Characterization of an orphan specialized metabolite biosynthetic gene cluster in the potato common scab pathogen *Streptomyces scabiei*

by

© Jingyu Liu

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

March 2021

St. John's Newfoundland and Labrador

ABSTRACT

Streptomyces scabiei is the main causative agent of potato common scab (CS) disease, which causes significant economic losses to potato growers worldwide. The ability of *S. scabiei* to cause CS is primarily based on the production of a phytotoxic specialized metabolite called thaxtomin A, which is an essential virulence factor for the organism. In addition, *S. scabiei* has the genetic potential to produce other specialized metabolites that might contribute to CS disease development; however, most of the genes predicted to be involved in production of these specialized metabolites are silent under laboratory conditions, and thus there is little known regarding the nature of these metabolites and their role in CS disease. The aim of this research was to characterize one cryptic specialized metabolite that is thought to resemble a phenazine. The putative biosynthetic genes that produce this cryptic metabolite are expressed at low levels or not at all under laboratory conditions, and therefore different strategies were employed to activate or enhance expression of these genes in *S. scabiei*.

Specifically, the first approach taken was to use different bioinformatics tools to predict the boundaries of the biosynthetic gene cluster (BGC), and it is proposed that the BGC is much larger than initially thought. Molecular and culture-based strategies were successfully employed to activate expression of the BGC, and this also enabled a better understanding of the role of predicted regulatory genes in controlling the expression of the BGC. Metabolomics analysis conducted in this study suggests that the BGC is involved in the production of phenazine-related metabolites. This analysis additionally provided new insights into the overall specialized metabolic potential of *S. scabiei* and enabled the prediction of several novel compounds that were not previously known to be produced by this organism. The phenazine-like BGC is conserved in the genomes of other Actinobacteria, including two phytopathogenic *Streptomyces* spp. Although the

role of the resulting metabolite(s) in the pathogenicity of *S. scabiei* remains unclear, evidence is presented in this thesis suggesting that the molecule(s) may exhibit antibacterial activity, and thus may have a function in inter-microbial interactions as described for other phenazine-producing organisms.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr. Dawn Bignell, for giving me the opportunity to work in her lab and for all her help and guidance throughout the whole period of my Ph.D. study. In particular, I thank her for providing the thorough suggestions for my thesis; I thank her for attending my wedding and for her generosity; I thank her for buying gifts for my baby David (especially the baby clothes she bought, which have newborn size, three months size, six months size, nine months size and 12 months size, so, I would not have to to buy baby clothes until he is one year old!); I thank her for her help and support during my pregnancy and after the delivery of my baby. Without her help and support, I could not have made it through all these challenging experiences.

I would also like to thank my committee members Dr. Kapil Tahlan and Dr. Linda Jewell, for all the suggestions they have provided me during this project. I also want to thank Dr. Mervyn Bibb for providing the *Streptomyces coelicolor* M1154 strain, Dr. Gerard Wright for providing the *Streptomyces albus* J1074 strain, and Dr. Kapil Tahlan for providing the *Streptomyces avermitilis* SUK17 strain. I would like to thank the Dorrestein lab and Louis-Felix Nothias in particular for conducting the LC-MS^2 analysis and for assisting with the data analyses.

I would like to thank a former lab member of the Bignell lab, Zhenlong (Lancy) Cheng, for guiding me during the first year of my Ph.D. I also want to thank a current lab member, Yuting (Phoebe) Li, for her help on my project and my life. I would like to thank Arshad Ali Shaikh and Brandon Piercey from the Tahlan lab, who helped me with metabolomics analysis and bioinformatics analysis. I would also like to thank my husband Xin Lin for assistance with the DeepBGC and BioDendro analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

Δ: deletion

λ: lambda bacteriophage

aa: amino acids

ABC: ATP-binding cassette

amp^r: ampicillin resistance

antiSMASH: antibiotics & Secondary Metabolite Analysis Shell

apra^r: apramycin resistance

ATP: adenosine triphosphate

BGC: biosynthetic gene cluster

BiG-SCAPE: biosynthetic gene similarity clustering and prospecting engine

BLAST: basic local alignment search tool

bp: base pair

cDNA: complementary DNA

CFA: coronafacic acid

CFA-*a*Ile: coronafacoyl-L-*allo*-isoleucine

CFA-Ile: coronafacoyl-L-isoleucine

CFA-Val: *N*-coronafacoyl-L-valine

CFM-ID: competitive fragmentation modeling for metabolite identification

CoA: coenzyme A

CORASON: core analysis of syntenic orthologues to prioritize natural product gene clusters

CS: common scab

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

egfp: enhanced green fluorescent protein gene

ESI: electrospray ionization

EtBr: ethidium bromide

FBMN: feature-based molecular networking

GBL: γ-butyrolactone

GC content: guanine-cytosine content

GCFs: gene cluster families

GNPS: global natural product social molecular networking

HRESIMS: high resolution electrospray ionization mass spectra

hyg^r: hygromycin B resistance

IPTG: isopropyl β-D-thiogalactopyranoside

ISP-4: international streptomyces project medium 4

ISR: induced systemic resistance

JA: jasmonic acid

JA-Ile: jasmonoyl-L-isoleucine

kan^r: kanamycin resistance

LB: luria-bertani

LC-MS^2: liquid chromatography–mass spectrometry–mass spectrometry

m/z: mass-to-charge ratio

MFs: molecular families

MIBiG: Minimum Information about a Biosynthetic Gene cluster

MYMm: Modified Maltose-Yeast Extract-Malt Extract agar

NA: nutrient agar

NAP: network annotation propagation

NCBI: national center for biotechnology information

NRP: nonribosomal peptide

NRPS: nonribosomal peptide synthetase

OBA: oat bran agar

OD: optical density

oriT: origin of transfer

PCR: polymerase chain reaction

PK: polyketide

PKS: polyketide synthase

PMA: potato mash agar

RiPP: ribosomally-synthesized and post-translationally modified peptides

RNA: ribonucleic acid

rpm: revolutions per minute

RT: reverse transcriptase or retention time

RT-PCR: reverse transcriptase PCR

SA: starch asparagine agar

SFM: soy flour mannitol agar

SMART: simple modular architecture research tool

SOB: super optimal broth

SOC: super optimal broth with catabolite repression

TBE: tris-borate-EDTA

TSB: trypticase soy broth

UV: ultraviolet

v/v: volume/volume

w/v: weight/volume

WT: wild type

YMSm: modified yeast extract -malt extract broth -soluble starch agar

YT: yesast extract-tryptone

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 General features of *Streptomyces*

Streptomyces are Gram-positive, filamentous Actinobacteria that are ubiquitous in soil (Harir et al. 2018) and can also be found in marine environments (Ward and Bora 2006). They are primarily saprophytes that can secrete various enzymes to decompose recalcitrant biological polymers (Hodgson 2000), such as starch, lignin and cellulose (Schlatter et al. 2009), and thus they play an important role in the recycling of nutrients in the environment (Hodgson 2000). *Streptomyces* species have a complex morphological life cycle that begins with a single spore, germinating under favorable conditions to form long, filamentous cells called hyphae (Figure 1.1) (Seipke et al. 2012). During vegetative growth, the hyphae elongate by tip extension and branch to form a complex, dense network called the substrate (or vegetative) mycelium (McCormick and Flärdh 2012). Then, in response to environmental stresses, for example nutrient limitation, *Streptomyces* switch from vegetative growth to the production of aerial hyphae and spores. At the same time, many specialized metabolites such as antibiotics are produced during this transition phase (Elliot et al. 2007).

Streptomyces are unusual among bacteria in that their chromosome is linear and exceptionally large, generally 8 Mb to 11 Mb, with more than 70% G+C content (Lin et al. 1993; Ventura et al. 2007). The chromosome is usually organized in a core/arm structure, where the core region (~6 Mb) contains unconditionally essential genes, and the two arm regions at both ends mainly carry conditionally adaptive genes (Bentley et al. 2002). The genomes of these organisms can also include large plasmids that are linear or circular in structure (Kinashi 2011), and it has been suggested that the exceptionally large genomes of *Streptomyces* are one of the reasons why

these organisms have a complex life cycle and can survive in variable soil conditions (Chen et al. 2002).

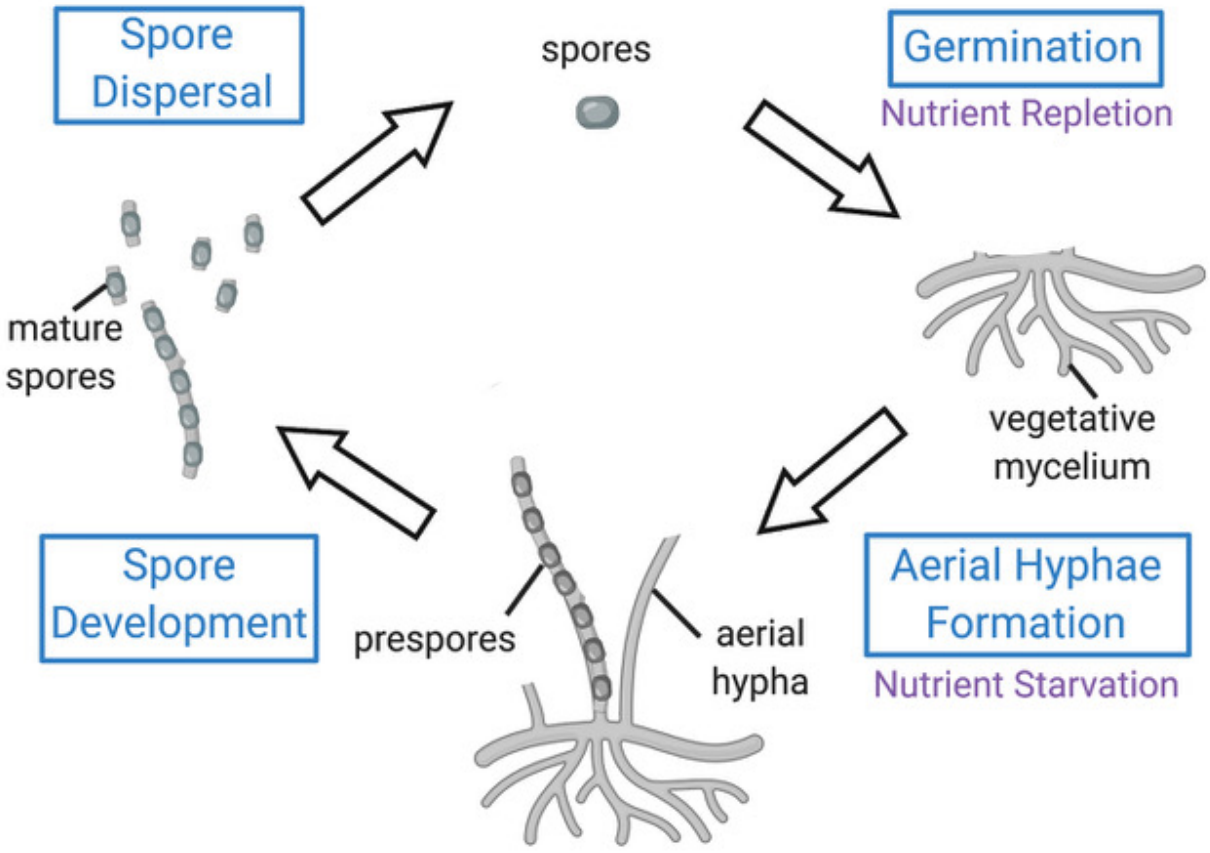


Figure 1.1 The life cycle of *Streptomyces* spp. Image provided by Yuting Li (Memorial University).

1.2 Secondary or specialized metabolites

The name "secondary metabolite" comes from the observation that these compounds are not essential for growth, development, and reproduction of the producing organism under laboratory conditions, in contrast to primary metabolites (e.g. amino acids, nucleic acids, lipids, and carbohydrates) (Mosunova et al. 2020). More recently, these molecules have been described as "specialized metabolites" since they are thought to play important roles in the survival of

producing organisms in their natural environments (Mosunova et al. 2020). Specialized metabolites are small organic natural products with varied biological functions and chemical structures. According to their chemical structure, function and biosynthesis, specialized metabolites can be classified into six main classes: fatty acid-derived substances and polyketides, phenylpropanoids, terpenoids and steroids, alkaloids, specialized amino acids and peptides, and specialized carbohydrates (Hanson 2003). Specialized metabolites are produced mainly by bacteria, fungi and plants (Thirumurugan et al. 2018), with microorganisms having a filamentous morphology and a complex life cycle being the most prolific producers. For example, about two-thirds of all known antibiotics are made by actinomycetes, particularly *Streptomyces* species (Takahashi and Nakashima 2018). *Streptomyces* are well-known for their ability to produce a diverse range of specialized metabolites that have useful applications in human medicine, including antibiotics, anti-cancer agents and immunosuppressants (Newman and Cragg 2007). In addition, many *Streptomyces* specialized metabolites exhibit herbicidal, insecticidal and fungicidal activities, and thus have useful applications in agriculture (Bérdy 2005; Sadeghi et al. 2012).

1.2.1 Metabolic pathways of specialized metabolite production in bacteria

Bacterial specialized metabolites are usually synthesized at the end of the exponential phase and at the beginning of the stationary phases of bacterial growth (idiophase) (Figure 1.2) (Gokulan et al. 2014). They are produced in response to environmental stress, limited nutrients and growth conditions (Nigam and Singh 2014). The metabolic pathways for biosynthesis of specialized metabolites start from primary metabolism (Andryukov et al. 2019). Among the key pathways responsible for the production of specialized metabolites with antibacterial activities, the best characterized are polyketide, nonribosomal peptide, hybrid (nonribosomal peptide-

polyketide), β-lactam, oligosaccharide and the pathways using shikimic acid (Figure 1.2) (Gokulan et al. 2014; Andryukov et al. 2019).

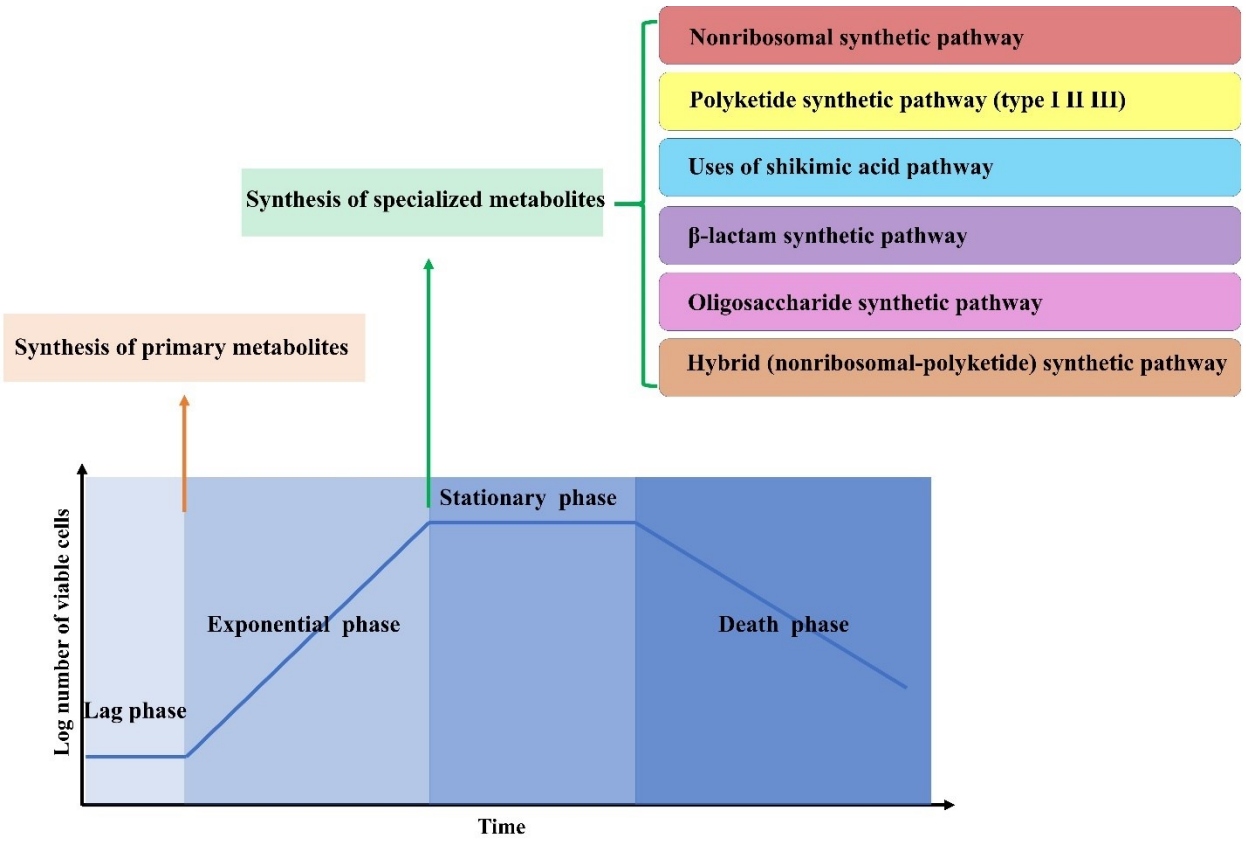


Figure 1.2 Phases of bacterial growth and various specialized metabolite biosynthetic pathways. The primary metabolites are often produced at the late lag phase and during the exponential phase. Most specialized metabolites are produced at the end of the exponential phase and at the beginning of the stationary phases of bacterial growth. The figure is based on that from (Harir et al. 2018).

1.2.1.1 Nonribosomal peptide biosynthesis pathways

Nonribosomal peptides (NRPs), such as bacitracin, polymyxin B and vancomycin, are a class of peptide specialized metabolites, which are synthesized by large multienzyme complexes called nonribosomal peptide synthetases (NRPSs) (Gokulan et al. 2014; Martínez-Núñez and López 2016). NRPS enzymes are composed of different modules called the initiation, elongation and termination modules (Figure 1.3), each of which carries out activation and incorporation of one amino acid into the growing peptide chain (Süssmuth and Mainz, 2017). The modules themselves are comprised of distinct catalytic domains, three of which are the core domains required for peptide biosynthesis: the adenylation domain (A), the peptidyl carrier protein (PCP) domain, and the condensation (C) domain, with A and PCP being essential to every module (Figure 1.3) (Desriac et al. 2013). The A domain is responsible for the selection and activation of an amino acid substrate. Once activated, the A domain loads the amino acid onto the 4′-phosphopantetheine prosthetic group of the PCP domain (also known as the thiolation domain). The tethered amino acid is then brought to the C domain, which catalyzes the formation of a peptide bond between the activated amino acid and the growing peptide chain attached to the upstream module. Optional catalytic domains that can be present within a module include epimerization (E), methylation (M), reduction (R), oxidation (Ox), formylation (F) and heterocyclization (Cy) domains, which catalyze modifications in the amino acid substrate and contribute to the large structural diversity of NRP products. The final product is then released by a thioesterase (TE) domain within the termination module, which can also mediate macrocyclization of the product during the release step (Süssmuth and Mainz 2017). In addition to the NRPS genes, BGCs that are responsible for producing NRPs often harbour genes involved in the synthesis of the building blocks, modification of the peptide product, the export of the product, self-resistance, and gene regulation (Süssmuth and Mainz 2017).

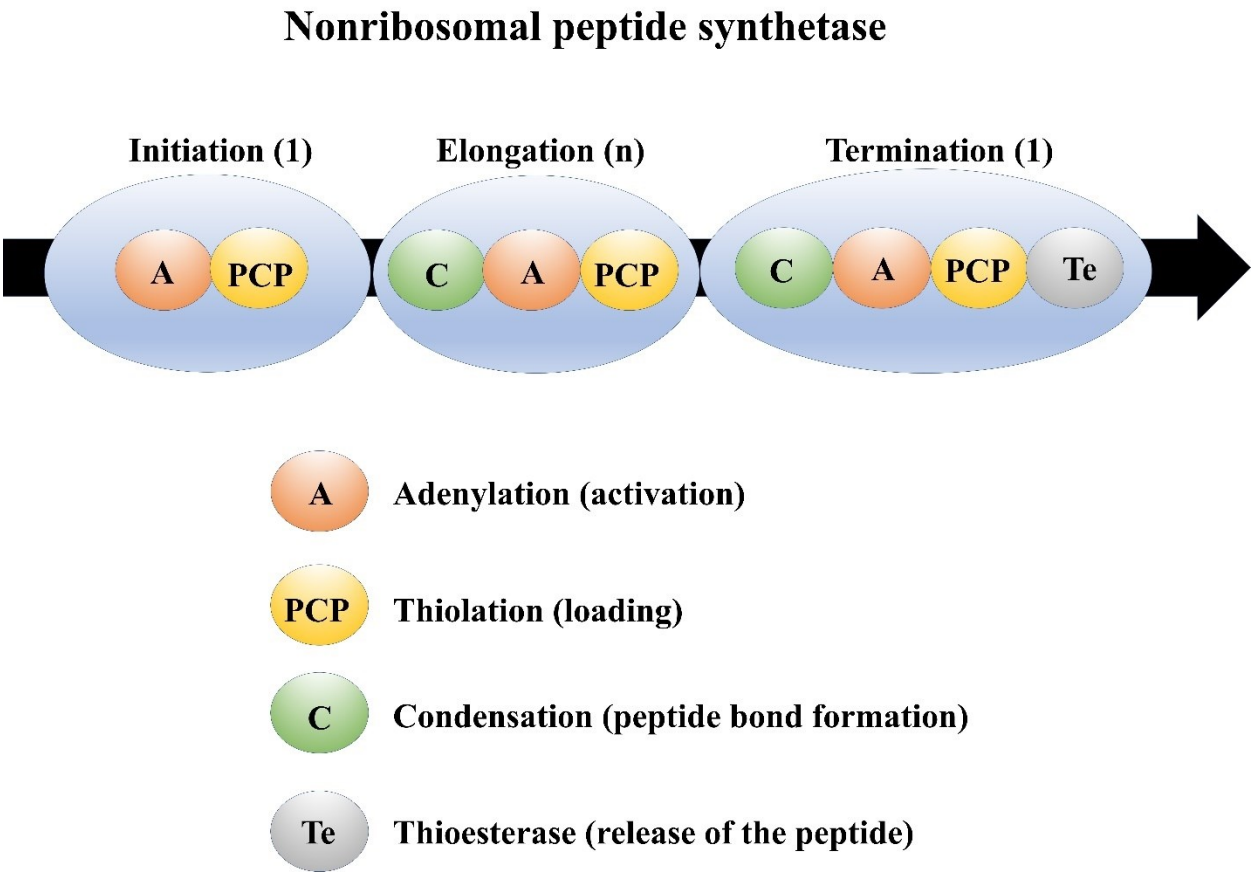


Figure 1.3 Minimal module and domain structure of nonribosomal peptide synthetases (NRPSs). A given NRPS will harbour one initiation module, one or more (n) elongation modules, and one termination module.

1.2.1.2 Polyketide biosynthesis pathways

Polyketide (PK) compounds, including erythromycin, tetracycline and avermectin, are synthesized from simple acyl precursors such as propionyl CoA, acetyl CoA, and methylmalonyl CoA (Thirumurugan et al. 2018). PKs are assembled by polyketide synthase (PKS) enzymes, which can be divided into three types: I - III, based on the structure and enzymatic mechanism (Gokulan et al. 2014). In addition to PKS genes, PK-associated BGCs also contain other genes involved in starter or extender unit synthesis, modification of the PK, export, self-resistance and regulation (Cummings et al. 2014).

There are two groups of Type I PKS: modular and iterative. Iterative Type I PKSs (e.g. PksA from the fungal genus *Aspergillus*, initiating aflatoxin B1 biosynthesis) utilize a single module iteratively to build the polyketide backbone, whereas the modular type I PKSs (e.g. 6-deoxyerythronolide B synthase from the bacterium *Saccharopolyspora erythraea*, producing the 6-DEB as precursor for the erythromycin) contain a sequence of separate modules, and each module has a particular function (Keatinge-Clay 2016; Grininger 2020). Bacterial type I PKSs use various starter units (often malonyl-CoA), which are the initiating precursors for polyketide synthesis in the assembly of their products (Moore and Hertweck 2002). They typically employ malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA or methoxymalonly-CoA as extender units (Chan et al. 2009). The modules responsible for the incorporation of extender units typically consist of three core domains: an acyl carrier protein (ACP) domain, a ketosynthase (KS) domain and an acyltransferase (AT) domain (Figure 1.4) (Guzmán-Chávez et al. 2018). Modification of the intermediate may require other optional domains, such as a dehydratase (DH), ketoreductase (KR), and/or enoyl reductase (ER), which result in different levels of reduction of the extender unit following incorporation into the growing PK backbone. This contributes to the structural diversity of PK molecules (Guzmán-Chávez et al. 2018). The final polyketide chain is released from the PKS by a thioesterase (TE) domain, which can also mediate hydrolysis and macrocyclization of the product to yield either a linear or a cyclized product (Hertweck 2009). Type II PKSs or discrete PKSs (e.g. tetracenomycin PKS for tetracenomycin C biosynthesis) consist of a series of discrete catalytic domains (Shen 2003), employ an ACP domain to transfer acyl-CoA substrates and use malonyl-CoA as an extender unit. Type III PKSs (e.g. RppA synthase for the biosynthesis of flavolin) catalyze various biochemical reactions within a single active site

(Moore and Hopke 2001). They function without ACP and they use malonyl-CoA or methylmalonyl-CoA as an extender unit (Harir et al. 2018; Nivina et al. 2019).

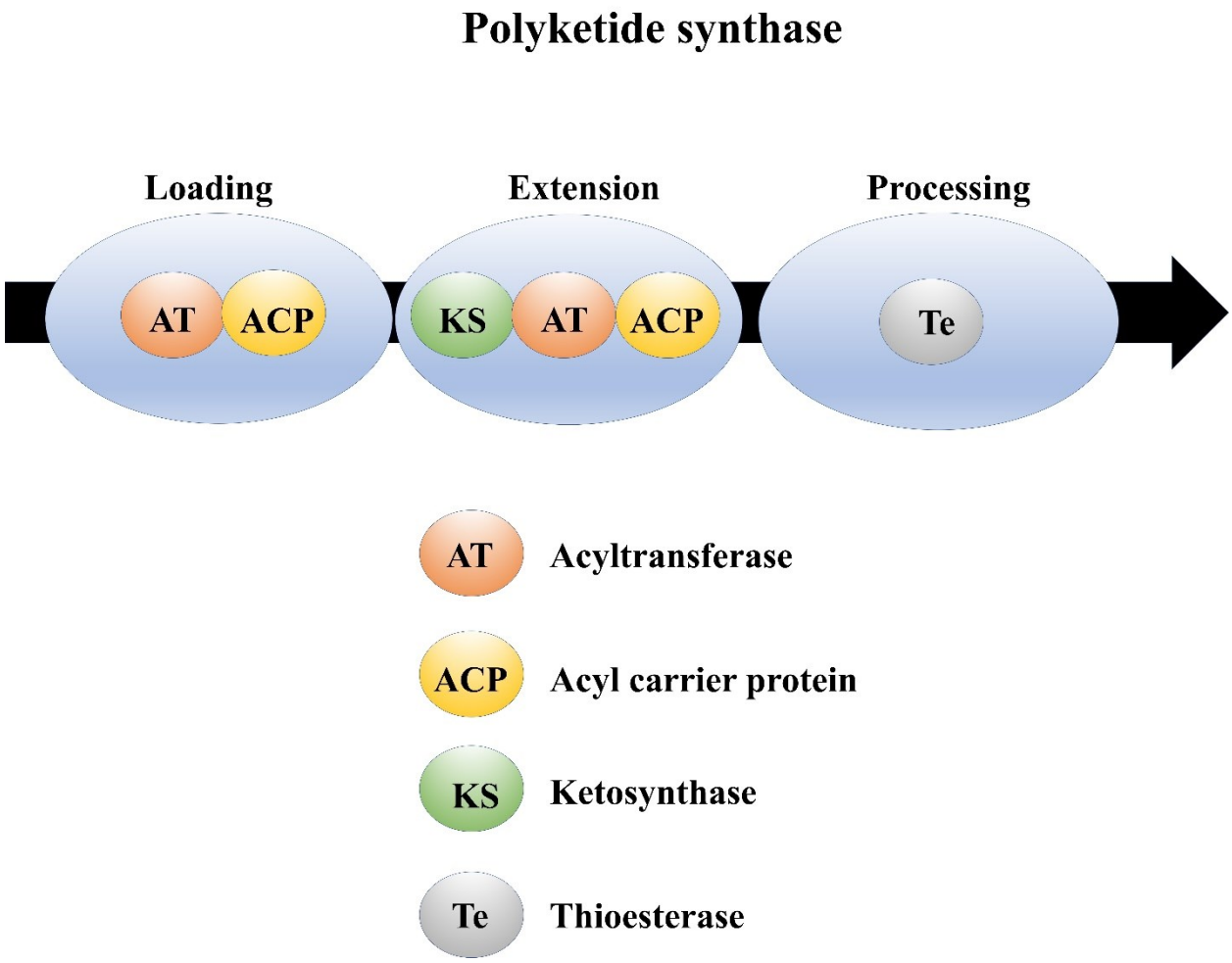


Figure 1.4 Minimal domain structure of a type I polyketide synthase (PKS). A given PKS will consist of one loading module, one or more (n) extension modules, and one processing module.

1.2.1.3 Shikimate and phenazine biosynthetic pathways

The shikimate biosynthesis pathway is a primary metabolic pathway ubiquitous in plants, algae, fungi, and bacteria, but not in animals (Francenia Santos-Sánchez et al. 2019). This pathway is composed of seven sequential enzymatic steps that convert the primary metabolites phosphoenolpyruvic acid (PEP) and D-erythrose-4-phosphate (E4P) to chorismate (CHR), the precursor for the aromatic amino acids phenylalanine, tyrosine and tryptophan, and for *p*-amino and *p*-hydroxy benzoate (Herrmann 1995). Various aromatic compounds with diverse biological activities can be biosynthesized using the intermediates and derivatives from this pathway. Engineering the shikimate pathway for enhancing the supply of precursor molecules can increase the production of various specialized molecules, such as salicylic acid, alkaloids, flavonoids, coumarin, violacein, and salvianic acid A (Figure 1.5) (Jiang and Zhang 2016). In addition, the biosynthesis of phenazine compounds by *Pseudomonas*, *Burkholderia*, *Brevibacterium*, and *Streptomyces* species also depends on the aromatic building blocks derived from the shikimate pathway, with chorismic acid as the most proximate branch point intermediate (Figure 1.5) (Laursen and Nielsen 2004). As the aim of this thesis was to characterize a cryptic specialized metabolite that is thought to resemble a phenazine, the biosynthesis of phenazines is reviewed in detail in the following paragraphs.

The biosynthesis of the tricyclic phenazines remained unclear until the 1990s, when the genome sequencing of *Pseudomonas aureofaciens* (Pierson III and Thomashow 1992; Pierson et al. 1995) and *Pseudomonas fluorescens* (Mavrodi et al. 1997) were completed. This showed that the phenazine gene cluster in each species contains seven genes, *phzABCDEFG*, which form a defined operon (Mavrodi et al. 1998). The operon is present in the genome of *P. aeruginosa* in two copies, and the expression of both copies of the operon is differentially regulated (Mavrodi et al.

2001). It has been reported that *phzC-G* are essential for phenazine biosynthesis, whereas *phzA* and *phzB* are important but non-essential for the biosynthesis of phenazines (McDonald et al. 2001; Guo et al. 2017). McDonald and colleagues found that the yield of phenazine-1-carboxylic acid (PCA) decreases 4-fold due to the absence of *phzA*, while the yield of PCA decreases 8-fold when both *phzA* and *phzB* are absent (McDonald et al. 2001). Guo and colleagues demonstrated that the production of PCA and phenazine-1-carboxamide (PCN) decreases 2-fold if *phzA* is disrupted (Guo et al. 2017). Ahuja et al. proposed that PhzA/B form a dimer that catalyzes a double condensation reaction between two molecules of 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid (Ahuja et al. 2008), but their roles need to be confirmed by further experiments.

The functions of PhzC, PhzD and PhzE are well established by experimental studies (Figure 1.5). PhzC, a 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase, catalyzes the first step of the shikimate pathway, redirecting intermediates from primary metabolism into phenazine biosynthesis (Mavrodi et al. 2006; Gross and Loper 2009). The chorismic acid produced in the process is then transformed into 2-amino-2-deoxy- isochorismic acid (ADIC) by PhzE, an anthranilate synthase homologue (McDonald et al. 2001). The isochorismatase PhzD is responsible for the conversion of ADIC to *trans*-2,3-dihydro-3-hydroxyanthranilic acid (DHHA) (Parsons et al. 2003). Presumably, the isomerase PhzF transforms DHHA to 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid by performing a [1,5]-prototropic shift (Blankenfeldt et al. 2004). PhzG has been suggested to be responsible for the final aromatization of phenazines (Parsons et al. 2004). In many strains, the phenazine BGCs also contain additional genes, such as *phzM, phzH*, *phzS* and *phz*O, which encode enzymes involved in phenazine decoration (Chin-A-Woeng et al. 2001; Delaney et al. 2001; Mavrodi et al. 2001; Parsons et al. 2007). The genes can be located near the core operon or somewhere else in the genome.

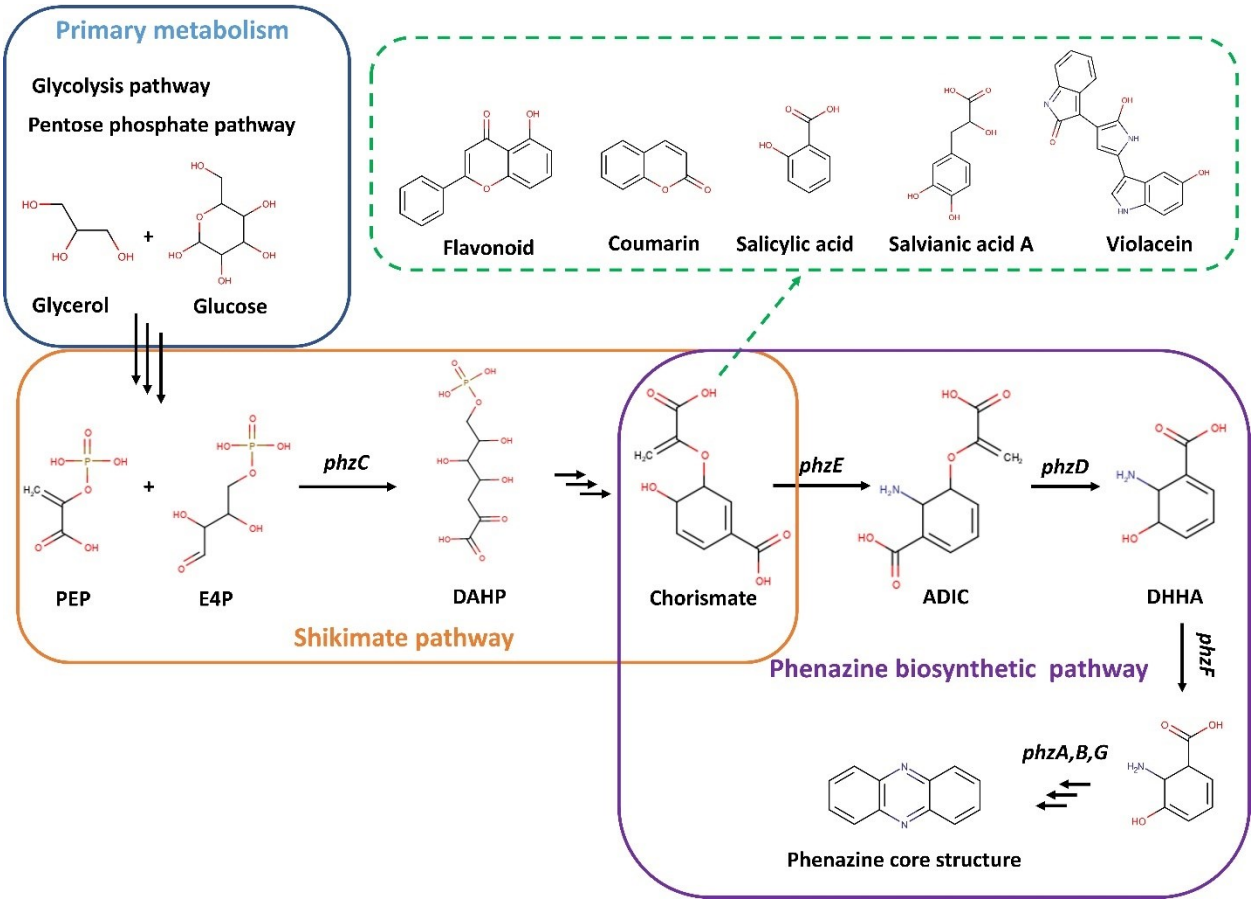


Figure 1.5 Shikimate and phenazine biosynthetic pathways for production of chorismate derivatives. PEP: phosphoenolpyruvate; E4P: D-erythrose 4-phosphate; DAHP: 3-deoxy-D-arabino-heptulosonate-7-phosphate; ADIC: 2-amino-2-desoxy-isochorismic acid; DHHA: *trans*-2,3-dihydro-3-hydroxyanthranilic acid.

1.2.2 Roles of specialized metabolites

From a broad perspective, the roles of specialized metabolites for the producing microorganism can be divided into two main categories: 1) providing protection from competition or predation; 2) improving the microbe's ability to grow, reproduce or disperse in a suitable environment. Most specialized metabolites fit into these two categories (Vining 1990).

Within natural environments that are nutrient limiting, specialized metabolites of particular relevance are those that can kill or slow the growth of competing microorganisms. For example,

antibiotics are thought to be important factors in the competition for space and resources (Hibbing et al. 2010). The role of antibiotic production in competition has been studied extensively. When antibiotic-producing marine bacteria and other non-producers were grown in seawater mixed cultures, the producer always became dominant, whereas the non-producer was competed out (Lemos et al. 1991). Ruiz-Barba et al. demonstrated that the bacteriocin-producing *Lactobacillus plantarum* LPCO10 strain could proliferate to dominate the microflora in natural Spanish-style green olive fermentations during the 12-week process, while the non-bacteriocin-producing *L. plantarum* 55-1 failed to persist after seven weeks (Ruiz-Barba et al. 1994). The above studies suggest that antibiotic production aids in the survival of the producing organisms by killing or inhibiting the growth of others.

Specialized metabolites produced by microorganisms are also thought to mediate symbiotic relationships with eukaryotes, including plants. For example, subinhibitory concentrations of 2,4-diacetylphloroglucinol (DAPG) produced by *P. fluorescens* triggers salicylic acid and ethylene signalling-dependent induced systemic resistance (ISR) against fungal and bacterial pathogens in *Arabidopsis* (Iavicoli et al. 2003; Weller et al. 2012). Plants can also benefit indirectly from microbial specialized metabolites that inhibit the growth of plant pathogens (Ortíz-Castro et al. 2009). On the contrary, some specialized metabolites produced by microbes can have negative effects on plants. For example, phytotoxins such as the thaxtomins produced by plant pathogenic *Streptomyces* species, play an essential role in the development of CS, acid scab and soil rot of sweet potato (Li et al. 2019b). In addition, many plant pathogens, including *Agrobacterium tumefaciens*, *Erwinia* spp. and *P. savastanoi*, produce phytohormones such as auxins, which induce tumor and gall formation in plant tissues (Lambrecht et al. 2000; Chalupowicz et al. 2006; Robert-seilaniantz et al. 2007).

