRASFQ VSSREEAGRSWVRHATGHVCSGQSSAVGALKEAPWEIQ RRCPSVLSSEALYPLLNEHALDYGPCFQGVEQ VWLGTGEVLGRVRLPGDMASSSGAYRIHPALLDACFQV LTALLTTPESIEIRRRLTDLHEPDLPRSRAPV NQAVSDTWLWDAALDGGRRQSASVPVDLVLGSFHAKW EVMERLAQAYIIGTLRIWNVFCAAGERHTIDEL LVRLQISVVYRKVIKRWMEHLVAIGILVGDGEHFVSSQP LPEPDLAAVLEEAGRVFADLPVLFEWCKFAG ERLADVLTGKTLALEILFPGGSFDMAERIYRDSPIARYSN GIVRGVVESAARVVAPSGMFSILEIGAGTG ATTAAVLPVLLPDRTEYHFTDVSPLFLARAEQRFRDYPF LKYGILDVDQEPAGQGYAHQRFDVIVAANVI HATRDIRATAKRLLSLLAPGGLLVLVEGTGHPIWFDITTG

		LIEGWQKYEDDLRIDHPLLPARTWCDVLRR VGFADAVSLPGDGSPAGILGQHVILSRAPGIAGAACDSS GESATESPAARAVRQEWADGSADVVHRMALE RMYFHRRPGRQVWVHGRLRTGGGAFTKALAGDLLLFE DTGQVVAEVQGLRLPQLEASAFAPRDPREEWLY ALEWQRKDPIPEAPAAASSSSAGAWLVLMDQGGTGAAL VSLLEGRGEACVRVIAGTAYACLAPGLYQVDP AQPDGFHTLLRDAFGEDRICRAVVHMWSLDATAAGERA TAESLQADQLLGSLSALSLVQALVRRRWRNMP RLWLLTRAVHAVGAEDAAASVAQAPVWGLGRTLALEH PELRCTLVDVNPAPSPEDAAALAVELGASDRED QVALRSDGRYVARLVRSSFSGKPATDCGIRADGSYVITD GMGRVGLSVAQWMVMQGARHVVLVDRGGASE ASRDALRSMAEAGAEVQIVEADVARRDDVARLLSKIEPS MPPLRGIVYVDGTFQGDSSMLELDARRFKEW MYPKVLGAWNLHALTRDRSLDFFVLYSSGTSLLGLPGQ GSRAAGDAFLDAIAHHRCKVGLTAMSINWGLL SEASSPATPNDGGARLEYRGMEGLTLEQGAAALGRLLA RPRAQVGVMRLNLRQWLEFYPNAARLALWAEL LKERDRADRGASNASNLREALQSARPEDRQLILEKHLSE LLGRGLRLPPERIERHVPFSNLGMDSLIGLE LRNRIEAALGITVPATLLWTYPNVAALSGSLLDILFPNAG ATHAPATEREKSFENDAADLEALRGMTDEQ KDALLAEKLAQLAQIVGE
EPOE	2439	>AAF26923.1 polyketide synthase (Sorangium cellulosum) MATTNAGKLEHALLLMDKLAKKNASLEQERTEPIAIVGI GCRFPGGADTPEAFWELLDSGRDAVQPLDRR WALVGVHPSEEVPRWAGLLTEAVDGFDAAFFGTSPREA RSLDPQQRLLLEVTWEGLEDAGIAPQSLDGSR TGVFLGACSSDYSHTVAQQRREEQDAYDITGNTLSVAA GRLSYTLGLQGPCLTVDTACSSSLVAIHLACR SLRARESDLALAGGVNMLLSSKTMIMLGRIQALSPDGHC RTFDASANGFVRGEGCGMVVLKRLSDAQRHG DRIWALIRGSAMNQDGRSTGLMAPNVLAQEALLREALQ SARVDAGAIGYVETHGTGTSLGDPIEVEALRA VLGPARADGSRCVLGAVKTNLGHLEGAAGVAGLIKAAL ALHHELIPRNLHFHTLNPRIRIEGTALALATE PVPWPRAGRPRFAGVSAFGLSGTNVHVVLEEAPATVLA PATPGRSAELLVLSAKSAAALDAQAARLSAHI AAYPEQGLGDVAFSLVSTRSPMEHRLAVAATSREALRS ALEVAAQGQTPAGAARGRAASSPGKLAFLFAG QGAQVPGMGRGLWEAWPAFRETFDRCVTLFDRELHQP LCEVMWAEPGSSRSSLLDQTAFTQPALFALEYA LAALFRSWGVEPELVAGHSLGELVAACVAGVFSLEDAV

		RLVVARGRLMQALPAGGAMVSIAAPEADVAAA VAPHAALVSIAAVNGPEQVVIAGAEKFVQQIAAAFAAR GARTKPLHVSHAFHSPLMDPMLEAFRRVTESV TYRRPSIALVSNLSGKPCTDEVSAPGYWVRHAREAVRFA DGVKALHAAGAGLFVEVGPKPTLLGLVPACL PDARPVLLPASRAGRDEAASALEALGGFWVVGGSVTWS GVFPSGGRRVPLPTYPWQRERYWIEAPVDREA DGTGRARAGGHPLLGEVFSVSTHAGLRLWETTLDRKRL PWLGEHRAQGEVVFPGAGYLEMALSSGAEILG DGPIQVTDVVLIETLTFAGDTAVPVQVVTTEERPGRLRF QVASREPGERRAPFRIHARGVLRRIGRVETP ARSNLAALRARLHAAVPAAAIYGALAEMGLQYGPALRG LAELWRGEGEALGRVRLPEAAGSATAYQLHPV LLDACVQMIVGAFADRDEATPWAPVEVGSVRLFQRSPG ELWCHARVVSDGQQASSRWSADFELMDGTGAV VAEISRLVVERLASGVRRRDADDWFLELDWEPAALGGP KITAGRWLLLGEGGGLGRSLCSALKAAGHVVV HAAGDDTSTAGMRALLANAFDGQAPTAVVHLSSLDGG GQLGPGLGAQGALDAPRSPDVDADALESALMRG CDSVLSLVQALVGMDLRNAPRLWLLTRGAQAAAAGDV SVVQAPLLGLGRTIALEHAELRCISVDLDPAEP EGEADALLAELLADDAEEEVALRGGDRLVARLVHRLPD AQRREKVEPAGDRPFRLEIDEPGALDQLVLRA TGRRAPGPGEVEISVEAAGLDSIDIQLALGVAPNDLPGEE IEPLVLGSECAGRIVAVGEGVNGLVVGQPV IALAAGVFATHVTTSATLVLPRPLGLSATEAAAMPLAYL TAWYALDKVAHLQAGERVLIHAEAGGVGLCA VRWAQRVGAEVYATADTPENRAYLESLGVRYVSDSRS GRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLM VLRACGRLVKLGRRDDCADTQPGLPPLLRNFSFSQVDLR GMMLDQPARIRALLDELFGLVAAGAISPLGS GLRVGGSLTPPPVETFPISRAAEAFRRMAQGQHLGKLVL TLDDPEVRIRAPAESSVAVRADGTYLVTGGL GGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAAVA ALEAHGARVTVAKADVADRSQIERVLREVTASG MPLRGVVHAAGLVDDGLLMQQTPARFRTVMGPKVQGA LHLHTLTREAPLSFFVLYASAAGLFGSPGQGNY AAANAFLDALSHHRRAQGLPALSIDWGMFTEVGMAVA QENRGARQISRGMRGITPDEGLSALARLLEGDR VQTGVIPITPRQWVEFYPATAASRRLSRLVTTQRAVADR TAGDRDLLEQLASAEPSARAGLLQDVVRVQV SHVLRLPEDKIEVDAPLSSMGMDSLMSLELRNRIEAALG VAAPAALGWTYPTVAAITRWLLDDALVVRLG GGSDTDESTASAGSFVHVLRFRPVVKPRARLFCFHGSGG SPEGFRSWSEKSEWSDLEIVAMWHDRSLASE DAPGKKYVQEAASLIQHYADAPFALVGFSLGVRFVMGT

		AVELASRSGAPAPLAVFTLGGSLISSSEITPE METDIIAKLFFRNAAGFVRSTQQVQADARADKVITDTM VAPAPGDSKEPPVKIAVPIVAIAGSDDVIVPP SDVQDLQSRTTERFYMHLLPGDHEFLVDRGREIMHIVDS HLNPLLAARTTSSGPAFEAK
EPOF	(419)	>AAF26924.1 cytochrome P450 (Sorangium cellulosum) MTQEQANQSETKPAFDFKPFAPGYAEDPFPAIERLREAT PIFYWDEGRSWVLTRYHDVSAVFRDERFAVS REEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKL VNPSFTSRAIDLLRAEIQRTVDQLLDARSGQ EEFDVVRDYAEGIPMRAISALLKVPAECDEKFRRFGSAT ARALGVGLVPQVDEETKTLVASVTEGLALLH DVLDERRRNPLENDVLTMLLQAEADGSRLSTKELVALV GAIIAAGTDTTIYLIAFAVLNLLRSPEALELV KAEPGLMRNALDEVLRFDNILRIGTVRFARQDLEYCGAS IKKGEMVFLLIPSALRDGTVFSRPDVFDVRR DTGASLAYGRGPHVCPGVSLARLEAEIAVGTIFRRFPEM KLKETPVFGYHPAFRNIESLNVILKPSKAG

Supplementary Table 2. Bacterial genera and their species and strains found to contain 1 or 2 the EAFHs.

Genus	the EAFHs	Protein encoded by ECSGs (Species or strain)	Identity	Accession
Acaryochloris				
	P	non-ribosomal peptide synthetase (Acaryochloris marina)	46%	WP_012166728.1
Actinoalloteichus				
	C	hypothetical protein (Actinoalloteichus hymeniacidonis)	41%	WP_069850034.1
	C	type I polyketide synthase (Actinoalloteichus cyanogriseus)	43%	WP_051314059.1

	D			
	D	hypothetical protein (Actinoalloteichus sp. ADI127-7)	44%	WP_083683266.1
	D	hypothetical protein (Actinoalloteichus sp. GBA129-24)	44%	WP_083691035.1
Actinokineospora				
	C	hypothetical protein (Actinokineospora bangkokensis)	43%	WP_084793822.1
Actinopolyspora				
	D	hypothetical protein (Actinopolyspora erythraea)	41%	WP_094904606.1
Actinosynnema				
	B	hypothetical protein (Actinosynnema sp. ALI-1.44)	44%	WP_076989535.1
	D	hypothetical protein (Actinosynnema sp. ALI-1.44)	44%	WP_076986437.1
Algicola				
	P	non-ribosomal peptide synthetase (Algicola sagamiensis)	41%	WP_018694106.1
Allokutzneria				
	C	polyketide synthase (Allokutzneria sp. NRRL B-24872)	43%	WP_086824400.1
Amycolatopsis				

	C	Acyl transferase domain-containing protein (Amycolatopsis tolypomycina)	44%	SED10139.1
	C	hypothetical protein (Amycolatopsis tolypomycina)	44%	WP_091314315.1
	C	hypothetical protein DV20_13005 (Amycolatopsis rifamycinica)	43%	KDN21843.1
	C	type I polyketide synthase (Amycolatopsis orientalis)	43%	WP_051173827.1
	C	Acyl transferase domain-containing protein (Amycolatopsis australiensis)	44%	SFW88505.1
	C	hypothetical protein (Amycolatopsis orientalis)	43%	WP_065912991.1
	C	type I polyketide synthase (Amycolatopsis vancoresmycina)	43%	WP_051767655.1
Anabaena				
	A	malonyl CoA-acyl carrier protein transacylase (Anabaena sp. UHCC 0451)	48%	ATX68115.1
	E	AnaE (Anabaena sp. 37)	40%	AEQ38170.1
	E	polyketide synthase (Anabaena sp. wa102)	40%	WP_053540260.1
	E	polyketide synthase (Anabaena sp. AL93)	40%	OBQ19025.1

Archangium				
	D	type I polyketide synthase (Archangium gephyra)	49%	WP_047858716.1
	D	type I polyketide synthase (Archangium violaceum)	48%	WP_052517812.1
	D	hypothetical protein (Archangium sp. Cb G35)	48%	WP_073564910.1
	F	cytochrome P450 (Archangium sp. Cb G35)	50%	WP_073558427.1
	F	cytochrome P450 (Archangium gephyra)	49%	WP_047858501.1
	F	cytochrome P450 (Archangium gephyra)	47%	WP_047856820.1
	F	cytochrome P450 (Archangium violaceum)	48%	WP_043398650.1
	F	cytochrome P450 (Archangium gephyra)	38%	WP_047858632.1
	F	cytochrome P450 (Archangium violaceum)	37%	WP_084609684.1
	F	hypothetical protein Q664_10145 (Archangium violaceum Cb vi76)	38%	KFA93270.1
	F	hypothetical protein BO221_49900 (Archangium sp. Cb G35)	37%	OJT16443.1
	F	cytochrome P450 (Archangium sp. Cb G35)	37%	WP_083682769.1
Beggiatoa				
	P	hypothetical protein	47%	OQY56275

		B6247_04555 (Beggiatoa sp. 4572_84)		.1
	P	hypothetical protein B6247_12265 (Beggiatoa sp. 4572_84)	54%	OQY54265.1
Blastopirellula				
	E	beta-ketoacyl synthase (Blastopirellula sp.)	38%	PHS11018.1
Brevibacillus				
	P	non-ribosomal peptide synthetase (Brevibacillus thermoruber)	46%	WP_051188156.1
Cellvibrio				
	P	hypothetical protein (Cellvibrio sp. PSBB006)	42%	WP_087465554.1
Chloroflexi				
	F	cytochrome (Chloroflexi bacterium RBG_16_68_14)	37%	OGO50980.1
	F	cytochrome P450 (Chloroflexi bacterium)	34%	PJF22864.1
Chroococcidiopsis				
	P	non-ribosomal peptide synthetase (Chroococcidiopsis thermalis)	49%	WP_015155649.1
Chroogloeocystis				
	A	short-chain dehydrogenase (Chroogloeocystis siderophila)	50%	WP_073549585.1

Clostridium				
	P	non-ribosomal peptide synthetase ((Clostridium) cellulolyticum)	44%	WP_01592 4496.1
Coleofasciculus				
	P	non-ribosomal peptide synthetase (Coleofasciculus chthonoplastes)	46%	WP_00610 4235.1
Crenothrix				
	A	hypothetical protein (Crenothrix polyspora)	51%	WP_08714 5075.1
	E	hypothetical protein (Crenothrix polyspora)	42%	WP_08714 5075.1
Crocosphaera				
	P	non-ribosomal peptide synthetase (Crocosphaera watsonii)	42%	WP_00730 8981.1
	P	non-ribosomal peptide synthetase (Crocosphaera watsonii)	42%	WP_00730 3375.1
	P	non-ribosomal peptide synthetase (Crocosphaera watsonii)	42%	WP_02183 1016.1
	P	non-ribosomal peptide synthetase (Crocosphaera watsonii)	42%	WP_02183 5974.1
Cuspidothrix				
	E	anatoxin-a synthetase E polyketide synthase (Cuspidothrix issatschenkoi	40%	AIU56833. 1

		LBRI48)		
	E	anatoxin-a synthetase E polyketide synthase (Cuspidothrix issatschenkoi RM-6)	40%	AIU56842.1
	E	anatoxin-a synthetase E polyketide synthase (Cuspidothrix issatschenkoi CHABD3)	40%	AIU56851.1
Cyanothece				
	E	type I polyketide synthase (Cyanothece sp. PCC 7822)	41%	WP_013334461.1
	P	MULTISPECIES: non-ribosomal peptide synthetase (Cyanothece)	44%	WP_009547877.1
Cylindrospermum				
	A	PuwE (Cylindrospermum alatosporum CCALA 988)	51%	AIW82282.1
	E	polyketide synthase (Cylindrospermum stagnale)	40%	WP_041233336.1
	E	polyketide synthase family protein (Cylindrospermum stagnale PCC 7417)	40%	AFZ28187.1
Desulfobacter				
	P	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Desulfobacter vibrioformis)	44%	
Enhygromyxa				
	F	putative cytochrome P450 hydroxylase (Enhygromyxa	36%	KIG15792.1

		salina)		
Fischerella				
	P	non-ribosomal peptide synthetase (Fischerella sp. PCC 9339)	47%	WP_017308082.1
	P	non-ribosomal peptide synthetase (Fischerella sp. PCC 9431)	47%	WP_035121683.1
	P	non-ribosomal peptide synthetase (Fischerella sp. PCC 9339)	47%	WP_017308480.1
	P	non-ribosomal peptide synthetase (Fischerella sp. NIES-4106)	45%	WP_096678113.1
	P	non-ribosomal peptide synthetase (Fischerella sp. PCC 9339)	45%	WP_017308928.1
	P	non-ribosomal peptide synthetase module-related protein (Fischerella sp. MV11)	51%	ACN96038.1
Gammaproteobacteria				
	F	hypothetical protein CBD96_02940 (Gammaproteobacteria bacterium TMED236)	33%	OUW93258.1
	F	hypothetical protein CBC64_04330 (Gammaproteobacteria bacterium TMED104)	32%	OUV31048.1
Gloeobacter				
	B	polyketide synthase	43%	WP_01114

		(Gloeobacter violaceus)		1952.1
Gloeocapsa				
	P	non-ribosomal peptide synthetase (Gloeocapsa sp. PCC 73106)	45%	WP_006529541.1
Hapalosiphon				
	P	non-ribosomal peptide synthetase (Hapalosiphon sp. MRB220)	47%	WP_053457693.1
	P	non-ribosomal peptide synthase (Hapalosiphon sp. MRB220)	45%	WP_053456024.1
Hassallia				
	A	short-chain dehydrogenase (Hassallia byssoidea VB512170)	50%	KIF28472.1
	P	amino acid adenylation protein (Hassallia byssoidea VB512170)	47%	KIF36761.1
Herbidospora				
	C	type I polyketide synthase (Herbidospora sakaeratensis)	40%	WP_062342604.1
	C	type I polyketide synthase (Herbidospora sakaeratensis)	41%	WP_062339551.1
Herpetosiphon				
	E	Beta-ketoacyl synthase (Herpetosiphon aurantiacus DSM 785)	42%	ABX06592.1
	E	Beta-ketoacyl synthase (Herpetosiphon aurantiacus	42%	ABX06591.1

		DSM 785)		
Hyalangium				
	B	polyketide synthase (Hyalangium minutum)	45%	WP_052420712.1
	E	type I polyketide synthase (Hyalangium minutum)	42%	WP_052420711.1
Hydrococcus				
	A	beta-ketoacyl synthase (Hydrococcus rivularis)	52%	WP_073601403.1
Hydrogenophaga				
	A	hypothetical protein (Hydrogenophaga sp. PBC)	53%	WP_069902798.1
	A	type I polyketide synthase (Hydrogenophaga intermedia)	53%	WP_009517177.1
	E	hypothetical protein (Hydrogenophaga sp. PML113)	44%	WP_070398025.1
	E	Zn-dependent oxidoreductase/polyketide synthase module (Hydrogenophaga intermedia)	41%	WP_009517173.1
	E	type I polyketide synthase (Hydrogenophaga intermedia)	41%	WP_061285380.1
	E	hypothetical protein (Hydrogenophaga sp. PBC)	41%	WP_069902800.1
Jahnella				

	B	polyketide synthase (Jahnella sp. MSr9139)	42%	AHB82062.1
Kamptonema				
	E	MULTISPECIES: polyketide synthase (Kamptonema)	40%	WP_007358366.1
	P	MULTISPECIES: non-ribosomal peptide synthetase (Kamptonema)	44%	WP_007358393.1
Kitasatospora				
	D	ebeC-type I polyketide synthase (Kitasatospora aburaviensis)	45%	SCN11951.1
	D	type I polyketide synthase (Kitasatospora griseola)	43%	WP_043908455.1
	D	hypothetical protein CG736_06105 (Kitasatospora sp. CB02891)	43%	PJN27782.1
Ktedonobacter				
	B	type I polyketide synthase (Ktedonobacter racemifer)	42%	WP_007919320.1
	F	cytochrome P450 (Ktedonobacter racemifer)	35%	WP_007918627.1
Leptolyngbya				
	A	hypothetical protein (Leptolyngbya sp. 'hensonii')	51%	WP_083636534.1
	A	hypothetical protein BST81_00790 (Leptolyngbya sp. 'hensonii')	51%	OLP20418.1
	A	hypothetical protein (Leptolyngbya sp. 'hensonii')	52%	WP_075596634.1

	A	hypothetical protein (Leptolyngbya sp. 'hensonii')	51%	WP_075596636.1
	E	hypothetical protein (Leptolyngbya sp. 'hensonii')	42%	WP_083636534.1
	E	hypothetical protein BST81_00790 (Leptolyngbya sp. 'hensonii')	42%	OLP20418.1
Lewinella				
	B	type I polyketide synthase (Lewinella cohaerens)	39%	WP_020539102.1
Limnoraphis				
	P	non-ribosomal peptide synthetase (Limnoraphis robusta)	44%	WP_046277879.1
Marinococcus				
	F	cytochrome P450 (Marinococcus halophilus)	35%	WP_079474565.1
Mastigocladus				
	P	non-ribosomal peptide synthetase (Mastigocladus laminosus)	45%	WP_052508374.1
Methylibium				
	E	type I polyketide synthase (Methylibium sp. YR605)	44%	WP_052204532.1
	E	type I polyketide synthase (Methylibium sp. CF468)	44%	WP_052211680.1
Methylomonas				
	P	non-ribosomal peptide synthetase (Methylomonas koyamae)	41%	

	P	non-ribosomal peptide synthetase (Methylomonas koyamae)	41%	
Microcoleus				
	P	non-ribosomal peptide synthetase (Microcoleus sp. PCC 7113)	51%	WP_015181766.1
Minicystis				
	F	putative cytochrome P450 hydroxylase (Minicystis rosea)	39%	APR82207.1
Mycobacterium				
	E	hypothetical protein (Mycobacterium sp. CECT 8778)	41%	WP_099038924.1
	E	type I polyketide synthase (Mycobacterium tusciae)	41%	WP_006245903.1
Nocardiopsis				
	D	polyketide synthase type I (Nocardiopsis sp. FU 40)	42%	AEP40936.1
Nodularia				
	A	short-chain dehydrogenase (Nodularia sp. NIES-3585)	49%	WP_089094281.1
	P	non-ribosomal peptide synthetase (Nodularia spumigena)	40%	WP_063872547.1
	P	non-ribosomal peptide synthetase (Nodularia spumigena)	41%	WP_006194152.1
Paenibacillus				

	F	cytochrome P450 (Paenibacillus fonticola)	36%	WP_019637875.1
	P	non-ribosomal peptide synthetase (Paenibacillus sp. P46E)	44%	WP_074112461.1
	P	non-ribosomal peptide synthetase (Paenibacillus sp. P46E)	42%	WP_074112459.1
Phormidesmis				
	A	beta-ketoacyl synthase (Phormidesmis priestleyi)	50%	WP_083582965.1
	E	beta-ketoacyl synthase (Phormidesmis priestleyi)	43%	WP_083582965.1
Planctomycetaceae				
	A	hypothetical protein CBB70_13645 (Planctomycetaceae bacterium TMED10)	50%	OUT63629.1
Planktothrix				
	A	short-chain dehydrogenase (Planktothrix serta)	49%	WP_083623701.1
Pleurocapsa				
	E	type I polyketide synthase (Pleurocapsa sp. PCC 7319)	37%	WP_019504788.1
Polyangium				
	A	polyketide synthase (Polyangium spumosum)	53%	ANI24099.1
	B	polyketide synthase (Polyangium spumosum)	60%	ANI24099.1

	D	polyketide synthase (Polyangium spumosum)	45%	ANI24099.1
Pseudoalteromonas				
	B	hypothetical protein (Pseudoalteromonas luteoviolacea)	39%	WP_063364358.1
	B	hypothetical protein (Pseudoalteromonas sp. HM-SA03)	39%	WP_095727193.1
	E	hypothetical protein (Pseudoalteromonas luteoviolacea)	39%	WP_063364354.1
	P	non-ribosomal peptide synthetase (Pseudoalteromonas luteoviolacea)	42%	WP_063379439.1
	P	non-ribosomal peptide synthetase (Pseudoalteromonas sp. HM-SA03)	42%	WP_095727196.1
Pseudonocardia				
	B	type I polyketide synthase (Pseudonocardia spinosispora)	43%	WP_028933378.1
	C	type I polyketide synthase (Pseudonocardia spinosispora)	41%	WP_051341806.1
Rhizobacter				
	E	MULTISPECIES: type I polyketide synthase (Rhizobacter)	44%	WP_056809349.1
Rhizobiales				