Some specialized metabolites, such as siderophores and ionophores, act as metal transport agents, conferring a growth advantage for the producing organism under iron-limiting conditions (Kramer et al. 2020). Siderophores are low molecular weight, iron-chelating compounds produced by bacteria and fungi (Neilands 1995). Iron exists predominantly in the ferric form (Fe^{3+}) under aerobic conditions, and it cannot be utilized by microorganisms due to the very low solubility of ferric ion in water at neutral pH (Kramer et al. 2020). Siderophores can chelate low abundance ferric ion to facilitate its transport into the cells (Guerinot 1994; Kramer et al. 2020). Ionophores, also called ion carriers, act by enhancing the ion permeability of the cell membranes (Senges et al. 2018). Iron-transport factors are sometimes considered agents of competition because these compounds can starve other species of iron when such species lack the ability to take up the siderophore-Fe^{3+} complexes (Guerinot 1994; Kramer et al. 2020).

Autoregulators, such as the γ-butyrolactones (GBLs), are specialized metabolites that serve as chemical signals regulating antibiotic production, differentiation, and sporulation in *Streptomyces* (Nodwell 2014). GBLs have a fully saturated five-membered heterocyclic ring containing four carbons and one oxygen, and they are widely distributed among streptomycetes (Willey and Gaskell 2011; Sidda and Corre 2012). So far, 14 different GBLs are known, including A-factor from *S. griseus*, VBs from *S. viriginae*, SCBs from *S. coelicolor,* IM-2 from *S. lavendulae,* and factor 1 from *S. viridochromogens*, all of which share a core 2,3-disubstituted GBL skeleton and differ in their fatty acid side chain (Niu et al. 2016). The effects of GBLs are transmitted via specific binding to their receptor proteins, for example, ScbR and ScbR2, and CprA and CprB in *S. coelicolor*; BarA and BarB in *S. virginiae*; and JadR2 and JadR3 in *S. venezuelae* (Cuthbertson and Nodwell 2013; Niu et al. 2016). These receptor proteins usually belong to the TetR family of transcriptional regulators and can control the expression of the cluster-situated regulators (CSRs)

and pleiotropic regulators (e.g. *adpA*), thereby influencing the production of other specialized metabolites (Horinouchi 2007).

1.3 Phytopathogenic *Streptomyces*

The ability to infect living plant tissues and to cause disease is a rare trait among bacteria belonging to the genus *Streptomyces,* with only a dozen or so species out of the hundreds described so far having this capability. The first described and best characterized phytopathogenic species is *Streptomyces scabiei* (syn. *S. scabies*), although there are other pathogenic species such as *S. acidiscabies, S. turgidiscabies, S. reticuliscabiei and S. aureofaciens*, *S. niveiscabiei, S. luridiscabiei, S. puniciscabiei, S. europaeiscabiei* and *S. stelliscabiei*, and some pathogenic strains of *S. bottropensis* have also been reported (Loria et al. 2006; Li et al. 2019b). These phytopathogenic *Streptomyces* are neither tissue nor host-specific and can cause scab disease on various crops such as potato, carrot, radish, beet, peanut and turnip (Bignell et al. 2010a). Scab disease is characterized by the formation of superficial, raised or deep-pitted scab-like lesions on the tuber surface, and these lesions negatively impact the quality and market value of table stock, processing (e.g. Fench fry, potato chip production) and seed potatoes (Figure 1.6) (Li et al. 2019b). There is also some evidence that CS can reduce the yield of the potato crop (Hiltunen et al. 2009) as well as the overall size of affected tubers (Wanner et al. 2014).

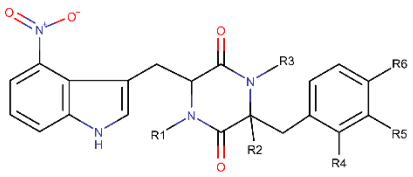
Control strategies for CS disease include crop rotation, irrigation during tuber formation, lowering the soil pH, biological control, using disease-free seed tubers, seed treatment with fungicide, sulphur fertilizer and chemical fumigation (Dees and Wanner 2012). However, the effects of these control methods are inconsistent, they can fail altogether, or they can promote the development of other diseases, or can be impractical and environmentally unfriendly (Loria et al. 1997; Dees and Wanner 2012; Li et al. 2019b). Thus, a thorough understanding of the molecular

mechanisms of *S. scabiei* plant pathogenicity is critical to the development of strategies that can effectively manage the disease.

Figure 1.6 CS disease symptoms. Image provided by Dawn Bignell (Memorial University).

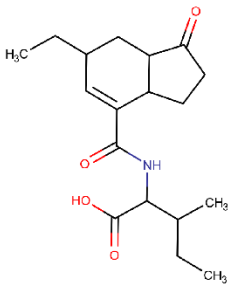
1.4 Specialized metabolites produced by *S. scabiei*

Under standard laboratory conditions, *S. scabiei* is known to produce at least five different types of bioactive specialized metabolites, of which three are known or suspected to play a role in mediating host-pathogen interactions (Figure 1.7).

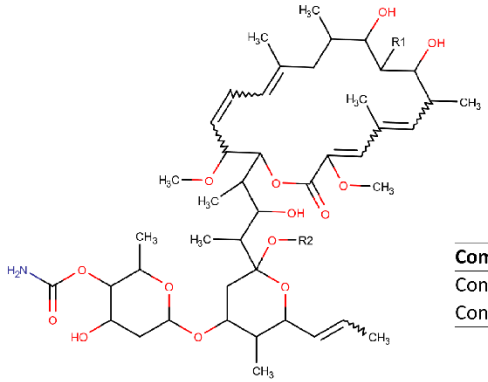


Compound	R1	R2	R3	R4	R5	R6
Thaxtomin A	Me	OH	Me	H	OH	H
Thaxtomin B	Me	OH	Me	H	H	H
Thaxtomin C	Me	H	H	H	H	H
Thaxtomin D	Me	H	Me	H	H	H

(i)

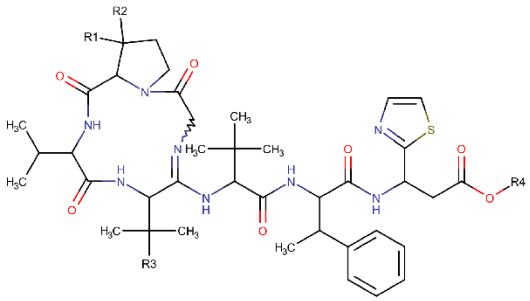


(ii)



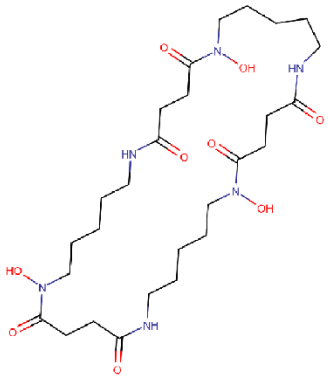
Compound	R1	R2
Concanamycin A	CH_2CH_3	H
Concanamycin B	CH_3	H

(iii)

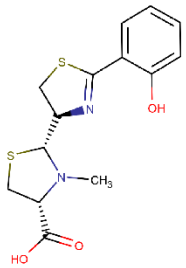


Compound	R1	R2	R3	R4
Bottromycin A2	Me	H	Me	Me
Bottromycin B2	H	H	Me	Me
Bottromycin C2	Me	Me	Me	Me

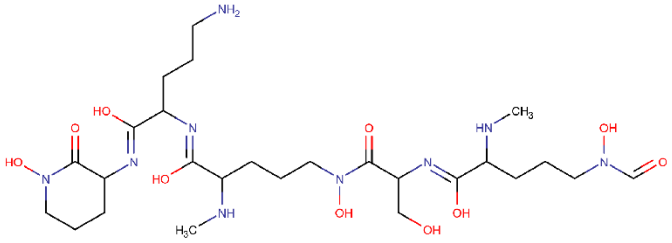
(iv)



Desferrioxamine E



Pyochelin



Scabichelin

(v)

Figure 1.7 Chemical structures of specialized metabolites produced by *S. scabiei*. (i) Thaxtomin A, B, C and D; (ii) *N*-coronafacoyl-L-isoleucine; (iii) Concanamycin A and B ; (iv) Bottromycins A2, B2 and C2; (v) siderophores including Desferrioxamine E, Pyochelin and Scabichelin.

1.4.1 Thaxtomins

The thaxtomins are a family of nitrated 2,5-diketopiperazines that exhibit phytotoxic activity and are the key pathogenicity determinants produced by *S. scabiei* and other scab-causing *Streptomyces* species (Figure 1.7) (Li et al. 2019a). Eleven thaxtomin analogues have been identified, of which thaxtomin A is the predominant analogue produced by *S. scabiei* (King and Calhoun 2009). Thaxtomin A is required for the pathogenicity of *S. scabiei* and other scab-causing *Streptomyces* species (Goyer et al. 1998; Healy et al. 2000; Joshi et al. 2007; King et al. 1989; King and Calhoun 2009; Loria et al. 1995). It functions as a cellulose biosynthesis inhibitor in higher plants and may facilitate the penetration of expanding plant tissues by bacteria during host colonization and infection (Loria et al. 2008).

The biosynthesis of thaxtomin A involves a highly conserved BGC that is composed of seven genes: *txtA, txtB, txtC, txtD, txtE, txtH* and *txtR* (Figure 1.8) (Li et al. 2019a). *txtD* encodes a nitric oxide synthase that produces nitric oxide (NO) from L-arginine, and *txtE* encodes a P450 monooxygenase that uses the NO to nitrate L-tryptophan, producing the intermediate L-4-nitro-tryptophan (Johnson et al. 2009; Barry et al. 2012). *txtA* and *txtB* encode non-ribosomal peptide synthetases (NRPSs) that are responsible for synthesizing the cyclic dipeptide backbone of thaxtomin from L-4-nitro-tryptophan and L-phenylalanine (Johnson et al. 2009). *txtC* encodes a cytochrome P450 monooxygenase that is required for post-cyclization hydroxylation steps (Alkhalaf et al., 2019; Healy et al. 2002). *txtH* encodes a MbtH-like protein (MLP), which is thought to function as a chaperone for enabling the proper folding of the thaxtomin NRPS

adenylation domains (Li et al. 2019a). *txtR* encodes a CSR that controls the expression of the biosynthetic genes within the thaxtomin BGC (Joshi et al. 2007b).

Gene	Gene name	Predicted protein product
1	*SCAB31761/txtC*	P450 monooxygenase
2	*SCAB31771/txtH*	MbtH-like protein
3	*SCAB31781/txtB*	NRPS
4	*SCAB31791/txtA*	NRPS
5	*SCAB31801/txtR*	Transcriptional regulator
6	*SCAB31831/txtE*	P450 monooxygenase
7	*SCAB31841/txtD*	Nitric oxide synthase

Figure 1.8 Organization and composition of the thaxtomin BGC in *S. scabiei.*

1.4.2 Coronafacoyl phytotoxins

Coronafacoyl phytotoxins are non-host-specific plant toxins known or predicted to be produced by several different plant pathogenic bacteria. They are structurally and functionally similar to the bioactive plant hormone jasmonoyl-L-isoleucine (JA-Ile), and it is thought that the production of these molecules enables the pathogen to manipulate jasmonate signalling in the plant host in order to overcome host defenses during infection (Bignell et al. 2018). *S. scabiei* produces *N*-coronafacoyl-L-isoleucine (CFA-Ile) (Figure 1.7) as the main coronafacoyl phytotoxin, though other minor compounds have also been detected, including one predicted to be CFA-Val (Bignell et al. 2018). Disruption of CFA-Ile production in *S. scabiei* results in reduced disease symptom development in tobacco seedlings (Bignell et al. 2010b), while elevated phytotoxin production has been correlated with increased necrosis and pitting of potato tuber tissue (Cheng et al. 2019),

suggesting that CFA-Ile enhances the virulence phenotype of *S. scabiei,* though it is not required for pathogenicity.

The CFA-Ile BGC in *S. scabiei* consists of at least 15 genes, nine of which are homologous to those found within the coronafacic acid (CFA) BGC in the bacterial plant pathogen *Pseudomonas syringae* (Figure 1.9) (Bignell et al. 2010b). Among the genes conserved in both organisms are the *cfa1-8* and *cfl* genes. *cfa1-5* encode enzymes believed to synthesize the 2-carboxy-2-cyclopentenone intermediate, *cfa6-7* encode the large, multi-modular PKS that generates the CFA backbone, and *cfa8* encodes a predicted crotonyl-CoA carboxylase/reductase that may produce the ethylmalonyl-CoA extender unit used for CFA polyketide biosynthesis (Rangaswamy et al. 1998a, b; Bender et al. 1999; Bignell et al. 2018). The *cfl* gene encodes an acyl-CoA ligase that catalyzes the ligation of CFA to its amino acid partner (isoleucine in the case of CFA-Ile) during the final stage of phytotoxin biosynthesis (Fyans et al. 2015).

Interestingly, the *S. scabiei* CFA-Ile BGC contains six additional genes that are absent from the *P. syringae* CFA gene cluster. Four of these genes (*SCAB79681/oxr*, *SCAB79691/CYP107AK1*, *SCAB79711* and *SCAB79721/sdr*) were predicted to encode biosynthetic enzymes and were shown to be co-transcribed with the *cfa* and *cfl*, while the remaining two genes (*SCAB79581/orf1, SCAB79591/cfaR*) are divergently co-transcribed and have been shown to be involved in regulation (Bignell et al. 2010b; Cheng et al. 2015, 2019). Recent studies from our lab have confirmed that *oxr*, *sdr* and *CYP107AK1* are required for the biosynthesis of CFA-Ile, with *CYP107AK1* being essential for metabolite production (Bown et al. 2016, 2017). As homologues of these genes are not found anywhere in the genome of *P. syringae*, it appears that *S. scabiei* and *P. syringae* use distinct biosynthetic pathways for producing the same family of phytotoxins (Bignell et al. 2018).

Gene	Gene name	Predicted protein product
1	*SCAB79581/orf1*	ThiF superfamily protein
2	*SCAB79591/cfaR*	PAS-LuxR DNA binding protein
3	*SCAB79601/cfa1*	Acyl carrier protein (ACP)
4	*SCAB79611/cfa2*	Type II fatty acid dehydratase (DH)
5	*SCAB79621/cfa3*	Type II β-ketoacyl synthase (KS)
6	*SCAB79631/cfa4*	Unknown
7	*SCAB79641/cfa5*	Acyl-CoA ligase
8	*SCAB79651/cfa6*	Type I PKS
9	*SCAB79661/cfa7*	Type I PKS
10	*SCAB79671/cfl*	Acyl-CoA ligase
11	*SCAB79681/oxr*	F_{420}-dependent oxidoreductase
12	*SCAB79691/CYP107AK1*	P450 monooxygenase
13	*SCAB79701/cfa8*	Crotonyl-CoA reductase/carboxylase (CCR)
14	*SCAB79711*	Hydroxybutyryl-CoA dehydrogenase
15	*SCAB79721/sdr*	Short chain dehydrogenase/reductase

Figure 1.9 Organization and composition of the coronafacoyl phytotoxin BGC in *S. scabiei*.

1.4.3 Concanamycins

Concanamycins are a family of specialized metabolites that are characterized by an 18-membered macrolide ring and a β-hydroxyhemiacetal side chain (Haydock et al. 2005) (Figure 1.7). They function as vacuolar-type ATPase inhibitors and are biologically active against fungi, plants and cancer cells (Kinashi et al. 1984; Seki-Asano et al. 1994). Several different *Streptomyces* spp. have been reported to produce concanamycins, including *S. diastatochromogenes, S. neyagawaensis,* and *S. scabiei* (Natsume et al. 1996, 1998, 2001, 2005; Haydock et al. 2005). *S. scabiei* produces two members of the concanamycin family – concanamycin A and B – both of which exhibit root growth inhibitory activity against different plant species (Li et al. 2019b). Natsume and colleagues suggested that concanamycin A and

thaxtomin A may have synergistic action in inducing necrosis of potato tuber tissue, and that concanamycin A may contribute to the development of deep-pitted lesions by strains of *S. scabiei* (Natsume et al. 2017).

The concanamycin BGC was previously identified in *S. scabiei* 87-22 by genomic analysis (Figure 1.10) (Bignell et al. 2010a) and is homologous to the concanamycin A BGC from *S. neyagawaensis* (Haydock et al. 2005). The *S. neyagawaensis* concanamycin BGC is 100 kb long and contains 28 genes, including six PKS genes that are predicted to be involved in the formation of the polyketide backbone. The 22 remaining genes are predicted to function in the modification of the polyketide backbone, biosynthesis of the sugar moiety, or regulation of metabolite production (Haydock et al. 2005).

Gene	Gene name	Predicted protein product
1	*SCAB83841*	Conserved hypothetical protein
2	*SCAB83861*	Conserved hypothetical protein
3	*SCAB83871*	Putative type I PKS
4	*SCAB83891*	Putative type I PKS
5	*SCAB83901*	Putative type I PKS
6	*SCAB83911*	Putative type I PKS
7	*SCAB83921*	Putative type I PKS
8	*SCAB83931*	Putative type I PKS
9	*SCAB83941*	Putative methoxymalonate biosynthesis protein
10	*SCAB83951*	Putative methoxymalonate biosynthesis protein
11	*SCAB83961*	Putative methoxymalonate biosynthesis protein
12	*SCAB83971*	Putative methoxymalonate biosynthesis protein
13	*SCAB83981*	Putative glycosyltransferase
14	*SCAB83991*	Putative NDP-hexose 2,3-dehydratase
15	*SCAB84001*	Putative NDP-hexose 3-ketoreductase
16	*SCAB84011*	Putative NDP-hexose 4-ketoreductase
17	*SCAB84021*	Putative carbomoyltransferase
18	*SCAB84031*	Putative dTDP-glucose synthase
19	*SCAB84041*	Putative dTDP-glucose 4,6-dehydratase
20	*SCAB84051*	Putative O-methyltransferase
21	*SCAB84061*	Conserved hypothetical protein
22	*SCAB84071*	Putative thioesterase
23	*SCAB84081*	Conserved hypothetical protein
24	*SCAB84091*	Putative phosphopantetheinyl transferase
25	*SCAB84101*	Putative AfsR-type regulatory protein
26	*SCAB84131*	Putative crotonyl-CoA reductase
27	*SCAB84141*	Putative hydroxybutyryl-CoA reductase

Figure 1.10 Organization and composition of the concanamycin BGC in *S. scabiei.*

1.4.4 Bottromycins

Bottromycins were first isolated from *S. bottropensis* (Waisvisz et al. 1957), and are ribosomally synthesized and post-translationally modified peptides (RiPPs) composed of a macrolactamidine ring and a thiazole heterocycle (Figure 1.7) (Crone et al. 2012; Gomez-

Escribano et al. 2012a). Bottromycins function as bacterial protein synthesis inhibitors by blocking the aminoacyl-tRNA binding to the A site of bacterial ribosomes (Otaka and Kaji 1976, 1981, 1983), and are biologically active against mycoplasmas and Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant Enterococci (VRE) (Kobayashi et al. 2010).

The bottromycin BGC has been characterized in *S. scabiei* in previous studies (Gomez-Escribano et al. 2012a; Crone et al. 2016), where the production of the metabolite was detected when the bacterium was grown on glucose-yeast extract-malt extract (GYM) medium (Crone et al. 2016). The *S. scabiei* bottromycin BGC contains 13 genes: *btmA-btmM. btmD* encodes a precursor peptide (BtmD) that undergoes a series of modifications catalyzed by enzymes encoded in the bottromycin BGC (Figure 1.11) (Vior et al. 2020). *btmL* encodes a potential regulator, which is conserved in all characterized bottromycin BGCs (Crone et al. 2012; Gomez-Escribano et al. 2012a; Huo et al. 2012). Interestingly, BtmL is not a master regulator of the BGC, but instead, it is a *btmD*-specific regulator (Vior et al. 2020).

Gene	Gene name	Predicted protein product
1	*SCAB56591/btmM*	Putative cytosolic aminopeptidase
2	*SCAB56601/btmL*	Hypothetical protein
3	*SCAB56611/btmK*	Conserved hypothetical protein
4	*SCAB56621/btmJ*	Conserved hypothetical protein
5	*SCAB56631/btmI*	Conserved hypothetical protein
6	*SCAB56641/btmH*	Conserved hypothetical protein
7	*SCAB56651/btmG*	Conserved hypothetical protein
8	*SCAB56661/btmF*	Conserved hypothetical protein
9	*SCAB56671/btmE*	Hypothetical protein
10	*SCAB56681/btmD*	Hypothetical protein
11	*SCAB56691/btmC*	Conserved hypothetical protein
12	*SCAB56701/btmB*	Conserved hypothetical protein
13	*SCAB56711/btmA*	Putative integral membrane protein

Figure 1.11 Organization and composition of the bottromycin BGC in *S. scabiei*.

1.4.5 Siderophores

By chelating iron with high affinity, siderophore-mediated iron uptake may play an important role in pathogen survival within the plant host (Luo et al. 2005; Miethke and Marahiel 2007). Siderophores are biosynthesised via two main pathways: NRPS-dependent and NRPS-independent. The production of siderophores, such as desferrioxamine, pyochelin and scabichelin has previously been confirmed in *S. scabiei,* and the corresponding BGCs have also been characterized (Figure 1.7; Figure 1.12; Figure 1.13; Figure 1.14), (Seipke et al. 2011; Bicz 2013; Kodani et al. 2013). The desferrioxamines are biosynthesised via a NRPS-independent pathway, while pyochelin and scabichelin are biosynthesised via a NRPS-dependent pathway (Seipke et al. 2011; Bicz 2013; Kodani et al. 2013).

Gene	Gene name	Predicted protein product
1	*SCAB57871*	Conserved hypothetical protein
2	*SCAB57891*	Putative calcineruin-like phosphoesterase
3	*SCAB57901*	Conserved hypothetical protein
4	*SCAB57911*	β-*N*-acetylhexosaminidase
5	*SCAB57921/desD*	Putative desferrioxamine biosynthesis protein
6	*SCAB57931/desC*	Putative desferrioxamine biosynthesis acyl-CoA-dependent acetyltranferase
7	*SCAB57941/desB*	Putative desferrioxamine biosynthesis monooxygenase
8	*SCAB57951/desA*	Putative siderophore biosynthesis pyridoxal-dependent decarboxylase
9	*SCAB57961*	Putative IS630 family insertion sequence

Figure 1.12 Organization and composition of the desferrioxamine BGC in *S. scabiei*.

Gene	Gene name	Predicted protein product
1	*SCAB1371*	Putative AfsR-family regulatory protein
2	*SCAB1381*	Putative salicylate synthase
3	*SCAB1391*	Conserved hypothetical protein
4	*SCAB1401*	Putative TetR-family regulator
5	*SCAB1411*	Putative AMP-binding NRPS ligase
6	*SCAB1421*	Putative thioesterase
7	*SCAB1431*	Putative ABC transporter ATP-binding subunit
8	*SCAB1441*	Putative ABC transporter ATP-binding subunit
9	*SCAB1451*	Putative transporter component
10	*SCAB1461*	Putative oxidoreductase
11	*SCAB1471*	Putative NRPS
12	*SCAB1481*	Putative NRPS

Figure 1.13 Organization and composition of the pyochelin BGC in *S. scabiei*.

Gene	Gene name	Predicted protein product
1	*SCAB85431*	Putative ABC transport system integral membrane component
2	*SCAB85441*	Putative ABC-transport system ATP-binding component
3	*SCAB85451*	Putative secreted siderophore-binding lipoprotein (transport system associated)
4	*SCAB85461*	Putative MbtH-like protein
5	*SCAB85471*	Putative NRPS/siderophore biosynthesis protein
6	*SCAB85481*	Putative siderophore transport system integral membrane component, FecCD family
7	*SCAB85491*	Putative siderophore transport system integral membrane component, FecCD family
8	*SCAB85501*	Putative siderophore transport system ATP- binding component
9	*SCAB85511*	Putative formyltransferase
10	*SCAB85521*	Putative peptide N-oxygenase

Figure 1.14 Organization and composition of the scabichelin BGC in *S. scabiei*.

1.4.6 Cryptic phenazine-like BGC

An orphan phenazine-like BGC in *S. scabiei* 87-22 was initially annotated more than ten years ago, and the annotation was based on the presence of genes encoding proteins that are similar to known phenazine biosynthetic enzymes (Dr. Bignell, personal communication) (Yaxley 2009). It was hypothesized that the resulting metabolite may be important for the pathogenic phenotype of *S. scabiei* during infection of the plant host, or that it may allow *S. scabiei* to compete for limited nutrients in the soil environment by killing or inhibiting the growth of other microorganisms. However, this phenazine-like BGC was determined to be silent under laboratory conditions (D. Bignell, personal communication), so the nature of the product and its function remains unknown.

1.5 Strategies to characterize orphan specialized metabolite BGCs

Orphan (cryptic) specialized metabolite BGCs are found within the genomes of many microorganisms. They are usually silent or produce no noticeable amounts of metabolite(s) under standard laboratory conditions, and thus the encoded metabolite(s) is unknown (Gross 2007). With increasing numbers of microbial genomes sequenced, a greater number of orphan BGCs have also been identified. The activation and characterization of these orphan BGCs are likely to allow for the discovery of many novel natural products. Different approaches have been used to discover new metabolites, and these approaches can be divided into two categories: (1) activation of silent BGCs in native hosts; (2) Activation of silent BGCs in heterologous hosts (Figure 1.15) (Reen et al. 2015; Ren et al. 2017; Zhang et al. 2019; Nguyen et al. 2020). The workflow for characterization of orphan specialized metabolite BGCs begins with the identification of BGCs using computational tools, followed by BGC activation and product detection (Figure 1.15).

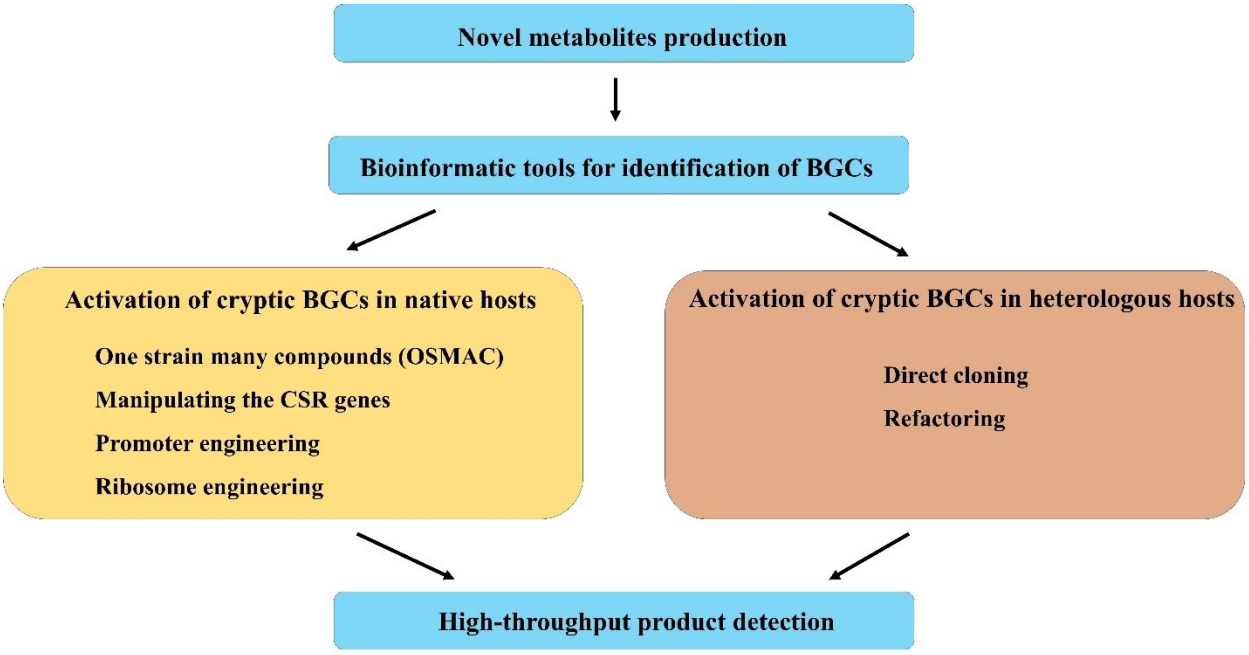


Figure 1.15 Overview of the workflow for the discovery of novel specialized metabolites.

1.5.1 Bioinformatic tools for identification of BGCs

Various bioinformatics tools and databases for the identification of BGCs have been developed in recent years, such as BAGEL, ClusterScan, CLUSEAN, NPsearcher, NaPDoS, SEARCHPKS, and SMURF (Weber and Kim 2016). A recently established website called "Specialized Metabolite Bioinformatics Portal" (SMBP) provides information on these mining computational tools and databases (Weber and Kim 2016). However, most of these tools are limited to mining specific classes of specialized metabolites, including RiPPs, PK and NRP.

The rule-based tools, such as PRISM (Skinnider et al. 2017) and antiSMASH (Blin et al. 2019b), are the most widely used genome mining tools for predicting various types of specialized metabolite BGCs. These tools use Hidden Markov Models (HMM) and a human defined rules based approach to specifically identify certain types of BGCs (Skinnider et al. 2017; Blin et al. 2019b). The PRISM version 3 can detect 22 distinct types of BGCs, and antiSMASH version 5 can identify 52 different types of BGCs. Both tools can provide rapid BGC identification and are very user-friendly, making them the most popular genome mining tools. However, the rule-based tools are limited to identifying BGCs that are similar to ones known to produce preexisting metabolites.

Accordingly, the rule-independent tools, such as ClusterFinder, DeepBGC and EvoMining, have been developed to allow for the detection of novel BGCs associated with the production of unknown specialized metabolite (Cimermancic et al. 2014; Cruz-Morales et al. 2016; Hannigan et al. 2019; Sélem-Mojica et al. 2019). ClusterFinder utilizes a machine learning approach employing a HMM method to detect BGCs (Cimermancic et al. 2014). DeepBGC employs a deep learning strategies using Recurrent Neural Networks (RNNs) and protein family (Pfam) domains to detect BGCs (Hannigan et al. 2019). By contrast, another genome-mining approach called EvoMining

complements traditional genome-mining approaches by incorporating phylogenomic analysis-based evolutionary principles (Cruz-Morales et al. 2016; Sélem-Mojica et al. 2019). EvoMining leads to the identification of paralogs of primary metabolic enzymes with functional divergence, and surmises that the additional copies are repurposed for specialized metabolism (Cruz-Morales et al. 2016; Sélem-Mojica et al. 2019).

1.5.2 Activation of cryptic BGCs in native hosts

1.5.2.1 One strain many compounds (OSMAC) approach

Cultivation-based approaches are an integral component of the OSMAC framework, which proposes that a single strain has the potential to produce different molecules under different environmental conditions (Romano et al. 2018). Zeeck and colleagues were the first to postulate the OSMAC approach; they found that a strain of *Aspergillus ochraceus* was able to produce up to 20 different metabolites by changing the cultivation parameters such as media composition, temperature, salinity, aeration and culture vessel (Bode et al. 2002). Subsequent studies conducted in recent years have revealed plentiful cultivation parameters affecting the production of specialized metabolites in microbes, including nutritional (carbon source, nitrogen source, sulfur and phosphorus sources and trace elements) and physical (temperature, vessel types, aeration, and shaking conditions, osmotic stress, salinity and pH) parameters (Romano et al. 2018). Differential production has also been reported between solid and liquid media (English et al. 2017). Moreover, co-cultivation of microbes (prokaryote-prokaryote, prokaryote-eukaryote, and eukaryote-eukaryote), and addition of external cues can also be framed within the OSMAC principle, though they were not directly considered when the OSMAC was first postulated (English et al. 2017). The cultivation-based approaches are ideal tools for microorganisms that are refractory to genetic manipulation, because these approaches do not require identification of the specialized metabolite

BGCs or manipulating the regulatory networks controlling their expression. On the other hand, these approaches are untargeted for specific molecules, and investigation of the various cultivation parameters that promote specialized metabolite production can be challenging and time-consuming from a practical perspective (Romano et al. 2018).

1.5.2.2 Manipulating the CSR genes

In *Streptomyces* species, the expression of specialized metabolite BGCs is tightly controlled and modulated by multi-level transcriptional regulation. The CSRs are usually regulators that are encoded within the BGC that they control, while the cluster-independent regulators, or pleiotropic regulators, govern the expression of multiple BGCs, usually by targeting the CSRs within gene clusters (Bibb 2005; van Wezel and McDowall 2011). The overexpression of activators or deletion of repressors are straightforward ways to activate or elevate the expression of specialized metabolite BGCs (Figure 1.15) (Zarins-Tutt et al. 2016). A repressor can bind to the operator, which obstructs the RNA polymerase from initiating gene transcription. By deletion of repressors, transcription can be activated. Production can also be achieved by overexpression of activators, which are DNA-binding proteins that bind to enhancers or promoter-proximal elements, facilitating the binding of the RNA polymerase to the promoter (Ptashne 2014). However, it is hard to predict what molecules will be biosynthesized or which BGCs will be involved by exploiting the activity of pleiotropic regulators. By contrast, manipulating CSRs represents a more targeted strategy to activate the expression of cryptic BGCs (Zhang et al. 2019). For example, in *Streptomyces lincolnensis*, overexpression of the newly discovered cluster-situated activator *lmbU* increased the production of the antibiotic lincomycin A. LmbU appears to belong to a new family of regulators as it shares little sequence or structural similarity with other transcription factors (Hou et al. 2018). In addition, the constitutive expression of *vemR* and *astG*, two LuxR family

transcriptional activator genes, led to the discovery of a new biaryl polyketide, venemycin, and two known ansatrienins in *S. venezuelae* and *Streptomyces* sp. XZQH13, respectively (Xie et al. 2015; Thanapipatsiri et al. 2016). However, overexpression of ARR (atypical response regulator) family activators, JadR1 and PgaR1, did not have significant effects on the production of jadomycin and gaudimycin in *Streptomyces venezuelae* and *Streptomyces* sp. PGA64, respectively (Guo et al. 2015), which suggests that the efficacy of this approach is pathway-specific and varies case by case.

1.5.2.3 Promoter engineering

Activation of silent gene clusters can also be achieved by promoter engineering. Promoter elements oversee the first stage of gene expression, and introducing a strong inducible or constitutive promoter upstream of biosynthetic genes within the target BGC allows for the natural regulatory system to be bypassed (Figure 1. 17) (Myronovskyi and Luzhetskyy 2016; Zarins-Tutt et al. 2016). Various promoters have been successfully applied to activate silent gene clusters, such as the constitutive promoters *ermE*p*, SF14p, SP44 and *kasO*p*, and the inducible promoters *tipA*p, *tcp830* and *nitA*p (Myronovskyi and Luzhetskyy 2016). Among these promoters, *ermE*p* is widely used for the overexpression of target genes. However, even *ermE*p* fails to give desirable results in some *Streptomyces* species due to the complicated regulatory networks that are present in streptomycetes, and this strategy requires prior knowledge of the transcriptional organization of a BGC so that the promoters can be introduced at the appropriate places to enable expression of all of the biosynthetic genes in the cluster (Zhou et al. 2011).

1.5.2.4 Ribosome engineering

Ribosome engineering was first developed by the Ochi group with the idea that silent genes may be activated by modulating the ribosomal proteins or rRNA (Figure 1.15). They found that production of actinorhodin in *S. coelicolor* was activated by a ribosomal mutation in the *rpsL* gene (encoding the ribosomal protein S12) that confers resistance to streptomycin (Shima et al. 1996). The mutation of ribosomal protein S12 leads to an increased stability of ribosome and enrichment of some translation factors (e.g. the ribosome recycling factor, RRF) (Ochi et al. 2004; Wang et al. 2009). Challenging *Streptomyces* with streptomycin can also introduce a mutation in the *rsmG* gene (encoding for 16S rRNA methyltransferase), which can also activate the expression of silent specialized metabolite pathways (Nishimura et al. 2007). Similarly, actinorhodin production was also activated by treating *S. coelicolor* and *S. lividans* with rifampicin, which can induce a mutation in *rpoB* gene (encoding the β-subunit of RNA polymerase) (Xu et al. 2002). Other drugs used to modulate the ribosome by introducing a mutation in a ribosomal component include kanamycin, chloramphenicol, lincomycin, gentamicin, paromomycin, thiostrepton, fusidic acid, spectinomycin, and neomycin (Ochi et al. 2004). However, similar to the cultivation-based approaches, ribosome engineering does not target specific BGCs.

1.5.3 Activation of cryptic BGCs in heterologous hosts

The direct cloning of a cryptic BGC into different heterologous hosts is another approach to relieve the tight regulation of the BGC in the native host (Figure 1.15). The heterologous hosts should be easy to culture, genetically amenable, and be capable of producing large amounts of specific compounds with minimal interference from host specialized metabolites (Baltz 2016). *Escherichia coli* has been successfully used as a heterologous host to biosynthesize several PK compounds, including erythromycin from *Saccharopolyspora erythraea,* 6-methylsalicylic acid from *Penicillium patulum*, and several flavonoids (Ahmadi and Pfeifer 2016). However, *E. coli* is

not a useful host for producing many other metabolites due to a lack of key metabolic enzymes for generating various precursor molecules (Yang et al. 2020). A series of *Streptomyces* strains have been developed as heterologous hosts, such as *S. lividans* K4-114/TK23/1326, *S. ambofaciens* BES2074*, S. avermitilis* SUKA5/17, *S. albus* J1074 and *S. coelicolor* M512/ M1154 (Baltz 2010). These strains have been successfully used as heterologous hosts to produce a variety of cryptic natural products including venemycin from *S. venezuelae*, fralnimycin from *Frankia* spp., and lavendiol from *S. lavendulae* (Zhang et al. 2019).

Direct cloning can also be coupled with refactoring to activate cryptic BGCs in heterologous hosts (Figure 1.15). For example, direct cloning of the taromycin BGC in *S. coelicolor* did not activate the expression of the BGC, but the production of taromycin was activated following deletion of the negative regulator encoded within the BGC (Yamanaka et al. 2014). The commonly used refactoring strategies include overexpression of transcriptional activators, deletion of negative repressors, and introducing a strong constitutive promoter upstream of the biosynthetic genes as mentioned above.