	F	cytochrome (Rhizobiales bacterium NRL2)	37%	ANK82883.1
	F	cytochrome P450 (Rhizobiales bacterium)	36%	PJK29078.1
Roseofilum				
	E	hypothetical protein BI308_07410 (Roseofilum reptotaenium AO1-A)	40%	OJJ26221.1
	E	hypothetical protein BI308_07390 (Roseofilum reptotaenium AO1-A)	40%	OJJ26217.1
Saccharibacillus				
	F	cytochrome P450 (Saccharibacillus sp. O16)	36%	WP_087797517.1
	F	cytochrome P450 (Saccharibacillus kuerlensis)	36%	WP_018976232.1
	F	cytochrome P450 (Saccharibacillus sacchari)	36%	WP_084778106.1
Saccharopolyspora				
	D	type I polyketide synthase (Saccharopolyspora erythraea)	42%	WP_050784278.1
	D	erythronolide synthase (Saccharopolyspora erythraea)	42%	WP_011873138.1
	D	6-deoxyerythronolide B synthase II (Saccharopolyspora erythraea NRRL 2338)	42%	CAA44448.1
	D	hypothetical protein (Saccharopolyspora	42%	WP_081494992.1

		erythraea)		
Scytonema sp. NIES-4073				
	A	beta-ketoacyl synthase (Scytonema sp. NIES-4073)	49%	WP_096563846.1
	A	short-chain dehydrogenase (Scytonema sp. NIES-4073)	50%	WP_096562237.1
Streptacidiphilus				
	C	type I polyketide synthase (Streptacidiphilus carbonis)	41%	WP_052433191.1
Synechococcus				
	E	type I polyketide synthase (Synechococcus sp. PCC 7335)	38%	WP_006453421.1
Teredinibacter				
	P	hybrid non-ribosomal peptide synthetase/type I polyketide synthase (Teredinibacter turnerae)	41%	WP_018015938.1
Thiothrix				
	A	hypothetical protein BWK73_04140 (Thiothrix lacustris)	51%	OQX16393.1
Thiotrichaceae				
	P	hypothetical protein BWK78_06215 (Thiotrichaceae bacterium IS1)	47%	OQW90774.1
Trichormus				

	A	Short-chain dehydrogenase/reductase SDR (Trichormus variabilis ATCC 29413)	51%	ABA23591.1
	A	Short-chain dehydrogenase/reductase SDR (Trichormus variabilis ATCC 29413)	49%	ABA23590.1
Xanthobacter				
	P	non-ribosomal peptide synthetase (Xanthobacter autotrophicus)	43%	WP_012113029.1

Supplementary Table 3. Initial antiSMASH analysis of the DNA sequence that contains the EAFHs in *C. parasitica* NIES-267.

ECSGs	**Location in Genome**	**Sequence length expected to contain ESGCs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* epothilone gene cluster**
EPOP	1426723-1434210	1426723-1434210	0

Supplementary Table 4. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *M. producens* PAL-8-15-08-1.

ECSGs	**Location in Genome**	**Sequence length expected to contain ESGCs**	**ESGC / Percent Homology of ESGC genes to *S. cellulosum's* epothilone gene cluster**
EPOA	6062944-6069605	4798098-6069605	0
EPOE	6062944-6069605		
EPOP	4798098-4802345		

Supplementary Table 5. Initial antiSMASH analysis of the DNA sequence that contains the the EAFHs in *N. punctiforme* PCC 73102.

ECSGs	Location in Genome	Sequence length expected to contain ESGCs	ESGC / Percent Homology of ESGC genes to *S. cellulosum's* epothilone gene cluster
EPOA	2532676-2538726	2532676-4268529	0
EPOA	4191201-4197833		
EPOB	2538829-2544381		
EPOE	4191201-4197833		
EPOP	4263100-4268529		

Supplementary Table 6. Other SMs' gene clusters found in the genome of *C. fuscus*-strain-DSM-52655.

Cluster number/ type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity to the identified BGC
Terpene	1399132-1420166	Carotenoid	90
Cf_Putative	1432215 -1460080	Xenocyloins	25
Other	2670595-2715004	VEPE/AEPE /TG-1	100
Cf_Putative	3137864-3173416	Lankacidin	13
Terpene	3174459-3201123	Geosmin	100
		2-methylisoborneol	66

Bacteriocin-Proteusin	3767187-3791403	Polytheonamides	16
		Kirromycin	16
Terpene-Cf_Putative	4256616-4289057	Roseoflavin	50
NRPS	5394494-5459784	Tubulysin	15
Cf_Putative	6823018-6834889	Melithiazol(s)	15
		Melithiazol	15
NRPS	9186414-9239907	Myxochelin (A)	58
		Paenibactin	33
		Bacillibactin	33
		Bacillibactin	33
		Griseobactin	33
		Fuscachelin	33
		Mirubactin	25
		Heterobactin(s)	25
		Nataxazole	33
		A33853	33
T3PKS-Amglyccycl	9319808-9368341	Cetoniacytone A	22

Supplementary Table 7. Other SMs' gene clusters found in the genome of *C. fuscus*-DSM-2262.

Cluster type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity to the identified BGC
Microviridin	559145-588800	Microviridin B	66
	1432215 -1460080	Microviridin B	66

Supplementary Table 8. Other SMs' gene clusters found in the genome of *C. crocatus*-strain Cm-c5.

Cluster type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity to the identified BGC
T3PKS	750571-791677	Leupyrrin	14
		Alkylresorcinol(s)	5
		Thuggacin (A1 and B2)	5
Terpene	1897735-1918673	Carotenoid	100
Terpene	2091670-2112647	Ambruticin	11
		Jerangolid	11

Thiopeptide-NRPS	9000891-9085911	Anabaenopeptin	42
		Puwainaphycins (F/G)	57
		Nostophycin	28
Terpene	9248706-9270988	Geosmin	100
		2-methylisoborneol	66

Supplementary Table 9. Other SMs' gene clusters found in the genome of *M. boletus*-DSM-14713.

Cluster number/ type	**Location in genome**	**Homologous known gene clusters**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
NRPS	1100466-1152656	Myxochelin (A)	66
		Paenibactin	33
		Bacillibactin	33
		Griseobactin	33
		Fuscachelin	33
		Mirubactin	25
		Heterobactin (A and B)	25
		Vibriobactin	25
		Nataxazole	33
Terpene	2394158-2416407	Geosmin	100
		2-methylisoborneo	66

		l	
T2PKS	3629267-3671767	Hexaricin (A-C)	27
		FD-594	24
		Griseorhodin	18
		Frankiamicin	21
		TLN-05220	21
		Xantholipin	18
		Pradimicin	18
		Fredericamycin	18
		Lysolipin	21
		Arixanthomycin	21
Terpene	5165021-5236367	Carotenoid	99
		Xenocyloins	18
Microviridin-Lantipeptide	6033461-6091352	Microviridin B	66
Other	6096741-6141150	VEPE / AEPE / TG-1	80
Butyrolactone-Lantipeptide-T1PKS-NRPS	8232812-8324573	Cystomanamides	60

Other	9858118-9899230	Pyrrolnitrin	100

Supplementary Table 10. Other SMs' gene clusters found in the genome of *M. stipitatus* DSM 14675.

Cluster type	**Location in genome**	**Homologous known gene clusters**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
T3PKS	273202-314284	Alkylresorcinol	66
Terpene	752121-774388	Geosmin	100
		2-methylisoborneol	66
Cf_Saccharide-OtherKS	4339321-4423322	Crocacin	14
Cf-Putative	4507024-4529896	Cystothiazole A	17
Other	6788075 - 6832487	VEPE / AEPE / TG-1	100
Terpene	7535752-7559391	Carotenoid	100

Supplementary Table 11. Other SMs' gene clusters found in the genome of *M. hansupus*-strain-mixupus.

Cluster number/ type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity to the identified BGC
T3PKS	273202-314284	Alkylresorcinol	66
Terpene	752121-774388	Geosmin	100
		2-methylisoborneol	66
Cf_Saccharide-OtherKS	4339321-4423322	Crocacin	14
Cf-Putative	4507024-4529896	Cystothiazole A	17
Other	6788075 - 6832487	VEPE / AEPE / TG-1	100
Terpene	7535752-7559391	Carotenoid	100

Supplementary Table 12. Other SMs' gene clusters found in the genome of *M. xanthus*-DK-1622.

Cluster number/ type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity

			to the identified BGC
Terpene	1011919-1037030	Carotenoid	100
Other	1770138-1814550	VEPE / AEPE / TG-1	100
Cf_Putative	3927178-3947221	Cystothiazole A	17
T1PKS-Cf_Saccharide	4012412-4088613	Myxalamid(s)	18
Transatpks-otherKS-NRPS	4710911-4820521	Myxovirescin (A)	100
Bacteriocin-Cf_Saccharide-NRPS	5715538 - 5825093	Microsclerodermins	14
Terpene	7706986 - 7729253	Geosmin	100
		2-methylisoborneol	66
52/T3PKS	8147550-8188632	Alkylresorcinol	66

Supplementary Table 13. Other SMs' gene clusters found in the genome of *N. exedens*-strain-DSM-71-utg0.

Cluster type	Location in genome	Homologous known gene clusters	Percent of genes of known gene clusters that showed similarity to the identified BGC
Terpene-Lantipeptide-T1PKS-NRPS	1833443-1941377	Nostopeptolide	37
		Microsclerodermins	37
		Epothilone	25
		Kedarcidin	25
		Jagaricin	25
		Tallysomycin(s)	25
		Bleomycin	25
		Sporolide	25
Phenazine	8153259-8177700	Lomofungin	21
		Pyocyanine	21
		Endophenazines (A-D)	21
		Marinophenazines	26
		Phenazine(s)	21
		Esmeraldin	17
		Calcimycin	8
		Nataxazole	8

Cf_Putative	10537642-10581788	Citrulline	18
Terpene	10628750-10651104	Geosmin	100
		2-methylisoborneol	66
		Myxothiazol	31
		Leupyrrin	21
		Antalid	21
		Anatoxin	15
		Cryptophycin	10
		Stigmatellin	47
Other	9858118-9899230	Pyrrolnitrin	100

Supplementary Table 14. Other SMs' gene clusters found in the genome of *S. aurantiaca*-DW4-3-1.

Cluster number/ type	**Location in genome**	**Homologous known gene clusters**	**Percent of genes of known gene clusters that showed similarity to the identified BGC**
Terpene	1320786-1343086	Geosmin	100
		2-methylisoborneol	66
T3PKS	1508976-1550061	Alkylresorcinol	100

T1PKS-Cf_Saccharide	1898779-1974399	Dawenol	100
Other	2808565 - 2852977	VEPE / AEPE / TG-1	100
Cf_Putative	3773993 - 3788568	Validamycin valienol/validamine-7-P	50
NRPS	4311702 - 4371327	Myxochelin (A)	100
		Paenibactin	33
		Bacillibactin	33
		Griseobactin	33
		Fuscachelin	33
		Mirubactin	33
		Heterobactin (A and B)	33
		Nataxazole	33
		A33853	33
T1PKS - Cf_Saccharide-NRPS	4751261-4894265	Myxothiazol	100
		Cystothiazole A	100
		Melithiazol	85
		Epothilone	71

		Epothilone	71
		Epothilone	71
		Epothilone	71
		Myxalamid	100
		Stigmatellin	157
Bacteriocin-Lantipeptide-T1PKS-NRPS	5000204-5088183	Nostopeptolide	37
		Cystobactamide(s)	25
Bacteriocin-T1PKS-NRPS		Dkxanthene	94
		Chlorizidine A	23
		Leupyrrin	23
		Ajudazol	35
		Microsclerodermins	17
		Myxalamid(s)	35
		Tubulysin	35
		Pyralomicin	17
Cf_Putative	6071450-6091878	Tubulysin	29
T1PKS	6406347-6494318	Aurafuron A (1) and B (2)	100
		Ajudazol	14

		E-492 / E-975	14
		E-837	21
		Stigmatellin	28
		Melithiazol	21
		Jerangolid	21
		Cystothiazole A	21
		Crocacin	14
T1PKS-Ectoine	6505436-6567622	Cystobactamide(s)	20
Cf_Fatty_Acid	6928075-6952441	Roseoflavin	50
Cf_Putative	8325875-8335655	Aurachin	16
Terpene	8649953-8670963	Carotenoid	100
Microviridin	8911421-8931618	Microviridin K	25
		Microviridin B	25

REFERENCES

1. Abhishek Sharma NK, Ekta Menghani. BIOACTIVE SECONDARY METABOLITES: AN OVERVIEW. *International Journal of Scientific & Engineering Research,*. 2014;5(4):1395-1407.

2. Brakhage AA. Regulation of fungal secondary metabolism. *Nature reviews Microbiology.* 2013;11(1):21-32.

3. Demain AL, Fang A. The natural functions of secondary metabolites. *Advances in biochemical engineering/biotechnology.* 2000;69:1-39.

4. Demain AL. Contribution of genetics to the production and discovery of microbial pharmaceuticals. *Pure and Applied Chemistry.* Vol 601988:833.

5. Gaudelli NM, Long DH, Townsend CA. beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature.* 2015;520(7547):383-387.

6. Hamad B. The antibiotics market. *Nature reviews Drug discovery.* 2010;9(9):675-676.

7. Demain AL, Elander RP. The beta-lactam antibiotics: past, present, and future. *Antonie van Leeuwenhoek.* 1999;75(1-2):5-19.

8. Paradkar AJ, SE.; Mosher, RH. *Biotechnology of antibiotics.* Vol 82. New York: Marcel Dekker, Inc.; 1997.

9. van Bambeke F, Mingeot-Leclercq M-P, Glupczynski Y, Tulkens PM. 137 - Mechanisms of Action A2 - Cohen, Jonathan. In: Powderly WG, Opal SM, eds. *Infectious Diseases (Fourth Edition)*: Elsevier; 2017:1162-1180.e1161.

10. Gupta A, Biyani M, Khaira A. Vancomycin nephrotoxicity: myths and facts. *The Netherlands journal of medicine.* 2011;69(9):379-383.

11. Hicks RW, Hernandez J. Perioperative pharmacology: a focus on vancomycin. *AORN journal.* 2011;93(5):593-596; quiz 597-599.

12. Kahne D, Leimkuhler C, Lu W, Walsh C. Glycopeptide and lipoglycopeptide antibiotics. *Chemical reviews.* 2005;105(2):425-448.

13. Badran EF, Shamayleh A, Irshaid YM. Pharmacokinetics of vancomycin in neonates admitted to the neonatology unit at the Jordan University Hospital. *International journal of clinical pharmacology and therapeutics.* 2011;49(4):252-257.

14. Roszell S, Jones C. Intravenous administration issues: a comparison of intravenous insertions and complications in vancomycin versus other antibiotics. *Journal of infusion nursing : the official publication of the Infusion Nurses Society.* 2010;33(2):112-118.

15. Dunbar LM, Milata J, McClure T, Wasilewski MM. Comparison of the efficacy and safety of oritavancin front-loaded dosing regimens to daily dosing: an analysis of the SIMPLIFI trial. *Antimicrobial agents and chemotherapy.* 2011;55(7):3476-3484.

16. Jeong H, Sim YM, Kim HJ, Lee D-W, Lim S-K, Lee SJ. Genome Sequence of the Vancomycin-Producing Amycolatopsis orientalis subsp. orientalis Strain KCTC 9412(T). *Genome Announcements.* 2013;1(3):e00408-00413.

17. Mc GJ, Bunch RL, Anderson RC, et al. Ilotycin, a new antibiotic. *Antibiotics & chemotherapy (Northfield, Ill).* 1952;2(6):281-283.

18. Gaynor M, Mankin AS. Macrolide antibiotics: binding site, mechanism of action, resistance. *Current topics in medicinal chemistry.* 2003;3(9):949-961.

19. Jacobsson S, Golparian D, Phan LT, et al. In vitro activities of the novel bicyclolides modithromycin (EDP-420, EP-013420, S-013420) and EDP-322 against MDR clinical Neisseria gonorrhoeae isolates and international reference strains. *The Journal of antimicrobial chemotherapy.* 2015;70(1):173-177.

20. Or YS, Clark RF, Wang S, et al. Design, Synthesis, and Antimicrobial Activity of 6-O-Substituted Ketolides Active against Resistant Respiratory Tract Pathogens. *Journal of Medicinal Chemistry.* 2000;43(6):1045-1049.

21. Reeves AR, English RS, Lampel JS, Post DA, Vanden Boom TJ. Transcriptional Organization of the Erythromycin Biosynthetic Gene Cluster of Saccharopolyspora erythraea. *Journal of Bacteriology.* 1999;181(22):7098-7106.

22. Parenti F, Beretta G, Berti M, Arioli V. Teichomycins, new antibiotics from Actinoplanes teichomyceticus Nov. Sp. I. Description of the producer strain, fermentation studies and biological properties. *The Journal of antibiotics.* 1978;31(4):276-283.

23. Walsh CT, Fisher SL, Park IS, Prahalad M, Wu Z. Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chemistry & biology.* 1996;3(1):21-28.

24. Murray BE. Vancomycin-resistant enterococcal infections. *The New England journal of medicine.* 2000;342(10):710-721.

25. Chopra I. New developments in tetracycline antibiotics: glycylcyclines and tetracycline efflux pump inhibitors. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy.* 2002;5(3-4):119-125.

26. Greer ND. Tigecycline (Tygacil): the first in the glycylcycline class of antibiotics. *Proceedings (Baylor University Medical Center).* 2006;19(2):155-161.

27. Hopwood DA. Genetic Contributions to Understanding Polyketide Synthases. *Chemical reviews.* 1997;97(7):2465-2498.

28. Felnagle EA, Jackson EE, Chan YA, et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Molecular pharmaceutics.* 2008;5(2):191-211.

29. Heifets L, Simon J, Pham V. Capreomycin is active against non-replicating M. tuberculosis. *Annals of clinical microbiology and antimicrobials.* 2005;4:6.

30. Johansen SK, Maus CE, Plikaytis BB, Douthwaite S. Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Molecular cell.* 2006;23(2):173-182.

31. Umezawa H. Bleomycin and other antitumor antibiotics of high molecular weight. *Antimicrobial agents and chemotherapy.* 1965;5:1079-1085.

32. Smith BL, Bauer GB, Povirk LF. DNA damage induced by bleomycin, neocarzinostatin, and melphalan in a precisely positioned nucleosome. Asymmetry in protection at the periphery of nucleosome-bound DNA. *The Journal of biological chemistry.* 1994;269(48):30587-30594.

33. Dawson S, Malkinson JP, Paumier D, Searcey M. Bisintercalator natural products with potential therapeutic applications: isolation, structure determination, synthetic and biological studies. *Natural product reports.* 2007;24(1):109-126.

34. Gause GG, Jr., Loshkareva NP, Zbarsky IB. Effect of olivomycin and echinomycin on initiation and growth of RNA chains catalyzed by RNA polymerase. *Biochimica et biophysica acta.* 1968;166(3):752-754.

35. Hammond SM, Kliger BN. Mode of Action of the Polyene Antibiotic Candicidin: Binding Factors in the Wall of Candida albicans. *Antimicrobial Agents and Chemotherapy.* 1976;9(4):561-568.

36. Tweit RC, Pandey RC, Rinehart KL, Jr. Characterization of the antifungal and antiprotozoal antibiotic partricin and structural studies on partricins A and B. *The Journal of antibiotics.* 1982;35(8):997-1012.

37. Clemons KV, Stevens DA. Efficacy of the partricin derivative SPA-S-753 against systemic murine candidosis. *The Journal of antimicrobial chemotherapy.* 2001;47(2):183-186.

38. Mozzi G, Benelli P, Bruzzese T, Galmozzi MR, Bonabello A. The use of lipid emulsions for the i.v. administration of a new water soluble polyene antibiotic, SPK-843. *The Journal of antimicrobial chemotherapy.* 2002;49(2):321-325.

39. Konishi M, Nishio M, Saitoh K, Miyaki T, Oki T, Kawaguchi H. Cispentacin, a new antifungal antibiotic. I. Production, isolation, physico-chemical properties and structure. *The Journal of antibiotics.* 1989;42(12):1749-1755.

40. Hasenoehrl A, Galic T, Ergovic G, et al. In vitro activity and in vivo efficacy of icofungipen (PLD-118), a novel oral antifungal agent, against the pathogenic yeast Candida albicans. *Antimicrobial agents and chemotherapy.* 2006;50(9):3011-3018.

41. Molina JM, Tourneur M, Sarfati C, et al. Fumagillin treatment of intestinal microsporidiosis. *The New England journal of medicine.* 2002;346(25):1963-1969.

42. Schindel L. Treatment of amebiasis with fumagillin. *Journal of the American Medical Association.* 1954;155(10):903-905.

43. Sin N, Meng L, Wang MQ, Wen JJ, Bornmann WG, Crews CM. The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proceedings of the National Academy of Sciences of the United States of America.* 1997;94(12):6099-6103.

44. Porrini MP, Audisio MC, Sabate DC, et al. Effect of bacterial metabolites on microsporidian Nosema ceranae and on its host Apis mellifera. *Parasitology research.* 2010;107(2):381-388.

45. Lin H-C, Chooi Y-H, Dhingra S, Xu W, Calvo AM, Tang Y. The Fumagillin Biosynthetic Gene Cluster in Aspergillus fumigatus Encodes a Cryptic Terpene Cyclase Involved in the Formation of β-trans-Bergamotene. *Journal of the American Chemical Society.* 2013;135(12):4616-4619.

46. Wiwanitkit V. Interest in paromomycin for the treatment of visceral leishmaniasis (kala-azar). *Therapeutics and clinical risk management.* 2012;8:323-328.

47. Fernández MM, Malchiodi EL, Algranati ID. Differential Effects of Paromomycin on Ribosomes of Leishmania mexicana and Mammalian Cells. *Antimicrobial Agents and Chemotherapy.* 2011;55(1):86-93.

48. Kino T, Hatanaka H, Hashimoto M, et al. FK-506, a novel immunosuppressant isolated from a Streptomyces. I. Fermentation, isolation, and physico-chemical and biological characteristics. *The Journal of antibiotics.* 1987;40(9):1249-1255.

49. Karpas A, Lowdell M, Jacobson SK, Hill F. Inhibition of human immunodeficiency virus and growth of infected T cells by the immunosuppressive drugs cyclosporin A and FK 506. *Proceedings of the National Academy of Sciences of the United States of America.* 1992;89(17):8351-8355.

50. Reis SA, Moussatché N, Damaso CRA. FK506, a secondary metabolite produced by Streptomyces, presents a novel antiviral activity against Orthopoxvirus infection in cell culture. *Journal of Applied Microbiology.* 2006;100(6):1373-1380.

51. Helynck G, Dubertret C, Mayaux JF, Leboul J. Isolation of RP 71955, a new anti-HIV-1 peptide secondary metabolite. *The Journal of antibiotics.* 1993;46(11):1756-1757.

52. Vezina C, Kudelski A, Sehgal SN. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *The Journal of antibiotics.* 1975;28(10):721-726.

53. Amedei A, D'Elios MM. New therapeutic approaches by using microorganism-derived compounds. *Current medicinal chemistry.* 2012;19(22):3822-3840.

54. Kasap B. Sirolimus in pediatric renal transplantation. *Pediatric transplantation.* 2011;15(7):673-685.

55. Prasad R, Singh UK, Mittal R, Mishra OP. Immunosuppressive therapy in children. *Journal of the Indian Medical Association.* 2011;109(2):101-102, 104-107.

56. Dreyfuss M, Härri E, Hofmann H, Kobel H, Pache W, Tscherter H. Cyclosporin A and C. *European journal of applied microbiology and biotechnology.* 1976;3(2):125-133.

57. White DJG, Calne RY. The Use of Cyclosporin A Immunosuppression in Organ Grafting. *Immunological Reviews.* 1982;65(1):115-131.

58. Borel JF, Feurer C, Gubler HU, Stahelin H. Biological effects of cyclosporin A: a new antilymphocytic agent. *Agents and actions.* 1976;6(4):468-475.

59. Giddings LA, Newman DJ. Microbial natural products: molecular blueprints for antitumor drugs. *Journal of industrial microbiology & biotechnology*. 2013;40(11):1181-1210.

60. de Vries JF, Zwaan CM, De Bie M, et al. The novel calicheamicin-conjugated CD22 antibody inotuzumab ozogamicin (CMC-544) effectively kills primary pediatric acute lymphoblastic leukemia cells. *Leukemia.* 2012;26(2):255-264.

61. Waksman SA, Woodruff HB. Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. *Proceedings of the Society for Experimental Biology and Medicine.* 1940;45(2):609-614.

62. Hackmann C. [Experimental investigations on the effects of actinomycin C (HBF 386) in malignancies]. *Zeitschrift fur Krebsforschung.* 1952;58(4-5):607-613.

63. Bitzer J, Gesheva V, Zeeck A. Actinomycins with altered threonine units in the beta-peptidolactone. *Journal of natural products.* 2006;69(8):1153-1157.