1.5.4 High-throughput product detection

In addition to the activation of silent BGCs, another challenge is how to decipher their product quickly. With the widespread application of mass spectrometry, many computational approaches have been developed to link BGCs to their products based on the metabolomics data. For example, the Genomes to Natural Products platform (GNP) was recently developed to predict and identify PKS and NRPS natural products (Johnston et al. 2015). These PKS and NRPS products can be further linked to their corresponding BGCs by the computational tools GRAPE (generalized retro-biosynthetic assembly prediction engine) and GARLIC (global alignment for natural products cheminformatics) (Dejong et al. 2016). Additionally, iSNAP (informatic search

strategy for natural products) can be used to determine the site-specific modifications of NRPs (Ibrahim et al. 2012). Web applications, such as NRPquest/RiPPquest/Cycloquest (http://cyclo.ucsd.edu) were developed to correlate the tandem MS data of NRPs, RiPPs and cyclopeptides with their corresponding gene cluster (Mohimani et al. 2011, 2014b, a). However, most of these bioinformatic tools are restricted to the analysis of peptide natural products. Recently, the Dorrestein group developed the Global Natural Products Social Molecular Networking (GNPS) platform, which allows researchers to share their metabolomics data online and will further accelerate the discovery of novel metabolites (Wang et al. 2016). Other computational tools, such as Network Annotation Propagation (NAP) (da Silva et al. 2018), SIRIUS (Böcker et al. 2009), MetWork (Beauxis and Genta-Jouve 2019), ChemDistiller (Laponogov et al. 2018) and Competitive Fragmentation Modeling for Metabolite Identification 3.0 (CFM-ID 3.0) (Allen et al. 2014) have also been used to predict the structures of unknown compounds based on the existing structural databases. Besides analyzing the BGCs and MS^2 data separately, a new strategy called metabologenomics has been devised to match genes to their product metabolites by grouping BGCs and MS^2 spectrums into gene cluster families (GCFs) and molecular families (MFs) and then correlating GCFs to MFs (Nguyen et al. 2013; Goering et al. 2016; Navarro-Muñoz et al. 2020).

1.6 Goals and objectives

Streptomyces are Gram-positive, filamentous soil-dwelling bacteria renowned for their ability to produce a wide array of medically and agriculturally important specialized metabolites, including antibiotics, anticancer agents, immunosuppressants, herbicides and insecticides. Under standard laboratory conditions, most *Streptomyces* species produce ~2-6 bioactive specialized metabolites, and yet many species are now known to have the genetic capacity to produce many

more novel natural products with potentially useful bioactivities. One such organism is *S. scabiei,* which is an important causative agent of potato CS. *S. scabiei* is known to produce at least five bioactive compounds in the laboratory, whereas the recently available genome sequence of this organism has revealed the presence of at least 30 different loci that may participate in the biosynthesis of natural products. Many of these gene clusters are cryptic, and some are silent under laboratory conditions, and so the nature of the product that is produced and its function for the producing organism remains unclear.

The overall goal of this thesis research was to investigate a cryptic and silent specialized metabolite BGC that is present on the chromosome of *S. scabiei* 87-22. The gene cluster contains open reading frames (ORFs) encoding proteins that are similar to those involved in the biosynthesis of phenazines, and I hypothesize that the resulting metabolite may be important for the pathogenic phenotype of *S. scabiei* during infection of the plant host, or that it may allow *S. scabiei* to compete for limited nutrients in the soil environment by killing or inhibiting the growth of other microorganisms.

The specific aims of the thesis were as follows:

1. To conduct a detailed bioinformatics analysis of the *S. scabiei* phenazine-like BGC;

2. To construct engineered strains in order to activate expression of the phenazine-like BGC;

3. To conduct gene expression analysis, morphological analysis, bioactivity analysis and pathogenicity testing of the engineered strains to determine the conditions that enable expression of the BGC and to characterize the function(s) of the resulting metabolite(s); and

4. To conduct an untargeted metabolomics analysis of wild type (WT) *S. scabiei* and an engineered strain that expresses the phenazine-like BGC in order to identify the molecule(s) that are produced by the BGC.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Enzymes, chemicals, and equipment

Enzymes used in this study (restriction enzymes, *Taq* DNA polymerase, Phusion High-Fidelity DNA polymerase, T4 DNA ligase, shrimp alkaline phosphatase, RNAse-free DNase I) were purchased from New England Biolabs, Canada. The kits used for DNA gel extraction (Wizard SV Gel and PCR Clean-Up kit) and cloning of PCR inserts (pGEM®-T Easy vector system) were purchased from Promega Inc., Canada. The kit used for plasmid DNA isolation from *Escherichia coli* strains (EZ-10 Spin Column Plasmid DNA kit BS614-250Preps) was obtained from Bio Basic Inc., Canada. The kit used for genomic DNA preparation from *Streptomyces* strains (QIAamp DNA Mini Kit) was purchased from Qiagen Inc. Canada. The kit for total RNA isolation from *Streptomyces* strains (innuSPEED Bacteria-Fungi RNA Kit) was purchased from Analytik Jena, USA. Ingredients for media, other reagents and chemicals were obtained from Fisher Scientific, Canada, VWR International (Canada) or Sigma-Aldrich, Canada. Oligonucleotide primers used in this study were purchased from Integrated DNA Technologies, USA.

A P300 Nanophotometer (Implen Inc., USA) was used to quantify DNA and RNA, and to measure the optical density of bacterial cultures. A C1000TM Thermal Cycler (Bio-Rad, Canada) was used for PCR reactions. Agarose gel electrophoresis was conducted using a Bio-Rad Power Pac 300. A Gene Pulser Xcell (Bio-Rad, Canada) was used for electroporation of DNA into microbial cells, and a SpeedMill PLUS tissue homogenizer (Analytik Jena, USA) was used for lysis of microbial cells. Agarose gel images were acquired using a Chemi-Imager Gel Documentation System with a Fluorchem HD2 upgrade (Alpha Innotech, USA) or a GelDoc-

It®TS2 310 Imager (UVP Analytik Jena, USA). All aseptic work was conducted in a class II biosafety cabinet (nuAire, Inc. USA).

2.1.2 General solutions, buffers, growth media, and antibiotics

All solutions, buffers, media, and antibiotics were prepared using ultrapure water. Solutions and buffers were sterilized either by autoclaving at 121°C for 20 min or by filtration through a 0.2 µm filter (Stericon). Antibiotic solutions (Table 2.1) were sterilized by filtration through a 0.2 µm filter. All media were sterilized by autoclaving at 121°C for 20 min. Recipes used for the preparation of buffers, solutions and media used in this study can be found in Appendix 1.

Table 2.1 Antibiotics used in this study.

Antibiotic	**Stock concentration (mg/mL)**	**Solvent**	**Working concentration for *Streptomyces* (µg/mL)**
Ampicillin	100	Water	100
Apramycin	50	Water	50
Chloramphenicol	25	Ethanol	25
Hygromycin	100	Water	100
Kanamycin	50	Water	50
Nalidixic acid	50	0.3M NaOH	50

2.1.3 Bacterial strains, cosmids and plasmids

Bacterial strains, cosmids and plasmids used in this study are listed in Tables 2.2 and 2.3.

Table 2.2 Bacterial strains used in this study.

Strain	Description	Resistance*	Reference or Source
***Escherichia coli* strains**			
NEB5α	DH5α derivative; high efficiency competent cells used for transformation	none	New England Biolabs, Canada
BW25113/pIJ790	Host for Redirect PCR targeting system	cam^r	(Gust et al. 2003a, b)
ET12567/pUZ8002	Non-methylating host (*dam⁻ dcm⁻ hsdM⁻*); carries the pUZ8002 plasmid that encodes the machinery for the conjugal transfer of DNA into *Streptomyces*	cam^r, tet^r, kan^r	(MacNeil et al. 1992)
ESS	Indicator strain for antimicrobial bioassays	none	(Wang et al. 2004)
***S. scabiei* strains**			
87-22	WT strain	none	(Loria 1995)
87-22/ pJL1	Harbours an integrative plasmid for overexpressing *egfp* from the *kasO*p*promoter	$apra^r$	This study
87-22/ pJL2	Harbours an integrative plasmid for overexpressing *egfp* from the SP44 promoter	$apra^r$	This study
87-22/ pJL3	Harbours an integrative plasmid for overexpressing *SCAB12111* from the *ermE*p* promoter	$apra^r$	This study
87-22/ pJL4	Harbours an integrative plasmid for overexpressing *SCAB12111* from the *kasO*p* promoter	$apra^r$	This study

87-22/ pJL5	Harbours an integrative plasmid for overexpressing *SCAB12111* from the SP44 promoter	apra^r	This study
Δ*SCAB12091*	*SCAB12091* null mutant derived from 87-22	hyg^r	This study
Δ*SCAB12101*	*SCAB12101* null mutant derived from 87-22	hyg^r	This study
Δ(*SCAB12091*+*SCAB12101*)	*SCAB12091*+ *SCAB12101* null mutant derived from 87-22	hyg^r	This study
Δ*SCAB12081*	*SCAB12081* null mutant derived from 87-22	hyg^r	This study
87-22/ pJL11	Harbours the *ermE*p* promoter upstream of the native *SCAB12021* gene in the 87-22 chromosome	apra^r	This study
87-22/ pJL12	Harbours the *kasO*p* promoter upstream of the native *SCAB12021* gene in the 87-22 chromosome	apra^r	This study
87-22/ pJL13	Harbours the SP44 promoter upstream of the native *SCAB12021* gene in the 87-22 chromosome	apra^r	This study
87-22/ pJL15	Vector control strain (without *egfp*), harbours an integrative plasmid with the *ermE*p* promoter	apra^r	This study
87-22/ pIJ8641	Vector control strain, harbours an integrative plasmid with the *ermE*p* promoter expressing *egfp*	apra^r	This study

***Streptomyces coelicolor* strains**

M1154	Engineered heterologous host for expression of specialized metabolites	none	(Gomez-Escribano and Bibb 2011)
M1154/ pJL14	M1154 containing the pJL14 cosmid with the phenazine-like biosynthetic genes integrated into the ΦC31 *attB* site	apra^r	This study
***Streptomyces avermitilis* strains**			
SUK17	Engineered heterologous host for expression of specialized metabolites	none	(Ikeda et al. 2003)
SUK17/pJL14	SUK17 containing the pJL14 cosmid with the phenazine-like biosynthetic genes integrated into the ΦC31 *attB* site	apra^r	This study
***Streptomyces albus* strains**			
J1074	Engineered heterologous host for expression of specialized metabolites	none	(Zaburannyi et al. 2014)
J1074/pJL14	J1074 containing the pJL14 cosmid with the phenazine-like biosynthetic genes integrated into the ΦC31 *attB* site	apra^r	This study
Saccharomyces cerevisiae	Indicator strain for antimicrobial bioassays	none	K. Tahlan, Memorial University of Newfoundland
Bacillus subtilis	Indicator strain for antimicrobial bioassays	none	K. Tahlan, Memorial University of Newfoundland

*cam^r, tet^r, kan^r, thio^r, apra^r, hyg^r and amp^r = chloramphenicol, tetracycline, kanamycin, thiostrepton, apramycin, hygromycin B and ampicillin resistance, respectively.

Table 2.3 Plasmids and cosmids used in this study.

Plasmid or Cosmid	Description	Resistance*	Reference or Source
pGEM-T® Easy	Cloning vector for PCR products	amp^r	Promega, Canada
pGEM/SP44	DNA fragment consisting of the SP44 promoter + 6 × HIS + tev cloned into the pGEM-T® Easy vector	amp^r	(Cheng 2018)
pGEM/*kasO*p*	DNA fragment consisting of the *kasO*p* promoter + 6 × HIS + tev cloned into the pGEM-T® Easy vector	amp^r	(Cheng 2018)
pIJ8641	Overexpresses the *egfp* gene from the *ermE*p* promoter, integrates into the *Streptomyces* ϕC31 *attB* site	apra^r	J. Sun, unpublished
pIJ8668	non-integrative plasmid lacks the ϕC31 attachment site (*attP*) and *int* gene.	apra^r	(Sun et al. 1999)
pIJ10700	Template for PCR amplification of the [*hyg*+*oriT*] cassette used for PCR targeting	amp^r, hyg^r	(Gust et al. 2003b)
pIJ10702 (pMJCOS1)	Harbours a ~5 kb SspI fragment containing *aac(3)IV, oriT, int* and *attP*	amp^r, apra^r	(Foulston and Bibb 2010; Yanai et al., 2006)
Cosmid 2194	SuperCos1 derivative containing genes from the *S. scabiei* 87-22 phenazine-like BGC	amp^r, kan^r	This study
pRLDB49-4	Cosmid 2194 derivative in which *SCAB12081* is replaced with the [*hyg*+*oriT*] cassette	hyg^r, amp^r, kan^r	D. Bignell, unpublished

pJL1	pIJ8641 derivative in which *ermE*p* is replaced with *kasO*p*	apra^r	This study
pJL2	pIJ8641 derivative in which *ermE*p* is replaced with SP44	apra^r	This study
pJL3	pIJ8641 derivative in which *egfp* is replaced with *SCAB12111*	apra^r	This study
pJL4	pJL1 derivative in which *egfp* is replaced with *SCAB12111*	apra^r	This study
pJL5	pJL2 derivative in which *egfp* is replaced with *SCAB12111*	apra^r	This study
pJL6	Cosmid 2194 derivative in which *SCAB12091* is replaced with the [*hyg*+*oriT*] cassette	hyg^r, amp^r, kan^r	This study
pJL7	Cosmid 2194 derivative in which *SCAB12101* is replaced with the [*hyg*+*oriT*] cassette	hyg^r, amp^r, kan^r	This study
pJL8	Cosmid 2194 derivative in which *SCAB12091* and *SCAB12101* are replaced with the [*hyg*+*oriT*] cassette	hyg^r, amp^r, kan^r	This study
pJL9	pJL16 derivative in which *ermE*p* is replaced with *kasO*p*	apra^r	This study
pJL10	pJL16 derivative in which *ermE*p* is replaced with SP44	apra^r	This study
pJL11	pJL16 derivative harbouring the *SCAB12021* gene downstream of *ermE*p*	apra^r	This study

pJL12	pJL9 derivative harbouring the *SCAB12021* gene downstream of *kasO*p*	apra^r	This study
pJL13	pJL10 derivative harbouring the *SCAB12021* gene downstream of SP44	apra^r	This study
pJL14	Cosmid 2194 derivative in which the kanamycin resistance gene was replaced with a fragment containing *aac(3)IV*, *oriT*, *int* and *attP*	apra^r, amp^r	This study
pJL15	pIJ8641 derivative lacking the *egfp* gene	apra^r	This study
pJL16	pIJ8668 derivative in which *ermE*p* is inserted in front of *egfp*	apra^r	This study
pJL17	Contains the *SCAB12111* gene cloned into the pGEM-T Easy vector	amp^r	This study
pJL18	Contains the *SCAB12021* gene cloned into the pGEM-T® Easy vector	amp^r	This study

*kan^r, thio^r, apra^r, hyg^r and amp^r = kanamycin, thiostrepton, apramycin, hygromycin B and ampicillin resistance, respectively.

2.1.4 Primers

Oligonucleotide primers used in this study are listed in Table 2.4. The primers were purified by standard desalting prior to shipment.

Table 2.4 Oligonucleotide primers used in this study.

Name	Primer Sequence (5'-3')*	Application	Size (bp)
TB1	CTATGCGACGGTGGGCACCG CCAGGACGATCGTGCTGGTT GTAGGCTGGAGCTGCTTC	Generation of *SCAB12081* deletion cosmid using Redirect PCR targeting (Forward)	1369
TB2	ATGCTCGAGAACGACGTGGA ACTGGAAGCCCCGCTTCCGAT TCCGGGGATCCGTCGACC	Generation of *SCAB12081* deletion cosmid using Redirect PCR targeting (Reverse)	1369
TB3	ATGGCACAGCAGGCGCGGGC AGTGGAGACGAGGCGGCTGA TTCCGGGGATCCGTCGACC	Generation of *SCAB12091* deletion cosmid using Redirect PCR targeting (Forward)	1369
TB4	TCACTGCCGGGCGAGCGCCG CCTCCTGCAGACGGGCGCCT GTAGGCTGGAGCTGCTTC	Generation of *SCAB12091* deletion cosmid using Redirect PCR targeting (Reverse)	1369
TB8	GCGC**CATATG**ACCAGCGACG CGATCGC	Cloning of *SCAB12021* gene (Forward)	1239
TB9	TATA**GCGGCCGC**TCAGTGCA GCAGCGACCGGC	Cloning of *SCAB12021* gene (Reverse)	1239
TB10	TACCGCTCGAACGTGTTCTC	Verification of *SCAB12091* deletion strain (Forward)	2358
TB11	TAGTGGTCGACTGCTCGTTC	Verification of *SCAB12091* deletion strain (Reverse)	2358
TB12	CCGAGCTGATCTGGGAGAAC	Verification of *SCAB12091* deletion strain (Reverse)	367
TB13	GCCCTGTACTTCCACTTCCC	Verification of *SCAB12091* deletion strain (Forward)	367
DRB648	CGACAGCCACTTCATTCTCA	Verification of *SCAB12081* deletion strain (Forward)	504
DRB649	AACGTGTTCTCGGTGGAGAT	Verification of *SCAB12081* deletion strain (Reverse)	504
*ermE*p*For	GCGATGCTGTTGTGGGC	Sequencing of genes cloned downstream of the *ermE*p* promoter	N/A
*kasO*p-SP44 For	GACT**GATATC**TGTTCACATTC GAAC	Cloning of *kasO*p* and SP44 promoters (Forward)	106
*kasO*p-SP44 Rev	GCGC**CATATG**GACACTCCTT ACTTAGACTG	Cloning of *kasO*p* and SP44 promoters (Reverse)	106

JL2	GTTGAGGCCCCAGTTGAC	Sequencing of *SCAB12111* cloned downstream of the *ermE*p* promoter and verification of the *SCAB12101* deletion cosmid (Reverse)	1457
JL3	ACTTCATCAGCGACTCCTC	Sequencing of *SCAB12111*	N/A
JL4	CTTTGACAGTTTCTTCCCCC	Sequencing of *SCAB12111*	N/A
JL5	GCGC**CATATG**GTCGATCTCG ACGTCCA	Cloning of *SCAB12111* gene (forward)	1248
JL6	ATAT**GCGGCCGC**TCAGGCGG ACAGTCCATGTG	Cloning of *SCAB12111* gene (reverse)	1248
JL7	GTGACCAGCGACGCGATC	Sequencing of *SCAB12021*	N/A
JL8	CTCACGGTCGGGGTCCAG	Sequencing of *SCAB12021*	N/A
JL13	GACCCGACCGGTATCAGC	RT-PCR analysis of *SCAB12021* (forward)	103
JL14	AAGGTCGAGGTTCTCCCACT	RT-PCR analysis of *SCAB12021* (reverse)	103
JL15	CTGGCCCATCTCGTCAAC	RT-PCR analysis of *SCAB12031*(forward)	120
JL16	GCGGGAGACGTGGAAGAC	RT-PCR analysis of *SCAB12031* (reverse)	120
JL17	AGTACTTCGTCGAGCCCTTTC	RT-PCR analysis of *SCAB12041*(forward)	145
JL18	CGCTGTTCCTCCGTCATC	RT-PCR analysis of *SCAB12041* (reverse)	145
JL19	TTCCTGTCGGTGTGTCTGAG	RT-PCR analysis of *SCAB12051* (forward)	142
JL20	AGGTGTTGTAGAAACCGACC A	RT-PCR analysis of *SCAB12051* (reverse)	142
JL23	GCAGAACTCGGTGACGAAG	Sequencing of the junction region between *SCAB12111* and *ermE*p*	N/A
JL24	CTTCGACACGAAGGGGTTC	Verification of the *SCAB12101* deletion cosmid (forward)	501
JL25	AGCAGGGACTCCCACACAT	Verification of the *SCAB12101* deletion cosmid (reverse)	501
JL26	ATGCCACGCCAGGAGCGAGC CGAACGCACTCGCGAGCAGT GTAGGCTGGAGCTGCTTC	Generation of *SCAB12091* and *SCAB12101* double deletion strain	1584
JL27	ACCCCTTACGCCGATTCAC	Verification of the *SCAB12101* deletion cosmid (forward)	1457

JL36	AGGTCATCTTCGCCGACTC	RT-PCR analysis of *SCAB12001* (forward)	134
JL37	GCCGTACAGGAAGGAGATCA	RT-PCR analysis of *SCAB12001* (reverse)	134
JL38	AGGCCGAACTGCCATACTT	RT-PCR analysis of *SCAB12011* (forward)	190
JL39	GGAACCACCTCGTAGTCCAG	RT-PCR analysis of *SCAB12011* (reverse)	190
JL40	TCATCCACCACTTCATCGTC	RT-PCR analysis of *SCAB12061* (forward)	154
JL41	CGGAAGTACACCTCCAGGAC	RT-PCR analysis of *SCAB12061* (reverse)	154
JL42	CTCAGGGAGTGGACTTCTCG	RT-PCR analysis of *SCAB12071* (forward)	145
JL43	CAGATCGGCCAGTTGTCTCT	RT-PCR analysis of *SCAB12071* (reverse)	145
JL44	AGAACGACGTGGAACTGGAA	RT-PCR analysis of *SCAB12081* (forward)	108
JL45	TGAAGCGTGACTGTCGTCTC	RT-PCR analysis of *SCAB12081* (reverse)	108
JL46	GTACTTCCACTTCCCGTCCA	RT-PCR analysis of *SCAB12091* (forward)	114
JL47	GTCAGGTCGATCACCTCCTG	RT-PCR analysis of *SCAB12091* (reverse)	114
JL48	GGCCCTGTACTTCCACTTCA	RT-PCR analysis of *SCAB12101* (forward)	138
JL49	GTCGTGTCGATCAGCACCT	RT-PCR analysis of *SCAB12101* (reverse)	138
JL50	ATCGTGATCGACGGTGTGT	RT-PCR analysis of *SCAB12121* (forward)	170
JL51	ATGGACTGTCCGCCTGAG	RT-PCR analysis of *SCAB12121* (reverse)	170
JL52	GAAGCAGTTCATCGACAGCA	RT-PCR analysis of *SCAB12131* (forward)	132
JL53	GTGTTGTTCAGTGCCAGCAG	RT-PCR analysis of *SCAB12131* (reverse)	132
JL54	ATCTACCGCTCCGTCATCC	RT-PCR analysis of *SCAB12141* (forward)	104
JL55	ATGGGGTAGTGGACGAAGTG	RT-PCR analysis of *SCAB12141* (reverse)	104
JL56	GGTGTGGAGTTCCGGTTCTA	RT-PCR analysis of *SCAB12151* (forward)	154
JL57	TGTACGGCTTGGAGACACAG	RT-PCR analysis of *SCAB12151* (reverse)	154
JL58	AGGGACTCATCCTCGATCTG	RT-PCR analysis of *SCAB12171* (forward)	118

JL59	ATGACGTCGTTGTTGAGACG	RT-PCR analysis of *SCAB12171* (reverse)	118
JL60	GTGCGACCTGTCTGACCTG	RT-PCR analysis of *SCAB12181* (forward)	127
JL61	GTCGGCAGCAGATGGTTC	RT-PCR analysis of *SCAB12181* (reverse)	127
JL80	GACCGGTGTCCGTCGCCT	Verification of *SCAB12081* deletion mutant (forward)	1108
JL81	AGCCGACAGGGGTTCGA	Verification of *SCAB12081* deletion mutant (reverse)	1108
JL82	CGATCCCGAAGAGGAAGACG	RT-PCR analysis of *SCAB11981* (forward)	144
JL83	TCGAAGACTCCAAGCCCAAC	RT-PCR analysis of *SCAB11981* (reverse)	144
JL84	CAGTACACGTCCTCCAGCAG	RT-PCR analysis of *SCAB11971* (forward)	149
JL85	TACGACGTGTCCTCGCTGAT	RT-PCR analysis of *SCAB11971* (reverse)	149
JL86	TGAAGGACGGGAAGAACAGC	RT-PCR analysis of *SCAB11961* (forward)	205
JL87	CAGTACTCCACCCGTATCGC	RT-PCR analysis of *SCAB11961* (reverse)	205
JL88	CGGCCGTTGATCTCGATGAT	RT-PCR analysis of *SCAB11921* (forward)	183
JL89	GCACTACCCGTACTTCGAGG	RT-PCR analysis of *SCAB11921* (reverse)	183
JL90	GTTGAGTCCGAAGAGCAGGG	RT-PCR analysis of *SCAB11871* (forward)	189
JL91	CCGGTGATCAGCCTGAAACT	RT-PCR analysis of *SCAB11871* (reverse)	189
JL92	AGCAGTTCGACGCTGTACTC	RT-PCR analysis of *SCAB11861* (forward)	199
JL93	CCGGATCGATGGTCCTCAAG	RT-PCR analysis of *SCAB11861* (reverse)	199
JL94	CCGTCCTCGTAGATGAAGGC	RT-PCR analysis of *SCAB11881* (forward)	222
JL95	GATGATCGGGATGGACTCCG	RT-PCR analysis of *SCAB11881* (reverse)	222

*Nonhomologous extensions are underlined, and the engineered restriction sites are indicated in bold.

2.2 General DNA methods

2.2.1 Digestion of DNA with restriction enzymes

Restriction digestion of DNA was performed according to the manufacturer's instructions. The volume of restriction enzyme always occupied no more than 10% v/v of the total reaction volume. Reactions were carried out in a total volume of 50 µL and were incubated at 37 °C for 2 hours or overnight (Table 2.5). When digesting plasmids with a single restriction enzyme, the 5′ ends of the digested DNA were dephosphorylated afterward by adding two units of shrimp alkaline phosphatase to prevent self-annealing during the ligation reactions. For subcloning, the digested DNA was recovered either by ethanol precipitation or by gel purification (Section 2.2.4).

Table 2.5 Composition of restriction digestion reactions performed in this study.

Reaction composition	**Volume**
10× Reaction Buffer (NEB)	5 µL
DNA	1 µg
Restriction enzyme (NEB)	1 µL (when multiple enzymes used, 1 µL of each was added, but the total volume of enzyme was ≤5 µL)
Water	To 50 µL final volume

2.2.2 Ligation of DNA into plasmids

T4 DNA ligase was used for ligation reactions as per the manufacturer's recommendations. Generally, the inserts to be ligated into the plasmid vectors were gel purified (See Sections 2.2.4) and quantified (Section 2.2.7). In instances where the sizes of the vector and insert were similar, the ligation reactions were set up using a 1:3 (vector: insert) molar ratio, and in all other instances,

a molar ratio of 1:1 was used. The ligation reactions were performed in a total volume of 10 µL (Table 2.6) and were incubated at 4°C overnight and then 16°C for 4 hours.

For ligation of DNA fragments into the pGEM-T Easy vector, the DNA fragment was first A-tailed using *Taq* DNA polymerase. A molar ratio of 1:3 (vector: insert) was used to clone the A-tailed DNA fragment into the vector following the manufacturer's instructions.

Table 2.6 Ligation reaction composition used in this study.

Reaction Composition	**Volume (µL)**
10× T4 DNA ligase Buffer (NEB)	1
Vector DNA	50 ng
Insert DNA	X
T4 DNA ligase (NEB)	1
Water	To 10 µL final volume

2.2.3 Agarose gel electrophoresis

Electrophoresis of DNA was performed using a 1-1.5% w/v agarose gel (depending on the size of the DNA), 1× TBE buffer (pH = 8.3) and a voltage of 90-130V for 40 min- 60 min. DNA samples (100 ng or 5 µL if the concentration was unknown) were mixed with 1 µL of 10× DNA gel loading dye and distilled water and were then loaded onto the gel. A 1kb DNA ladder (FroggaBio Inc., Toronto, ON) was used as a standard for band size estimation. DNA was visualized following electrophoresis; the gel was stained with ethidium bromide (2.0 µg/mL in 1× TBE buffer) for 30 min and was then destained for 10 min in 1× TBE buffer before being exposed to UV light and photographed. For RNA visualization, the gel was stained using GelRed® stain (Biotium), which was added to the molten agarose at 1× final concentration before gel casting (Section 2.3.2).

2.2.4 Gel purification of DNA

DNA was subjected to agarose gel electrophoresis as described in Section 2.2.3. After staining the gel with ethidium bromide, the gel was exposed to UV light, and a small slice of the gel containing the desired DNA fragment was excised using a scalpel and was transferred into a 1.5 mL microcentrifuge tube. The DNA was then purified from the gel slice using a commercial gel extraction kit (Section 2.1.1) and following the manufacturer's instructions.

2.2.5 Polymerase Chain Reaction (PCR)

PCR for routine cloning was performed using the proofreading enzyme Phusion DNA polymerase (Table 2.7), while *Taq* DNA polymerase was used for other routine PCR (Table 2.8). PCR used for the construction of mutant cosmids with the Redirect PCR targeting system were conducted using either enzyme (Table 2.9; Table 2.10), depending on which one worked better. A water control was included when performing every PCR.

Table 2.7 Typical PCR conditions using NEB Phusion High-Fidelity DNA polymerase.

Reaction Composition	**Volume**	**Final concentrations**
5× Phusion GC Buffer (containing $MgCl_2$)	10 µL	1×
dNTPs (10 mM each)	1 µL	200 µM
Forward primer (10 pmol/µL)	2.5 µL	0.5 µM
Reverse primer (10 pmol/µL)	2.5 µL	0.5 µM
DMSO	2.5 µL	5%
DNA template	50-100 ng	50-100 ng
Phusion Polymerase (2U/µL)	0.5 µL	1.0 units/50 µl PCR
Water	To 50 µL final volume	

Thermal cycling conditions:

Initial denaturation step: 98°C, 30 s (3 min for genomic DNA)
Then 30-35 cycles of:
    Denaturation: 98°C, 10 s
    Annealing: 60°C, 30 s
    Extension: 72°C, X s (30 s/kb)
Final extension: 72°C, 5 min
Hold: 4°C

Table 2.8 Typical PCR conditions using NEB *Taq* DNA polymerase.

Reaction Composition	Volume	Final concentrations
10× Taq Buffer (containing $MgCl_2$)	2.5 µL	1×
dNTP mix (10 mM each)	0.5 µL	200 µM
Forward primer (10 pmol/µL)	0.5 µL	0.2 µM
Reverse primer (10 pmol/µL)	0.5 µL	0.2 µM
DMSO	1.25 µL	5%
DNA template	50-100 ng	50-100 ng
Taq Polymerase (5U/µL)	0.125 µL	0.625 units/25 µl PCR
Water	To 25 µL final volume	

Thermal cycling conditions:

Initial denaturation step: 95°C, 30 s (3 min for genomic DNA)
Then 30-35 cycles of:
    Denaturation: 95°C, 30 s
    Annealing: 55°C, 30 s
    Extension: 68°C, X s (1 min/kb)
Final extension: 68°C, 5 min
Hold: 4°C

Table 2.9 Redirect PCR conditions using NEB Phusion High-Fidelity DNA polymerase.

Reaction Composition	Volume	Final concentrations
5× Phusion GC Buffer (containing $MgCl_2$)	10 µL	1×
dNTPs (10 mM each)	1 µL	200 µM
Forward primer (10 pmol/µL)	2.5 µL	0.5 µM
Reverse primer (10 pmol/µL)	2.5 µL	0.5 µM
DMSO	2.5 µL	5%
DNA template	50-100 ng	50-100 ng
Phusion Polymerase (2U/µL)	0.5 µL	1.0 units/50 µl PCR
Water	To 50 µL final volume	

Thermal cycling conditions:

Initial denaturation step: 98°C, 2 min
Then 10 cycles of:
    Denaturation: 98°C, 45 s
    Annealing: 50°C, 45 s
    Extension: 72°C, X s (30 s/kb)
Then 15 cycles of:
    Denaturation: 98°C, 45 s
    Annealing: 55°C, 45 s
    Extension: 72°C, X s (30 s/kb)
Final extension: 72°C, 5 min
Hold: 4°C

Table 2.10 Redirect PCR conditions using NEB *Taq* DNA polymerase.

Reaction Composition	Volume	Final concentrations
10× *Taq* Buffer (containing $MgCl_2$)	5 µL	1×
dNTP mix (10 mM each)	1 µL	200 µM
Forward primer (10 pmol/µL)	1 µL	0.2 µM
Reverse primer (10 pmol/µL)	1 µL	0.2 µM
DMSO	2.5 µL	5%
DNA template	50-100 ng	50-100 ng
Taq Polymerase (5U/µL)	0.25 µL	1.25 units/50 µl PCR
Water	To 50 µL final volume	

Thermal cycling conditions:

Initial denaturation step: 95°C, 2 min
Then 10 cycles of:
    Denaturation: 95°C, 45 s
    Annealing: 50°C, 45 s
    Extension: 68°C, X s (1 min/kb)
Then 15 cycles of:
    Denaturation: 95°C, 45 s
    Annealing: 55°C, 45 s
    Extension: 68°C, X s (1 min/kb)
Final extension: 68°C, 5 min
Hold: 4°C

2.2.6 DNA quantification and sequencing

DNA samples were quantified using a nanophotometer (Section 2.1.1) following the manufacturer's instructions. DNA sequencing was conducted at The Centre for Applied Genomics (TCAG) in Toronto, Canada. The DNA samples for sequencing were prepared following Sanger Sequencing Sample Submission Guidelines posted on their website (http://www.tcag.ca/facilities/dnaSequencingSynthesis.html#3).

2.3 RNA methods

2.3.1 Growth and sampling

Seed cultures for RNA extraction were prepared by inoculating 50 µL of a *Streptomyces* spore stock (see Section 2.5.2) into 5 mL of TSB (Appendix 1) followed by incubation with shaking at 28 °C for 24-48 hours until dense mycelial growth was obtained. An aliquot of the seed cultures (50 µL) was then spread onto the surface of ISP-4, SFM, SA, MYM_m and YMS_m agar media (Appendix 1) overlaid with sterile cellophane disks. The plates were incubated at 28°C for 44 hours, after which the mycelia (100-200 mg) were collected from the cellophane disks using a sterile flat-edge spatula. The mycelia were transferred and to 2 mL microcentrifuge tubes containing beads (Analytik Jena AG, Germany) and were flash-frozen in a dry ice-ethanol bath prior to storage at -80°C.

2.3.2 RNA extraction and DNA removal

Total RNA was extracted from frozen *Streptomyces* mycelia using a commercial kit and tissue homogenizer (Section 2.1.1) as per manufacturer's protocol. After the extraction, the total RNA was treated with RNAse-free DNase I as per the manufacturer's instructions except that the RNA was treated with DNase I twice. The RNA was then precipitated as described before (Sambrook and Russell, 2001) using NaOAc and ethanol and was kept overnight at -20°C. The RNA was pelleted by centrifugation at 14,000 rpm for 10 min, and the pellet was rinsed with 500 µL of 70% ethanol, centrifuged at 14,000 rpm for 1 min, and allowed to air dry for 15 min at room temperature. Finally, the RNA was resuspended in 50 µL of RNase-free water. The concentration of RNA was determined using a nanophotometer as per the manufacturer's instructions, and the

RNA integrity was confirmed by electrophoresis using a 1.2% w/v agarose gel. The RNA was stored at -80°C for future use.

For electrophoresis of RNA, the gel tank was first cleaned with 0.5% w/v SDS (sodium dodecyl sulfate), rinsed with RNAse-free water, and then rinsed with 100% v/v ethanol and allowed to air dry. The 1×TBE buffer (Tris-borate-EDTA) buffer and 1.2% w/v RNase-free agarose gel was prepared using RNAse-free water, and the RNA samples were prepared by adding 1 µL of 6× purple gel loading dye. The RNA was visualized following electrophoresis using GelRed® stain (Biotium), which was added to the molten agarose at 1× final concentration before gel casting.

2.3.3 Reverse Transcriptase PCR (RT-PCR)

The SuperScript™ III or IV First-Strand Synthesis System (Invitrogen, Canada) was used to synthesize complementary DNA (cDNA) as per the manufacturer's instructions using 2 µg of DNase-treated total RNA and 50 ng of random hexamers per reaction. The cDNA was then used as a template for PCR amplification using *Taq* DNA polymerase (New England Biolabs, Canada) and using primers listed in Table 2.4. A negative control (no RT) reaction in which no reverse transcriptase enzyme was added was included to verify the absence of genomic DNA in each of the RNA samples. The *murX* gene was used as a reference gene in this study. Typical PCR reaction conditions used are shown in Table 2.11.

Table 2.11 Typical PCR conditions using cDNA as template.

Solution	Volume (µL)	Final concentrations
10× Taq Buffer	2	1×
dNTP mix (10 mM each)	0.5	200 µM
Forward primer (10 pmol/µL)	1	0.5 µM
Reverse primer (10 pmol/µL)	1	0.5 µM
DMSO	1	5%

cDNA template	2	-
Taq Polymerase (NEB) 5U/µL	0.125	0.625 units/20 µl PCR
Water	To 20 µL final volume	

Thermal cycling conditions:

Initial denaturation step: 95°C, 2 min
Then 18-30* cycles of:
Denaturation: 95°C, 15 s
Annealing: 60°C, 30 s
Extension: 68°C, 15 s
Hold: 4°C

* The appropriate cycle number was determined for each primer set used

2.4 Growth, manipulation and storage of *E. coli*

2.4.1 Growth conditions and storage of *E. coli* strains

E. coli strains were cultured at 37°C unless otherwise indicated. Liquid cultures were routinely grown with shaking (180-200 rpm) in LB broth, low salt LB broth, SOB or SOC, while solid cultures were grown on LB or low salt LB agar plates (Appendix 1). When necessary, liquid and solid media were supplemented with antibiotics at the concentrations listed in Table 2.1. Low salt LB medium was used exclusively with hygromycin B since the activity of this antibiotic is negatively affected by salt concentration. All *E. coli* strains were kept at 4°C on agar media for short-term storage, or were maintained as stocks in 20% v/v glycerol at -80°C for long-term storage (Sambrook and Russell, 2001).

2.4.2 Isolation of plasmid and cosmid DNA from *E. coli*

The alkali lysis method (Sambrook and Russell 2001) was routinely used to isolate the plasmid and cosmid DNA from overnight liquid cultures of *E. coli.* When highly pure DNA was required, a commercial kit (Section 2.1.1) was used for DNA extraction following the manufacturer's protocol.

2.4.3 Preparation of chemically competent *E. coli* cells

Competent cells were prepared according to the protocol described by Inoue and colleagues (Inoue et al. 1990). Briefly, a seed culture was prepared by inoculating *E. coli* cells from a glycerol stock into 3 mL of LB broth in a 15 mL test tube (with appropriate antibiotics if necessary) and then incubating overnight. The overnight culture (500 µL) was sub-cultured into 50 mL of SOB in a 250 mL flask, and the culture was incubated until the OD_{600} was ~ 0.4-0.6. The culture was chilled on ice for at least 10 min, after which the cells were pelleted at 4000rpm for 10 min at 4° C. The pellet was resuspended gently in 20 mL of ice-cold TB buffer (Appendix 1) and was kept on ice for 10 min. Next, the cells were pelleted at 4000 rpm for 10 min at 4°C, and were then resuspended in 2.33 mL of TB buffer + 175 µL of DMSO. The suspension was kept on ice for another 10 min, and was then aliquoted into ice-cold 1.5 mL microcentrifuge tubes (100 µL per tube). The tubes were frozen immediately in liquid N_2 and stored at -80°C.