64. Chen C, Song F, Wang Q, et al. A marine-derived Streptomyces sp. MS449 produces high yield of actinomycin X2 and actinomycin D with potent anti-tuberculosis activity. *Applied microbiology and biotechnology.* 2012;95(4):919-927.

65. Mu X, Song L, Li Q, Yin R, Zhao X, Wang D. Comparison of pulsed actinomycin D and 5-day actinomycin D as first-line chemotherapy for low-risk gestational trophoblastic neoplasia. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics.* 2018;143(2):225-231.

66. Prouvot C, Golfier F, Massardier J, et al. Efficacy and Safety of Second-Line 5-Day Dactinomycin in Case of Methotrexate Failure for Gestational Trophoblastic Neoplasia. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society.* 2018;28(5):1038-1044.

67. Sehgal SN, Baker H, Vezina C. Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *The Journal of antibiotics.* 1975;28(10):727-732.

68. Eng CP, Sehgal SN, Vezina C. Activity of rapamycin (AY-22,989) against transplanted tumors. *The Journal of antibiotics.* 1984;37(10):1231-1237.

69. Uehara Y, Hori M, Takeuchi T, Umezawa H. Phenotypic change from transformed to normal induced by benzoquinonoid ansamycins accompanies

inactivation of p60src in rat kidney cells infected with Rous sarcoma virus. *Molecular and cellular biology.* 1986;6(6):2198-2206.

70. Whitesell L, Mimnaugh EG, De Costa B, Myers CE, Neckers LM. Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. *Proceedings of the National Academy of Sciences of the United States of America.* 1994;91(18):8324-8328.

71. Stebbins CE, Russo AA, Schneider C, Rosen N, Hartl FU, Pavletich NP. Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell.* 1997;89(2):239-250.

72. Tsuji N, Kobayashi M, Nagashima K, Wakisaka Y, Koizumi K. A new antifungal antibiotic, trichostatin. *The Journal of antibiotics.* 1976;29(1):1-6.

73. Ueda H, Nakajima H, Hori Y, et al. FR901228, a novel antitumor bicyclic depsipeptide produced by Chromobacterium violaceum No. 968. I. Taxonomy, fermentation, isolation, physico-chemical and biological properties, and antitumor activity. *The Journal of antibiotics.* 1994;47(3):301-310.

74. Cheng Y-Q, Yang M, Matter AM. Characterization of a Gene Cluster Responsible for the Biosynthesis of Anticancer Agent FK228 in Chromobacterium violaceum No. 968. *Applied and Environmental Microbiology.* 2007;73(11):3460-3469.

75. Kollar P, Rajchard J, Balounova Z, Pazourek J. Marine natural products: bryostatins in preclinical and clinical studies. *Pharmaceutical biology.* 2014;52(2):237-242.

76. Reichenbach H, Hofle G. Discovery and development of the epothilones : a novel class of antineoplastic drugs. *Drugs in R&D.* 2008;9(1):1-10.

77. Bollag DM, McQueney PA, Zhu J, et al. Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer research.* 1995;55(11):2325-2333.

78. Ferrandina G, Mariani M, Andreoli M, Shahabi S, Scambia G, Ferlini C. Novel drugs targeting microtubules: the role of epothilones. *Current pharmaceutical design.* 2012;18(19):2793-2803.

79. Kowalski RJ, Giannakakou P, Hamel E. Activities of the microtubule-stabilizing agents epothilones A and B with purified tubulin and in cells resistant to paclitaxel (Taxol(R)). *The Journal of biological chemistry.* 1997;272(4):2534-2541.

80. Chou TC, Zhang XG, Balog A, et al. Desoxyepothilone B: an efficacious microtubule-targeted antitumor agent with a promising in vivo profile relative to epothilone B. *Proceedings of the National Academy of Sciences of the United States of America.* 1998;95(16):9642-9647.

81. Blum W, Aichholz R, Ramstein P, et al. In vivo metabolism of epothilone B in tumor-bearing nude mice: identification of three new epothilone B metabolites by capillary high-pressure liquid chromatography/mass spectrometry/tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2001;15(1):41-49.

82. Lee FY, Borzilleri R, Fairchild CR, et al. BMS-247550: a novel epothilone analog with a mode of action similar to paclitaxel but possessing superior antitumor efficacy. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2001;7(5):1429-1437.

83. Lee FY, Smykla R, Johnston K, et al. Preclinical efficacy spectrum and pharmacokinetics of ixabepilone. *Cancer chemotherapy and pharmacology*. 2009;63(2):201-212.

84. Gottfried K, Klar U, Platzek J, Zorn L. Biocatalysis at Work: Applications in the Development of Sagopilone. *ChemMedChem.* 2015;10(7):1240-1248.

85. Beer TM, Higano CS, Saleh M, et al. Phase II study of KOS-862 in patients with metastatic androgen independent prostate cancer previously treated with docetaxel. *Invest New Drugs.* 2007;25(6):565-570.

86. Trivedi M, Budihardjo I, Loureiro K, Reid TR, Ma JD. Epothilones: a novel class of microtubule-stabilizing drugs for the treatment of cancer. *Future Oncol.* 2008;4(4):483-500.

87. Welker M, Dittmann E, von Dohren H. Cyanobacteria as a source of natural products. *Methods in enzymology.* 2012;517:23-46.

88. Baltz RH. Antimicrobials from actinomycetes: back to the future. *Microbe.* 2007;2(3):125–131.

89. Ikeda H, Ishikawa J, Hanamoto A, et al. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nature biotechnology.* 2003;21(5):526-531.

90. Bentley SD, Chater KF, Cerdeno-Tarraga AM, et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature.* 2002;417(6885):141-147.

91. Paterson J, Jahanshah G, Li Y, Wang Q, Mehnaz S, Gross H. The contribution of genome mining strategies to the understanding of active principles of PGPR strains. *FEMS microbiology ecology.* 2017;93(3).

92. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes - a review. *Natural product reports.* 2016;33(8):988-1005.

93. Kang HS. Phylogeny-guided (meta)genome mining approach for the targeted discovery of new microbial natural products. *Journal of industrial microbiology & biotechnology.* 2017;44(2):285-293.

94. Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature reviews Microbiology.* 2007;5(3):175-186.

95. Thaker MN, Wang W, Spanogiannopoulos P, et al. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat Biotechnol.* 2013;31(10):922-927.

96. Wolf T, Shelest V, Nath N, Shelest E. CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics (Oxford, England).* 2016;32(8):1138-1143.

97. Weber T, Blin K, Duddela S, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research.* 2015;43(W1):W237-W243.

98. Chen LY, Wang XQ, Wang YM, et al. Genome mining of Streptomyces xinghaiensis NRRL B-24674(T) for the discovery of the gene cluster involved in anticomplement activities and detection of novel xiamycin analogs. *Applied microbiology and biotechnology.* 2018;102(22):9549-9562.

99. Chen H, Du L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Applied microbiology and biotechnology.* 2016;100(2):541-557.

100. Tian J, Chen H, Guo Z, et al. Discovery of pentangular polyphenols hexaricins A-C from marine Streptosporangium sp. CGMCC 4.7309 by genome mining. *Applied microbiology and biotechnology.* 2016;100(9):4189-4199.

101. Paulus C, Rebets Y, Zapp J, Ruckert C, Kalinowski J, Luzhetskyy A. New Alpiniamides From Streptomyces sp. IB2014/011-12 Assembled by an Unusual Hybrid Non-ribosomal Peptide Synthetase Trans-AT Polyketide Synthase Enzyme. *Frontiers in microbiology.* 2018;9:1959.

102. Micallef ML, D'Agostino PM, Sharma D, Viswanathan R, Moffitt MC. Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC genomics.* 2015;16:669.

103. Pistorius D, Muller R. Discovery of the rhizopodin biosynthetic gene cluster in Stigmatella aurantiaca Sg a15 by genome mining. *Chembiochem : a European journal of chemical biology.* 2012;13(3):416-426.

104. Dalby SM, Goodwin-Tindall J, Paterson I. Total synthesis of (-)-rhizopodin. *Angewandte Chemie (International ed in English).* 2013;52(25):6517-6521.

105. Clark AM. Natural products as a resource for new drugs. *Pharmaceutical research.* 1996;13(8):1133-1144.

106. Kmietowicz Z. Few novel antibiotics in the pipeline, WHO warns. *BMJ.* 2017;358:j4339.

107. Littler E, Oberg B. Achievements and challenges in antiviral drug discovery. *Antiviral chemistry & chemotherapy.* 2005;16(3):155-168.

108. Brown GD, Denning DW, Levitz SM. Tackling human fungal infections. *Science (New York, NY).* 2012;336(6082):647.

109. Brown GD, Denning DW, Gow NA, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Science translational medicine.* 2012;4(165):165rv113.

110. Roemer T, Krysan DJ. Antifungal drug development: challenges, unmet clinical needs, and new approaches. *Cold Spring Harbor perspectives in medicine.* 2014;4(5).

111. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians.* 2018.

112. Alfarouk KO, Stock CM, Taylor S, et al. Resistance to cancer chemotherapy: failure in drug response from ADME to P-gp. *Cancer cell international.* 2015;15:71.

113. Undevia SD, Gomez-Abuin G, Ratain MJ. Pharmacokinetic variability of anticancer agents. *Nature reviews Cancer.* 2005;5(6):447-458.

114. Thorn M, Finnstrom N, Lundgren S, Rane A, Loof L. Cytochromes P450 and MDR1 mRNA expression along the human gastrointestinal tract. *British journal of clinical pharmacology.* 2005;60(1):54-60.

115. Sparreboom A, van Asperen J, Mayer U, et al. Limited oral bioavailability and active epithelial excretion of paclitaxel (Taxol) caused by P-glycoprotein in the intestine. *Proceedings of the National Academy of Sciences of the United States of America.* 1997;94(5):2031-2035.

116. He K, Iyer KR, Hayes RN, Sinz MW, Woolf TF, Hollenberg PF. Inactivation of cytochrome P450 3A4 by bergamottin, a component of grapefruit juice. *Chemical research in toxicology.* 1998;11(4):252-259.

117. McFadyen MC, Melvin WT, Murray GI. Cytochrome P450 enzymes: novel options for cancer therapeutics. *Molecular cancer therapeutics.* 2004;3(3):363-371.

118. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. *Oncogene.* 2006;25(11):1679-1691.

119. Kauvar LM, Morgan AS, Sanderson PE, Henner WD. Glutathione based approaches to improving cancer treatment. *Chemico-biological interactions.* 1998;111-112:225-238.

120. Schuetz EG, Schinkel AH, Relling MV, Schuetz JD. P-glycoprotein: a major determinant of rifampicin-inducible expression of cytochrome P4503A in mice and humans. *Proceedings of the National Academy of Sciences of the United States of America.* 1996;93(9):4001-4005.

121. Cole SP, Bhardwaj G, Gerlach JH, et al. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science (New York, NY).* 1992;258(5088):1650-1654.

122. Sauna ZE, Smith MM, Muller M, Kerr KM, Ambudkar SV. The mechanism of action of multidrug-resistance-linked P-glycoprotein. *Journal of bioenergetics and biomembranes.* 2001;33(6):481-491.

123. Sakata K, Kwok TT, Murphy BJ, Laderoute KR, Gordon GR, Sutherland RM. Hypoxia-induced drug resistance: comparison to P-glycoprotein-associated drug resistance. *British journal of cancer.* 1991;64(5):809-814.

124. Fu Y, Lee AS. Glucose regulated proteins in cancer progression, drug resistance and immunotherapy. *Cancer biology & therapy.* 2006;5(7):741-744.

125. Lyon RC, Cohen JS, Faustino PJ, Megnin F, Myers CE. Glucose metabolism in drug-sensitive and drug-resistant human breast cancer cells monitored by magnetic resonance spectroscopy. *Cancer research.* 1988;48(4):870-877.

126. Lu D, Hong Chen E, Lu T-R, Ding J, Xu B. *anticancer drug combination; the next medical challenge.* Vol 32015.

127. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-

label, randomised controlled trial. *Lancet (London, England).* 2010;376(9742):687-697.

128. Bozic I, Reiter JG, Allen B, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife.* 2013;2:e00747.

129. Carrick S, Parker S, Thornton CE, Ghersi D, Simes J, Wilcken N. Single agent versus combination chemotherapy for metastatic breast cancer. *The Cochrane database of systematic reviews.* 2009(2):Cd003372.

130. Krysan DJ. The unmet clinical need of novel antifungal drugs. *Virulence.* 2017;8(2):135-137.

131. Clancy CJ, Nguyen MH. Emergence of Candida auris: An International Call to Arms. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2017;64(2):141-143.