2.4.4 Transformation of DNA into chemically competent *E. coli* cells

Transformations were performed according to the protocol described by Inoue and colleagues (Inoue et al. 1990). Briefly, a tube of chemically competent *E. coli* cells (100 µL per tube) was thawed on ice for 10 min, and 50 µL aliquots of cells were then transferred to pre-chilled tubes. Pure plasmid DNA (5 ng) or a ligation reaction mix (5 µL) was added to each tube and was gently mixed. The cell/DNA mixtures were incubated on ice for 30 min and then heat-shocked at 42°C for 45 sec. Immediately, each tube was placed back on ice for at least 2 min, after which SOC medium (800 µL) was added to each tube, which were incubated with shaking for 1 hour at 37°C. One hundred microliters of each cell suspension was spread onto an LB agar plate containing the appropriate antibiotic(s). For transformations involving the pGEM®-T Easy plasmid, X-Gal (40 µL of a 2% w/v solution) and IPTG (100 µL of a 100 mM solution) were each spread onto the

LB agar plates prior to inoculation with the transformed cells to enable blue-white selection. The remaining cells in the tubes were pelleted and resuspended in 100 ul of SOC medium, and the cells were spread onto a second LB agar plate (with appropriate antibiotics). The plates were then incubated at 37°C for 12-16 h.

2.4.5 Preparation of electrocompetent *E. coli* cells

Electrocompetent cells of *E. coli* BW25113/pIJ790 (Table 2.2) containing Cosmid 2194 (Table 2.3) were prepared according to the protocol described by Fu and colleagues with slight modifications (Fu et al. 2010). Briefly, an overnight culture (500 µL) was inoculated into 50 mL of SOB (without added $MgSO_4$) containing chloramphenicol and kanamycin (Table 2.1) in a 250 mL flask. L-Arabinose (500 µL of a 1 M solution; final concentration is 10 mM) was then added to induce the expression of the λRed recombinase enzymes, and the culture was incubated with shaking at 28°C until the OD_{600} was ~ 0.4-0.6. The cells were immediately transferred to a 50 mL ice-cold centrifuge tube and were recovered by centrifugation at 4000 rpm for 5 min at 4°C. After decanting the medium, the cell pellet was washed twice with ice-cold 10% v/v glycerol (50 mL for the first wash, 25 mL for the second wash) and was resuspended in the remaining 100 µL of 10% v/v glycerol. The cells were then used immediately for electroporation.

2.4.6 Electroporation of DNA into electrocompetent *E. coli* cells

Electroporation was performed following the protocol described by Fu and colleagues with slight modifications (Fu et al. 2010). Briefly, about 100 ng (0.5-2 µL) of DNA was added to a pre-chilled electroporation cuvette (1 mm gap, VWR International) and 100 µL of electrocompetent *E. coli* cells were transferred into the cuvette. The cells and DNA were mixed gently by pipetting up and down or by lightly shaking the cuvette. Electroporation was performed using an

electroporator (Section 2.1.1) set to 200 Ω, 25 μF and 1.4 kV, with an expected time constant of 4.5 – 4.9 ms. One milliliter of ice-cold SOC (without $MgSO_4$) was immediately added to the shocked cells, and the cells were transferred to an ice-cold 1.5 mL microcentrifuge tube and were incubated at 37°C for 1h with shaking. Following incubation, the cells were plated onto LB agar plates containing appropriate antibiotics and were incubated at 37°C overnight.

2.5 Growth, manipulation and storage of *Streptomyces*

2.5.1 Growth condition and storage of *Streptomyces*

Streptomyces strains were routinely cultured at 28 °C on PMA plates or in TSB (Appendix 1) with shaking (200 rpm) in stainless steel spring flasks. Appropriate antibiotics were added to the medium when necessary (Table 2.1). Strains were maintained on agar plates at 4 °C for short-term storage, or as spore suspensions in 20% v/v glycerol at – 80 °C for long-term storage (Section 2.5.2).

For metabolite analysis, 100 μL of a spore stock was inoculated into 5 mL of TSB, and the culture was incubated for 24 hrs with shaking. Then, 100 μL of the TSB culture was spread onto YMSm), MYMm, and OBA (Appendix 1), and the plates were incubated at 28 °C for 14 days.

2.5.2 Preparation of *Streptomyces* spore stocks

For storage purposes (no quantification), a fast preparation procedure was used. Briefly, strains were grown on PMA plates (with appropriate antibiotics when necessary) for 7-10 days or until well sporulated. The spores were then scraped from the plates using a sterile spatula and were transferred into sterile 1.5 mL microcentrifuge tubes. Then, 0.5-1 mL of sterile 20% v/v glycerol (0.5-1 mL) was added to each tube. The spores were resuspended by vortexing for 1 min, and were then frozen at -80 °C.

When quantification of the spore stock was necessary, 3 mL of a sterile 0.01% v/v Tween-20 solution (in water) was added to the well-sporulated plate. The spores were gently scraped from the plate surface using a sterile wire loop to produce a suspension, and the suspension was then transferred to a sterile conical tube (15 mL). Another 3 mL of the Tween-20 solution was added to rinse the plate surface and the solution was then transferred to the conical tube. The tube was placed into a sonication bath for 5 minutes. The spore suspension was filtered through sterile cotton in a 10 mL syringe to remove the mycelia. The cotton was rinsed with sterile water until the filtrate was clear. The filtrate was centrifuged for 10 min at 3000 rpm to pellet the spores, and after decanting the supernatant, the spore pellet was resuspended in sterile 20% v/v glycerol (0.5-1 mL). The spore suspension was quantified by determining the number of colonies forming units (cfu) per mL of the suspension using the standard plate count method using NA (Appendix 1). The spore suspension was stored at -20°C (short term) or -80°C (long term).

2.5.3 Extraction of genomic DNA from *Streptomyces*

Strains were cultured in 10 mL of NB (with appropriate antibiotics when necessary) (Appendix 1) at 25 ºC for 2-4 days. The cultures were pelleted by centrifugation at 4500 rpm for 10 min at room temperature, and the mycelial pellets were used for genomic DNA extraction using a commercial kit (Section 2.1.1). The genomic DNA was quantified and then stored at – 20 °C.

2.5.4 Introduction of DNA from *E. coli* into *Streptomyces* by conjugation

Conjugations were performed following the protocol described by Kieser et al. (2000) with some modifications. Briefly, *E. coli* ET12567/pUZ8002, containing the plasmid/cosmid to be transferred, was incubated overnight at 37 °C in 5 mL of LB broth containing chloramphenicol and the plasmid/cosmid-specific antibiotic(s). Five hundred microliters of the overnight culture

were then sub-cultured into 50 mL of LB broth with the same antibiotics, and the culture was incubated at 37 °C until the OD_{600} reached 0.4-0.6. The cells were washed twice with 50 mL of LB broth to remove the antibiotics, after which they were resuspended in 500 µl of LB broth.

Streptomyces spores were collected from a single, well-sporulated PMA plate, and the spores were washed twice with LB broth. The spores (500 µL suspension in LB broth) were then mixed with an equal volume of *E. coli* cells, and the suspension was subjected to centrifugation at 4000 rpm for 5 min. The pellet was resuspended in ca. 500 µL of remaining LB, and the suspension was then plated on two MS agar + 10mM $MgCl_2$ plates (without antibiotics) (Appendix 1). The plates were incubated at 28°C for 16 h, after which they were overlaid with 1 mL of sterile water containing nalidixic acid (to kill *E. coli* selectively) and appropriate antibiotics (Table 2.1) to select for the incoming plasmid/cosmid.

2.5.5 Construction of *S. scabiei* engineered strains

2.5.5.1 Construction of *S. scabiei* gene deletion mutants

The Δ*SCAB12091,* Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*) and Δ*SCAB12081* mutant strains were constructed by replacing the target gene in the *S. scabiei* chromosome with a DNA cassette [*hyg*+*oriT*] conferring resistance to hygromycin B. This was accomplished using the previously described Redirect PCR targeting system (Gust et al. 2003a, 2004). The [*hyg* + *oriT*] extended resistance cassette was amplified from pIJ10700 by PCR using primers that contained 39 nt extensions that would allow replacement of the target *S. scabiei* gene(s) with the resistance cassette. The PCR products were gel-purified and then electroporated into *E. coli* BW25113/pIJ790 containing Cosmid 2194 to afford pJL6 (Δ*SCAB12091*/Cosmid 2194), pJL7 (Δ*SCAB12101*/Cosmid 2194) and pJL8 [Δ(*SCAB12091*+Δ*SCAB12101*)/Cosmid 2194],

respectively. The Δ*SCAB12081* mutant cosmid pRLDB49-4 (Table 2.3) was previously constructed by D. Bignell (unpublished results).

The mutant cosmids were introduced into *S. scabiei* 87-22 by intergeneric conjugation with *E. coli* (Section 2.4.4). Exconjugants that arose, which were hygromycin resistant, were then patched onto NA plates containing kanamycin in order to screen for those exconjugants that were kanamycin sensitive (and therefore had undergone a double crossover recombination event). Next, up to four hygromycin resistant and kanamycin sensitive exconjugants per mutant strain were streaked for single colonies onto ISP-4 containing hygromycin B and nalidixic acid and were incubated for 7- 10 days or until sporulating colonies were observed. Finally, spores from a single colony of each mutant isolate were used to prepare spore stocks for storage, as described in Section 2.5.2.

2.5.5.2 Construction of *S. scabiei* overexpression strains

Overexpression of *SCAB12111* was accomplished by constructing plasmids pJL3, pJL4, and pJL5 (Table 2.3), which express the *SCAB12111* gene from the strong, constitutive *ermE*p*, *kasO*p* and SP44 promoter, respectively, and integrate into the ΦC31 *attB* site within the *S. scabiei* chromosome. Briefly, the *SCAB12111* gene was PCR-amplified from Cosmid 2194 (Table 2.3) and cloned into the pGEM®-T Easy vector. The recombinant vector was confirmed by enzyme digestion and DNA sequencing. pGEM-T/*SCAB12111* was then digested with *Nde*I and *Not*I to release the *SCAB12111* gene, and the resulting DNA fragment was cloned into similarly digested pIJ8641, pJL1 and pJL2 to afford pJL3, pJL4, and pJL5, respectively. pIJ8641 carries the constitutive *ermE*p* promoter in front of the *egfp* gene, while pJL1 and pJL2 were constructed by replacing the *ermE*p* promoter with *kasO*p* and SP44, respectively (Table 2.3). A pIJ8641 derivative lacking the *egfp* gene (pJL15) was also constructed by digestion of pIJ8641 with *Nde*I

and *Not*I to release the *egfp* gene. The backbone of pIJ8641 without *egfp* was then gel-purified and blunted using the Quick Blunting™ Kit (New England Biolabs) following the manufacture's instructions. The blunted backbone of pIJ8641 was then re-ligated using T4 ligase to afford pJL15. The pJL3, pJL4 and pJL5 constructs were transferred into *S. scabiei* 87-22 by intergeneric conjugation with *E. coli*, and exconjugants were selected using apramycin. Control strains of *S. scabiei* were also generated by introducing the pIJ8641, pJL15, pJL1 and pJL2 plasmids by intergeneric conjugation.

Overexpression of the *SCAB12021-SCAB12071* genes was achieved by constructing three plasmids, pJL11, pJL12, and pJL13 (Table 2.3), which contained the *SCAB12021* coding sequence downstream of the *ermE*p*, *kasO*p* and SP44 promoter, respectively. Briefly, the *SCAB12021* gene was PCR-amplified and cloned into the pGEM®-T Easy vector. The *SCAB12021* gene was then released from the plasmid by digestion with *Nde*I and *Not*I and was cloned into similarly digested pJL16, pJL9, and pJL10 to afford pJL11, pJL12 and pJL13, respectively (Table 2.3). pJL16 was constructed by releasing the *ermE*p* from pIJ8641 and cloning it into the *Nde*I and *Eco*RV sites of pIJ8668, a non-integrative plasmid containing *egfp* (Table 2.3). pJL9 and pJL10 were constructed from pJL16 by replacing the *ermE*p* promoter with *kasO*p* and SP44, respectively. The recombinant vectors were then transferred into *S. scabiei* 87-22 by conjugation with *E. coli*.

Exconjugants were selected using apramycin (Table 2.1). A total of three apramycin-resistant exconjugants per engineered strain were streaked for single colonies onto ISP-4 containing apramycin and nalidixic acid and were incubated for 7-10 days or until sporulating colonies were observed. Finally, spores from a single colony of each mutant isolate were used to prepare spore stocks for storage, as described in Section 2.5.2.

2.5.5.3 Construction of heterologous BGC expression strains

Heterologous expression of BGCs was achieved by introducing a 5,247 bp DNA fragment containing the apramycin resistance gene (*aac(3)IV*), an origin of transfer (*oriT*), and the ΦC31 integrase (*int)* and attachment site (*attP*) into Cosmid 2194, which contains the phenazine-like BGC (Table 2.3). The [*aac(3)IV*+ *oriT*+ *int*+ *attP*] fragment was used to replace the kanamycin resistance gene on the cosmid backbone by Redirect PCR targeting and enables the cosmid to integrate into the ΦC31 *attB* site within the chromosome of the heterologous hosts *S. avermitilis, S. coelicolor* and *S. albus* (Table 2.2). To construct the integrative cosmid (pJL14; Table 2.3), pIJ10702 (also known as pMJCOS1) was digested with *Ssp*I to release the 5,247 bp [*aac(3)IV*+ *oriT*+ *int*+ *attP*] fragment. The fragment was gel-purified and then electroporated into *E. coli* BW25113/pIJ790 containing Cosmid 2194 to afford pJL14. Transformants were selected on LB agar containing ampicillin (to select for Cosmid 2194) and apramycin (to select for the integration of the [*oriT-attP-int-aac(3)IV*] fragment into the cosmid). pJL14 was then transferred into the heterologous hosts by intergeneric conjugation with *E. coli.* Apramycin resistant exconjugants (three per engineered strain) were streaked for single colonies onto ISP-4 containing apramycin and nalidixic acid and were incubated for 7-10 days or until sporulating colonies were observed. Finally, spores from a single colony of each mutant isolate were used to prepare spore stocks for storage, as described in Section 2.5.2.

2.5.5.4 Verification of constructed plasmids/cosmids and engineered strains

All constructed plasmids/cosmids were verified by PCR, restriction digestion, and/or DNA sequencing, and constructed engineered strains were all verified by PCR. The primers used for PCR verification are listed in Table 2.4.

2.5.6 Morphological analysis of engineered strains

To investigate the morphological phenotype of the constructed engineered *Streptomyces* strains, they (along with the corresponding controls) were cultured on different agar media (YMS/MYM/OBA/SFM) for 8 days at 28°C. Three replicate plates per strain per medium were prepared and incubated. Photographs of the plates were taken every day from Day 2 to Day 8.

2.6 Bioinformatics analysis

2.6.1 DNA and protein sequence analysis

DNA and predicted protein sequences of the putative phenazine-like BGC in *S. scabiei* were obtained from the StrepDB website (http://strepdb.*Streptomyces*.org.uk). DNA sequences and gene annotations were visualized using Geneious Pro software version 6.1.2 (Biomatters Ltd.). NCBI BLASTP (National Center for Biotechnology Information Basic Local Alignment Search Tool for proteins http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to identify the most similar proteins in the non-redundant (nr) database. The protein sequences of the biosynthetic genes from the putative phenazine-like BGC were used as query sequences in the BLASTP search. Protein domain analysis was conducted using the Pfam (https://pfam.xfam.org/), SMART (http://smart.embl-heidelberg.de/), and Prosite (https://prosite.expasy.org/) programs.

2.6.2 Phylogenetic methods

Amino acid sequence alignments were conducted using MUSCLE within the Geneious version 6.1.2 software. Alignments were subsequently concatenated for phylogenetic analysis of the phenazine-like biosynthetic genes. A model selection was performed for each partition within the concatenated alignment using MEGA 10.0.5 to identify the most suited model for distance estimation (*SCAB12021*: LG+G+I+F; *SCAB12031*: JTT+G; *SCAB12041*: LG+G+F; *SCAB12051*:

LG+G+F; *SCAB12061*: WAG+G+F; *SCAB12071*: LG+G+F; *SCAB12081*: LG+G+I+F; *SCAB12091*: LG+G+I+F; *SCAB12101*: LG+G+I+F; *SCAB12091+SCAB12101*: LG+G+I+F; *SCAB12111*: LG+G+F). Subsequently, MEGA 10.0.5 was employed to build phylogenetic trees using the concatenated gene alignments by the maximum likelihood method. The significance of the branching order in each tree was tested using the bootstrapping method with 1000 repetitions.

2.6.3 Annotation of specialized metabolite BGCs

The whole-genome sequence of *S. scabiei* 87-22 (NC_013929.1) was uploaded to antiSMASH 5.0 (Blin et al. 2019b) and DeepBGC 1.0 (Hannigan et al. 2019) to identify the specialized metabolite BGCs present within the genome using the default parameters.

2.6.4 Large-scale network analysis, classification, and phylogenetic analysis of BGCs

A total of 5890 BGCs and 188 genome sequences (187 from *Streptomycetaceae* family and one from *Pseudomonas aeruginosa* PAO1), were downloaded as .gbk files from the antiSMASH 5.0 database (Blin et al. 2019a). Sequence similarity networks and gene cluster families (GCFs) of these 5890 BGCs were generated using BiG-SCAPE with the default parameters (https://git.wur.nl/medema-group/BiG-SCAPE) (Navarro-Muñoz et al. 2020). Singletons, that is the BGCs with distances lower than the default cut off distance of 0.3, were also included in the networks. The Minimum Information about a Biosynthetic Gene cluster (MIBiG) database version 1.4 of annotated BGCs was included in this analysis (Kautsar et al. 2019). Network files generated by BiG-SCAPE were visualized using Cytoscape version 3.8.0 (Shannon et al. 2003), and BGC annotations were incorporated into Cytoscape networks by importing curated tables created by BiG-SCAPE.

The evolutionary relationships between these 5890 BGCs within and across GCFs were analyzed using CORASON software (https://github.com/nselem/corason) (Navarro-Muñoz et al. 2020). *SCAB12051* was used as a query gene and NC_013929.1_region.008.gbk (the butyrolactone BGC with the phenazine genes predicted by antiSMASH) was used as a query BGC. The multi-locus phylogenetic tree was constructed using CORASON with default parameters.

2.6.5 Whole-genome alignment using Mauve

Genome sequences used in this work were obtained from the Joint Genome Institute Integrated Microbial Genomes database (https://img.jgi.doe.gov/cgi-bin/er/main.cgi). The whole-genome alignment was conducted using the Progressive Mauve algorithm with default parameters to estimate the boundary of BGCs (Darling et al. 2010).

2.6.6 MultiGeneBlast against MIBiG database

MultiGeneBlast with default parameters was used to blast the unknown BGC against the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database to predict the closest compound encoded by the unknown BGC (Kautsar et al. 2019).

2.7 Metabolomics analysis of *Streptomyces* strains

2.7.1 Extraction of metabolites from *S. scabiei*

Specialized metabolites were extracted from whole plates of WT *S. scabiei* 87-22 and the Δ*SCAB12101* mutant cultured on YMSm, MYMm and OBA for 14 days. Briefly, each agar plate was cut into small pieces using a sterile pipette tip, and the agar pieces were transferred to a clean 250 mL flask. HPLC-grade ethyl acetate (20 mL) was then added to each flask, and the agar pieces were soaked in the ethyl acetate overnight. The ethyl acetate extracts were then transferred to clean evaporation flasks, and the agar pieces were washed with 10 mL of fresh ethyl acetate. The ethyl

acetate washes were then transferred to the corresponding evaporation flask, and the extracts were dried by rotary evaporation. The residual material was redissolved in 1 mL of HPLC grade methanol, and 500 µL of each extract was transferred to a 96-well plate (BD Biosciences, Canada) for LC-MS^2 analysis.

2.7.2 LC-MS^2 analysis of extracts

LC-MS^2 analysis of culture extracts was performed by the Dorrestein lab at the University of California San Diego. A 10 µL aliquot of each extract was analyzed using a Thermo Fisher Scientific Vanquish UHPLC System coupled to a Thermo Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer. Separation was conducted on a Scherzo SM-C18 column (2 × 250 mm, 3 µm, 130 Å; Imtakt, United States) maintained at 40°C using a water/acetonitrile gradient with 0.1% formic acid following the program from AbuSara et al. 2019 (AbuSara et al. 2019). Mass spectra were recorded in mixed mode by also following the MS setting from AbuSara et al. 2019 (AbuSara et al. 2019). The raw LC-MS^2 data files were converted into mzXML format using MSConvert for further analysis (Chambers et al. 2012)

2.7.3 Annotation and analysis of MS data from WT *S. scabiei* 87-22

The LC-MS^2 data from WT *S. scabiei* 87-22 in positive and negative ionization mode were first analyzed with MZmine2 (v2.53) for feature-based molecular networking (FBMN) analysis to generate three MS^2 mgf files and three quantification csv files: one in positive ionization mode, one in negative ionization mode, and another in mixed ionization mode (Nothias et al. 2020). The MZmine2 parameter settings are outlined in Table 2.12. The peak areas of the control (uninoculated medium) samples were manually subtracted from the corresponding test sample data before uploading to GNPS for FBMN (Pluskal et al. 2010; Wang et al. 2016; Nothias et al. 2020).

Cytoscape 3.7.2 (Shannon et al. 2003) was used to visualize the resulting molecular networks, and known metabolites were annotated by comparing the precursor ion mass and mass fragmentation pattern with the GNPS library (Wang et al. 2016), and by cross-checking with published results. Molecules were designated as isomers if they exhibited the same precursor ion mass and mass fragmentation patterns but differed in their retention times. Network Annotation Propagation (NAP) (da Silva et al. 2018) and the CFM-ID web server (Allen et al. 2014) were used to putatively annotate known compounds. SIRIUS (version 4.0.1) (Böcker et al. 2009) was used for molecular formula prediction, and MetWork (Beauxis and Genta-Jouve 2019) was used to predict the structures of unknown metabolites that were detected by spectral similarity analysis. Parameter settings used within SIRIUS, MetWork and CFM-ID are outlined in Table 2.13.

Recently, a new workflow for hierarchical clustering of MS^2 spectra in BioDendro was developed, which was also used to identify novel compounds (Rawlinson et al. 2020). The MS^2 mgf in mixed ionization mode and the quantification csv files exported from MZmine2 (v2.53) following the FBMN method were manually edited to meet the requirements for BioDendro, and the csv file was then converted to a txt file containing the feature list. The edited mgf file and the converted txt file were then submitted to BioDendro using a distance threshold of 0.6. The detailed parameter settings are outlined in Table 2.14. The resulting trees were visualized using Plotly (Plotly Technologies Inc., Dendrograms in Python).

The LC-MS^2 data of WT *S. scabiei* 87-22 are available on the MassIVE public repository under the accession number MSV000085858. The job links for WT *S. scabiei* 87-22 are available in Appendix 4.

Table 2.12 Parameter settings used within MZmine2 (v2.53) to generate the feature list for both BioDendro and FBMN.

Parameter	**Setting**
Mass detection	
Noise MS^1	Centroid; noise level 1.0E3
Noise MS^2	Centroid; noise level 1.0E2
Chromatogram builder	
Scans	MS level: 1
Min time span (min)	0.05
Min height	3.00E+03
m/z tolerance	0.01 *m/z* or 20.0 ppm
Chromatogram deconvolution	
Algorithm	Baseline cut-off
Min peak height	1.00E+03
Peak duration range (min)	0.00–2.00
Baseline level	5.00E+03
m/z center calculation	MEDIAN
Isotope Grouper	
m/z tolerance	0.01 *m/z* or 20.0 ppm
RT tolerance	0.25 absolute (min)
Monotonic shape	Checked
Maximum charge	4
Representative isotope	Most intense
Join aligner	
Peak list name	Aligned peak list
m/z tolerance	0.01 *m/z* or 20.0 ppm
Weight for *m/z*	0.8
Retention time tolerance	0.5 absolute (min)
Weight for RT	0.2
Require same charge state	Checked
Peak finder (multithreaded)	
Intensity tolerance	10.00%
m/z tolerance	0.01 *m/z* or 20.0 ppm
Retention time tolerance	0.5 absolute (min)
Peak list rows filter	
Minimum peaks in a row	3
Peak duration range	Checked 0.00–2.0
Keep only peaks with MS^2 scan (GNPS)	Checked
Reset the peak number ID	Checked
Duplicate peak filter	
Filter mode	OLD AVERAGE
m/z tolerance	0.005 *m/z* or 10.0 ppm
RT tolerance	0.5 absolute (min)
Export for/submit to GNPS	ONLY WITH MS^2

Table 2.13 Parameter settings used within SIRIUS (version 4.0.1), MetWork and CFM-ID (version 3.0).

Parameter	Setting
SIRIUS (version 4.0.1)	
Instrument	Orbitrap
ppm	5
Isotope Handling	Score
Candidates	10
Consider	Formulas from Bio databases
MetWork	
Depth limit	10
Cosine_mz_tolerance	0.02
Cosine_min_matched_peaks	2
Cosine_threshold	0.18
CFM-ID (version 3.0) (compound identification)	
Spectra Type	ESI
Scoring Function	Jaccard
Number of Results	10
Mass Tolerance	10 ppm

Table 2.14 Parameter settings used within BioDendro.

Parameter	Setting
*m/z*_tol	0.002
Retention_tol	5
Bin_threshold	0.0008
Scaling	False
Filtering	False
Eps	0.6
Neutral	False
Clustering method	Jaccard
Cutoff	0.6
Width	900
Height	1200

2.7.4 Combined analysis and annotation of MS data from WT *S. scabiei* 87-22 and the Δ*SCAB12101* mutant strain

Classical molecular networking was employed to generate molecular networks using positive and negative ionization mode data in GNPS (Wang et al. 2016). NAP was used to putatively annotate known compounds with default parameters (da Silva et al. 2018). Mass2Motifs (conserved fragments and neutral losses) were annotated using MS2LDA in GNPS platform (Van Der Hooft et al. 2016). Peptidic natural products were annotated using DEREPLICATOR in GNPS (Mohimani et al. 2017). The outputs from classical molecular networking, MS2LDA, NAP and DEREPLICATOR were combined using MolNetEnhancer (Ernst et al. 2019). Parameter settings used within classical molecular networking, MS2LDA and DEREPLICATOR are outlined in Table 2.15. The resulting networks were visualized in Cytoscape 3.8.0 (Shannon et al. 2003). BioDendro workflow was also employed in this analysis to create the hierarchical clustering tree of MS^2 spectra using the same parameters as described in Section 2.7.4 (Rawlinson et al. 2020).

The LC-MS^2 data of Δ*SCAB12101* and WT *S. scabiei* 87-22 are available on the MassIVE public repository under the accession number MSV000086630. The job links for Δ*SCAB12101* and WT *S. scabiei* 87-22 are available in Appendix 4.

Table 2.15 Parameter settings used within classical molecular networking, MS2LDA and DEREPLICATOR.

Parameter	Setting
Classical molecular networking	
Precursor Ion Mass Tolerance	0.02 Da
Fragment Ion Mass Tolerance	0.02 Da
Min Pairs Cos	0.6
Minimum Matched Fragment Ions	3
Network TopK	10
Minimum Cluster Size	1
Run MSCluster	Yes
Maximum Connected Component Size	100
Library Search Min Matched	3
Score Threshold	0.6
Search Analogs	Don't search
MS2LDA	
Bin Width:	0.005
Number of LDA Iterations	1000
Minimum MS^2 Intensity	100
LDA Free Motifs	300
Overlap score threshold	0.3
Probability value threshold	0.1
TopX in node	5
DEREPLICATOR	
Search Analogs (VarQuest)	Yes
Precursor Ion Mass Tolerance	0.02 Da
Fragment Ion Mass Tolerance	0.02 Da
PNP database	PNPdatabase
Max Charge	2
Min Number of AA	5
Max Allowed Modification Mass	150 Da
Min Matched Peaks with Known Compound	5

2.8 Bioassays

2.8.1 Potato tuber bioassays

The virulence phenotype of *Streptomyces* strains was assessed using a potato tuber bioassay, as described before (Loria 1995). *Streptomyces* strains were cultured on YMSm and MYMm agar for 7 days. Agar plugs from well-sporulated plates were inverted onto the tuber slices. The tuber

slices were incubated in a moist chamber at 22-25°C in the dark and were photographed after 10 days. The assay was conducted three times using three biological replicates per strain.

A potato tuber bioassay was also used to detect phytotoxic specialized metabolites from the organic culture extract of *Streptomyces* strains. The potato tuber slices were prepared as described before (Loria 1995). The specialized metabolites of *Streptomyces* were extracted as described in Section 2.7.1. Sterile Whatman paper disks (6 mm diameter) were placed onto each potato tuber slice, and then filter-sterilized extract (25 µL) was added to each disk. A positive control (where applicable) and a negative control (organic solvent) were included on each tuber slice. The tuber slices were incubated as described above. The paper disks were removed and the phytotoxic effects (e.g. tissue necrosis, pitting, hypertrophy, etc.) were compared and recorded. The assay was conducted three times using three biological replicates per extract.

2.8.2 Antimicrobial bioassays

The antimicrobial activity of the *Streptomyces* strains were tested using three methods: agar plug assay (Balagurunathan et al. 2020), disk diffusion assay (Balagurunathan et al. 2020) and the soft agar overlay assay (Zhang et al. 2020). Indicator strains included a Gram-positive bacterium, *Bacillus subtilis,* and a Gram-negative bacterium, *E. coli* ESS, as well as a fungal indicator strain, *Saccharomyces cerevisiae.*

In the agar plug method, the engineered *Streptomyces* strains along with their control strains were cultured on three different media, MYMm, YMSm and OBA, at 28 °C for 7 days. Then, triplicate agar plugs per strain/culture medium (6 mm diameter) were removed from the plates and were placed onto NA medium inoculated with the different indicator strains. Zones of inhibition around the agar plugs were measured after 24 h of incubation at 28 °C for *B. subtilis* and *S. cerevisiae* / 37 °C for *E. coli*.

In the disk diffusion method, the engineered *Streptomyces* strains along with their control strains were cultured on three different media, MYMm, YMSm and OBA , at 28 °C for 14 days, after which metabolites were extracted from the agar plates using ethyl acetate as described in Section 2.7.1. The ethyl acetate extracts were dried overnight in the fume hood, and the residual material was redissolved in 2 mL of methanol. The extracts (25 µL) were each added to a sterile filter paper disk (6 mm diameter) and the solvent was allowed to evaporate off. Then, the disks were placed onto the surface of NA medium previously inoculated with different indicator strains. Zones of inhibition around the paper disks were measured after 24 h of incubation at 28 °C for *B. subtilis* and *S. cerevisiae* / 37 °C for *E. coli*.

The soft agar overlay method was performed as describe by Zhang and colleagues (Zhang et al. 2020) with some modifications. Briefly, a small aliquot of *Streptomyces* spores was spotted onto YMSm and MYMm agar, and the plates were incubated for 4 days at 28 °C. Then, 15 mL of soft NA (0.75 % w/v agar) inoculated with 300 µL of an overnight culture of an indicator organism was overlayed over the YMSm and MYMm plates. The overlaid plates were incubated for 24 hours at 28 °C, and zones of inhibition were subsequently photographed.

CHAPTER 3: RESULTS

3.1 Bioinformatic analysis

3.1.1 Genomic analysis of specialized metabolite BGCs in *S. scabiei*

To examine the *S. scabiei* genome for specialized metabolite BGCs, two different genome mining tools were employed: antiSMASH 5.0, and DeepBGC 1.0. AntiSMASH predicted 34 putative specialized metabolite BGCs in the genome sequence, including eight terpene, six polyketide (PK), six nonribosomal peptide (NRP), five ribosomally-synthesized and post-translationally modified peptides (RiPP) and one hybrid PK-NRP BGC (Figure 3.1; Table 3.1). Ten of the predicted BGCs displayed high levels of similarity (≥70%) to BGCs in the MIBiG database, while five displayed moderate similarity (30-70%) and the remaining 19 showed low similarity (<30%) to known BGCs (Figure 3.1; Table 3.1). The combined length of the predicted BGCs is ca. 1,167 kb, accounting for ~11.5% of the *S. scabiei* genome. In contrast, DeepBGC identified 146 putative specialized metabolites BGCs for *S. scabiei*, of which 112 were not detected by antiSMASH (Figure 3.2; Table 3.2). The predicted BGCs include 15 PK, 11 RiPP, five NRP, four PK-Terpene, three terpene and one hybrid PK-NRP BGC, as well as 99 unclassified BGCs (Figure 3.2; Table 3.2). In addition, the predicted products of the BGCs include metabolites with antibacterial, cytotoxic, antibacterial-cytotoxic activities. The total length of these predicted BGCs according to DeepBGC is ca. 2,540 kb, making up ~25.03% of the genome.

It is noteworthy that neither antiSMASH nor DeepBGC classified the cryptic BGC that is the focus of this thesis as a phenazine BGC: antiSMASH predicted that the BGC is a butyrolactone BGC from *SCAB12041* to *SCAB12121* (region 8 in antiSMASH), and DeepBGC predicted it as an "Other" BGC from *SCAB12081* to *SCAB12101* (region 23 in DeepBGC) (Table 3.1; Table 3.2).

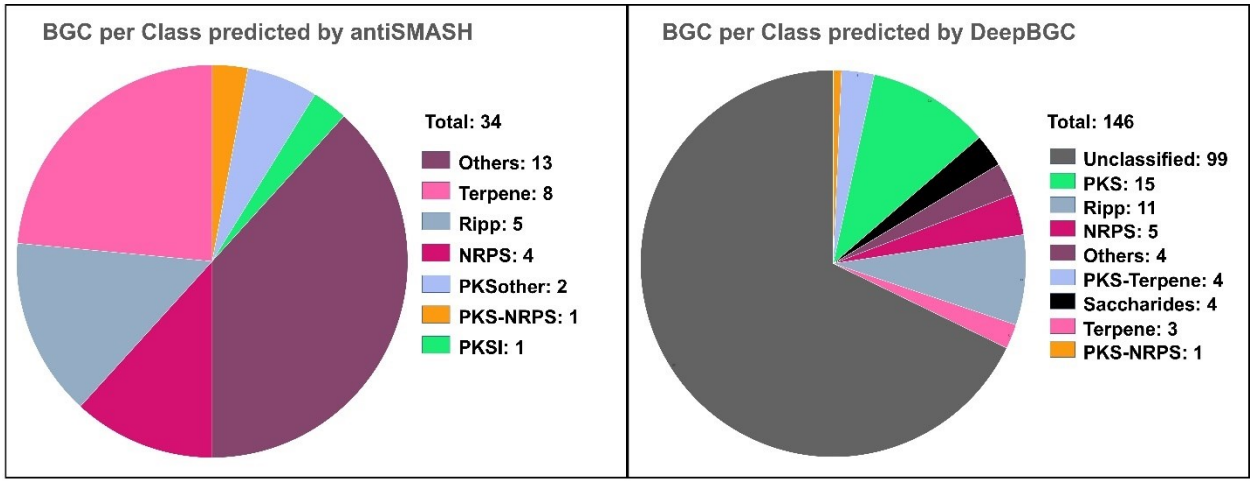


Figure 3.1 Classification of specialized metabolite BGCs predicted in the *S. scabiei* 87-22 genome using antiSMASH 5.0 (left) and DeepBGC (right).

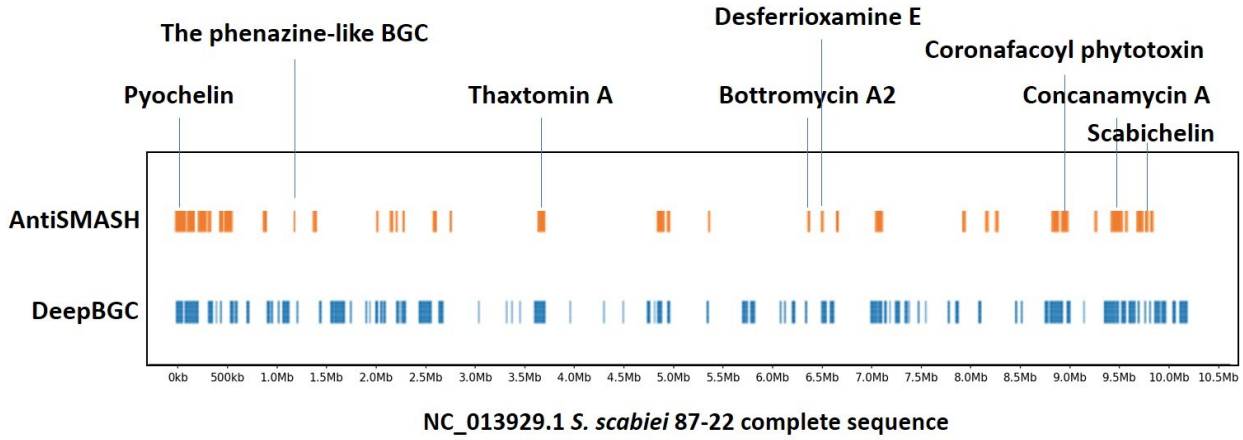


Figure 3.2 Comparison of specialized metabolite BGCs predicted in the genome of *S. scabiei* 87-22 using antiSMASH 5.0 (top) and DeepBGC (bottom). The relative location of the predicted BGCs on the chromosome by each tool is shown, and known BGCs that have been characterized and the phenazine-like BGC are labeled.

Table 3.1 Specialized metabolite BGCs predicted in the genome of *S. scabiei* 87-22 using antiSMASH 5.0.

BGC	BGC Type††	Start	Stop	Length (bp)	Most similar known cluster (% Similarity)†	MIBiG ID
1	NRPS/	122,258	207,140	84,883	Pyochelin (73%)	BGC0001801
2	**Betalactone**	210,499	245,699	35,201	Esmeraldin (8%)	BGC0000935
3	**NRPS**	331,948	400,458	68,511	Cadaside A/B (28%)	BGC0001968
4	**Lanthipeptide**	402,211	422,773	20,563	Labyrinthopeptin A1/A2/A3 (60%)	BGC0000519
5	**Terpene**	561,366	581,352	19,987	Ebelactone (5%)	BGC0001580
6	**Terpene**	602,733	628,185	25,453	Isorenieratene (100%)	BGC0001456
7	**Lanthipeptide, bacteriocin**	958,523	983,885	25,363	Informatipeptin (100%)	BGC0000518
8	**Butyrolactone**	1,354,553	1,365,035	10,483	Lactonamycin (5%)	BGC0000238
9	**Terpene**	1,445,117	1,469,780	24,664	Hopene (92%)	BGC0000663
10	**Siderophore**	2,067,195	2,079,101	11,907	Grincamycin (8%)	BGC0000229
11	**NRPS-like**	2,214,965	2,255,782	40,818	s56-p1 (11%)	BGC0001764
12	**Terpene**	2,273,305	2,294,341	21,037	Geosmin (100%)	BGC0001181
13	Bacteriocin /	2,343,879	2,354,292	10,414	Unknown	
14	Terpene /	2,615,998	2,636,826	20,829	FD-594 (8%)	BGC0000222
15	Siderophore/	2,789,747	2,799,722	9,976	Unknown	
16	**NRPS**	3,581,513	3,645,548	64,036	Thaxtomin A (50%)	BGC0000444
17	**T2PKS**	4,831,093	4,903,608	72,516	Spore Pigment (83%)	BGC0000271
18	**T1PKS, NRPS**	4,907,595	4,958,875	51,281	Herboxidiene (2%)	BGC0001065
19	**Lanthipeptide**	5,327,434	5,350,019	22,586	Unknown	
20	**Bottromycin, bacteriocin**	6,301,273	6,323,906	22,634	Bottromycin A2 (54%)	BGC0000469
21	Siderophore/	6,440,983	6,452,205	11,223	Desferrioxamine B/E (66%)	BGC0000940
22	**Melanin**	6,584,191	6,592,550	8,360	Melanin (80%)	BGC0000909
23	**T1PKS, Butyrolactone**	6,949,013	7,031,953	82,941	4-hexadecanoyl-3-hydroxy-2-(hydroxymethyl)-2H-furan-5-one (54%)	BGC0000140
24	**Ectoine**	7,828,037	7,838,435	10,399	Ectoine (100%)	BGC0000853
25	**NRPS-like**	8,050,515	8,093,153	42,639	Granaticin (5%)	BGC0000227
26	Terpene /	8,147,349	8,166,320	18,972	Unknown	
27	T1PKS, Indole/	8,700,909	8,763,617	62,709	5-isoprenylindole-3-carboxylate/β-D-glycosyl ester (28%)	BGC0001483

28	**T1PKS**	8,793,628	8,848,502	54,875	Lasalocid (22%)	BGC0001648
29	T3PKS /	8,870,608	8,911,792	41,185	Daptomycin (10%)	BGC0000336
30	Terpene /	9,109,516	9,130,190	20,675	Unknown	
31	**T1PKS**	9,288,913	9,413,896	124,984	Concanamycin A (89%)	BGC0000040
32	Siderophore /	9,430,360	9,443,981	13,622	Unknown	
33	**NRPS**	9,526,250	9,591,840	65,591	Scabichelin (100%)	BGC0000423
34	**Terpene**	9,618,596	9,639,636	21,041	Formicamycins A-M (4%)	BGC0001590

† % Similarity represents the percentage of genes in the query cluster which are present in the hit BGC from MIBiG

†† BGCs that were detected by both DeepBGC and AntiSMASH are indicated in bold font

Table 3.2 Specialized metabolite BGCs of *S. scabiei* 87-22 predicted by DeepBGC.

BGC	start	Stop	Length (bp)	Score	Activity	Class
1	18190	89325	71135	0.87099		
2	102567	242501	139934	0.90774	antibacterial	NRP
3	339862	386694	46832	0.9223	antibacterial	NRP
4	419778	424811	5033	0.57375	antibacterial	
5	460242	460860	618	0.53264	antibacterial	
6	464781	471708	6927	0.61423		
7	478224	478917	693	0.52238	antibacterial	
8	558035	559412	1377	0.55673		
9	560471	596943	36472	0.75848	antibacterial	RiPP
10	608874	622284	13410	0.79677	cytotoxic	Terpene
11	625737	631685	5948	0.75319	antibacterial	
12	722767	752342	29575	0.8646		
13	928431	959788	31357	0.78182	antibacterial	
14	968477	975355	6878	0.55394		RiPP
15	976085	988233	12148	0.697	antibacterial	RiPP
16	1041388	1043271	1883	0.56056	antibacterial	
17	1046732	1049002	2270	0.6328	antibacterial	Polyketide–Terpene
18	1069665	1070460	795	0.60303	antibacterial	
19	1082780	1151089	68309	0.88331	antibacterial	
20	1227544	1236579	9035	0.73594		
21	1262030	1262993	963	0.53046	antibacterial	
22	1271835	1273077	1242	0.64703	antibacterial	Polyketide–Terpene
23	1359263	1361084	1821	0.54029		Other
24	1452530	1462947	10417	0.76992		Terpene
25	1467705	1469780	2075	0.59669	antibacterial	Polyketide
26	1564735	1565578	843	0.50268	antibacterial	
27	1566887	1594637	27750	0.66358		
28	1599054	1615820	16766	0.67619	antibacterial	
29	1616648	1705872	89224	0.8619	antibacterial	
30	1761221	1772801	11580	0.64237		
31	1773733	1775613	1880	0.57162	antibacterial	
32	1909359	1911186	1827	0.53966	antibacterial	
33	1914029	1927879	13850	0.60088		
34	1951578	1956395	4817	0.66251		
35	2010412	2014771	4359	0.68677	antibacterial	Polyketide–Terpene
36	2016579	2035817	19238	0.75465	cytotoxic	
37	2057524	2079875	22351	0.73168		
38	2091914	2110560	18646	0.63221		

39	2217546	2219808	2262	0.51772	antibacterial	
40	2221477	2246953	25476	0.78499	antibacterial	
41	2254597	2264378	9781	0.62484	antibacterial	
42	2274616	2313073	38457	0.86707	antibacterial	Terpene
43	2438774	2440007	1233	0.52837	antibacterial	
44	2442681	2572684	130003	0.89718	antibacterial	
45	2642695	2690183	47488	0.86657		Saccharide
46	3040727	3049599	8872	0.71861	antibacterial	
47	3318981	3323908	4927	0.66275	antibacterial	
48	3372834	3377166	4332	0.59259	antibacterial	Other
49	3413899	3414799	900	0.51194	antibacterial	
50	3416297	3417527	1230	0.54599	antibacterial	Polyketide–Terpene
51	3419308	3419899	591	0.50372	antibacterial	
52	3447615	3448005	390	0.5786	antibacterial	
53	3450468	3454029	3561	0.648	antibacterial	
54	3455498	3457620	2122	0.65236	antibacterial	RiPP
55	3596329	3691476	95147	0.95036		NRP
56	3693118	3700314	7196	0.60355	antibacterial	Polyketide
57	3705494	3706967	1473	0.52383	antibacterial	
58	3956656	3957649	993	0.51556	antibacterial	
59	3984801	3987063	2262	0.57988	antibacterial	
60	4293306	4294014	708	0.50679		
61	4374160	4375063	903	0.52399	antibacterial	RiPP
62	4483338	4489440	6102	0.63171	antibacterial	RiPP
63	4558759	4559626	867	0.61354		Other
64	4726396	4757122	30726	0.87095	antibacterial	
65	4801003	4805993	4990	0.55578	antibacterial	
66	4824780	4874503	49723	0.87102	antibacterial–cytotoxic	Polyketide
67	4927379	4955605	28226	0.70212		NRP–Polyketide
68	5324269	5335478	11209	0.68262	antibacterial	RiPP
69	5676081	5710451	34370	0.84785	antibacterial	
70	5719497	5720355	858	0.58089	antibacterial	
71	5726193	5727498	1305	0.53724	antibacterial	
72	5729370	5735805	6435	0.55593	antibacterial	
73	5737051	5738020	969	0.50873	antibacterial	
74	5739707	5740640	933	0.53197	antibacterial	
75	5754384	5804300	49916	0.87958		
76	6055637	6066801	11164	0.68095		
77	6095670	6109369	13699	0.77797	antibacterial	Saccharide
78	6169809	6196846	27037	0.76409		Saccharide
79	6199153	6203896	4743	0.61044	antibacterial	Saccharide
80	6306672	6320207	13535	0.76854	antibacterial	RiPP

81	6354936	6355164	228	0.52503	antibacterial	
82	6467153	6472060	4907	0.57358		
83	6477734	6516964	39230	0.91582	cytotoxic	
84	6552902	6594640	41738	0.88769		
85	6886059	6887568	1509	0.53641	antibacterial	
86	6954173	7021172	66999	0.88977	antibacterial	Polyketide
87	7023681	7033123	9442	0.57824	antibacterial	
88	7038004	7071560	33556	0.74199	cytotoxic	
89	7073244	7074099	855	0.6867	antibacterial	
90	7094754	7099945	5191	0.5733		
91	7101471	7115200	13729	0.81514		Polyketide
92	7150674	7155389	4715	0.57942	antibacterial	
93	7201872	7202214	342	0.50319	antibacterial	
94	7203566	7249274	45708	0.80748		
95	7285360	7286879	1519	0.60602	antibacterial	
96	7287949	7288135	186	0.51186	antibacterial	
97	7299011	7321110	22099	0.71974	antibacterial	
98	7325660	7327918	2258	0.53438	antibacterial	
99	7340516	7340729	213	0.5655	antibacterial	
100	7341719	7346984	5265	0.52757		RiPP
101	7431017	7434375	3358	0.6133		
102	7437983	7443275	5292	0.53605	antibacterial	
103	7503514	7506157	2643	0.64196		
104	7508209	7509127	918	0.52999	antibacterial	RiPP
105	7670866	7672052	1186	0.65159	antibacterial	
106	7728476	7747536	19060	0.81518	antibacterial	Polyketide
107	7804675	7834761	30086	0.72796	antibacterial	
108	8033373	8059403	26030	0.64025		
109	8067096	8067822	726	0.5089	antibacterial	
110	8069773	8073649	3876	0.64876	antibacterial	NRP
111	8075685	8076998	1313	0.60864	antibacterial	
112	8404995	8416822	11827	0.62234	antibacterial	Polyketide
113	8459791	8466375	6584	0.5752	antibacterial	Polyketide
114	8467352	8471630	4278	0.66105	antibacterial	
115	8472529	8473180	651	0.52078	antibacterial	
116	8649331	8649556	225	0.69521	antibacterial	
117	8696279	8734257	37978	0.7999		Polyketide
118	8740827	8741916	1089	0.50152	antibacterial	
119	8744689	8759111	14422	0.63354	antibacterial	
120	8762953	8879380	116427	0.79835		Polyketide
121	8915934	8952852	36918	0.81443	cytotoxic	
122	9087945	9089987	2042	0.63454		
123	9287306	9294006	6700	0.64786		
124	9296256	9400948	104692	0.94098	antibacterial	Polyketide

125	9408135	9425419	17284	0.72038	antibacterial	
126	9427211	9438981	11770	0.60212		Other
127	9459239	9507403	48164	0.87056		
128	9537940	9538945	1005	0.50465		
129	9539837	9594734	54897	0.78684	antibacterial	NRP
130	9599393	9602653	3260	0.62584	antibacterial	
131	9628595	9629636	1041	0.55857	antibacterial	
132	9637730	9641105	3375	0.55909	antibacterial	
133	9696492	9703933	7441	0.67841		RiPP
134	9743520	9755172	11652	0.66772	cytotoxic	
135	9755757	9757020	1263	0.56562	antibacterial	
136	9783986	9787421	3435	0.55378	antibacterial	
137	9790145	9791327	1182	0.52215	antibacterial	Polyketide
138	9793130	9795526	2396	0.54378	antibacterial	Polyketide
139	9795992	9802438	6446	0.6239	antibacterial	
140	9807829	9848772	40943	0.78177		
141	9855232	9856093	861	0.55074	antibacterial	
142	9861950	9883242	21292	0.66367	antibacterial	
143	9884940	9908330	23390	0.72213		
144	9974070	10007752	33682	0.80345	antibacterial	Polyketide
145	10010729	10012040	1311	0.51486	antibacterial	Polyketide
146	10044582	10123583	79001	0.94919	antibacterial–cytotoxic	

3.1.2 Large-scale network analysis, classification, and phylogenetic analysis of BGCs

To investigate the evolution of the gene clusters within and across the gene cluster families (GCFs) in different organisms, a sequence similarity network consisting of 5890 BGCs as unique nodes from 188 genome sequences (187 from *Streptomycetaceae* family and one from *Pseudomonas aeruginosa* PAO1) was generated using BIG-SCAPE. BIG-SCAPE-predicted BGC classes included 915 RiPPs, 849 terpene, 818 NRPS, 445 PKS other (non-modular categories of PKSs), 373 PKS-NRPS hybrids, 364 PKSI (Type I PKS), 27 saccharides, and 2099 others (not belonging to any of the above-mentioned classes) (Figure 3.3). Of the 5890 BGCs analyzed, 1788 possess sequence similarities ≥75% with annotated BGCs in the MIBiG database, and 4102 possess sequence similarities <75% with annotated BGCs, which suggests that ~70% of these 5890

BGCs may produce metabolites that remain to be discovered. These 5890 BGCs were grouped into 431 GCFs (≥ three nodes), including 126 others GCFs, 90 NRPS GCFs, 60 RiPPs GCFs, 50 PKSother GCFs, 48 terpene GCFs, 36 PKSI GCFs, 14 PKS-NRPS hybrids and seven saccharides GCFs. The cryptic phenazine-like BGC from *S. scabiei* 87-22 (NC_013929.1 region.008.gbk) was classified in the "others" category. The sequence similarity network of BGCs classified as "others" is shown in Figure 3.4. The predicted product types within this category include siderophore, butyrolactone, melanin, ectoine, β-lactone, ladderane, arylpolyene, β-lactam, phenazine and other.

Out of 34 BGCs identified in *S. scabiei* 87-22 by antiSMASH, nine BGCs displayed similarity ≥75% with annotated BGCs in the MIBiG database, 26 BGCs associated with other BGCs, and eight BGCs are singletons (Figure 3.5). Region 24 (ectoine), 21 (desferrioxamine E), 15, 9 (hopene), 13 and 17 (spore pigment) form six large networks (>100 nodes) with other BGCs. *S. scabiei* 87-22 Region 8 (the phenazine-like BGC), *S. bottropensis* region 023, *S. stelliscabiei* region 001, and *Streptomyces* sp.1222.2 region 004 were grouped into one GCF and the phylogenetic relationships of these four BGCs within the GCF are shown in Figure 3.6.

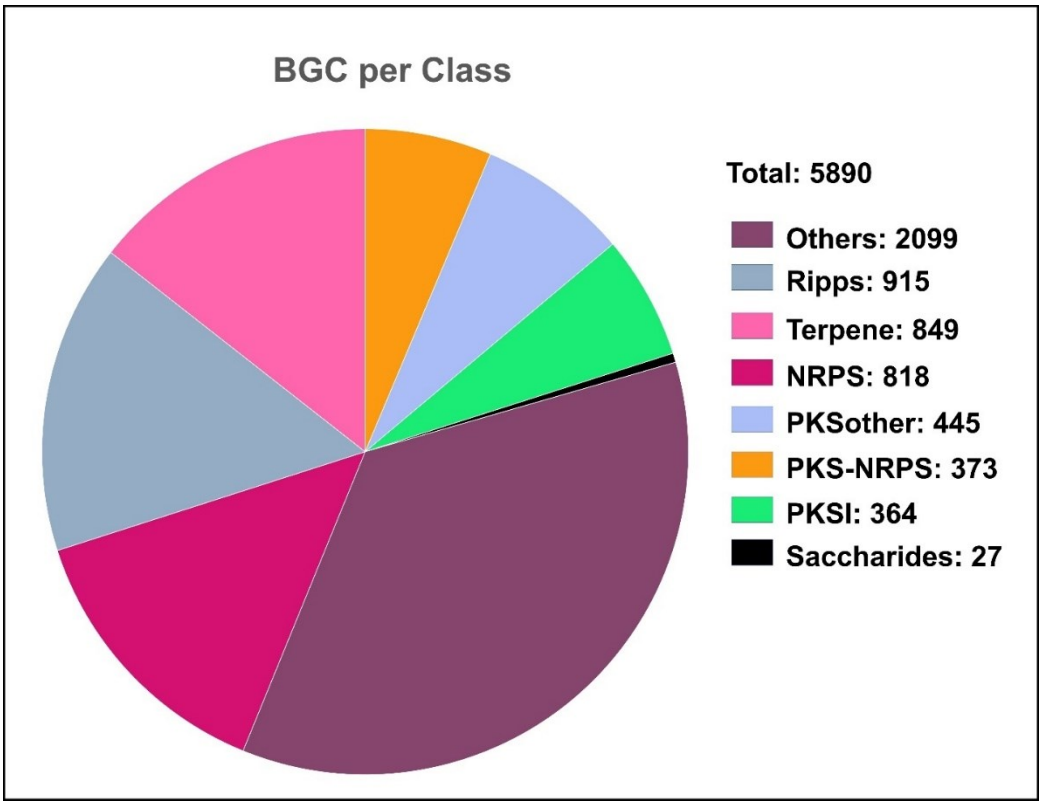


Figure 3.3 Classification of 5890 specialized metabolite BGCs from 188 genomes using BiG-SCAPE.

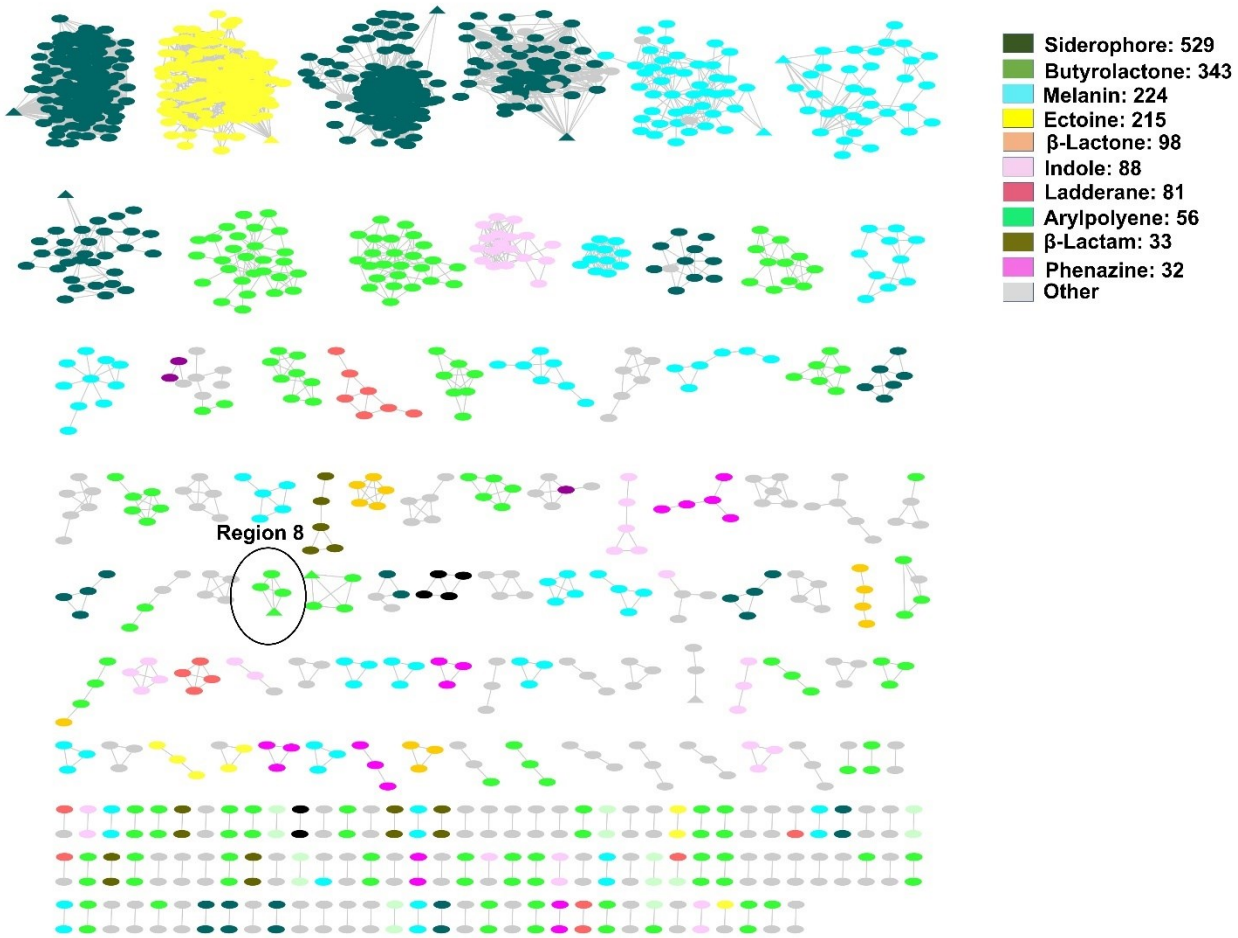


Figure 3.4 Sequence similarity network of BGCs classified as "Others" in the antiSMASH 5.0 database. Predicted product types were annotated with different colors, and the total number of BGCs for each product type is indicated. Triangular nodes represent BGCs from *S. scabiei*, and circular nodes represent MIBiG and other BGCs from the antiSMASH database. The cluster with the phenazine-like BGC from *S. scabiei* 87-22 (NC_013929.1 region.008.gbk) is circled.

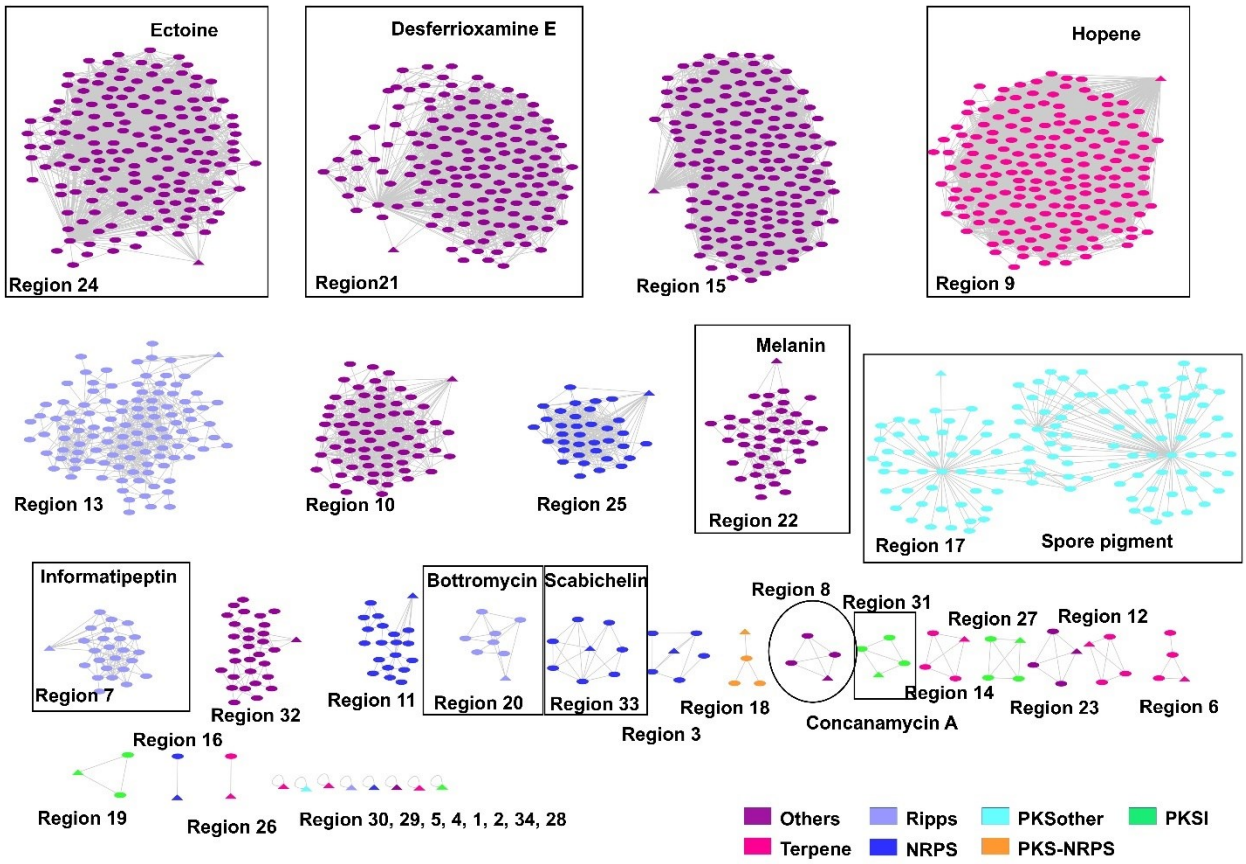


Figure 3.5 Sequence similarity network of the 34 BGCs identified in *S. scabiei* 87-22 by antiSMASH 5.0 compared against BGCs in the MIBiG database and in other strains. Triangular nodes represent BGCs from *S. scabiei*, and circular nodes represent MIBiG and other *Streptomyces* BGCs. Boxed clusters represent clusters containing at least one BGC with sequence similarity ≥ 75% to a characterized BGC deposited in the MIBiG database. The cluster with the phenazine-like BGC from *S. scabiei* 87-22 (NC_013929.1 region.008.gbk) is circled. Different BGC families were annotated by different colors.

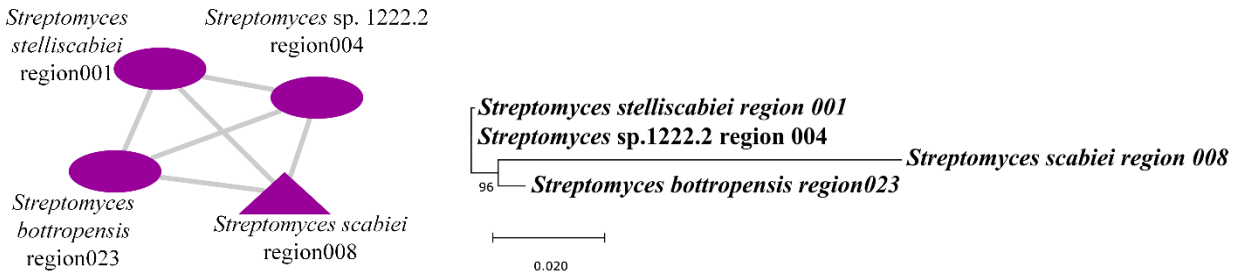


Figure 3.6 The similarity network of *S. scabiei* 87-22 region 008 (the phenazine-like BGC) with BGCs from other *Streptomyces* species (left) and their evolutionary relationships within the GCF (right).

To study the evolutionary relationships of *S. scabiei* 87-22 region 008 (the phenazine-like BGC) across different GCFs, CORASON was employed using *SCAB12051* as a query gene and NC_013929.1_region.008.gbk as a query BGC. A multi-locus phylogenetic tree was constructed, which showed the presence of *SCAB12051* in 52 BGCs across different GCFs with ~33% of these BGCs being phenazines (Figure 3.7). *S. scabiei* 87-22 region 008 formed a distinct clade with *S. bottropensis* region 023, *S. stelliscabiei* region 001, *Streptomyces* sp. 1222.2 region 004, *Kitasatospora* sp. Root107 region 004 and *Kitasatospora* sp. MMS16-BH015 region 001.

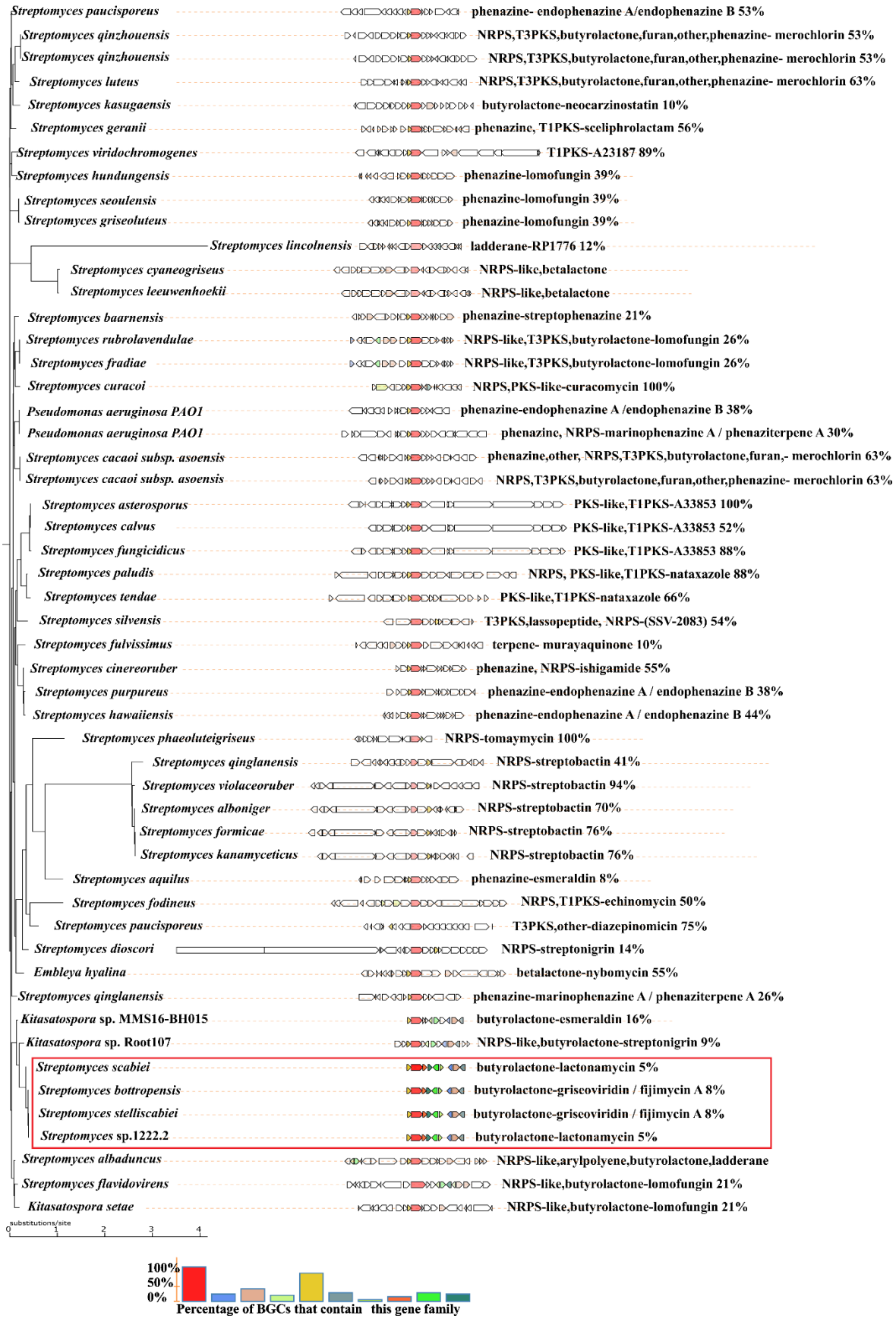

Streptomyces paucisporeus
phenazine- endophenazine A/endophenazine B 53%
Streptomyces qinzhouensis
NRPS,T3PKS,butyrolactone,furan,other,phenazine- merochlorin 53%
Streptomyces qinzhouensis
NRPS,T3PKS,butyrolactone,furan,other,phenazine- merochlorin 53%
Streptomyces luteus
NRPS,T3PKS,butyrolactone,furan,other,phenazine- merochlorin 63%
Streptomyces kasugaensis
butyrolactone-neocarzinostatin 10%
Streptomyces geranii
phenazine, T1PKS-sceliphrolactam 56%
Streptomyces viridochromogenes
T1PKS-A23187 89%
Streptomyces hundungensis
phenazine-lomofungin 39%
Streptomyces seoulensis
phenazine-lomofungin 39%
Streptomyces griseoluteus
phenazine-lomofungin 39%
Streptomyces lincolnensis
ladderane-RP1776 12%
Streptomyces cyaneogriseus
NRPS-like,betalactone
Streptomyces leeuwenhoekii
NRPS-like,betalactone
Streptomyces baarnensis
phenazine-streptophenazine 21%
Streptomyces rubrolavendulae
NRPS-like,T3PKS,butyrolactone-lomofungin 26%
Streptomyces fradiae
NRPS-like,T3PKS,butyrolactone-lomofungin 26%
Streptomyces curacoi
NRPS,PKS-like-curacomycin 100%
Pseudomonas aeruginosa PAO1
phenazine-endophenazine A /endophenazine B 38%
Pseudomonas aeruginosa PAO1
phenazine, NRPS-marinophenazine A / phenaziterpene A 30%
Streptomyces cacaoi subsp. asoensis
phenazine,other, NRPS,T3PKS,butyrolactone,furan,- merochlorin 63%
Streptomyces cacaoi subsp. asoensis
NRPS,T3PKS,butyrolactone,furan,other,phenazine- merochlorin 63%
Streptomyces asterosporus
PKS-like,T1PKS-A33853 100%
Streptomyces calvus
PKS-like,T1PKS-A33853 52%
Streptomyces fungicidicus
PKS-like,T1PKS-A33853 88%
Streptomyces paludis
NRPS, PKS-like,T1PKS-nataxazole 88%
Streptomyces tendae
PKS-like,T1PKS-nataxazole 66%
Streptomyces silvensis
T3PKS,lassopeptide, NRPS-(SSV-2083) 54%
Streptomyces fulvissimus
terpene- murayaquinone 10%
Streptomyces cinereoruber
phenazine, NRPS-ishigamide 55%
Streptomyces purpureus
phenazine-endophenazine A / endophenazine B 38%
Streptomyces hawaiiensis
phenazine-endophenazine A / endophenazine B 44%
Streptomyces phaeoluteigriseus
NRPS-tomaymycin 100%
Streptomyces qinglanensis
NRPS-streptobactin 41%
Streptomyces violaceoruber
NRPS-streptobactin 94%
Streptomyces alboniger
NRPS-streptobactin 70%
Streptomyces formicae
NRPS-streptobactin 76%
Streptomyces kanamyceticus
NRPS-streptobactin 76%
Streptomyces aquilus
phenazine-esmeraldin 8%
Streptomyces fodineus
NRPS,T1PKS-echinomycin 50%
Streptomyces paucisporeus
T3PKS,other-diazepinomicin 75%
Streptomyces dioscori
NRPS-streptonigrin 14%
Embleya hyalina
betalactone-nybomycin 55%
Streptomyces qinglanensis
phenazine-marinophenazine A / phenaziterpene A 26%
Kitasatospora sp. MMS16-BH015
butyrolactone-esmeraldin 16%
Kitasatospora sp. Root107
NRPS-like,butyrolactone-streptonigrin 9%
Streptomyces scabiei
butyrolactone-lactonamycin 5%
Streptomyces bottropensis
butyrolactone-griseoviridin / fijimycin A 8%
Streptomyces stelliscabiei
butyrolactone-griseoviridin / fijimycin A 8%
Streptomyces sp.1222.2
butyrolactone-lactonamycin 5%
Streptomyces albaduncus
NRPS-like,arylpolyene,butyrolactone,ladderane
Streptomyces flavidovirens
NRPS-like,butyrolactone-lomofungin 21%
Kitasatospora setae
NRPS-like,butyrolactone-lomofungin 21%
substitutions/site
0
1
2
3
4
100%
50%
0%
Percentage of BGCs that contain this gene family

Figure 3.7 CORASON phylogeny of phenazine-like BGCs (NC_013929.1_region.008.gbk) from *S. scabiei* 87-22. CORASON phylogenetic reconstruction with *SCAB12051* (*phzE*) as the query gene and the NC_013929.1_region.008.gbk as query cluster. The type of the BGC and the percentage of similarity to known BGC in MIBiG database were shown at the end of each BGCs. Boxed sections on the tree correspond to BiG-SCAPE-defined families (Figure 3.4; Figure 3.5).

3.1.3 Whole–genome alignment using Mauve

To predict the boundaries of the cryptic phenazine-like BGC in *S. scabiei*, whole–genome alignment by the progressive Mauve algorithm was employed (Darling et al. 2010). Genome sequences used in this work were obtained from the Joint Genome Institute Integrated Microbial Genomes database (Table 3.3). These species were chosen based on the BiG-SCAPE and CORASON results (see Section 3.1.2). Whole–genome alignment of *S. scabiei* with *S. stelliscabiei*, *Streptomyces* sp. 1222.2 and *S. bottropensis* revealed the presence of a 183 kb conserved region from *SCAB11061* to *SCAB12211* (Figure 3.8). However, there was a non-homologous gap in this conserved region, which separates the region into two parts: one is from *SCAB11061* to *SCAB11841* and the other is from *SCAB11861* to *SCAB12211*. The region from *SCAB11861* to *SCAB12211* contains the phenazine biosynthetic genes, which could be the boundaries of the phenazine-like BGC. Therefore, it was predicted that the phenazine–like BGC in *S. scabiei* is a 42,991 bp region, from *SCAB11861* to *SCAB12211*, including 34 genes (Figure 3.9). The 42.9 kb phenazine–like BGC is located in the arm region of the chromosome of *S. scabiei* and *Streptomyces* sp. 1222.2 (Figure 3.8). However, it is hard to predict the location of the BGC in *S. stelliscabiei* and *S. bottropensis*, as their *dnaA* gene (chromosomal replication initiator, located around the *oriC* region in the center of the *Streptomyces* linear chromosome) is not located in the center of the respective chromosome, which suggests that the genome sequences were not properly assembled (Figure 3.8). A comparison of the phenazine–like BGC from *S. scabiei* with that from *S.*

stelliscabiei, *Streptomyces* sp. 1222.2 and *S. bottropensis* revealed that these genes are of high similarity and the gene arrangements of these phenazine–like biosynthetic genes from the above seven species are all identical (Figure 3.9).

Table 3.3 Genome sequences used in this study.