132. Singh R, Kumar M, Mittal A, Mehta PK. Microbial metabolites in nutrition, healthcare and agriculture. *3 Biotech.* 2017;7(1):15.

133. Stegall MD, Morris RE, Alloway RR, Mannon RB. Developing New Immunosuppression for the Next Generation of Transplant Recipients: The Path Forward. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons.* 2016;16(4):1094-1101.

134. Jayaraj S, Suresh S, Kadeppagari R-K. Amylase inhibitors and their biomedical applications. *Starch - Stärke.* 2013;65(7-8):535-542.

135. Adrio J-L, Demain AL. Recombinant organisms for production of industrial products. *Bioengineered bugs.* 2010;1(2):116-131.

136. James T. Staley RWC, Rita R. Colwell, John G. Holt, Matthew D. Kane,Norman R. Pace, Abigail A. Salyers, James M. Tiedje. The Microbial World: Foundation of the Biosphere. *American society for microbiology.* 1997;63:6-7.

137. Floss HG. Combinatorial biosynthesis--potential and problems. *Journal of biotechnology.* 2006;124(1):242-257.

138. Lau J, Frykman S, Regentin R, Ou S, Tsuruta H, Licari P. Optimizing the heterologous production of epothilone D in Myxococcus xanthus. *Biotechnology and bioengineering.* 2002;78(3):280-288.

139. Newman GMCaDJ. Biodiversity: A continuing source of novel drug leads. *Pure Appl Chem.* 2005;77:7-24.

140. Thakkar SNPaD. combinatorial chemistry: A novel method in drug discovery and its application. *Indian Journal of Chemistry.* 2005;44B:335-348.

141. Nicolaou KC, Kim S, Pfefferkorn J, et al. Synthesis and Biological Activity of Sarcodictyins. *Angewandte Chemie (International ed in English).* 1998;37(10):1418-1421.

142. Du L, Lou L. PKS and NRPS release mechanisms. *Natural product reports.* 2010;27(2):255-278.

143. Gomes ES, Schuch V, de Macedo Lemos EG. Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology].* 2013;44(4):1007-1034.

144. Desriac F, Jegou C, Balnois E, Brillet B, Le Chevalier P, Fleury Y. Antimicrobial peptides from marine proteobacteria. *Marine drugs.* 2013;11(10):3632-3660.

145. Cacho RA, Jiang W, Chooi YH, Walsh CT, Tang Y. Identification and characterization of the echinocandin B biosynthetic gene cluster from Emericella rugulosa NRRL 11440. *J Am Chem Soc.* 2012;134(40):16781-16790.

146. Sussmuth RD, Mainz A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Ed Engl.* 2017;56(14):3770-3821.

147. Rachid S, Revermann O, Dauth C, Kazmaier U, Muller R. Characterization of a novel type of oxidative decarboxylase involved in the biosynthesis of the styryl moiety of chondrochloren from an acylated tyrosine. *The Journal of biological chemistry.* 2010;285(17):12482-12489.

148. Moore BS, Hopke JN. Discovery of a new bacterial polyketide biosynthetic pathway. *Chembiochem : a European journal of chemical biology.* 2001;2(1):35-38.

149. Funa N, Ohnishi Y, Fujii I, Shibuya M, Ebizuka Y, Horinouchi S. A new pathway for polyketide synthesis in microorganisms. *Nature.* 1999;400(6747):897-899.

150. Bangera MG, Thomashow LS. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from Pseudomonas fluorescens Q2-87. *Journal of bacteriology.* 1999;181(10):3155-3163.

151. Yu D, Xu F, Zeng J, Zhan J. Type III polyketide synthases in natural product biosynthesis. *IUBMB life.* 2012;64(4):285-295.

152. Mach B, Reich E, Tatum EL. Separation of the biosynthesis of the antibiotic polypeptide tyrocidine from protein biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America.* 1963;50(1):175-181.

153. Borowska ZK, Tatum EL. Biosynthesis of edeine by Bacillus brevis Vm4 in vivo and in vitro. *Biochim Biophys Acta.* 1966;114(1):206-209.

154. Spaeren U, Froholm LO, Laland SG. Further studies on the biosynthesis of gramicidin S and proteins in a cell-free system from Bacillus brevis. *The Biochemical journal.* 1967;102(2):586-592.

155. Berg TL, Froholm LO, Laland SG. The biosynthesis of gramicidin s in a cell-free system. *The Biochemical journal.* 1965;96:43-52.

156. Daniels MJ. Studies of the biosynthesis of polymyxin B. *Biochimica et biophysica acta.* 1968;156(1):119-127.

157. Joyce SA, Lango L, Clarke DJ. The Regulation of Secondary Metabolism and Mutualism in the Insect Pathogenic Bacterium Photorhabdus luminescens. *Advances in applied microbiology.* 2011;76:1-25.

158. Flardh K, Buttner MJ. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature reviews Microbiology.* 2009;7(1):36-49.

159. Ruiz B, Chavez A, Forero A, et al. Production of microbial secondary metabolites: regulation by the carbon source. *Critical reviews in microbiology.* 2010;36(2):146-167.

160. Vining LC. Functions of secondary metabolites. *Annual review of microbiology.* 1990;44:395-427.

161. Kawamura C, Tsujimoto T, Tsuge T. Targeted disruption of a melanin biosynthesis gene affects conidial development and UV tolerance in the Japanese pear pathotype of Alternaria alternata. *Molecular plant-microbe interactions : MPMI.* 1999;12(1):59-63.

162. Bills G, Li Y, Chen L, Yue Q, Niu XM, An Z. New insights into the echinocandins and other fungal non-ribosomal peptides and peptaibiotics. *Natural product reports.* 2014;31(10):1348-1375.

163. Johnson L. Iron and siderophores in fungal-host interactions. *Mycological research.* 2008;112(Pt 2):170-183.

164. Haas H. Iron - A Key Nexus in the Virulence of Aspergillus fumigatus. *Frontiers in microbiology.* 2012;3:28.

165. Blatzer M, Schrettl M, Sarg B, Lindner HH, Pfaller K, Haas H. SidL, an Aspergillus fumigatus transacetylase involved in biosynthesis of the siderophores ferricrocin and hydroxyferricrocin. *Applied and environmental microbiology.* 2011;77(14):4959-4966.

166. Yin WB, Baccile JA, Bok JW, Chen Y, Keller NP, Schroeder FC. A nonribosomal peptide synthetase-derived iron(III) complex from the pathogenic fungus Aspergillus fumigatus. *Journal of the American Chemical Society.* 2013;135(6):2064-2067.

167. Haselwandter K. *Structure and function of siderophores produced by mycorrhizal fungi.* Vol 722008.

168. Haselwandter K, Passler V, Reiter S, et al. Basidiochrome -- a novel siderophore of the Orchidaceous Mycorrhizal Fungi Ceratobasidium and Rhizoctonia spp. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine.* 2006;19(3):335-343.

169. Johnson LJ, Koulman A, Christensen M, et al. An Extracellular Siderophore Is Required to Maintain the Mutualistic Interaction of Epichloë festucae with Lolium perenne. *PLOS Pathogens.* 2013;9(5):e1003332.

170. Hansen FT, Droce A, Sorensen JL, Fojan P, Giese H, Sondergaard TE. Overexpression of NRPS4 leads to increased surface hydrophobicity in Fusarium graminearum. *Fungal biology.* 2012;116(8):855-862.

171. Turgeon BG, Oide S, Bushley K. Creating and screening Cochliobolus heterostrophus non-ribosomal peptide synthetase mutants. *Mycological research.* 2008;112(Pt 2):200-206.

172. Kim KH, Cho Y, M LAR, Cramer RA, Jr., Lawrence CB. Functional analysis of the Alternaria brassicicola non-ribosomal peptide synthetase gene AbNPS2 reveals a role in conidial cell wall construction. *Molecular plant pathology.* 2007;8(1):23-39.

173. Gonzalez O, Ortiz-Castro R, Diaz-Perez C, et al. Non-ribosomal Peptide Synthases from Pseudomonas aeruginosa Play a Role in Cyclodipeptide Biosynthesis, Quorum-Sensing Regulation, and Root Development in a Plant Host. *Microbial ecology.* 2017;73(3):616-629.

174. Spraker JE, Sanchez LM, Lowe TM, Dorrestein PC, Keller NP. Ralstonia solanacearum lipopeptide induces chlamydospore development in fungi and facilitates bacterial entry into fungal tissues. *The ISME journal.* 2016;10(9):2317-2330.

175. Li P, Yin W, Yan J, et al. Modulation of Inter-kingdom Communication by PhcBSR Quorum Sensing System in Ralstonia solanacearum Phylotype I Strain GMI1000. *Frontiers in microbiology.* 2017;8:1172.

176. Schroth MN, Hancock JG. Disease-suppressive soil and root-colonizing bacteria. *Science (New York, NY).* 1982;216(4553):1376-1381.

177. Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M. Natural functions of lipopeptides from Bacillus and Pseudomonas: more than surfactants and antibiotics. *FEMS Microbiology Reviews.* 2010;34(6):1037-1062.

178. Raaijmakers JM, de Bruijn I, de Kock MJ. Cyclic lipopeptide production by plant-associated Pseudomonas spp.: diversity, activity, biosynthesis, and regulation. *Molecular plant-microbe interactions : MPMI.* 2006;19(7):699-710.

179. Du L, Shen B. Biosynthesis of hybrid peptide-polyketide natural products. *Current opinion in drug discovery & development.* 2001;4(2):215-228.

180. Pavlidou M, Pross EK, Musiol EM, Kulik A, Wohlleben W, Weber T. The phosphopantetheinyl transferase KirP activates the ACP and PCP domains of the kirromycin NRPS/PKS of Streptomyces collinus Tu 365. *FEMS microbiology letters.* 2011;319(1):26-33.

181. Du L, Sanchez C, Shen B. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic engineering.* 2001;3(1):78-95.

182. Hertweck C. The Biosynthetic Logic of Polyketide Diversity. Vol 48. Weinheim: WILEY-VCH Verlag; 2009:4688-4716.

183. Zhang W, Tang Y. In vitro analysis of type II polyketide synthase. *Methods in enzymology.* 2009;459:367-393.

184. Lim Y, Go M, Yew W. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. *Molecules.* 2016;21(6):806.

185. Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. *Natural product reports.* 2001;18(4):380-416.

186. Stewart C, Jr., Vickery CR, Burkart MD, Noel JP. Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. *Current opinion in plant biology.* 2013;16(3):365-372.

187. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G. NORINE: a database of nonribosomal peptides. *Nucleic acids research.* 2008;36(Database issue):D326-331.

188. Caboche S, Pupin M, Leclere V, Jacques P, Kucherov G. Structural pattern matching of nonribosomal peptides. *BMC structural biology.* 2009;9:15.

189. Caboche S, Leclere V, Pupin M, Kucherov G, Jacques P. Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *Journal of bacteriology.* 2010;192(19):5143-5150.

190. Gao X, Haynes SW, Ames BD, et al. Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. *Nature chemical biology.* 2012;8(10):823-830.

191. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences of the United States of America.* 2014;111(25):9259-9264.

192. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews.* 2006;106(8):3468-3496.

193. Keating TA, Walsh CT. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current opinion in chemical biology.* 1999;3(5):598-606.

194. Singh R, Singh P, Sharma R, Mohapatra T. *NONRIBOSOMAL PEPTIDE SYNTHESIS IN MICROBES.* 2012.

195. Finking R, Marahiel MA. Biosynthesis of nonribosomal peptides1. *Annual review of microbiology.* 2004;58:453-488.

196. Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *The Journal of biological chemistry.* 1998;273(35):22773-22781.

197. Konz D, Marahiel MA. How do peptide synthetases generate structural diversity? *Chemistry & biology.* 1999;6(2):R39-48.

198. Walsh CT, Chen H, Keating TA, et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current opinion in chemical biology.* 2001;5(5):525-534.

199. Dieckmann R, Lee YO, van Liempt H, von Dohren H, Kleinkauf H. Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS letters.* 1995;357(2):212-216.

200. Lautru S, Challis GL. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology (Reading, England).* 2004;150(Pt 6):1629-1636.

201. May JJ, Kessler N, Marahiel MA, Stubbs MT. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America.* 2002;99(19):12120-12125.

202. Stachelhaus T, Huser A, Marahiel MA. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chemistry & biology.* 1996;3(11):913-921.

203. Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT. The EntF and EntE adenylation domains of Escherichia coli enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences of the United States of America.* 2000;97(6):2509-2514.

204. Bergendahl V, Linne U, Marahiel MA. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European journal of biochemistry.* 2002;269(2):620-629.