Organism	**Sequence Status**	**IMG Genome ID**	**Assembled Genome Size (bp)**
Streptomyces scabiei 87-22	Finished	646564576	10148695
Streptomyces sp. 1222.2	Permanent Draft	2728369725	9514697
Streptomyces bottropensis ATCC 25435	Permanent Draft	2551306667	8913629
Streptomyces stelliscabiei P3825	Permanent Draft	2645727861	10334041

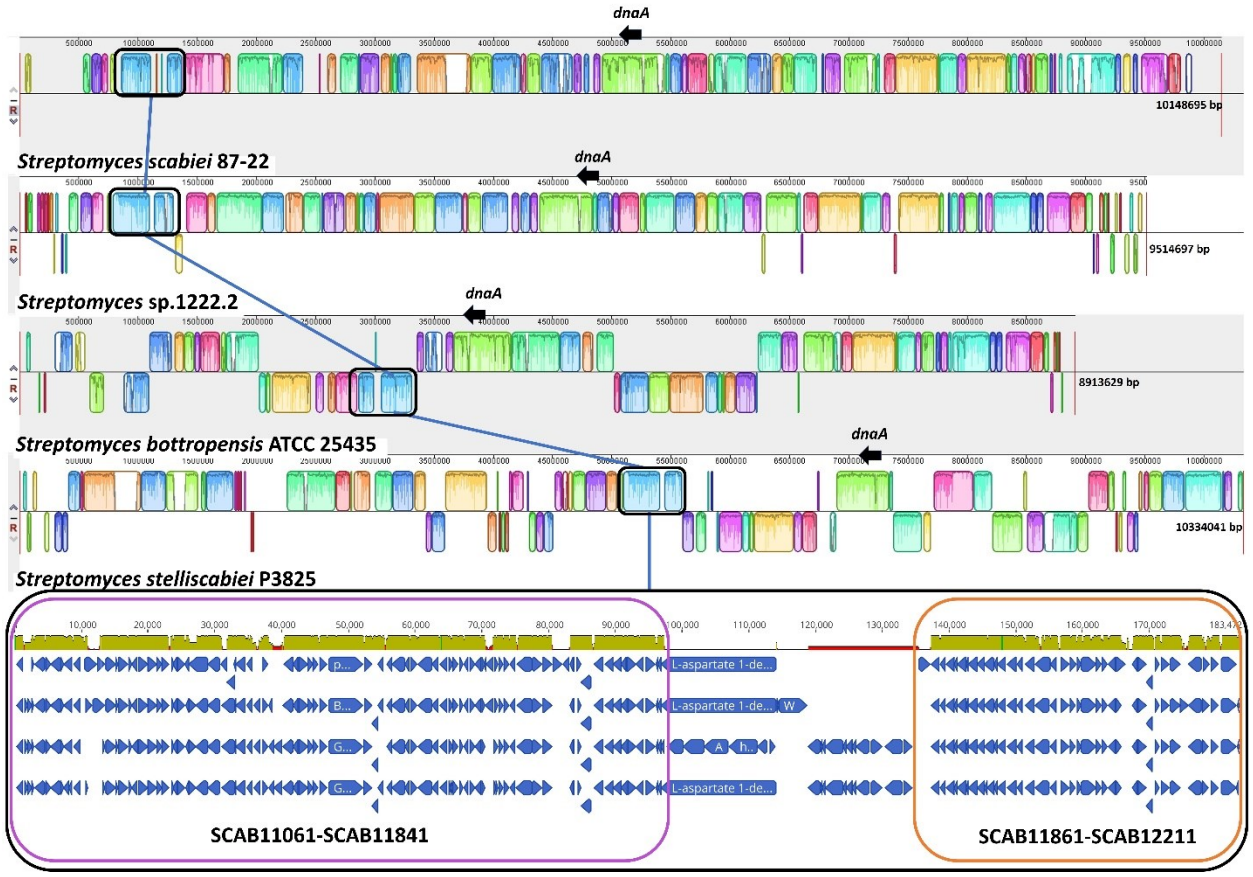


Figure 3.8 Whole–genome alignment of *S. scabiei*, *S. stelliscabiei*, *Streptomyces* sp. 1222.2 and *S. bottropensis* using Mauve. Colored blocks represent the conserved regions across the four species and the boundaries of colored blocks indicate the breakpoints of genome rearrangement. The homologous regions are indicated with the same color. The regions containing the phenazine-like BGC in different species are connected by lines and are boxed with thick black. The organization of the genes in the regions containing the phenazine-like BGC is shown at the bottom. The location of *dnaA* (chromosomal replication initiator) in each chromosome is indicated by the black arrows. The direction of the arrows indicates the direction of transcription of *dnaA.*

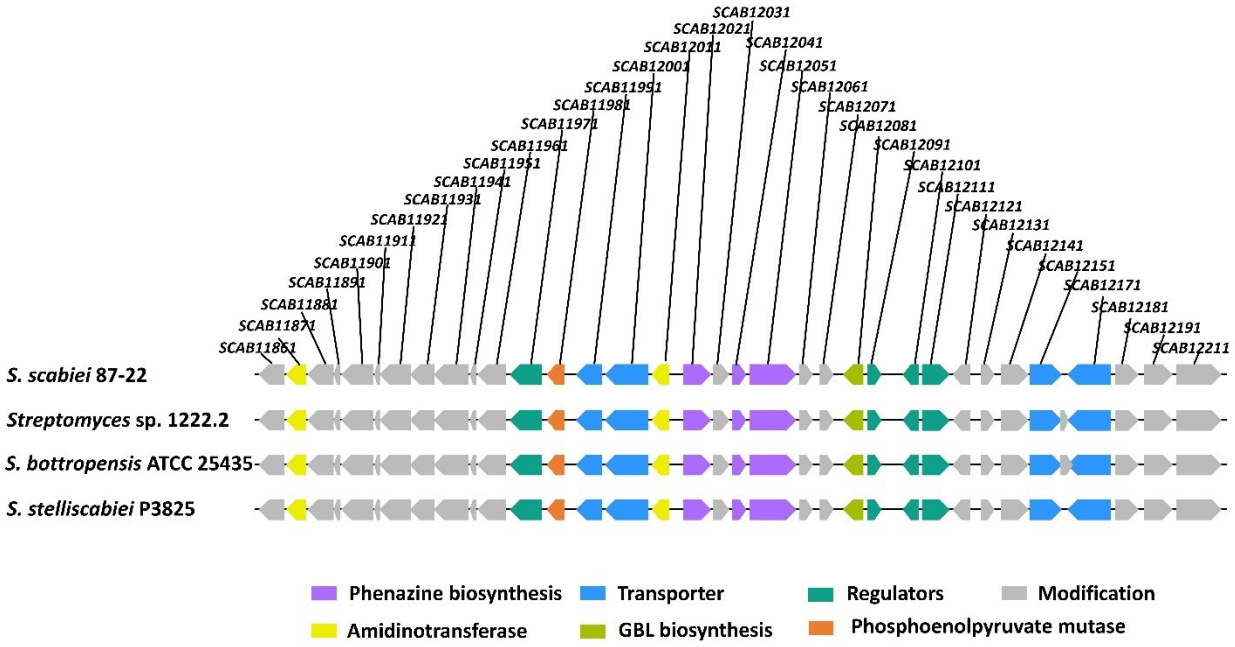


Figure 3.9 Comparison of the phenazine–like BGC from *S. scabiei*, *S. stelliscabiei*, *Streptomyces* sp. 1222.2, and *S. bottropensis*. The direction of the arrows indicates the predicted direction of transcription of each gene. Homologous genes in each cluster are indicated by arrows that are the same color. GBL: γ-Butyrolactone.

3.1.4 Predicted functions for genes in the putative phenazine–like BGC from *S. scabiei*

The function of each gene in the phenazine–like BGC was predicted using BlastP, Pfam, SMART and Prosite database searches (Table 3.4). A comparison of the amino acid sequences of each protein revealed that the *S. scabiei* proteins are most similar to the corresponding homologues from *S. bottropensis, S. stelliscabiei*, and *Streptomyces* sp. 1222.2 (Figure 3.10). Blast analysis of SCAB12021, SCAB12041, and SCAB12051 demonstrated that homologous proteins (Coverage ≥ 90%, Identity ≥ 57%) are present in other organisms: eight from *Streptomyces* species and seven from non–*Streptomyces* species (Table 3.5). Phylogenetic analysis revealed that SCAB12021, SCAB12041, and SCAB12051 and their homologues from *S. stelliscabiei*, *Streptomyces* sp. 1222.2, *S. bottropensis*, *S. ipomoeae*, *Kitasatospora* sp. Root107, *Kitasatospora* sp. MMS16–

BH015, *Saccharothrix* sp. ST–888, and *Streptacidiphilus pinicola* formed a distinct clade in each case (Figure S1, S3 and S4). Furthermore, this clade contained two subclades, with one consisting of *Streptomyces* homologues and the other consisting of the non–S*treptomyces* homologues.

Table 3.4 Predicted functions for proteins encoded in the putative phenazine–like BGC from *S. scabiei* 87-22.

Protein name and length	Closest Homologue in Database NR (%Cover/Identity/Similarity)	Domains Present (SMART/Pfam/Prosite)	Predicted Function
SCAB11861 414aa	ATP–grasp domain–containing protein [*Streptomyces stelliscabiei*], 100/95/97	Pfam: ATP–grasp_4 (PF13535) Prosite: PS50975 ATP_GRASP	ATP–dependent carboxylate–amine ligase
SCAB11871 280aa	Hypothetical protein [*Streptomyces bottropensis*], 100/97/98	Pfam: PF02274: Amidinotransf	Amidinotransferase
SCAB11881 395aa	Hypothetical protein [*Streptomyces bottropensis*], 100/99/100	Pfam: PF08028: Acyl–CoA_dh_2	Acyl–CoA dehydrogenase
SCAB11891 168aa	Flavin reductase family protein [*Streptomyces* spp.] 100/95/97	SMART: smart00903, Flavin reductase like domain Pfam: PF01613 Flavin reductase like domain	Flavin reductase
SCAB11901 451aa	FAD–binding oxidoreductase [*Streptomyces stelliscabiei*], 100/95/97	Pfam: FAD_binding_4, PF01565 Prosite: PS51387 FAD_PCMH PCMH–type FAD–binding domain	FAD–binding oxidoreductase
SCAB11911 114aa	Hypothetical protein [*Streptomyces* spp.] 100/99/99	No putative conserved domains have been detected	Unknown
SCAB11921 426aa	ATP–grasp domain–containing protein [*Streptomyces* spp.] 100/96/98	Pfam: PF13535: ATP–grasp 4 Prosite: PS50975 ATP_GRASP	ATP–dependent carboxylate–amine ligase
SCAB11931 357aa	Oxidoreductase [*Streptomyces stelliscabiei*] 100/99/100	Pfam: PF01266 FAD dependent oxidoreductase	FAD-dependent oxidoreductase
SCAB11941 512aa	Class I tRNA ligase family protein [*Streptomyces bottropensis*] 100/ 96/98	Pfam: PF09334: tRNA–synt_1g Prosite: PS00178 Aminoacyl–transfer RNA synthetases class–I signature	Unknown
SCAB11951 123aa	Cupin domain–containing protein [*Streptomyces* spp.] 100/98/99	SMART: smart00835: Cupin_1 Pfam: PF07883: Cupin_2	Unknown
SCAB11961 426aa	SidA/IucD/PvdA family monooxygenase [*Streptomyces* spp.] 100/99/99	Pfam: pfam13434: K_oxygenase L–lysine 6–monooxygenase (NADPH–requiring)	NADPH–dependent hydroxylase
SCAB11971 460aa	Helix–turn–helix transcriptional regulator, AfsR family [*Streptomyces bottropensis*], 100/91/94	SMART: smart00530: HTH_XRE Pfam: PF13560: HTH_31	DNA binding transcriptional regulator
SCAB11981 275aa	Isocitrate lyase/phosphoenolpyruvate mutase	Pfam: PF13714: PEP_mutase	Phosphoenolpyruvate mutase

	family protein [*Streptomyces*], 100/95/97		
SCAB11991 424aa	MFS transporter [*Streptomyces* spp.], 100/95/98	Pfam: PF07690: MFS_1 Prosite: PS50850, MFS, Major facilitator superfamily (MFS) profile	MFS transporter
SCAB12001 656aa	ABC transporter [*Streptomyces bottropensis*], 94/75/80	Pfam: PF04234: CopC PF05425: CopD PF05751: FixH	Transporter
SCAB12011 269aa	Hypothetical protein [*Streptomyces bottropensis*], 100/91/97	Pfam: pfam02274: Amidinotransf	Amidinotransferase
SCAB12021 412aa	Phospho–2–dehydro–3–deoxyheptonate aldolase, [*Streptomyces stelliscabiei*], 99/94/98	Pfam:DAHP_synth_2, PF01474	3–deoxy–7–phosphoheptulonate synthase; PhzC homologue
SCAB12031 266aa	2,3–dihydro–2,3–dihydroxybenzoate dehydrogenase, *Streptomyces bottropensis*, 100/91/94	Pfam: adh_short, PF00106 Prosite: ADH_SHORT, PS00061	2,3–dihydro–2,3–dihydroxybenzoate dehydrogenase
SCAB12041 225aa	Isochorismatase [*Streptomyces stelliscabiei*], 100/90/93	Pfam: Isochorismatase, PF00857 Prosite: no hit	Isochorismatase, also known 2,3 dihydro–2,3 dihydroxybenzoate synthase; PhzD homologue
SCAB12051 732aa	Phenazine–specific anthranilate synthase component I [*Streptomyces bottropensis*], 100/84/88	Pfam: Chorismate_bind PF00425 GATase PF00117 Prosite: Glutamine amidotransferase type–1	Chorismate binding enzyme anthranilate synthase; PhzE homologue
SCAB12061 234aa	Hypothetical protein [*Streptomyces bottropensis*], 100/94/96	No putative conserved domains have been detected	Unknown
SCAB12071 257aa	Oxidoreductase [*Streptomyces bottropensis* ATCC 25435], 100/93/97	Pfam: adh_short_C2, Enoyl–(Acyl carrier protein) reductase Prosite: ADH_SHORT, PS00061	Dehydrogenases/reductases
SCAB12081 333aa	AfsA/ScbA–like protein [*Streptomyces bottropensis*], 100/90/93	Pfam:AfsA PF03756	GBL biosynthesis
SCAB12091 213aa	γ–butyrolactone–binding protein [*Streptomyces stelliscabiei*], 100/98/100	Pfam: TetR_N, PF00440 Prosite: HTH_TETR_2, PS50977	Transcriptional regulator; GBL receptor
SCAB12101 240aa	γ–butyrolactone–binding protein [*Streptomyces bottropensis*], 100/90/93	Pfam: TetR_N, PF00440 Prosite: HTH_TETR_2, PS50977	Transcriptional regulator; Pseudo GBL receptor
SCAB12111 415aa	Secondary metabolite regulator [*Streptomyces bottropensis*], 100/95/96	Pfam: DDE_5, PF13546	Transcriptional Regulator
SCAB12121 303aa	Shikimate dehydrogenase [*Streptomyces bottropensis*], 100/86/90	Pfam: Shikimate_dh_N, PF08501	Shikimate 3–dehydrogenase (NADP+)

SCAB12131 204aa	NAD(P)H dehydrogenase [*Streptomyces stelliscabiei*], 100/93/97	Pfam: FMN_red, PF03358 Prosite: FLAVODOXIN_LIKE, PS50902	FMN–dependent oxidase
SCAB12141 401aa	Oxidoreductase [*Streptomyces bottropensis*], 100/90/93	Pfam: FAD_binding_3, PF01494	Oxidoreductase
SCAB12151 496aa	Peptide ABC transporter [*Streptomyces bottropensis*], 100/95/97	Pfam: SBP_bac_5, PF00496	Putative transport system peptide–binding protein
SCAB12171 630aa	ABC transporter ATP-binding protein [*Streptomyces stelliscabiei*] 99/95/98	Pfam: ABC_tran PF00005 Prosite: PS50893 ABC_TRANSPORTER_2	Transporter
SCAB12181 305aa	Flippase-like domain-containing protein [*Streptomyces europaeiscabiei*] 99/87/92	Pfam: LPG_synthase_TM PF03706	Channel
SCAB12191 360aa	Mechanosensitive ion channel family protein [*Streptomyces* sp. GY16] 100/96/98	Pfam: MS_channel PF00924	Channel
SCAB12211 739aa	NADP-dependent isocitrate dehydrogenase [*Streptomyces galbus*] 100/97/98	Pfam: Isocitrate dehydrogenase PF03971	Isocitrate dehydrogenase

S. scabiei	*S. bottropensis*	*S. stelliscabiei*	*S. sp. 1222.2*
SCAB11861	93.2	95.4	95.2
SCAB11871	96.8	96.4	96.4
SCAB11881	98.5	98.2	98.2
SCAB11891	94	94.6	94.6
SCAB11901	95.1	95.3	95.1
SCAB11911	99.1	99.1	99.1
SCAB11921	94.2	94.9	94.9
SCAB11931	96.6	94.9	96.9
SCAB11941	96.3	96.3	96.1
SCAB11951	97.6	97.6	97.6
SCAB11961	98.8	98.8	98.8
SCAB11971	91.1	91.8	91.8
SCAB11981	94.9	95.3	95.3
SCAB11991	95.3	95.3	93.9
SCAB12001	72.1	72.3	70
SCAB12011	91.4	91.4	91.4
SCAB12021	94.4	94.6	94.2
SCAB12031	91	89.6	88
SCAB12041	88.4	89.8	89.8
SCAB12051	83.9	84.1	81.6
SCAB12061	94	93.6	93.6
SCAB12071	93	92.2	88.8
SCAB12081	89.8	89.8	88.2
SCAB12091	95.3	97.6	95.3
SCAB12101	90	89.6	89.2
SCAB12111	94.5	94.3	91.6
SCAB12121	86.4	85.7	85.5
SCAB12131	92.6	93.1	92.6
SCAB12141	90.3	90.3	90.3
SCAB12151	95.4	95	95
SCAB12171	95.4	95.4	94.8
SCAB12181	84	38	40
SCAB12191	95	96	96
SCAB12211	96	97	97

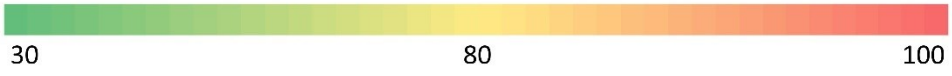


Figure 3.10 Heat map showing the protein BLAST identity levels to the *S. scabiei* 87-22 SCAB11861 – SCAB12151 proteins in different bacterial genomes. The numbers indicate the percentage of identity.

Table 3.5 Sequence homology of *S. scabiei* SCAB12021, SCAB12041, and SCAB12051 with proteins present in other systems revealed by BLAST searches (Coverage ≥ 90%, Identity ≥ 57%). Numbers denoted in each column indicate the number of amino acids (AA), % coverage (C)/% identity (I), compared with *S. scabiei* SCAB12021, SCAB12041, and SCAB12051.

Organism	**SCAB12021 (PhzC, 412AA) homologue**		**SCAB12041 (PhzD, 225AA) homologue**		**SCAB12051 (PhzE, 732AA) homologue**	
	AA	% C/%I	AA	%C/%I	AA	%C/%I
Other *Streptomyces* species						
Streptomyces stelliscabiei	412	99/94.4	225	100/89.78	676	96/83.24
Streptomyces sp. 1222.2	412	99/94.4	225	100/89.78	676	96/84.11
Streptomyces bottropensis	412	99/94.16	225	100/88.44	676	96/83.69
Streptomyces ipomoeae	412	99/86.86	219	97/83.11	667	92/75.73
Streptomyces eurocidicus	390	92/67.28	218	95/64.65	632	91/58.69
Streptomyces albireticuli	390	92/66.23	229	94/64.62	627	91/58.93
Streptomyces katrae	389	92/62.4	226	96/63.18	627	92/58.22
Streptomyces glaucescens	388	92/63.71	225	96/66.36	628	90/60.09
Streptomyces sp. KS 21	389	92/62.4	226	96/63.64	627	92/58.22
Non–*Streptomyces* species						
Gram positive species						
Kitasatospora sp. Root107	415	98/70.27	211	93/71.43	659	92/64.45
Kitasatospora sp. MMS16–BH015	409	98/70.66	211	93/72.51	643	91/63.54
Saccharothrix sp. ST–888	407	98/71.43	211	93/72.04	649	92/65.4
Streptacidiphilus pinicola	396	92/68.15	211	93/72.99	657	91/64.3
Actinobacteria bacterium OV450	389	92/63.71	226	96/63.64	627	92/58.22
Kutzneria buriramensis	386	92/61.78	223	93/60.48	636	92/57.52
Gram negative species						
Pseudomonas aeruginosa	405	94/57.44	207	92/59.9	624	92/55.6

The SCAB12021 amino acid sequence was most closely related to a predicted PhzC (3–deoxy–d–arabino–heptulosonate 7–phosphate synthases, DAHP7PS) from *S. stelliscabiei* (KND42883.1, coverage 99%, identity 94.4%), *Streptomyces* sp. 1222.2 (SOD67797.1, coverage 99%, identity 94.4%), *S. bottropensis* (EMF52115.1, coverage 99%, identity 94.16%) and *S. ipomoeae* (EKX64370.1, coverage 99%, identity 86.86%) (Table 3.5). A domain search of SCAB12021 revealed the presence of a Class–II DAHP synthetase domain (PF01474, amino acid residues 25 to 402), which catalyses the first step of the shikimate pathway for aromatic amino acid biosynthesis. Protein sequence alignment of SCAB12021 with other PhzC proteins from the above *Streptomyces* species and from *Pseudomonas aeruginosa* revealed the presence of the conserved motif (R/K)xxxxxxKPR(T/S) for binding phosphorylated monosaccharides (Figure 3.11), which is common to all DAHP7PS (Schofield et al. 2005). This suggests that SCAB12021 has a similar function as PhzC.

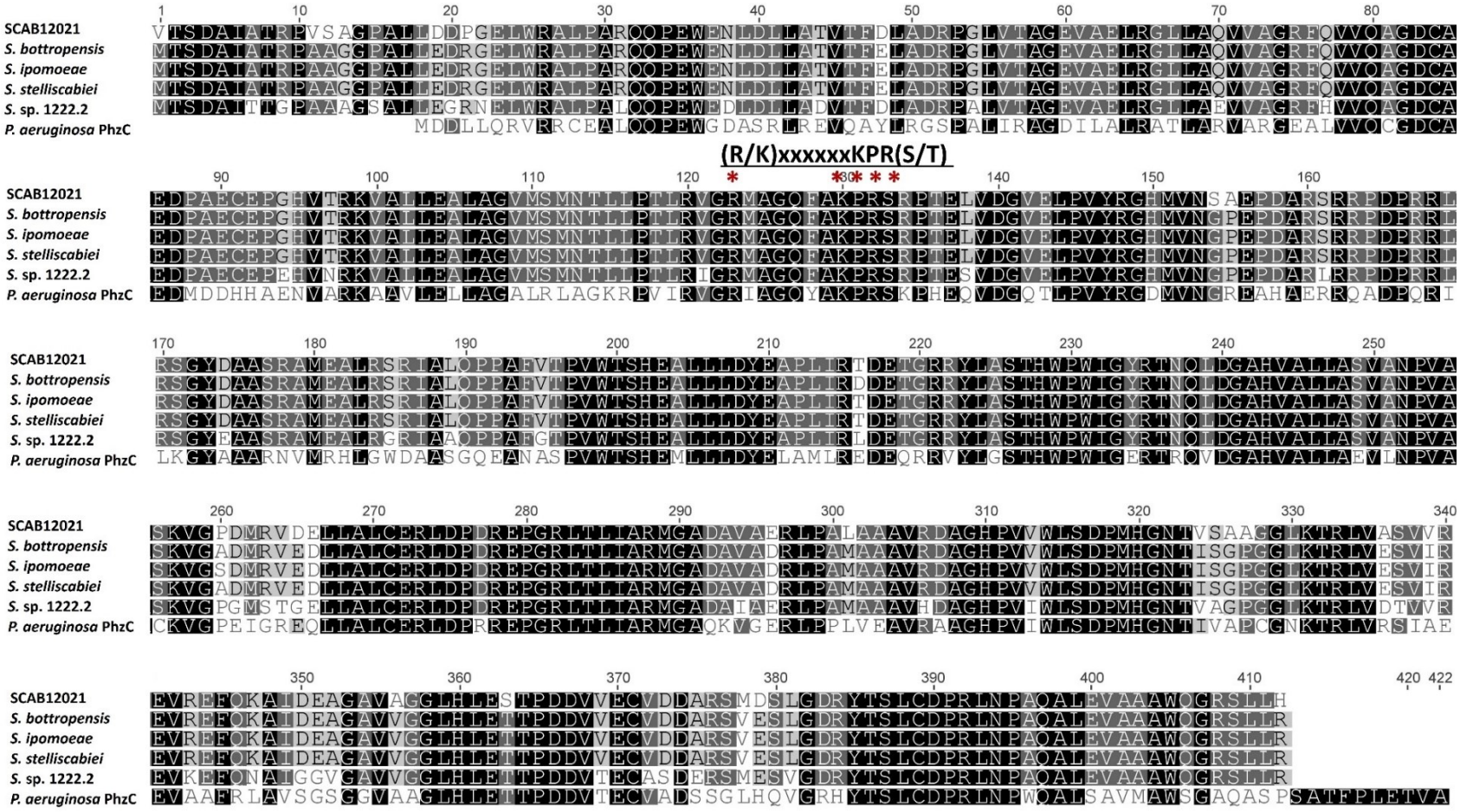

(R/K)xxxxxxKPR(S/T)
SCAB12021
S. bottropensis
S. ipomoeae
S. stelliscabiei
S. sp. 1222.2
P. aeruginosa PhzC

Figure 3.11 Amino acid sequence alignment of SCAB12021 with other PhzC homologues from *S. stelliscabiei, Streptomyces* sp. 1222.2*, S. bottropensis, S. ipomoeae,* and *P. aeruginosa*. The conserved (R/K)xxxxxxKPR(T/S) substrate–binding motif is labeled and indicated with asterisks (Schofield et al. 2005). Highly conserved amino acids in the alignment are highlighted as follows: black, 100% identity; dark grey, 80–99% identity; light grey, 60–79% identity; white, <60% identity.

The SCAB12041 amino acid sequence was determined to be most similar to a predicted PhzD (isochorismatase) from *S. stelliscabiei* (KND42881.1, coverage 100%, identity 89.78%), *Streptomyces* sp. 1222.2 (SOD67799.1, coverage 100%, identity 89.78%.), *S. bottropensis* (EMF52117.1, coverage 100%, identity 88.44%) and *S. ipomoeae* (WP 009321183.1, coverage 97%, identity 83.11%) (Table 3.5). A search of the Pfam database revealed the presence of an isochorismatase monooxygenase domain (PF00857, amino acid residues 31 to 204). Isochorismatases are also known as 2,3 dihydro–2,3 dihydroxybenzoate synthases, which catalyze the conversion of isochorismate, in the presence of water, to 2,3–dihydroxybenzoate and pyruvate. Protein sequence alignment of SCAB12041 with other PhzD homologues from the above *Streptomyces* species and *P. aeruginosa* revealed that the active sites and the substrate–binding sites based on the published structural analysis of PhzD (Parsons et al. 2003) are all conserved in these sequences, suggesting that these proteins are all functional PhzD homologues (Figure 3.12).

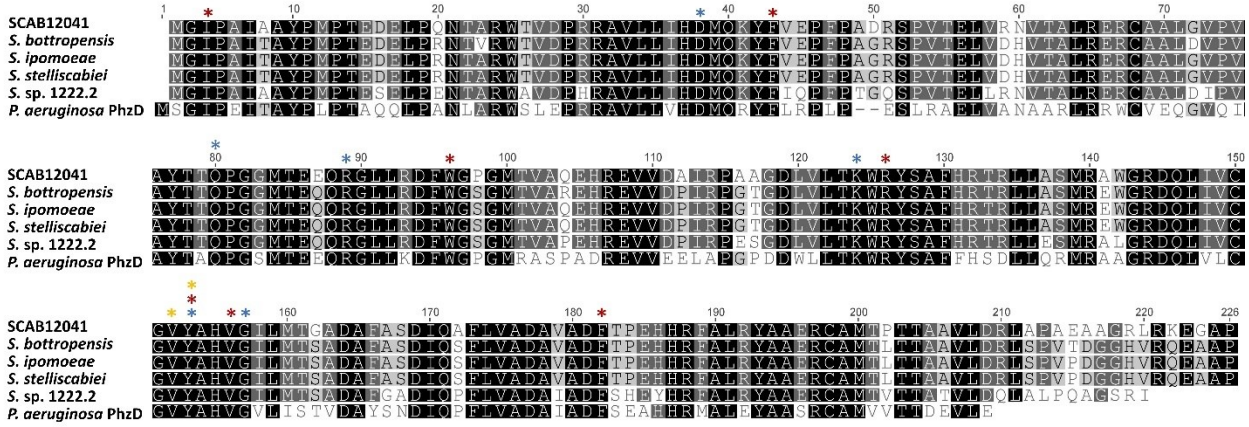


Figure 3.12 Amino acid sequence alignment of SCAB12041 with other PhzD homologues from *S. stelliscabiei, Streptomyces* sp. 1222.2*, S. bottropensis, S. ipomoeae,* and *P. aeruginosa*. Based on the published structural analysis of PhzD, the amino acids that are predicted to be involved in hydrophobic substrate binding are indicated with the red asterisks; the active sites are indicated with blue asterisks, the conserved cis–peptide bond is indicated with yellow asterisks (Parsons et al. 2003).

The amino acid sequence of SCAB12051 was most closely related to a predicted PhzE (phenazine–specific anthranilate synthase) from *S. stelliscabiei* (KND42880.1, coverage 96%, identity 83.24%), *Streptomyces* sp. 1222.2 (SOD67800.1, coverage 96%, identity 84.11), *S. bottropensis* (EMF52118.1, coverage 96%, identity 83.69%) and *S. ipomoeae* (EKX64378.1, coverage 92%, identity 75.73%) (Table 3.5). A domain search of SCAB12051 revealed the presence of two domains: a chorismate converting menaquinone, siderophore, tryptophan biosynthesis domain (MST) (PF00425, amino acid residues 156 to 415) and a type 1 glutamine amidotransferase (GATase1) domain (PF00117, amino acid residues 506 to 703) connected by a 45–residue linker. According to the PROSITE database search results, the GATase1 containing a well–conserved putative active site **PFLAV<u>C</u>LSHQ**VL (letters in bold indicate highly conserved residues, and the essential cysteine is underlined). Protein sequence alignment of SCAB12051 with other PhzE homologues from the above *Streptomyces* species and from *P. aeruginosa* revealed

that the active sites of PhzE (Li et al. 2011) are all conserved in these sequences, suggesting that these proteins are all functional PhzE homologues (Figure 3.13).

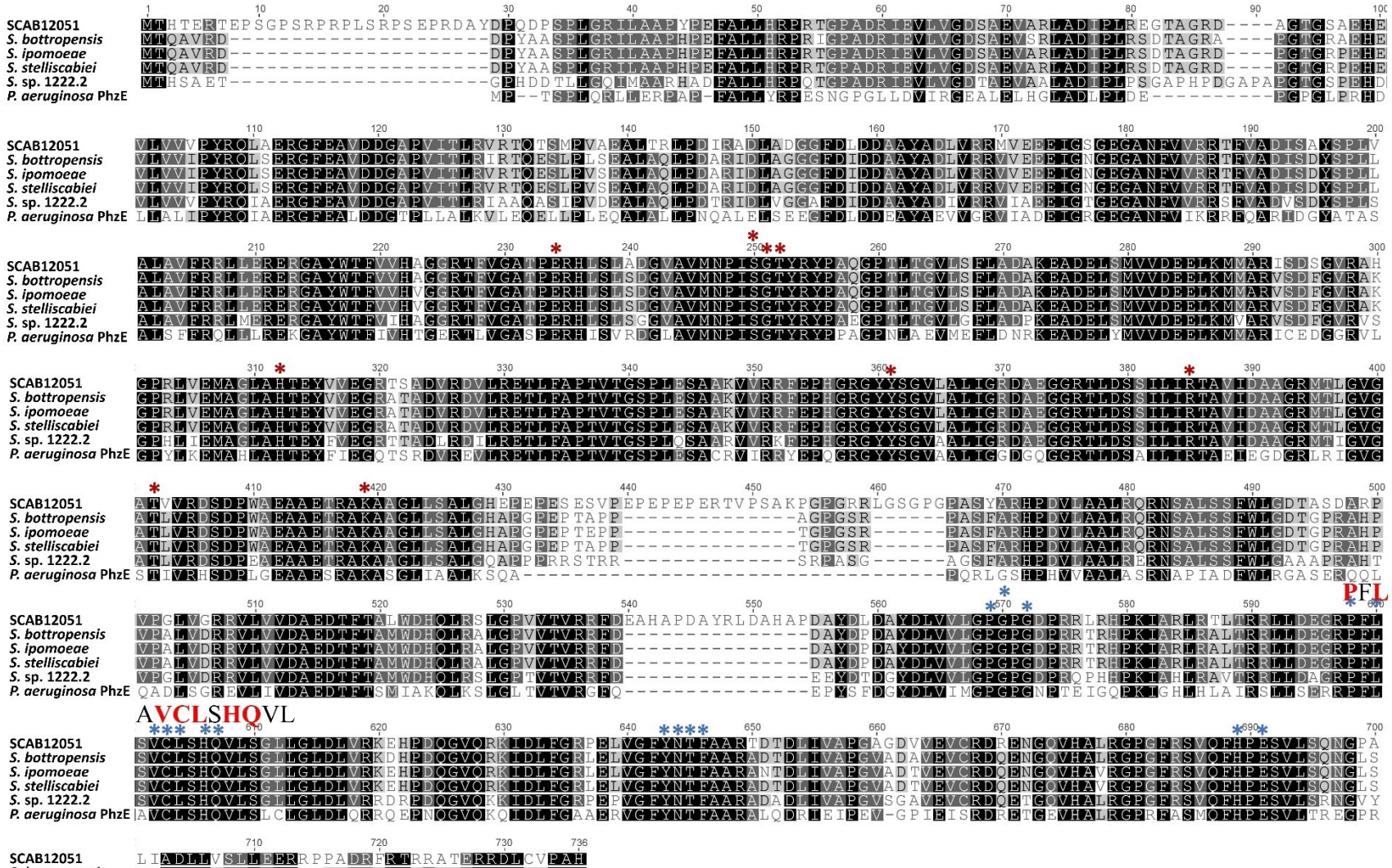

SCAB12051
S. bottropensis
S. ipomoeae
S. stelliscabiei
S. sp. 1222.2
P. aeruginosa PhzE
PFL
AVCLSHQVL

Figure 3.13 Amino acid sequence alignment of SCAB12051 with other PhzE homologues from *S. stelliscabiei, Streptomyces* sp. 1222.2*, S. bottropensis, S. ipomoeae,* and *P. aeruginosa*. Based on the published structural analysis of PhzE, the active sites of the MST domain are indicated with red asterisks, the active sites of the GATase1 domain are indicated with blue asterisks, and the well–conserved putative active site **P**F**L**A**V**<u>**C**</u>**L**S**HQ**VL (letters in bold indicate highly conserved residues, and the essential cysteine is underlined) is labeled and indicated with the asterisks.

SCAB12031 is predicted to be a 2,3–dihydro–2,3–dihydroxybenzoate dehydrogenase. A domain search of SCAB12031 revealed the presence of a short chain dehydrogenase domain (PF00106, amino acid residues 16 to 206). Remali and colleagues predicted that the 2,3–dihydro–2,3–dihydroxybenzoate dehydrogenase from *S. kebangsaanensis* is a PhzA homolog (Remali et al. 2017). SCAB12031 has 64.1% of identity and 74.5% similarity to PhzA from *S. kebangsaanensis* (Remali et al. 2017). Protein sequence alignment of SCAB12031 with other homologues from *S. stelliscabiei, Streptomyces* sp. 1222.2*, S. bottropensis, S. ipomoeae,* and with PhzA from *S. kebangsaanensis* and *P. aeruginosa* revealed that the conserved residues involved in substrate binding and catalysis in PhzA from *P. aeruginosa* (Ahuja et al. 2008) are not all present in the SCAB12031 homologues, and the chemical properties of the amino acid residues present in the SCAB12031 homologues are different from those present in PhzA from *P. aeruginosa.* In addition, SCAB12031 only has 26.0% amino acid identity and 31.5% similarity to PhzA from *P. aeruginosa.* Together, this suggests that SCAB12031 may not be a functional PhzA homolog (Figure 3.14).

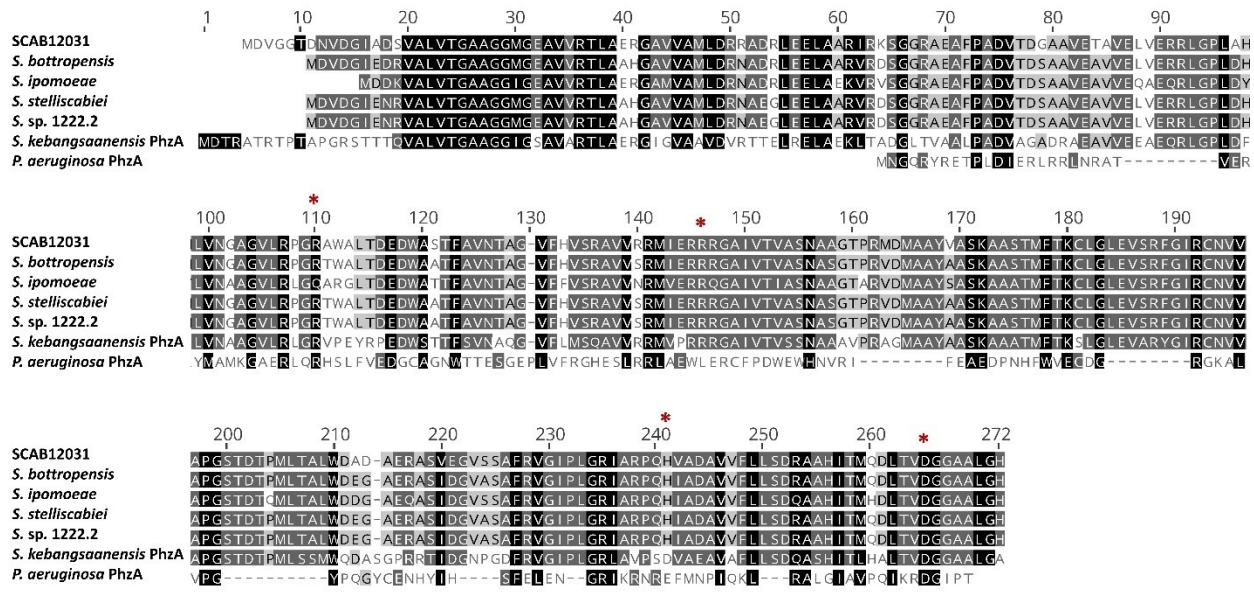


Figure 3.14 Amino acid sequence alignment of SCAB12031 with other homologues from *S. stelliscabiei, Streptomyces* sp. 1222.2*, S. bottropensis, S. ipomoeae,* and with PhzA from *S. kebangsaanensis* and *P. aeruginosa*. The conserved residues involved in substrate binding and catalysis in PhzA from *P. aeruginosa* are labeled and indicated with asterisks (Ahuja et al. 2008). Highly conserved amino acids in the alignment are highlighted as follows: black, 100% identity; dark grey, 80–99% identity; light grey, 60–79% identity; white, <60% identity.

Phylogenetic analysis of other proteins encoded near the three phenazine biosynthetic genes, such as SCAB12031, SCAB12061, SCAB12071, SCAB12081, SCAB12091, SCAB12101, and SCAB12111 (Figure 3.15; Figure S2 and Figures S5-S8), demonstrated that these proteins along with their homologues from *S. stelliscabiei*, *Streptomyces* sp. 1222.2, *S. bottropensis*, *S. ipomoeae*, *Kitasatospora* sp. Root107, *Kitasatospora* sp. MMS16–BH015, *Saccharothrix* sp. ST–888 and *Streptacidiphilus pinicola* form a distinct clade in each case, similar to what was observed for the phenazine biosynthetic proteins (Figure S1, S3 and S4).

Four putative transporters are found within the BGC: *SCAB12001*, *SCAB12151* and *SCAB12171* encode ABC transporters, and *SCAB11991* encodes an MFS transporter. Two amidinotransferases genes, *SCAB11871* and *SCAB12011*, are found within the BGC. The

streptomycin BGC from *S. griseus* also contains two amidinotransferases, *strB1* and *strB2*, which may be the result of a gene duplication event (Pissowotzki et al. 1991). *SCAB11861* and *SCAB11921* are predicted to encode two ATP–grasp domain–containing proteins, which possess the ATP–dependent carboxylate–amine ligase activity. *SCAB11981* is predicted to encode a phosphoenolpyruvate mutase, which suggests that the phenazine–like BGC may produce a phosphonate or phosphinate (C–P bond)–containing natural product (Tietz and Mitchell 2016)

Four putative CSR genes are found within the phenazine–like BGC. *SCAB12111* encodes a protein that is similar (41% amino acid identity) to TylR, a positive activator of tylosin production from *Streptomyces fradiae* (Bate et al. 1999, 2006; Stratigopoulos et al. 2004), and *SCAB11971* encodes a homologue of AfsR, which is a conserved pleiotropic regulator in *Streptomyces* spp. AfsR-like proteins are known to be associated with specialized metabolite BGCs and function as transcriptional activators (Horinouchi 2003), and thus it is predicted that *SCAB11971* may encode an activator of the phenazine-like BGC. Both *SCAB12091* and *SCAB12101* are predicted to encode GBL receptor proteins, which typically function as transcriptional repressors in the absence of the cognate GBL (Willey and Gaskell 2011; Nodwell 2014; Zhang et al. 2016). Phylogenetic analysis of SCAB12091 and SCAB12101 revealed that SCAB12091 has a close relationship with genuine GBL receptors, such as BarA, FarA and TylP, while *SCAB12101* forms a large clade with proteins such as BarB, JadR2 and TylQ, which resemble GBL receptors but do not bind GBLs and are thus referred to as “pseudo GBL receptors” (Figure 3.15) (Xu and Yang 2019). This suggests that SCAB12091 is a genuine GBL receptor and that SCAB12101 is a pseudo GBL receptor. *SCAB12081* is predicted to encode an AfsA–like protein that is required for the biosynthesis of GBLs. Both pseudo GBL receptor and genuine GBL receptor proteins can bind to partially palindromic DNA sequences (Kinoshita et al. 1997)

within the promoter regions of target genes and thereby repress the latter. Thus, the binding sites of SCAB12091 and SCAB12101 were predicted by searching for palindromic DNA sequences (EMBOSS palindrome, https://www.bioinformatics.nl/cgi-bin/emboss/palindrome) in the entire BGC (*SCAB11861-SCAB12211*). Ten potential binding sites were found in the entire region of the BGC (Figure 3.16). Notably, the binding sites upstream of *SCAB11981, SCAB12021*, and *SCAB12041* show a high degree of similarity, as they contain an identical 7 bp palindromic sequence (GTGGCGC) (Figure 3.16).

Other genes in the phenazine-like BGC encode predicted enzymes (Table 3.4) that may be responsible for the modification of the product. Notably, no homologues of the phenazine biosynthetic genes *phzF* and *phzG* are found within the BGC; however, two genes, *SCAB43981* and *SCAB51331,* are located in the core region of the *S. scabiei* 87-22 chromosome and are predicted to encode proteins that are similar to PhzF from *S. stelliscabiei* (83% identity) and PhzG from *P. aeruginosa* PAO1 (33.68% identity), respectively.

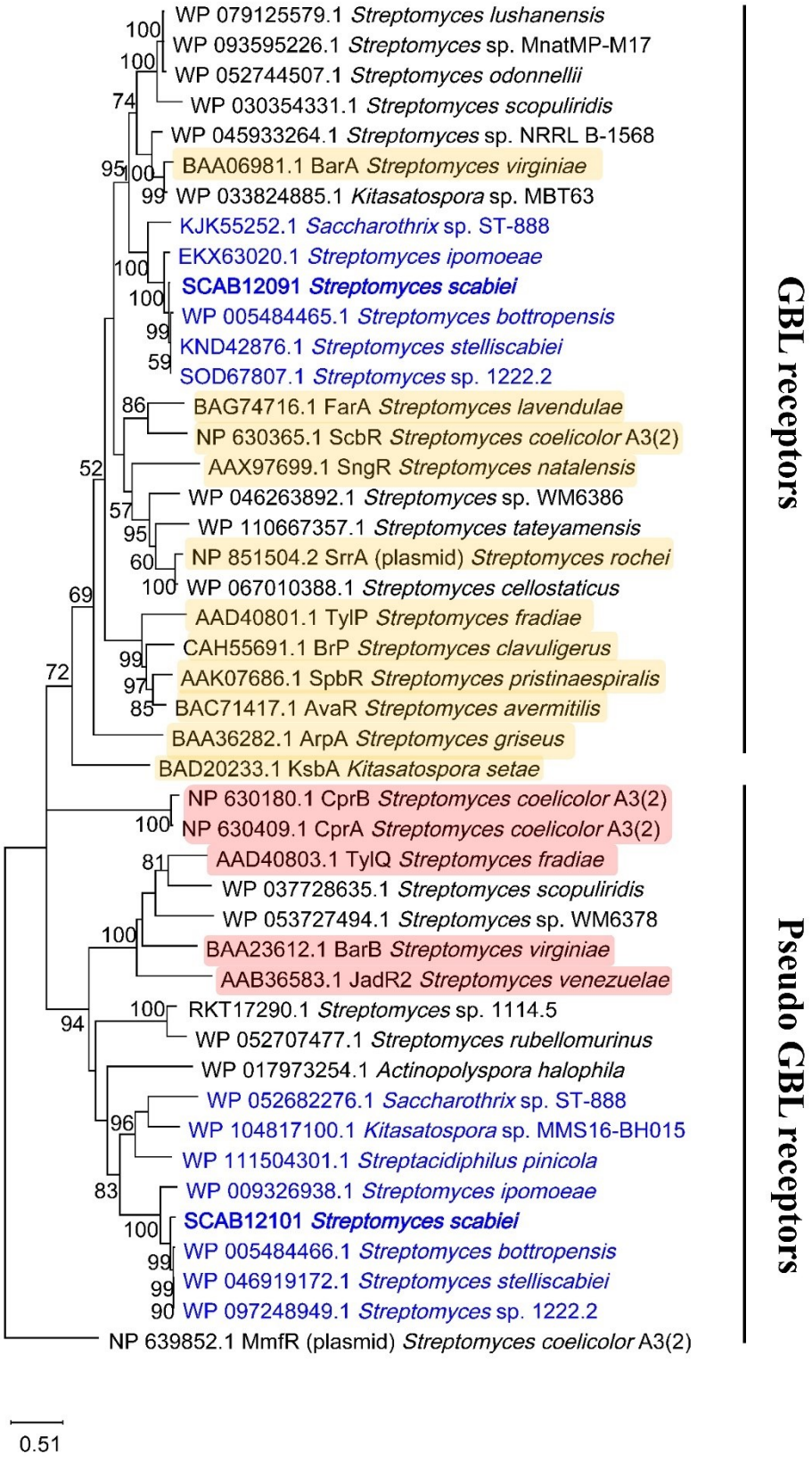

WP 079125579.1 Streptomyces lushanensis
WP 093595226.1 Streptomyces sp. MnatMP-M17
WP 052744507.1 Streptomyces odonnellii
WP 030354331.1 Streptomyces scopuliridis
WP 045933264.1 Streptomyces sp. NRRL B-1568
BAA06981.1 BarA Streptomyces virginiae
WP 033824885.1 Kitasatospora sp. MBT63
KJK55252.1 Saccharothrix sp. ST-888
EKX63020.1 Streptomyces ipomoeae
SCAB12091 Streptomyces scabiei
WP 005484465.1 Streptomyces bottropensis
KND42876.1 Streptomyces stelliscabiei
SOD67807.1 Streptomyces sp. 1222.2
BAG74716.1 FarA Streptomyces lavendulae
NP 630365.1 ScbR Streptomyces coelicolor A3(2)
AAX97699.1 SngR Streptomyces natalensis
WP 046263892.1 Streptomyces sp. WM6386
WP 110667357.1 Streptomyces tateyamensis
NP 851504.2 SrrA (plasmid) Streptomyces rochei
WP 067010388.1 Streptomyces cellostaticus
AAD40801.1 TylP Streptomyces fradiae
CAH55691.1 BrP Streptomyces clavuligerus
AAK07686.1 SpbR Streptomyces pristinaespiralis
BAC71417.1 AvaR Streptomyces avermitilis
BAA36282.1 ArpA Streptomyces griseus
BAD20233.1 KsbA Kitasatospora setae
NP 630180.1 CprB Streptomyces coelicolor A3(2)
NP 630409.1 CprA Streptomyces coelicolor A3(2)
AAD40803.1 TylQ Streptomyces fradiae
WP 037728635.1 Streptomyces scopuliridis
WP 053727494.1 Streptomyces sp. WM6378
BAA23612.1 BarB Streptomyces virginiae
AAB36583.1 JadR2 Streptomyces venezuelae
RKT17290.1 Streptomyces sp. 1114.5
WP 052707477.1 Streptomyces rubellomurinus
WP 017973254.1 Actinopolyspora halophila
WP 052682276.1 Saccharothrix sp. ST-888
WP 104817100.1 Kitasatospora sp. MMS16-BH015
WP 111504301.1 Streptacidiphilus pinicola
WP 009326938.1 Streptomyces ipomoeae
SCAB12101 Streptomyces scabiei
WP 005484466.1 Streptomyces bottropensis
WP 046919172.1 Streptomyces stelliscabiei
WP 097248949.1 Streptomyces sp. 1222.2
NP 639852.1 MmfR (plasmid) Streptomyces coelicolor A3(2)
GBL receptors
Pseudo GBL receptors
0.51

Figure 3.15 Phylogenetic relationships among homologues of SCAB12091 and SCAB12101 from the nr database. Bootstrap values are shown for the branch points supported in ≥ 50% out of 1000 repetitions. The scale bar indicates the number of amino acid substitutions per site. The known genuine GBL receptors are highlighted in yellow, and the known pseudo GBL receptors are highlighted in pink. The clades containing SCAB12091 and SCAB12101 are indicated in blue font.

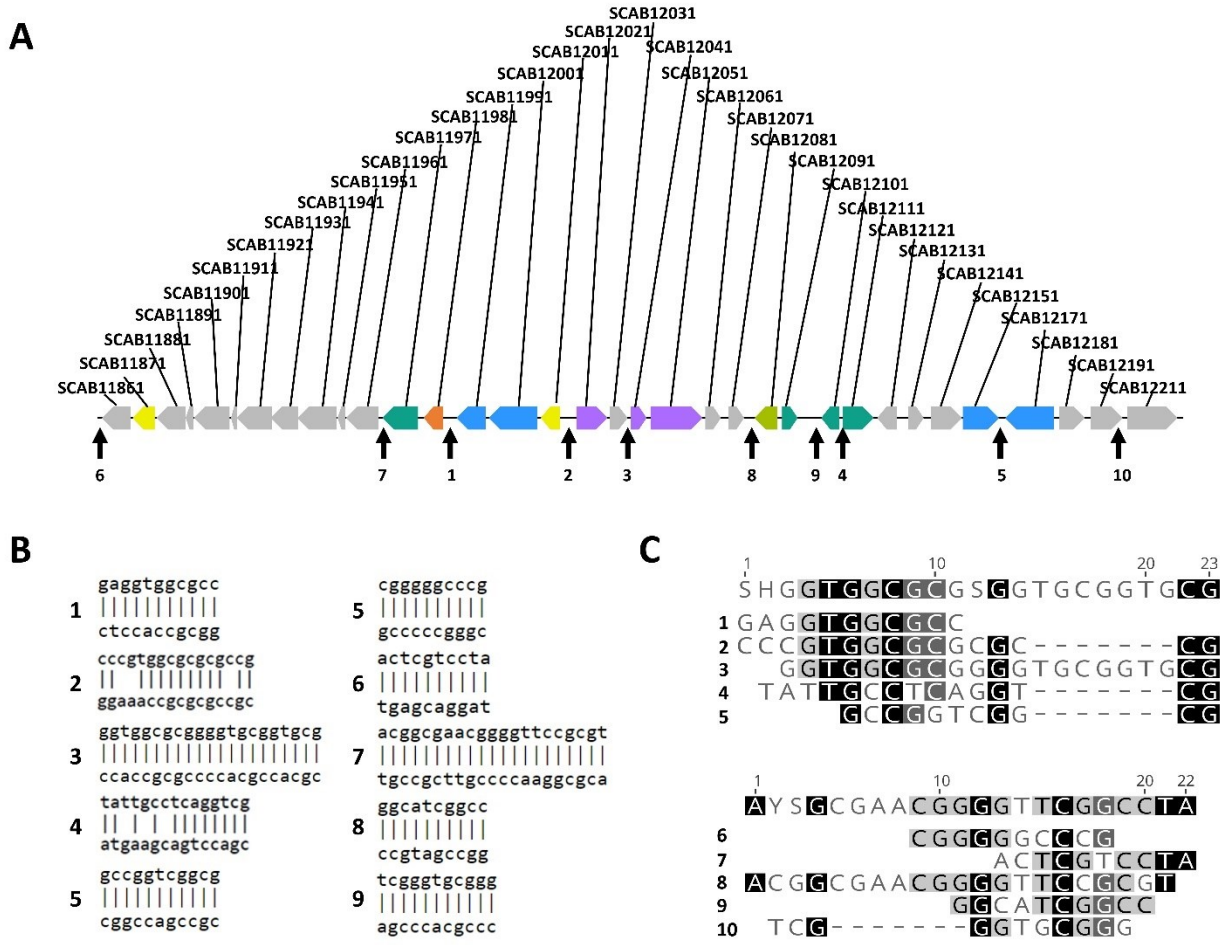


Figure 3.16 Predicted binding sites of the putative GBL receptor protein SCAB12091 and the putative pseudo GBL receptor protein SCAB12101. The locations of the predicted sites are shown in panel (A), the sequences of the predicted sites are shown in panel (B), and the alignment of the binding sites is shown in panel (C).

3.1.5 GC content of the putative phenazine–like BGC

GC content analysis for individual genes from the predicted phenazine-like BGC in *S. scabiei*, *S. stelliscabiei*, *Streptomyces* sp. 1222.2 and *S. bottropensis* shows that in general, the GC content is in the range of the average GC content of the corresponding genome (Figure 3.17). This suggests that if the phenazine biosynthetic genes in these species originate from another source, their GC content has ameliorated (so that it now resembles the host genomes) over time (Lawrence and Ochman 1997). A small number of genes from the phenazine–like BGC are exceptions, however. For example, *SCAB11911, SCAB11951, SCAB11961, SCAB12071, SCAB12081, SCAB12091, SCAB12131, SCAB12151, SCAB12191*, and *SCAB12211* have a GC content ~2% lower than the average % GC content of the corresponding genome (Figure 3.17).

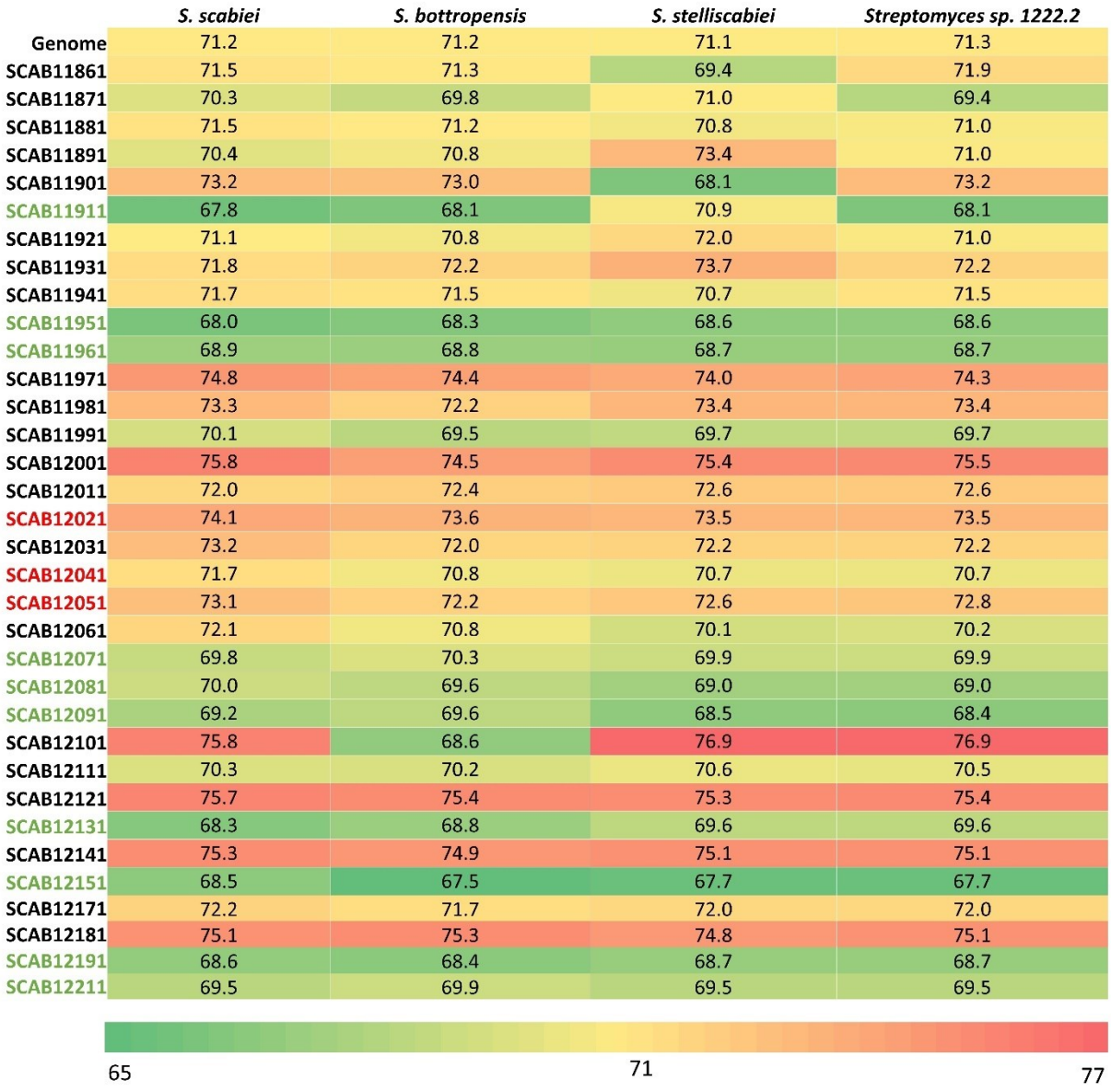


Figure 3.17 Heat map showing the % GC content of the genes within the predicted phenazine-like BGC from different organisms. Phenazine biosynthetic gene homologues (*phzC*, *phzD*, and *phzE*) are indicated in red font, and genes with a GC content ~2% lower than the average % GC content of the corresponding genome are indicated in green.

3.1.6 MultiGeneBlast MIBiG

To predict the closest known BGC related to the phenazine–like BGC, MultiGeneBlast was used to search for homologous BGCs in the MIBiG database (Kautsar et al. 2019). MultiGeneBlast analysis revealed the presence of 24 known BGCs sharing the core genes (*phzC, phzD, phzE*) with the unknown phenazine–like BGC. The unknown phenazine–like BGC was most similar to the BGC of diazaquinomycin A from *Streptomyces* sp. F001 (Figure 3.18). Among the 24 compounds produced by the known BGCs, some are NRP–type and PK–type natural products, some are phenazine–type, and some are other shikimate–derived compounds. The shared core genes (*phzC, phzD, phzE*) suggest that the unknown phenazine–like BGC may share several early pathway steps with these 24 known BGCs and then branch into different pathways.

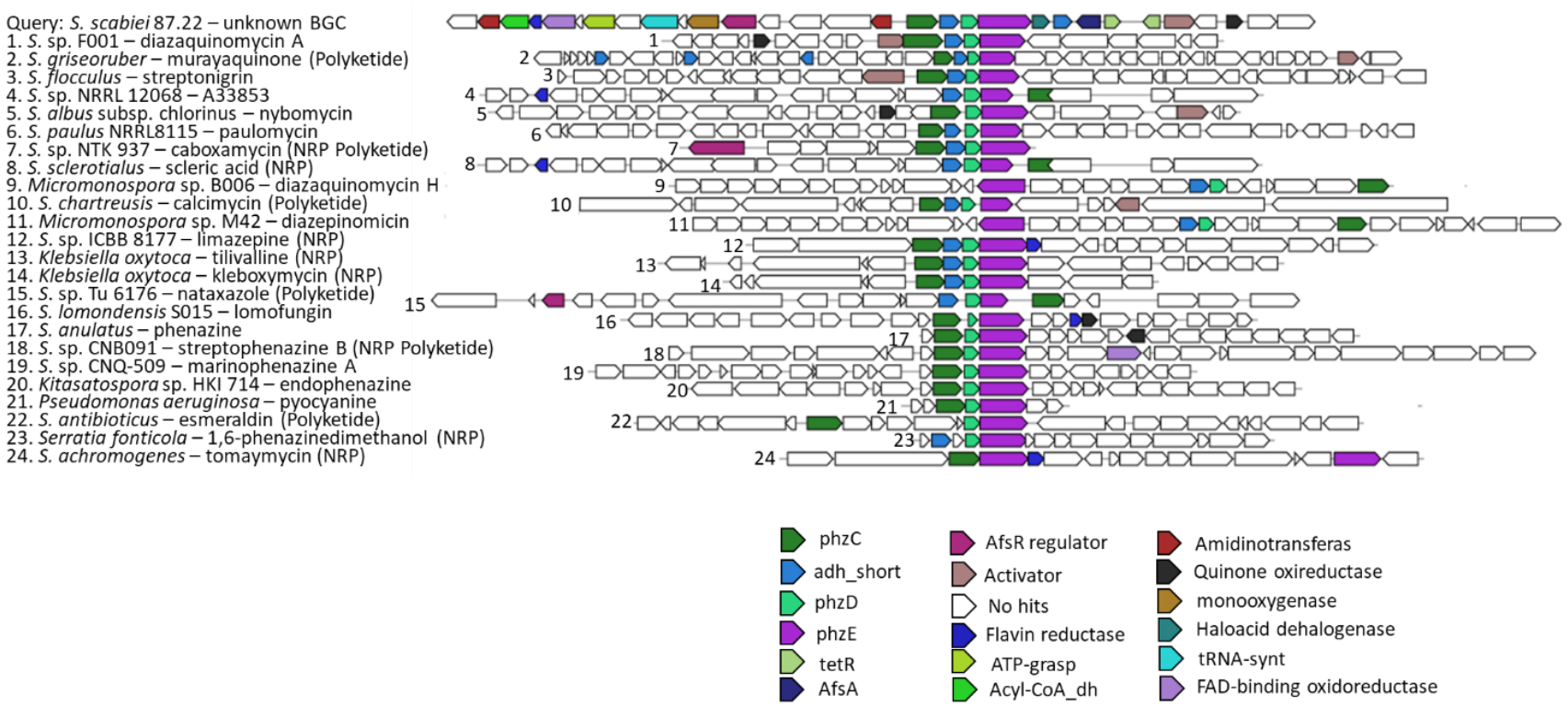

Query: S. scabiei 87.22 – unknown BGC
1. S. sp. F001 – diazaquinomycin A
2. S. griseoruber – murayaquinone (Polyketide)
3. S. flocculus – streptonigrin
4. S. sp. NRRL 12068 – A33853
5. S. albus subsp. chlorinus – nybomycin
6. S. paulus NRRL8115 – paulomycin
7. S. sp. NTK 937 – caboxamycin (NRP Polyketide)
8. S. sclerotialus – scleric acid (NRP)
9. Micromonospora sp. B006 – diazaquinomycin H
10. S. chartreusis – calcimycin (Polyketide)
11. Micromonospora sp. M42 – diazepinomicin
12. S. sp. ICBB 8177 – limazepine (NRP)
13. Klebsiella oxytoca – tilivalline (NRP)
14. Klebsiella oxytoca – kleboxymycin (NRP)
15. S. sp. Tu 6176 – nataxazole (Polyketide)
16. S. lomondensis S015 – lomofungin
17. S. anulatus – phenazine
18. S. sp. CNB091 – streptophenazine B (NRP Polyketide)
19. S. sp. CNQ-509 – marinophenazine A
20. Kitasatospora sp. HKI 714 – endophenazine
21. Pseudomonas aeruginosa – pyocyanine
22. S. antibioticus – esmeraldin (Polyketide)
23. Serratia fonticola – 1,6-phenazinedimethanol (NRP)
24. S. achromogenes – tomaymycin (NRP)
phzC
adh_short
phzD
phzE
tetR
AfsA
AfsR regulator
Activator
No hits
Flavin reductase
ATP-grasp
Acyl-CoA_dh
Amidinotransferas
Quinone oxireductase
monooxygenase
Haloacid dehalogenase
tRNA-synt
FAD-binding oxidoreductase

Figure 3.18 Result of MultiGeneBlast performed against the MIBiG database to predict the closest related compound produced by the phenazine–like BGC.

3.2 Construction of *Streptomyces* engineered strains

Preliminary gene expression analysis indicated that genes within the phenazine-like BGC are silent in *S. scabiei* when it is cultured in the lab (D. Bignell, unpublished), and thus one of the goals of this work was to try and activate the expression of the BGC. In order to activate gene expression, a number of engineered *S. scabiei* strains were constructed: a *SCAB12091* (GBL receptor) deletion strain; a *SCAB12101* (pseudo GBL receptor) deletion strain; a *SCAB12091* + *SCAB12101* double deletion strain; a *SCAB12081* (GBL synthase) deletion strain, which was initially constructed as part of a collaboration with another research group; and two *SCAB12111* (transcriptional activator) overexpression strains (one using the WT strain and one using the *SCAB12101* deletion mutant). In addition, strong, constitutive promoters were introduced upstream of the phenazine biosynthetic genes in *S. scabiei*, and some of the genes from the phenazine-like BGC were moved into three different heterologous hosts to determine if this would enable activation of gene expression.

3.2.1 Construction of the Δ*SCAB12081*, Δ*SCAB12091*, Δ*SCAB12101* and Δ(*SCAB12091*+ *SCAB12101*) mutant strains

Bioinformatics analyses suggested that *SCAB12091* and *SCAB12101* encode a GBL receptor and a pseudo GBL receptor, and these proteins have been shown in other *Streptomyces* spp. to function as repressors of gene expression; *SCAB12081* encodes a GBL synthase, which has been reported to be involved in the regulation of gene expression. To construct the regulatory gene deletion mutants of *S. scabiei,* the target genes were each replaced with a DNA cassette conferring

resistance to the antibiotic hygromycin B using the Redirect PCR targeting system (Gust et al. 2003a, 2004). First, mutant cosmids were constructed in which the target gene(s) was replaced with a [*hyg* + *oriT*] extended resistance cassette, and verification of gene replacement in each cosmid was performed using PCR (Figure 3.19 – Figure 3.22). The Δ*SCAB12081* mutant cosmid used was previously constructed by D. Bignell (unpublished results). Then, one mutant cosmid clone for each gene deletion was conjugated into WT *S. scabiei,* and hyg^r transconjugants were screened for kan^s (kanamycin sensitivity) to identify isolates that resulted from a double cross-over recombination event. In total, three *SCAB12091* mutant isolates, four *SCAB12101* mutant isolates, three *SCAB12091* + *SCAB12101* double deletion mutants and four *SCAB12081* mutant isolates were obtained, and all of the mutant isolates were verified by PCR (Figure 3.19 – Figure 3.22).

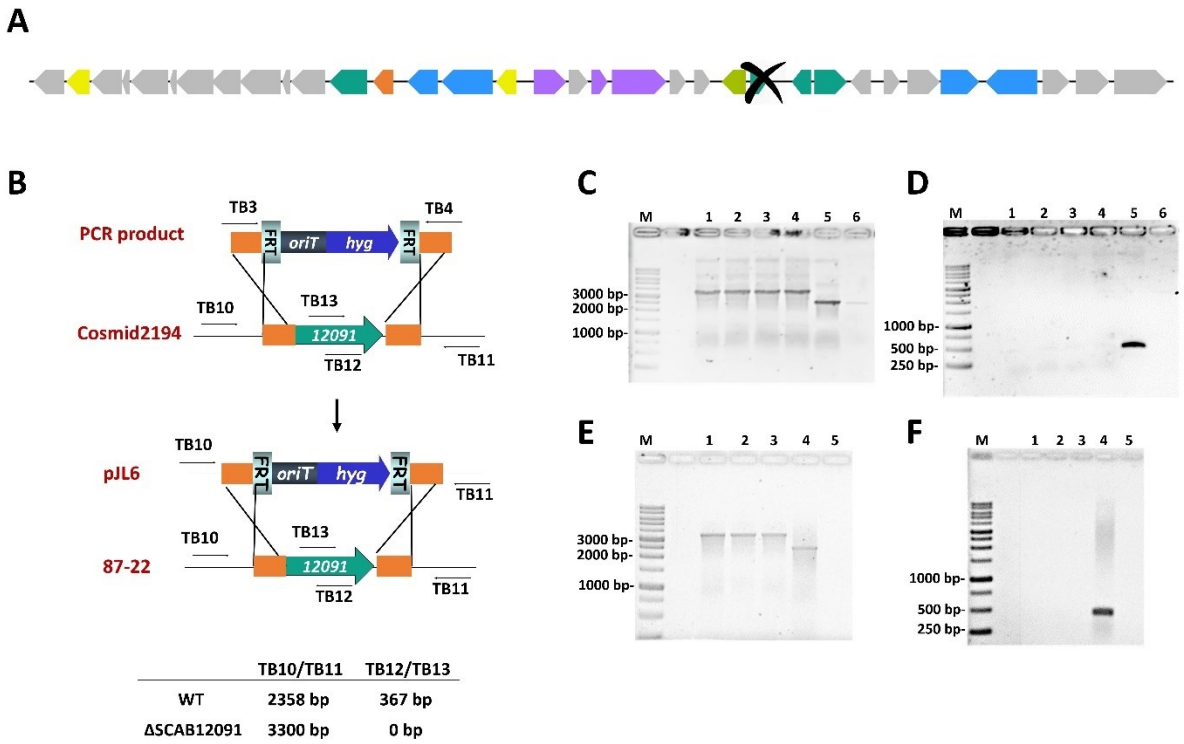
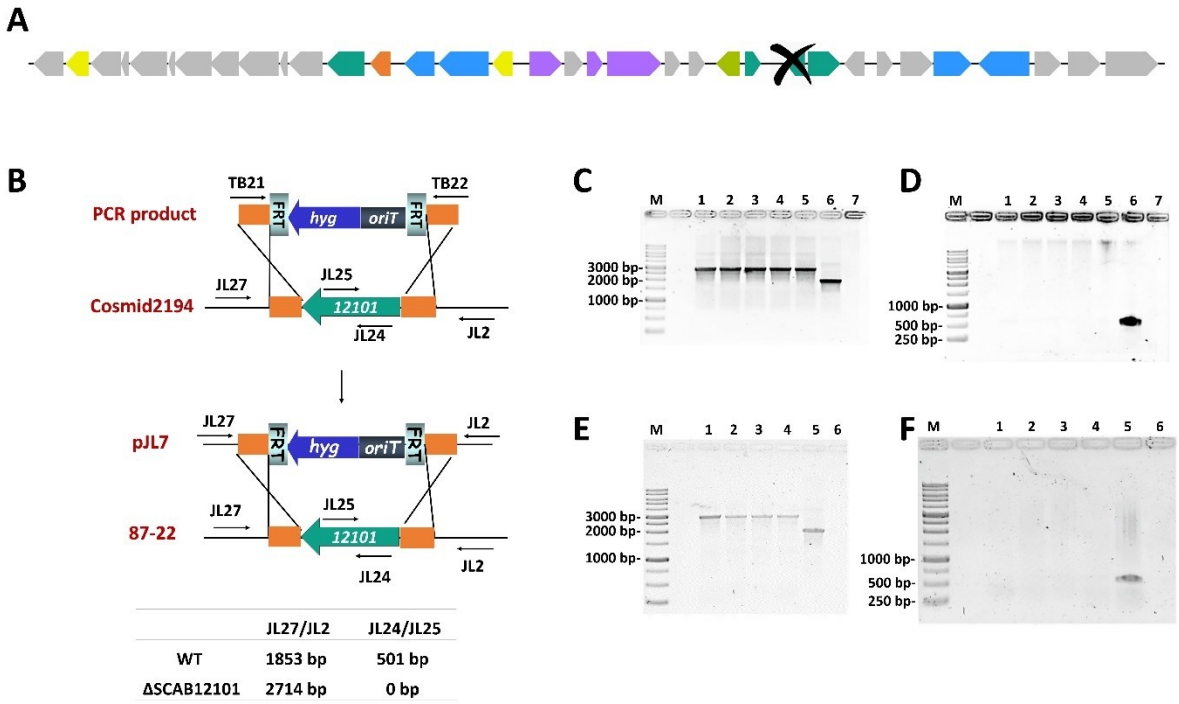


Figure 3.19 Construction and verification of the *S. scabiei SCAB12091* deletion mutant. **A**: Diagram of the phenazine-like BGC and the location of the *SCAB12091* gene that was targeted for deletion. **B**: Diagram illustrating the construction of the *S. scabiei* Δ*SCAB12091* mutant cosmid and the strategy used for cosmid verification by PCR. The primers used for cosmid verification are represented by the small arrows, and the expected product sizes are indicated. **C and D**: Verification of the Δ*SCAB12091* mutant cosmids (pJL6) using primers TB10/TB11 and TB12/TB13, respectively. Lane M: 1kb DNA ladder; Lanes 1–4: pJL6 clones #1-4; Lane 5: Cosmid 2194; Lane 6: water. **E and F**: Verification of the *SCAB12091* deletion mutant isolates using primers TB10/TB11 and TB12/TB13, respectively. Lane M: 1kb DNA ladder; Lanes 1–3: *S. scabiei* Δ*SCAB12091* isolates # 1-3; Lane 4: *S. scabiei* 87–22; Lane 5: water.



	JL27/JL2	JL24/JL25
WT	1853 bp	501 bp
ΔSCAB12101	2714 bp	0 bp

Figure 3.20 Construction and verification of the *S. scabiei SCAB12101* deletion mutant. **A**: Diagram of the phenazine-like BGC and the location of the *SCAB12101* gene that was targeted for deletion. **B**: Diagram illustrating the construction of the *S. scabiei* Δ*SCAB12101* mutant cosmid and the strategy used for cosmid verification by PCR. The primers used for cosmid verification are represented by the small arrows, and the expected product sizes are indicated. **C and D**: Verification of the Δ*SCAB12101* mutant cosmids (pJL7) using primers JL27/JL2 and JL24/JL25, respectively. Lane M: 1kb DNA ladder; Lanes 1–5: pJL7 clones #1-5; Lane 6: Cosmid 2194; Lane 7: water. **E and F**: Verification of the *SCAB12101* deletion mutant isolates using primers JL27/JL2 and JL24/JL25, respectively. Lane M: 1kb DNA ladder; Lanes 1–4: *S. scabiei* Δ*SCAB12101* isolates # 1-4; Lane 5: *S. scabiei* 87–22; Lane 6: water.

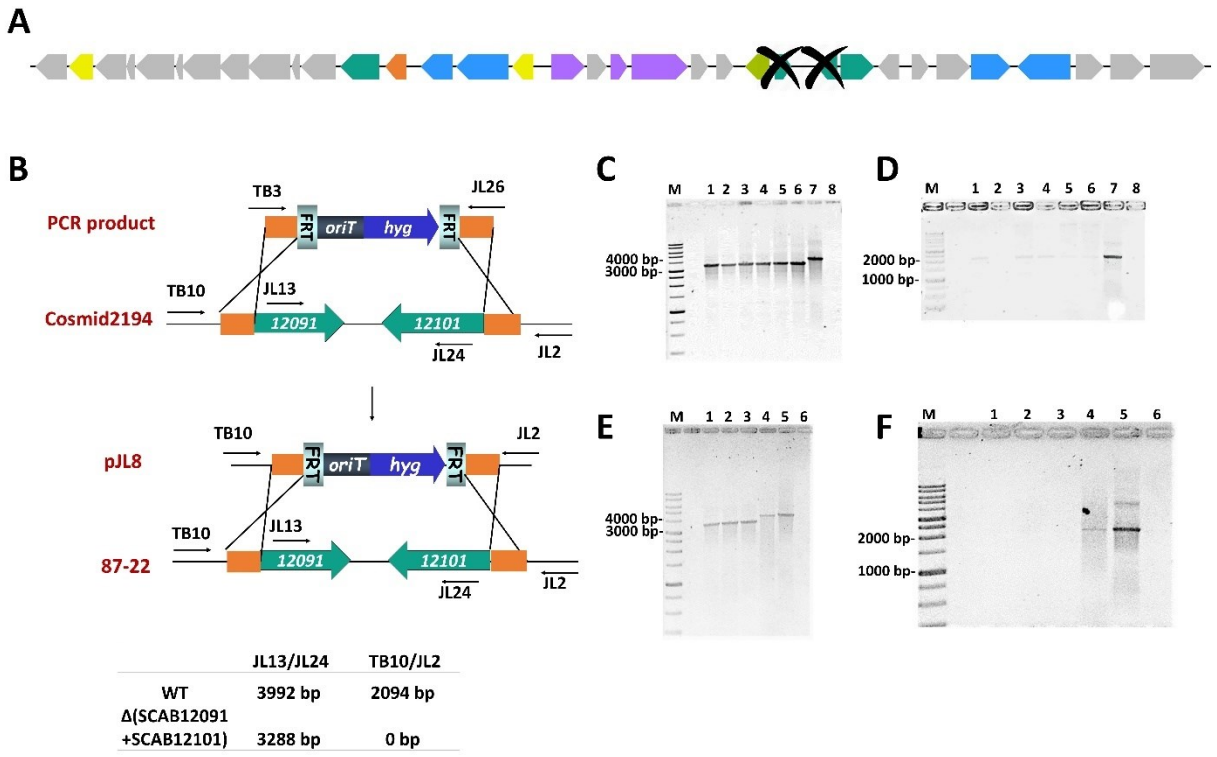


Figure 3.21 Construction and verification of the *S. scabiei SCAB12101+SCAB12091* deletion mutant. **A**: Diagram of the phenazine-like BGC and the location of the *SCAB12101* and *SCAB12091* genes that were targeted for deletion. **B**: Diagram illustrating the construction of the *S. scabiei* Δ(*SCAB12091+SCAB12101*) mutant cosmid and the strategy used for cosmid verification by PCR. The primers used for cosmid verification are represented by the small arrows, and the expected product sizes are indicated. **C and D**: Verification of the Δ(*SCAB12091+SCAB12101*) mutant cosmids (pJL8) using primers TB10/JL2 and JL13/JL24, respectively. Lane M: 1kb DNA ladder; Lanes 1–6: pJL8 clones #1-6; Lane 7: Cosmid 2194; Lane 8: water. **E and F**: Verification of the *SCAB12091+SCAB12101* deletion mutant isolates using primers TB10/JL2 and JL13/JL24, respectively. Lane M: 1kb DNA ladder; Lanes 1–3: *S. scabiei* Δ(*SCAB12091+SCAB12101*) isolates # 1-3; Lane 4: Cosmid 2194; Lane 5: *S. scabiei* 87-22; Lane 6: water.