205. Wang H, Sivonen K, Fewer DP. Genomic insights into the distribution, genetic diversity and evolution of polyketide synthases and nonribosomal peptide synthetases. *Current opinion in genetics & development.* 2015;35:79-85.

206. Lambalot RH, Gehring AM, Flugel RS, et al. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & biology.* 1996;3(11):923-936.

207. Walsh CT, Gehring AM, Weinreb PH, Quadri LE, Flugel RS. Post-translational modification of polyketide and nonribosomal peptide synthases. *Current opinion in chemical biology.* 1997;1(3):309-315.

208. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & biology.* 1999;6(8):493-505.

209. Lau J, Fu H, Cane DE, Khosla C. Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry.* 1999;38(5):1643-1651.

210. Khosla C. Harnessing the Biosynthetic Potential of Modular Polyketide Synthases. *Chemical reviews.* 1997;97(7):2577-2590.

211. Butler AR, Bate N, Cundliffe E. Impact of thioesterase activity on tylosin biosynthesis in Streptomyces fradiae. *Chemistry & biology.* 1999;6(5):287-292.

212. Konz D, Klens A, Schorgendorfer K, Marahiel MA. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chemistry & biology.* 1997;4(12):927-937.

213. Patel HM, Tao J, Walsh CT. Epimerization of an L-cysteinyl to a D-cysteinyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from Pseudomonas aeruginosa. *Biochemistry.* 2003;42(35):10514-10527.

214. Weber G, Schorgendorfer K, Schneider-Scherzer E, Leitner E. The peptide synthetase catalyzing cyclosporine production in Tolypocladium niveum is encoded by a giant 45.8-kilobase open reading frame. *Current genetics.* 1994;26(2):120-125.

215. Tillett D, Dittmann E, Erhard M, von Dohren H, Borner T, Neilan BA. Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system. *Chemistry & biology.* 2000;7(10):753-764.

216. Stachelhaus T, Walsh CT. Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry.* 2000;39(19):5775-5787.

217. Miller BR, Gulick AM. Structural Biology of Non-Ribosomal Peptide Synthetases. *Methods in molecular biology (Clifton, NJ).* 2016;1401:3-29.

218. Gaitatzis N, Kunze B, Muller R. In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America.* 2001;98(20):11136-11141.

219. Mariappan A, Makarewicz O, Chen XH, Borriss R. Two-component response regulator DegU controls the expression of bacilysin in plant-growth-promoting bacterium Bacillus amyloliquefaciens FZB42. *Journal of molecular microbiology and biotechnology.* 2012;22(2):114-125.

220. Inaoka T, Wang G, Ochi K. ScoC regulates bacilysin production at the transcription level in Bacillus subtilis. *Journal of bacteriology.* 2009;191(23):7367-7371.

221. Wang X, Luo C, Liu Y, et al. Three non-aspartate amino acid mutations in the ComA Response regulator receiver motif severely decrease surfactin production,

competence development and spore formation in Bacillus subtilis. *Journal of microbiology and biotechnology.* 2010;20(2):301-310.

222. Roongsawang N, Washio K, Morikawa M. Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. *International journal of molecular sciences.* 2010;12(1):141-172.

223. Nakano MM, Zuber P. Mutational analysis of the regulatory region of the srfA operon in Bacillus subtilis. *Journal of bacteriology.* 1993;175(10):3188-3191.

224. Hayashi K, Ohsawa T, Kobayashi K, Ogasawara N, Ogura M. The H2O2 stress-responsive regulator PerR positively regulates srfA expression in Bacillus subtilis. *Journal of bacteriology.* 2005;187(19):6659-6667.

225. Chaudhary AK, Singh B, Maharjan S, Jha AK, Kim BG, Sohng JK. Switching antibiotics production on and off in actinomycetes by an IclR family transcriptional regulator from Streptomyces peucetius ATCC 27952. *Journal of microbiology and biotechnology.* 2014;24(8):1065-1072.

226. Strieker M, Tanovic A, Marahiel MA. Nonribosomal peptide synthetases: structures and dynamics. *Current opinion in structural biology.* 2010;20(2):234-240.

227. Stachelhaus T, Marahiel MA. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *The Journal of biological chemistry.* 1995;270(11):6163-6169.

228. Conti E, Stachelhaus T, Marahiel MA, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo j.* 1997;16(14):4174-4183.

229. Weber T, Baumgartner R, Renner C, Marahiel MA, Holak TA. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure (London, England : 1993).* 2000;8(4):407-418.

230. Quadri LE, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT. Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry.* 1998;37(6):1585-1595.

231. Condurso HL, Bruner SD. Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery. *Natural product reports.* 2012;29(10):1099-1110.

232. Frueh DP, Arthanari H, Koglin A, et al. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature.* 2008;454(7206):903-906.

233. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure (London, England : 1993).* 2007;15(7):781-792.

234. Tanovic A, Samel SA, Essen LO, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (New York, NY).* 2008;321(5889):659-663.

235. Keating TA, Marshall CG, Walsh CT, Keating AE. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature structural biology.* 2002;9(7):522-526.

236. Kajimura Y, Kaneda M. Fusaricidin A, a new depsipeptide antibiotic produced by Bacillus polymyxa KT-8. Taxonomy, fermentation, isolation, structure elucidation and biological activity. *The Journal of antibiotics.* 1996;49(2):129-135.

237. Han JW, Kim EY, Lee JM, Kim YS, Bang E, Kim BS. Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnology letters.* 2012;34(7):1327-1334.

238. Kim SY, Park SY, Choi SK, Park SH. Biosynthesis of Polymyxins B, E, and P Using Genetically Engineered Polymyxin Synthetases in the Surrogate Host Bacillus subtilis. *Journal of microbiology and biotechnology.* 2015;25(7):1015-1025.

239. Nielsen ML, Isbrandt T, Petersen LM, et al. Linker Flexibility Facilitates Module Exchange in Fungal Hybrid PKS-NRPS Engineering. *PloS one.* 2016;11(8):e0161199.

240. Goering AW, Li J, McClure RA, Thomson RJ, Jewett MC, Kelleher NL. In Vitro Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free Protein Synthesis. *ACS synthetic biology.* 2017;6(1):39-44.

241. Prasad C. Bioactive cyclic dipeptides. *Peptides.* 1995;16(1):151-164.

242. Bellezza I, Peirce MJ, Minelli A. Cyclic dipeptides: from bugs to brain. *Trends in molecular medicine.* 2014;20(10):551-558.

243. Molnar I, Schupp T, Ono M, et al. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chemistry & biology.* 2000;7(2):97-109.

244. Skinnider MA, Dejong CA, Rees PN, et al. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Research.* 2015;43(20):9645-9662.

245. Sun Y-B. FasParser: a package for manipulating sequence data. *Zoological research.* 2017;38(2):110-112.

246. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution.* 1993;10(3):512-526.

247. Williams DR, Patnaik S, Clark MP. Total Synthesis of Cystothiazoles A and C. *The Journal of Organic Chemistry.* 2001;66(25):8463-8469.

248. Owen JG, Charlop-Powers Z, Smith AG, et al. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proceedings of the National Academy of Sciences of the United States of America.* 2015;112(14):4221-4226.

249. Czesny B, Goshu S, Williamson KC, Cook JL. The proteasome inhibitor epoxomicin has potent Plasmodium falciparum gametocytocidal activity. *Antimicrobial Agents and Chemotherapy.* 2009;53(10):4080-4085.

250. Khalil MW, Sasse F, Lunsdorf H, Elnakady YA, Reichenbach H. Mechanism of action of tubulysin, an antimitotic peptide from myxobacteria. *Chembiochem : a European journal of chemical biology.* 2006;7(4):678-683.

251. Young J, Stevens DC, Carmichael R, et al. Elucidation of Gephyronic Acid Biosynthetic Pathway Revealed Unexpected SAM-Dependent Methylations. *Journal of Natural Products.* 2013;76(12):2269-2276.

252. Meiser P, Bode HB, Muller R. The unique DKxanthene secondary metabolite family from the myxobacterium Myxococcus xanthus is required for developmental sporulation. *Proceedings of the National Academy of Sciences of the United States of America.* 2006;103(50):19128-19133.

253. Kunze B, Jansen R, Hofle G, Reichenbach H. Ajudazols, new inhibitors of the mitochondrial electron transport from Chondromyces crocatus. Production, antimicrobial activity and mechanism of action. *The Journal of antibiotics.* 2004;57(2):151-155.

254. Alvarez-Mico X, Jensen PR, Fenical W, Hughes CC. Chlorizidine, a Cytotoxic 5H-Pyrrolo[2,1-a]isoindol-5-one-Containing Alkaloid from a Marine Streptomyces sp. *Organic Letters.* 2013;15(5):988-991.

255. Flatt PM, Wu X, Perry S, Mahmud T. Genetic Insights into Pyralomicin Biosynthesis in Nonomuraea spiralis IMC A-0156. *Journal of Natural Products.* 2013;76(5):939-946.

256. Gaitatzis N, Silakowski B, Kunze B, et al. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *The Journal of biological chemistry.* 2002;277(15):13082-13090.

257. Fujii K, Sivonen K, Kashiwagi T, Hirayama K, Harada K-i. Nostophycin, a Novel Cyclic Peptide from the Toxic Cyanobacterium Nostoc sp. 152. *The Journal of Organic Chemistry.* 1999;64(16):5777-5782.

258. Hoffmann T, Müller S, Nadmid S, Garcia R, Müller R. Microsclerodermins from Terrestrial Myxobacteria: An Intriguing Biosynthesis Likely Connected to a Sponge Symbiont. *Journal of the American Chemical Society.* 2013;135(45):16904-16911.

259. Holmes TC, May AE, Zaleta-Rivera K, et al. Molecular Insights into the Biosynthesis of Guadinomine: A Type III Secretion System Inhibitor. *Journal of the American Chemical Society.* 2012;134(42):17797-17806.

260. Steinmetz H, Irschik H, Kunze B, Reichenbach H, Hofle G, Jansen R. Thuggacins, macrolide antibiotics active against Mycobacterium tuberculosis: isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chemistry (Weinheim an der Bergstrasse, Germany).* 2007;13(20):5822-5832.

261. Huss M, Ingenhorst G, Konig S, et al. Concanamycin A, the specific inhibitor of V-ATPases, binds to the V(o) subunit c. *The Journal of biological chemistry.* 2002;277(43):40544-40548.

262. Liaimer A, Helfrich EJN, Hinrichs K, et al. Nostopeptolide plays a governing role during cellular differentiation of the symbiotic cyanobacterium Nostoc punctiforme. *Proceedings of the National Academy of Sciences.* 2015;112(6):1862-1867.

263. Chang Z, Sitachitta N, Rossi JV, et al. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula. *Journal of natural products.* 2004;67(8):1356-1367.

264. Gisi U, Ziegler H. Fungicides, Phenylamides/Acycalanines. *Encyclopedia of Agrochemicals*: John Wiley & Sons, Inc.; 2003.

265. Trowitzsch-Kienast W, Forche E, Wray V, et al. Antibiotika aus Gleitenden Bakterien, 45. Phenalamide, neue HIV-1-Inhibitoren aus Myxococcus stipitatus Mx s40. *Liebigs Annalen der Chemie.* 1992;1992(7):659-664.

266. Sekiya S, Okumura M, Kubota K, Nakamura T, Sekine D, Hosokawa S. Remote Asymmetric Bromination Reaction with Vinylketene Silyl N,O-Acetal and Its Application to Total Synthesis of Pellasoren A. *Organic Letters.* 2017;19(9):2394-2397.

267. von Jagow G, Ljungdahl PO, Graf P, Ohnishi T, Trumpower BL. An inhibitor of mitochondrial respiration which binds to cytochrome b and displaces quinone from the iron-sulfur protein of the cytochrome bc1 complex. *The Journal of biological chemistry.* 1984;259(10):6318-6326.

268. Lohman JR, Huang SX, Horsman GP, et al. Cloning and sequencing of the kedarcidin biosynthetic gene cluster from Streptoalloteichus sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics. *Mol Biosyst.* 2013;9(3):478-491.

269. Fleta-Soriano E, Smutná K, Martínez JP, et al. The Myxobacterial Metabolite Soraphen A Inhibits HIV-1 by Reducing Virus Production and Altering Virion Composition. *Antimicrobial Agents and Chemotherapy.* 2017;61(8):e00739-00717.

270. Söker U, Sasse F, Kunze B, Höfle G. Synthesis of Melithiazol B and Related Compounds via Oxidative Degradation of Myxothiazol A and Z. *European Journal of Organic Chemistry.* 2000;2000(8):1497-1502.

271. Wen Y, Wu X, Teng Y, et al. Identification and analysis of the gene cluster involved in biosynthesis of paenibactin, a catecholate siderophore produced by Paenibacillus elgii B69. *Environmental microbiology.* 2011;13(10):2726-2737.

272. Herkommer D, Thiede S, Wosniok PR, et al. Stereochemical determination of the leupyrrins and total synthesis of leupyrrin A1. *Journal of the American Chemical Society.* 2015;137(12):4086-4089.

273. Tautz T, Hoffmann J, Hoffmann T, et al. Isolation, Structure Elucidation, Biosynthesis, and Synthesis of Antalid, a Secondary Metabolite from Polyangium species. *Organic Letters.* 2016;18(11):2560-2563.

274. Rellán S, Osswald J, Saker M, Gago-Martinez A, Vasconcelos V. First detection of anatoxin-a in human and animal dietary supplements containing cyanobacteria. *Food and Chemical Toxicology.* 2009;47(9):2189-2195.

275. Hrouzek P, Kuzma M, Černý J, et al. The Cyanobacterial Cyclic Lipopeptides Puwainaphycins F/G Are Inducing Necrosis via Cell Membrane Permeabilization

and Subsequent Unusual Actin Relocalization. *Chemical Research in Toxicology*. 2012;25(6):1203-1211.

276. Edwards DJ, Marquez BL, Nogle LM, et al. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chemistry & biology*. 2004;11(6):817-833.

277. Lee W, Schaefer K, Qiao Y, et al. The Mechanism of Action of Lysobactin. *Journal of the American Chemical Society.* 2016;138(1):100-103.

278. Baldeweg F, Kage H, Schieferdecker S, Allen C, Hoffmeister D, Nett M. Structure of Ralsolamycin, the Interkingdom Morphogen from the Crop Plant Pathogen Ralstonia solanacearum. *Organic letters.* 2017;19(18):4868-4871.

279. Eto K, Yoshino M, Takahashi K, Ishihara J, Hatakeyama S. Total Synthesis of Oxazolomycin A. *Organic Letters.* 2011;13(19):5398-5401.

280. Peypoux F, Besson F, Michel G, Delcambe L. Structure of bacillomycin D, a new antibiotic of the iturin group. *European journal of biochemistry*. 1981;118(2):323-327.

281. Romero D, de Vicente A, Rakotoaly RH, et al. The iturin and fengycin families of lipopeptides are key factors in antagonism of Bacillus subtilis toward Podosphaera fusca. *Molecular plant-microbe interactions : MPMI.* 2007;20(4):430-440.

282. Rex JH, Stevens DA. 39 - Drugs Active against Fungi, Pneumocystis, and Microsporidia. In: Bennett JE, Dolin R, Blaser MJ, eds. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)*. Philadelphia: Content Repository Only!; 2015:479-494.e474.

283. Miyanaga S, Sakurai H, Saiki I, Onaka H, Igarashi Y. Synthesis and evaluation of myxochelin analogues as antimetastatic agents. *Bioorganic & Medicinal Chemistry*. 2009;17(7):2724-2732.

284. Patzer SI, Braun V. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of Streptomyces sp. ATCC 700974. *Journal of bacteriology*. 2010;192(2):426-435.

285. Li K, Chen W-H, Bruner SD. Structure and Mechanism of the Siderophore-Interacting Protein from the Fuscachelin Gene Cluster of Thermobifida fusca. *Biochemistry*. 2015;54(25):3989-4000.

286. Giessen TW, Franke KB, Knappe TA, et al. Isolation, structure elucidation, and biosynthesis of an unusual hydroxamic acid ester-containing siderophore from Actinosynnema mirum. *Journal of natural products.* 2012;75(5):905-914.

287. Bosello M, Zeyadi M, Kraas FI, Linne U, Xie X, Marahiel MA. Structural characterization of the heterobactin siderophores from Rhodococcus erythropolis PR4 and elucidation of their biosynthetic machinery. *Journal of natural products.* 2013;76(12):2282-2290.

288. Sommer PS, Almeida RC, Schneider K, Beil W, Sussmuth RD, Fiedler HP. Nataxazole, a new benzoxazole derivative with antitumor activity produced by Streptomyces sp. Tu 6176. *The Journal of antibiotics.* 2008;61(11):683-686.

289. Lv M, Zhao J, Deng Z, Yu Y. Characterization of the Biosynthetic Gene Cluster for Benzoxazole Antibiotics A33853 Reveals Unusual Assembly Logic. *Chemistry & biology.* 2015;22(10):1313-1324.

290. Wu X, Flatt PM, Xu H, Mahmud T. Biosynthetic gene cluster of cetoniacytone A, an unusual aminocyclitol from the endosymbiotic Bacterium Actinomyces sp. Lu 9419. *Chembiochem : a European journal of chemical biology.* 2009;10(2):304-314.

291. Marquez BL, Watts KS, Yokochi A, et al. Structure and Absolute Stereochemistry of Hectochlorin, a Potent Stimulator of Actin Assembly. *Journal of Natural Products.* 2002;65(6):866-871.

292. van der Merwe D. Chapter 31 - Freshwater cyanotoxins. In: Gupta RC, ed. *Biomarkers in Toxicology*. Boston: Academic Press; 2014:539-548.

293. Reinert T, #xe1, Baldotto CSdR, Nunes FAP, Scheliga AAdS. Bleomycin-Induced Lung Injury. *Journal of Cancer Research.* 2013;2013:9.

294. Dineshkumar K, Aparna V, Madhuri KZ, Hopper W. Biological activity of sporolides A and B from Salinispora tropica: in silico target prediction using ligand-based pharmacophore mapping and in vitro activity validation on HIV-1 reverse transcriptase. *Chemical biology & drug design.* 2014;83(3):350-361.

295. Hoskins JW, Ofori LO, Chen CZ, et al. Lomofungin and dilomofungin: inhibitors of MBNL1-CUG RNA binding with distinct cellular effects. *Nucleic acids research.* 2014;42(10):6591-6602.

296. Hassan HM, Fridovich I. Mechanism of the antibiotic action pyocyanine. *Journal of bacteriology.* 1980;141(1):156-163.

297. Krastel P, Zeeck A, Gebhardt K, Fiedler HP, Rheinheimer J. Endophenazines A-D, new phenazine antibiotics from the athropod associated endosymbiont Streptomyces anulatus II. Structure elucidation. *The Journal of antibiotics.* 2002;55(9):801-806.

298. Rui Z, Ye M, Wang S, et al. Insights into a Divergent Phenazine Biosynthetic Pathway Governed by a Plasmid-Born Esmeraldin Gene Cluster. *Chemistry & Biology.* 2012;19(9):1116-1125.

299. Zhu Y, Zhang Q, Li S, et al. Insights into Caerulomycin A Biosynthesis: A Two-Component Monooxygenase CrmH-Catalyzed Oxime Formation. *Journal of the American Chemical Society.* 2013;135(50):18750-18753.

300. Leet JE, Schroeder DR, Hofstead SJ, et al. Kedarcidin, a new chromoprotein antitumor antibiotic: structure elucidation of kedarcidin chromophore. *Journal of the American Chemical Society.* 1992;114(20):7946-7948.

301. Sanchez LM, Dorrestein PC. Analytical chemistry: Virulence caught green-handed. *Nature chemistry.* 2013;5(3):155-157.

302. Miyaki T, Numata K, Nishiyama Y, et al. Tallysomycin, a new antitumor antibiotic complex related to bleomycin. V. Production, characterization and antitumor activity of tallysomycin S10b, a new biosynthetic tallysomycin derivative. *The Journal of antibiotics.* 1981;34(6):665-674.

303. Burlison JA, Blagg BSJ. Synthesis and Evaluation of Coumermycin A1 Analogues that Inhibit the Hsp90 Protein Folding Machinery. *Organic Letters.* 2006;8(21):4855-4858.

304. Eustáquio AS, Gust B, Luft T, Li S-M, Chater KF, Heide L. Clorobiocin Biosynthesis in Streptomyces: Identification of the Halogenase and Generation of Structural Analogs. *Chemistry & Biology.* 2003;10(3):279-288.

305. Gebhardt K, Meyer SW, Schinko J, Bringmann G, Zeeck A, Fiedler HP. Phenalinolactones A-D, terpenoglycoside antibiotics from Streptomyces sp. Tu 6071. *The Journal of antibiotics.* 2011;64(3):229-232.

306. Ding W, Lei C, He Q, Zhang Q, Bi Y, Liu W. Insights into bacterial 6-methylsalicylic acid synthase and its engineering to orsellinic acid synthase for spirotetronate generation. *Chemistry & biology.* 2010;17(5):495-503.

307. Etzbach L, Plaza A, Garcia R, Baumann S, Muller R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Organic letters.* 2014;16(9):2414-2417.

308. Boettger D, Hertweck C. Molecular diversity sculpted by fungal PKS-NRPS hybrids. *Chembiochem : a European journal of chemical biology.* 2013;14(1):28-42.

309. Vicente CM, Thibessard A, Lorenzi JN, et al. Comparative Genomics among Closely Related Streptomyces Strains Revealed Specialized Metabolite Biosynthetic Gene Cluster Diversity. *Antibiotics (Basel, Switzerland).* 2018;7(4).

310. Naughton LM, Romano S, O'Gara F, Dobson ADW. Identification of Secondary Metabolite Gene Clusters in the Pseudovibrio Genus Reveals Encouraging Biosynthetic Potential toward the Production of Novel Bioactive Compounds. *Frontiers in Microbiology.* 2017;8(1494).

311. Kundim BA, Itou Y, Sakagami Y, Fudou R, Yamanaka S, Ojika M. Novel antifungal polyene amides from the myxobacterium Cystobacter fuscus: isolation, antifungal activity and absolute structure determination. *Tetrahedron.* 2004;60(45):10217-10221.

312. Feng Z, Qi J, Tsuge T, Oba Y, Sakagami Y, Ojika M. Biosynthesis of 2'-O-methylmyxalamide D in the myxobacterium Cystobacter fuscus: a polyketide synthase-nonribosomal peptide synthetase system for the myxalamide D skeleton and a methyltransferase for the final O-methylation. *Bioscience, biotechnology, and biochemistry.* 2006;70(3):699-705.

313. Park S, Hyun H, Lee JS, Cho K. Identification of the Phenalamide Biosynthetic Gene Cluster in Myxococcus stipitatus DSM 14675. *Journal of microbiology and biotechnology.* 2016;26(9):1636-1642.

314. Buntin K, Weissman KJ, Muller R. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis. *Chembiochem : a European journal of chemical biology.* 2010;11(8):1137-1146.

315. Beyer S, Kunze B, Silakowski B, Muller R. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of Stigmatella aurantiaca Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain Sorangium cellulosum So ce90. *Biochimica et biophysica acta.* 1999;1445(2):185-195.

316. Feng Z, Qi J, Tsuge T, et al. Construction of a bacterial artificial chromosome library for a myxobacterium of the genus Cystobacter and characterization of an antibiotic biosynthetic gene cluster. *Bioscience, biotechnology, and biochemistry.* 2005;69(7):1372-1380.

317. Müller S, Rachid S, Hoffmann T, et al. Biosynthesis of Crocacin Involves an Unusual Hydrolytic Release Domain Showing Similarity to Condensation Domains. *Chemistry & Biology.* 2014;21(7):855-865.

318. Rachid S, Krug D, Kunze B, et al. Molecular and Biochemical Studies of Chondramide Formation—Highly Cytotoxic Natural Products from Chondromyces crocatus Cm c5. *Chemistry & Biology.* 2006;13(6):667-681.

319. Jansen R, Kunze B, Reichenbach H, Höfle G. Chondrochloren A and B, New β-Amino Styrenes from Chondromyces crocatus (Myxobacteria). *European Journal of Organic Chemistry*. 2003;2003(14):2684-2689.

320. Cortina NS, Krug D, Plaza A, Revermann O, Müller R. Myxoprincomide: A Natural Product from Myxococcus xanthus Discovered by Comprehensive Analysis of the Secondary Metabolome. *Angewandte Chemie International Edition.* 2012;51(3):811-816.

321. Deng Q, Zhou L, Luo M, Deng Z, Zhao C. Heterologous expression of Avermectins biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers. *Synthetic and Systems Biotechnology*. 2017;2(1):59-64.

322. Johnson LJ, Koulman A, Christensen M, et al. An extracellular siderophore is required to maintain the mutualistic interaction of Epichloe festucae with Lolium perenne. *PLoS pathogens.* 2013;9(5):e1003332.