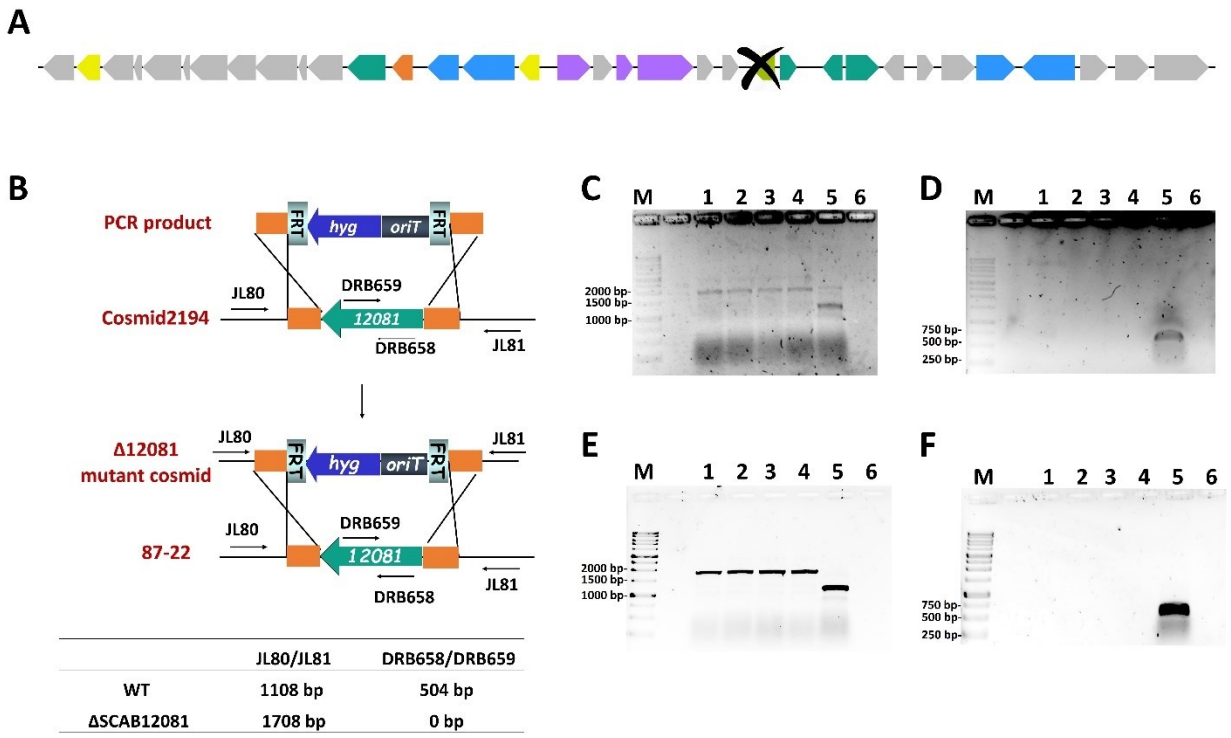


Figure 3.22 Construction and verification of the *S. scabiei SCAB12081* deletion mutant. **A**: Diagram of the phenazine-like BGC and the location of the *SCAB12081* gene that was targeted for deletion. **B**: Diagram illustrating the construction of the *S. scabiei* Δ*SCAB12081* mutant cosmid and the strategy used for cosmid verification by PCR. The primers used for cosmid verification are represented by the small arrows, and the expected product sizes are indicated. **C and D**: Verification of the Δ*SCAB12081* mutant cosmids using primers JL80/JL81 and DRB658/ DRB659, respectively. Lane M: 1kb DNA ladder; Lanes 1–4: Δ*SCAB12081* mutant cosmid clones #1-4; Lane 5: Cosmid 2194; Lane 6: water. **E and F**: Verification of the Δ*SCAB12081* deletion mutant isolates using primers JL80/JL81 and DRB658/ DRB659, respectively. Lane M: 1kb DNA ladder; Lanes 1–4: *S. scabiei* Δ*SCAB12081* isolates #1-4; Lane 5: *S. scabiei* 87–22; Lane 6: water.

3.2.2 Construction of the *S. scabiei SCAB12111* overexpression strain

The bioinformatic analysis conducted here suggested that overexpression of *SCAB12111* may lead to activation of the phenazine-like BGC. Thus, the gene was cloned into the plasmid pIJ8641 as described in Chapter 2 and as illustrated in Figure 3.30. pIJ8641 harbours the strong, constitutive *ermE*p* promoter in front of the *egfp* gene, and can integrate into the ΦC31 *attB* site of *Streptomyces* chromosomes (Sun et al. 1999). pJL3 was constructed by replacing *egfp* in pIJ8641 with the *SCAB12111* coding sequence and was verified by restriction digestion and DNA sequencing. Two other expression plasmids, pJL1 and pJL2, were also constructed by replacing the *ermE*p* promoter in pIJ8641 with the *kasO*p* and SP44 promoters, respectively. The reason why we chose these two promoters is that *kasO*p* and SP44 have both been reported to be stronger promoters than the *ermE*p* promoter, with SP44 being the strongest of the three (Wang et al. 2013; Bai et al. 2015; Cheng 2018). The sequences of *kasO*p* and SP44 used in this study are shown in Figure S9. pJL1 and pJL2, in turn, were used to construct pJL4 and pJL5, respectively, where the *egfp* gene in each was replaced with the *SCAB12111* gene. The resulting recombinant plasmids were confirmed by restriction digestion (Figure 3.23) and DNA sequencing, after which pJL3, pJL4 and pJL5 along with the corresponding control plasmids (pIJ8641, pJL1 and pJL2, respectively) were each moved into *S. scabiei* 87-22 by intergeneric conjugation with *E. coli*. A pIJ8641 derivative lacking the *egfp* gene (pJL15) was also constructed and moved into *S. scabiei* 87-22 as an empty vector control strain to eliminate any potential effects of *egfp* overexpression on *S. scabiei*.

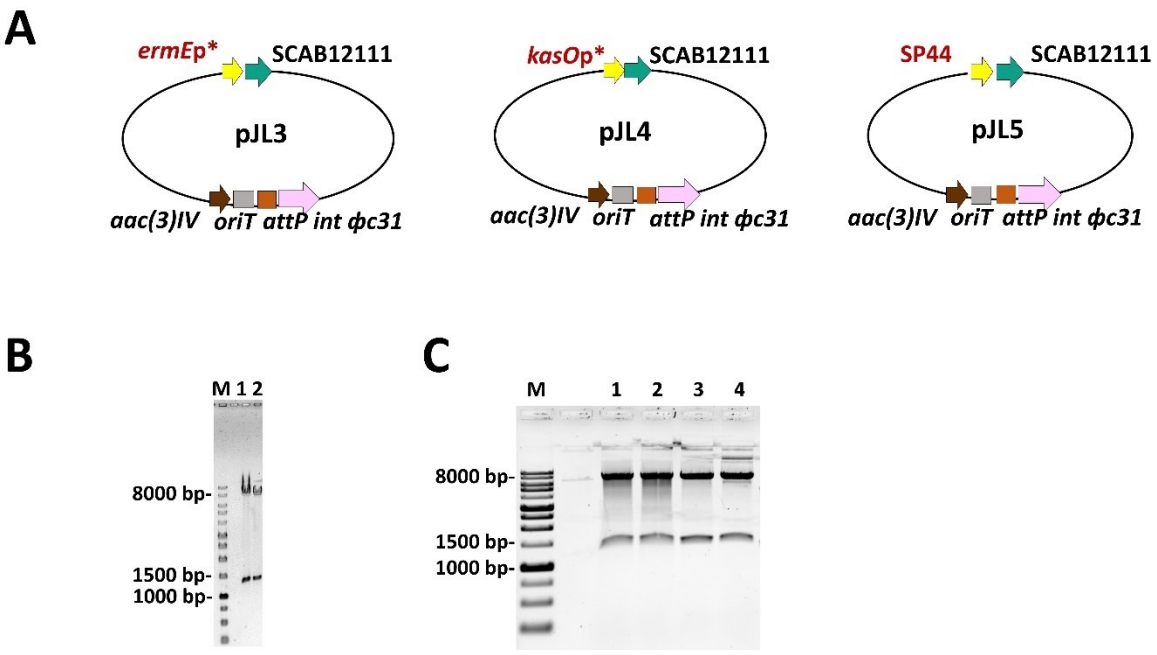


Figure 3.23 Construction of the *S. scabiei SCAB12111* overexpression strains. **A:** Diagram illustrating the map of *S. scabiei SCAB12111* overexpression plasmids pJL3, pJL4 and pJL5. **B:** Verification of pJL3 by digestion with *Nde*I and *Not*I. Lane M: 1kb DNA ladder; Lane 1–2: pJL3 clones #1-2; **C**: Verification of pJL4 and pJL5 by digestion with *Nde*I and *Not*I. Lane M: 1kb DNA ladder; Lane 1–2: pJL4 clones #1-2; Lane 3–4: pJL5 clones #1-2.

3.2.3 Overexpression of the phenazine biosynthetic genes using the *ermE*p*, *kasO*p* and SP44 promoters

Based on the genetic organization of *SCAB12021-SCAB12071* (Figure 3.9), it was predicted that these genes are co-transcribed. Plasmids were therefore constructed that enabled the introduction of the *ermE*p*, *kasO*p* and SP44 promoters upstream of the *SCAB12021-SCAB12071* genes within the phenazine-like BGC in *S. scabiei*. The *SCAB12021* coding sequence was cloned into three different plasmids: pJL16, pJL9, and pJL10, and this generated pJL11, pJL12 and pJL13, respectively, which were all verified by restriction digestion (Figure 3.24) and DNA sequencing. All three plasmids are non-replicating in *Streptomyces* (Sun et al. 1999) but can integrate into the *S. scabiei* chromosome via homologous recombination between the chromosomal and plasmid copies of *SCAB12021*. This, in turn, would result in expression of one copy of the *SCAB12021* gene along with any downstream co-transcribed genes from the *ermE*p*, *kasO*p* and SP44 promoters (Figure 3.24).

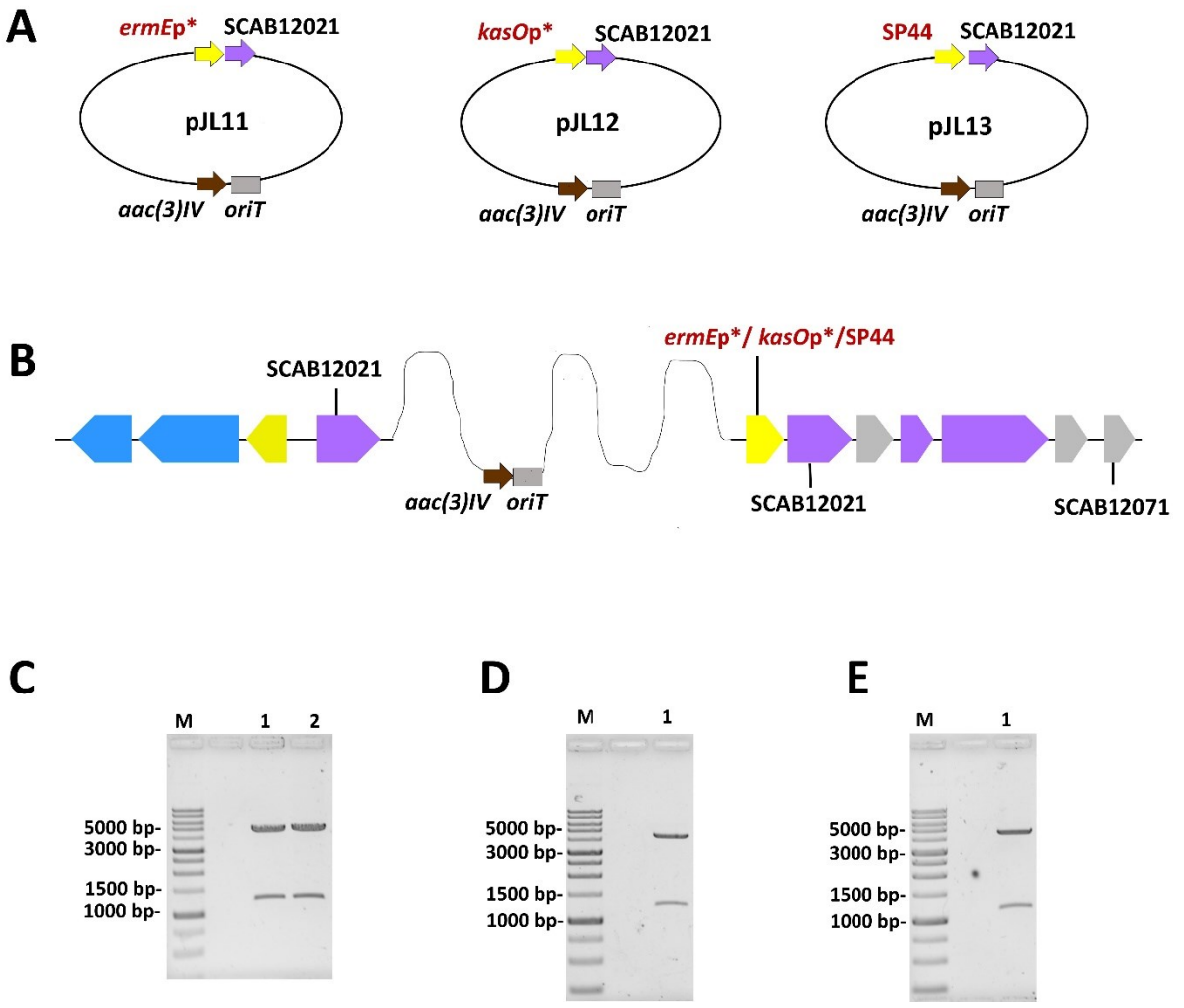
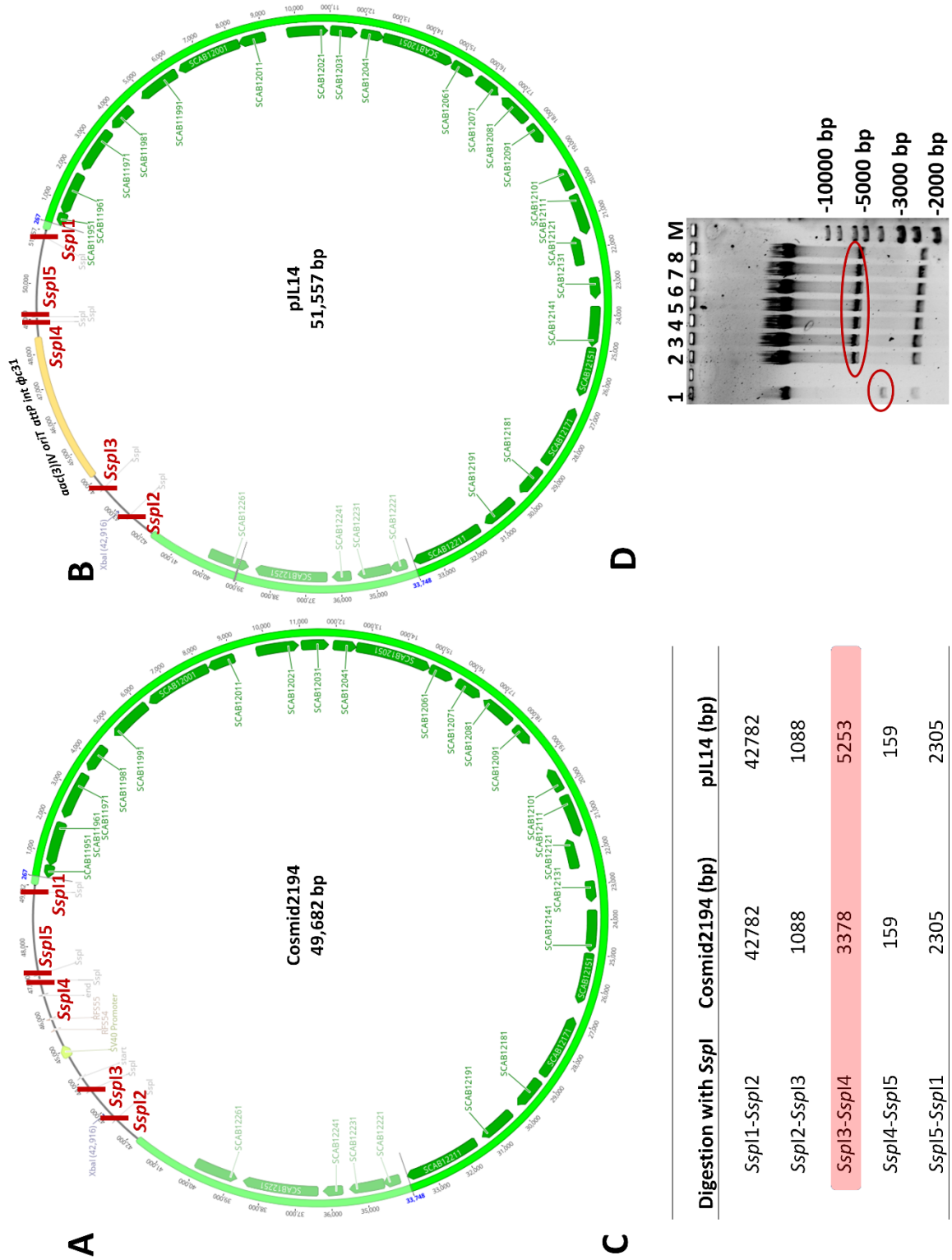


Figure 3.24 Construction of plasmids for introducing the *ermE*p*, *kasO*p* and SP44 promoters into the phenazine-like BGC in *S. scabiei*. **A**: Diagram illustrating the map of pJL11 (*ermE*p*-*SCAB12021*), pJL12 (*kasO*p*-*SCAB12021*) and pJL13 (SP44-*SCAB12021*). **B**: Diagram illustrating what the phenazine-like BGC look like upon integration of the plasmids: two copies of *SCAB12021* separated by the vector backbone, and the copy with *ermE*p*/*kaspO*p*/SP44 upstream of it will be followed by the other phenazine biosynthetic genes. **C**: Verification of pJL11 by digestion with *Nde*I and *Not*I. Lane M: 1kb DNA ladder; Lane 1–2: pJL11 clones #1-2; **D**: Verification of pJL12 by digestion with *Nde*I and *Not*I. Lane M: 1kb DNA ladder; Lane 1: pJL12. **E**: Verification of pJL13 by digestion with *Nde*I and *Not*I. Lane M: 1kb DNA ladder; Lane 1: pJL13.

3.2.4 Heterologous gene expression in *S. avermitilis* SUK17, *S. coelicolor* M1154 and *S. albus* J1074

S. avermitilis SUK17, *S. albus* J1074 and *S. coelicolor* M1154 have been engineered and successfully used as heterologous hosts to produce a variety of cryptic natural products (Baltz 2010; Zhang et al. 2019). In this study, these three strains were selected and used as heterologous hosts for the expression of the phenazine-like BGC. Briefly, Cosmid 2194 harbours a 41.7 kb fragment of the *S. scabiei* 87-22 chromosome, including 33.5 kb (*SCAB11951-SCAB12211*) of the predicted phenazine-like BGC (*SCAB11861-SCAB12211*) (Figure 3.25). The Redirect PCR targeting system was used to introduce a DNA fragment into the backbone of Cosmid 2194 in order to allow the cosmid to integrate into the chromosome of the heterologous hosts *S. avermitilis* SUK17, *S. coelicolor* M1154 and *S. albus* J1074. pIJ10702 (also known as pMJCOS1) was digested with *Ssp*I to release a 5,247 bp fragment containing *oriT*, *attP*, integrase (*int)* and the apramycin resistance gene *aac(3)IV*. The fragment was gel–purified and then electroporated into *E. coli* BW25113/pIJ790 containing Cosmid 2194, and transformants were selected using ampicillin (to select for Cosmid 2194) and apramycin (to select for integration of the [*oriT–attP–int–aac(3)IV*] fragment into the cosmid). The resulting cosmid, pJL14, was verified by restriction enzyme digestion with *Ssp*I to confirm that the fragment had been successfully introduced into the backbone by replacing the neomycin resistance gene; the expected banding pattern of the digestion of Cosmid 2194 and pJL14 was shown in Figure 3.25. Then, pJL14 was introduced into the three heterologous hosts by conjugation with *E. coli*, and the resulting exconjugants were screened using apramycin.



C

Digestion with *Sspl*	Cosmid2194 (bp)	pJL14 (bp)
*Sspl*1-*Sspl*2	42782	42782
*Sspl*2-*Sspl*3	1088	1088
*Sspl*3-*Sspl*4	3378	5253
*Sspl*4-*Sspl*5	159	159
*Sspl*5-*Sspl*1	2305	2305

Figure 3.25 Heterologous expression of the gene cluster in *S. avermitilis* SUK17, *S. coelicolor* M1154 and *S. albus* J1074. **A and B**: Diagram illustrating the map of Cosmid 2194 and pJL14.

Cosmid 2194 harbours a 41.7 kb fragment of the *S. scabiei* 87-22 chromosome (*SCAB11951-SCAB12261*), including 33.5 kb (*SCAB11951-SCAB12211,* highlighted) of the predicted phenazine-like BGC (*SCAB11861-SCAB12211*). The restriction enzyme digestion sites of *Ssp*I are labelled in red font. **C**: The expected banding pattern of the digestion of Cosmid 2194 and pJL14 with *Ssp*I. The expected difference was highlighted in red. **D**: Verification of pJL14 by digestion with *Ssp*I. Lane M: 1kb DNA ladder; Lane 1: Cosmid 2194; Lanes 1–6: pJL14 clones#1-6. The expected bands showing the difference are circled in red.

3.3 Gene expression, morphological and bioactivity analysis of the engineered *Streptomyces* strains

3.3.1 Gene expression analysis of the *Streptomyces* engineered strains

Previous gene expression analysis in WT *S. scabiei* revealed that the predicted core phenazine biosynthetic genes from the phenazine-like BGC were silent when the organism was cultured in oat bran medium (D. Bignell, unpublished). To determine whether deletion of the putative GBL receptor- and/or the pseudo GBL receptor-encoding genes can activate expression of the genes, the Δ*SCAB12091*, Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*) mutant strains as well as the WT were cultured for 44 hours on five different agar media (MYMm, SFM, YMSm, ISP-4, SA agar), after which total RNA was isolated from mycelia scraped from the plates. Semi-quantitative RT-PCR was then performed to analyze the relative expression of *SCAB12021* (*phzC*), *SCAB12041* (*phzD*) and *SCAB12051* (*phzE*) in each strain, with the *murX* gene serving as a positive control. As shown in Figure 3.26, all three phenazine biosynthetic genes were expressed at low levels or not at all in the WT strain on all five of the media tested. Deletion of *SCAB12091* did not affect the expression of the genes on the five media, but deletion of *SCAB12101* induced the expression of all three genes when the mutant was cultured on MYMm and YMSm. The double deletion mutant showed a similar gene expression profile as the Δ*SCAB12101* mutant (Figure 3.26). Together, these expression results show that *SCAB12101* but not *SCAB12091* affects the

expression of the core phenazine biosynthetic genes, and that YMSm and MYMm are the best media for activation of gene expression.

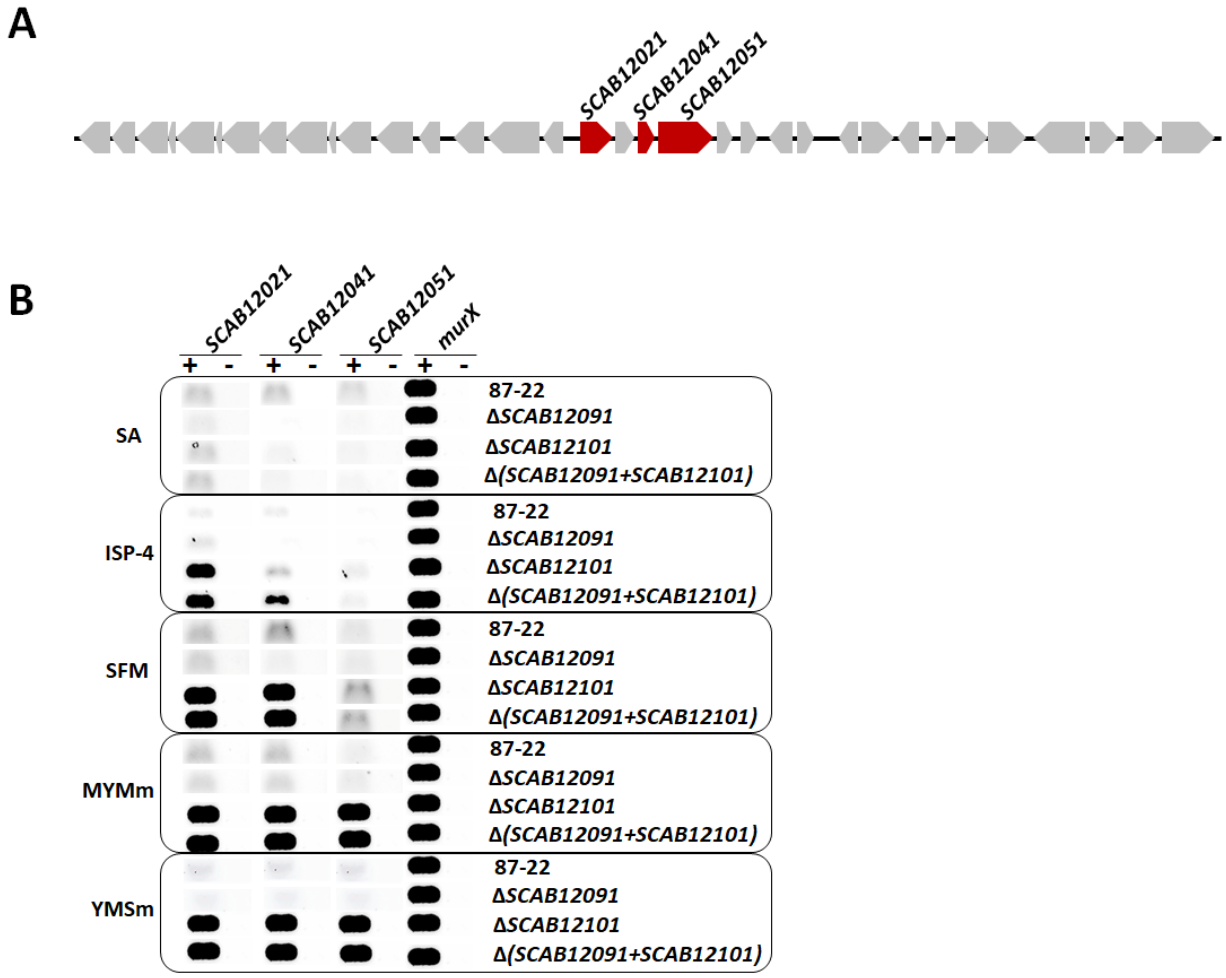


Figure 3.26 RT-PCR analysis of the putative phenazine biosynthetic genes within the phenazine-like BGC. **A:** The *SCAB12021*, *SCAB12041* and *SCAB12051* genes whose expression was assessed are indicated in red. **B:** Gene expression was assessed in WT *S. scabiei* 87-22 and in the Δ*SCAB12091*, Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*) mutant strains cultured on five different media (MYMm, SFM, YMSm, ISP-4, and SA agar). The *murX* gene was used as a positive control. Reactions were conducted with cDNA template that was prepared using reverse transcriptase (+), and no reverse transcription (-) control templates were included. Primers targeting the *SCAB12021*, *SCAB12041*, *SCAB12051* and *murX* genes were used in PCR reactions that were performed with 25 cycles for all of the genes.

Semi-quantitative RT-PCR was also performed on the Δ*SCAB12101* and WT strains to analyze the expression of the genes upstream of the core phenazine biosynthetic genes. As shown in Figure 3.27, deletion of *SCAB12101* increased the expression of all of these genes, though expression could also be readily detected in the WT strain cultured on YMSm. Thus, *SCAB12101* appears to influence the expression of multiple genes in the phenazine-like BGC.

The expression of the core phenazine biosynthetic genes was additionally assessed in the Δ*SCAB12081* mutant, which is predicted to be unable to make GBL signalling molecules. As shown in Figure 3.27, deletion of *SCAB12081* activated the expression of *SCAB12021*, *SCAB12041* and *SCAB12051* when the strain was cultured on MYMm and YMSm, similar to what was observed in the Δ*SCAB12101* mutant. Interestingly, expression of the predicted transcriptional regulatory gene *SCAB12111* was also activated in the Δ*SCAB12081* mutant compared to the WT strain, and similar results were observed in the Δ*SCAB12101* and Δ(*SCAB12091*+*SCAB12101*) mutants but not in the Δ*SCAB12091* mutant. Therefore, both *SCAB12081* and *SCAB12101* appear to affect the expression of the core phenazine biosynthetic genes as well as the putative transcriptional regulatory gene *SCAB12111*.

The effect of *SCAB12111* overexpression on transcription of the core phenazine biosynthetic genes was examined next. As shown in Figure 3.27, expression of *SCAB12021, SCAB12041* and *SCAB12051* was slightly elevated in the vector control strain (WT containing pJL15) compared to WT, whereas expression was much higher for all three genes in the *SCAB12111* overexpression strain (WT containing plasmid pJL3). When *SCAB12111* was overexpressed in the Δ*SCAB12101* mutant background, expression of *SCAB12021, SCAB12041* and *SCAB12051* was noticeably higher than in the vector control strain (Δ*SCAB12101* containing plasmid pJL15) and in the Δ*SCAB12101* mutant (Figure 3.28). This was observed when the strains

were cultured on MYMm and YMSm as well as on an additional medium, OBA, which is known to support the production of thaxtomins, concanamycins and coronafacoyl phytotoxins in *S. scabiei* (Johnson et al. 2007; Bignell et al. 2010b; Fyans et al. 2016).

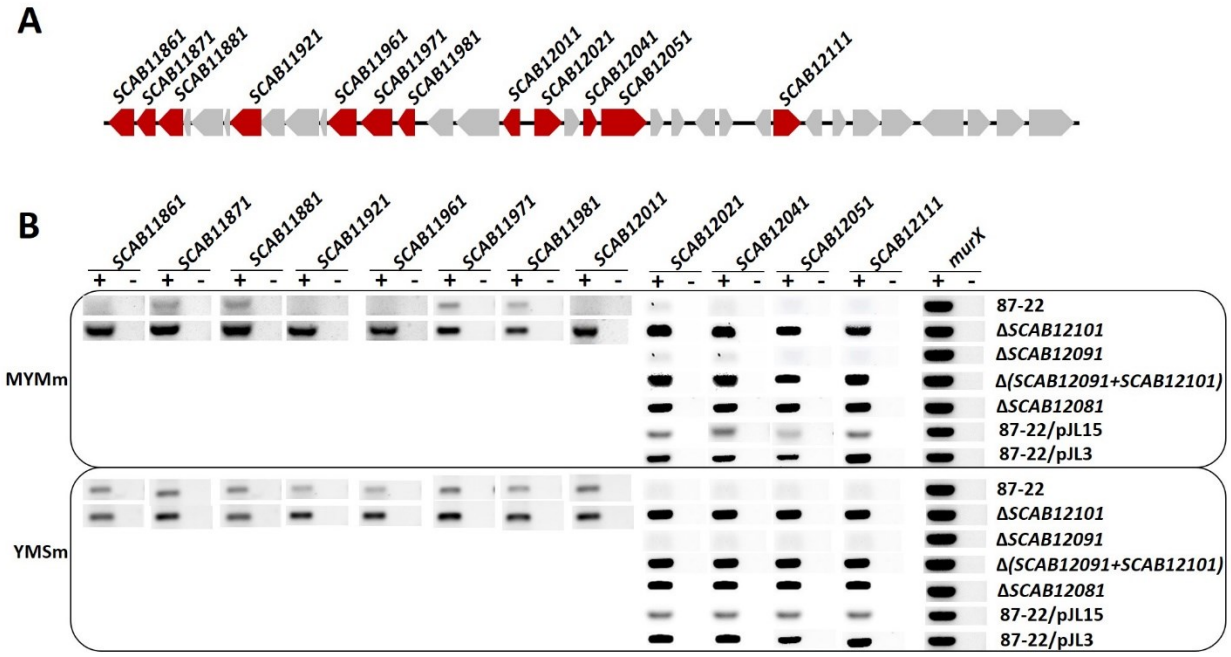


Figure 3.27 RT-PCR analysis of the putative phenazine biosynthetic genes within the phenazine-like BGC. **A:** The genes (*SCAB11861*, *SCAB11871, SCAB11881, SCAB11921, SCAB11961, SCAB11971, SCAB11981, SCAB12011, SCAB12021, SCAB12041, SCAB12051* and *SCAB12111*) whose expression was assessed are indicated in red. **B:** Gene expression was assessed in WT *S. scabiei* 87-22, Δ*SCAB12101*, Δ*SCAB12091*, Δ(*SCAB12091*+*SCAB12101*), Δ*SCAB12081,* 87-22 harbouring the empty vector (87-22/pJL15), and 87-22 harbouring the *SCAB12111* overexpression plasmid (87-22/pJL3) following cultivation on two different media (MYMm and YMSm). The *murX* gene was used as a positive control. Reactions were conducted with cDNA template that was prepared using reverse transcriptase (+), and no reverse transcription (-) control templates were included. Primers targeting the *SCAB11861*, *SCAB11871, SCAB11881, SCAB11921, SCAB11961, SCAB11971, SCAB11981, SCAB12011, SCAB12021, SCAB12041, SCAB12051* and *SCAB12111* and *murX* genes were used in PCR reactions that were performed with 25 cycles for *murX, SCAB12021*, *SCAB12041*, *SCAB12051, SCAB11921, SCAB11961*; 23 cycles for *SCAB11861*, *SCAB11871, SCAB11881* and *SCAB12011*; 21 cycles for *SCAB11971, SCAB11981* and *SCAB12111*.

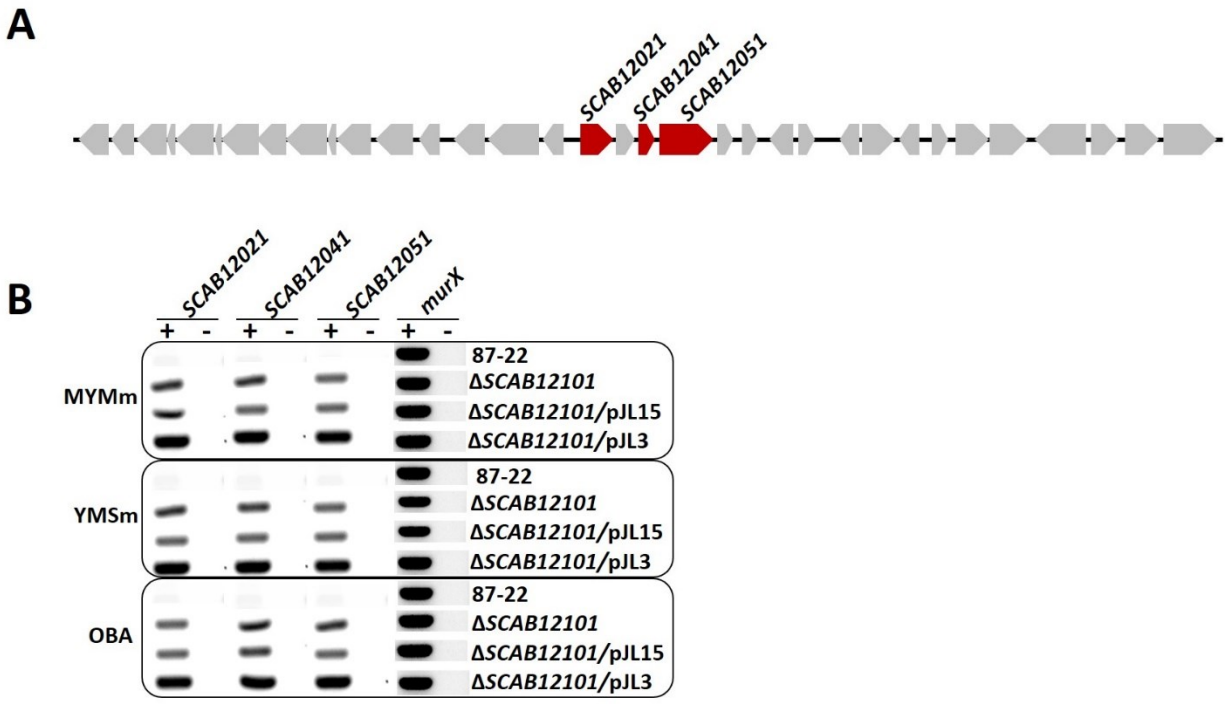


Figure 3.28 RT-PCR analysis of the putative phenazine biosynthetic genes within the phenazine-like BGC. **A:** The genes (*SCAB12021*, *SCAB12041*, *SCAB12051*) whose expression was assessed are indicated in red. **B:** Gene expression was assessed in WT *S. scabiei* 87-22, the Δ*SCAB12101* mutant strain, the Δ*SCAB12101* mutant harbouring the empty vector (Δ*SCAB12101*/pJL15), and the Δ*SCAB12101* mutant harbouring the *SCAB12111* overexpression plasmid (Δ*SCAB12101*/pJL3) following cultivation on MYMm, YMSm and OBA media. The *murX* gene was used as a positive control. Reactions were conducted with cDNA template that was prepared using reverse transcriptase (+), and no reverse transcription (-) control templates were included. Primers targeting the *SCAB12021, SCAB12041, SCAB12051* and *murX* genes were used in PCR reactions that were performed with 20 cycles for *SCAB12021*, and 19 cycles for *SCAB12041* and *SCAB12051.*

Overall, the RT-PCR results indicate that the predicted GBL biosynthetic gene *SCAB12081*, the predicted pseudo-GBL receptor gene *SCAB12101,* and the predicted transcriptional activator gene *SCAB12111* all contribute to the regulation of expression of the core phenazine biosynthetic genes within the phenazine-like BGC, and *SCAB12101* also influences the expression of other genes within the BGC. In contrast, the predicted GBL receptor gene *SCAB12091* appears to have no effect on expression of genes within the BGC. The results also show that expression of genes within the BGC can be enhanced by deleting *SCAB12081* or *SCAB12101* or by overexpressing *SCAB12111*, and thus the manipulation of regulatory genes within the BGC is a potentially useful strategy for studying the metabolite(s) produced by this gene cluster.

3.3.2 Bioactivity analysis of the engineered *Streptomyces* strains

3.3.2.1 Potato tuber slice bioassay for assessing phytotoxic activity and virulence

The phytotoxic activity of the engineered strains was tested using a potato tuber disk bioassay, which can detect the tissue necrosis and hypertrophy-inducing activity of phytotoxins within culture extracts. Assays using ethyl acetate extracts prepared from the various engineered *Streptomyces* strains cultured on three different media (MYMm, YMSm and OBA) demonstrated that the extracts prepared from uninoculated media and the heterologous expression strains had little or no effect on the tuber tissue (Figure 3.29). In contrast, extracts from WT *S. scabiei* 87-22 and the engineered *S. scabiei* strains were all able to cause necrosis and pitting of the tuber tissue, with the most severe effects observed with the OBA extracts. Compared to the WT extracts, many of the extracts from the engineered strains [e.g. Δ*SCAB12081,* Δ*SCAB12091,* Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*) and pJL3/87-22] showed more severe effects on the potato tuber tissue. However, the severity of the observed necrosis/pitting did not appear to correlate with the

phenazine-like BGC gene expression profiles observed (Figure 3.29).

The potato tuber tissue bioassay was also performed to assess the virulence phenotype of some of the engineered strains. As shown in Figure 3.30, when cultured on MYMm and YMSm, the Δ*SCAB12081* and Δ*SCAB12091* mutants caused similar necrosis and pitting as the WT strain, whereas the Δ*SCAB12101* and Δ*SCAB12091*/Δ*SCAB12101* mutants caused reduced necrosis and pitting of the tuber tissue compared to the WT strain. The *SCAB12111* overexpression strain was also found to cause less necrosis and pitting of the potato tuber compared with WT *S. scabiei* and the empty vector control (WT/pJL15) (Figure 3.30). In contrast, the heterologous expression strains, *S. avermitilis* SUK17/pJL14, *S. coelicolor* M1154/pJL14 and *S. albus* J1074/pJL14, did not cause any significant necrosis or pitting compared to the corresponding control strains (Figure 3.40).

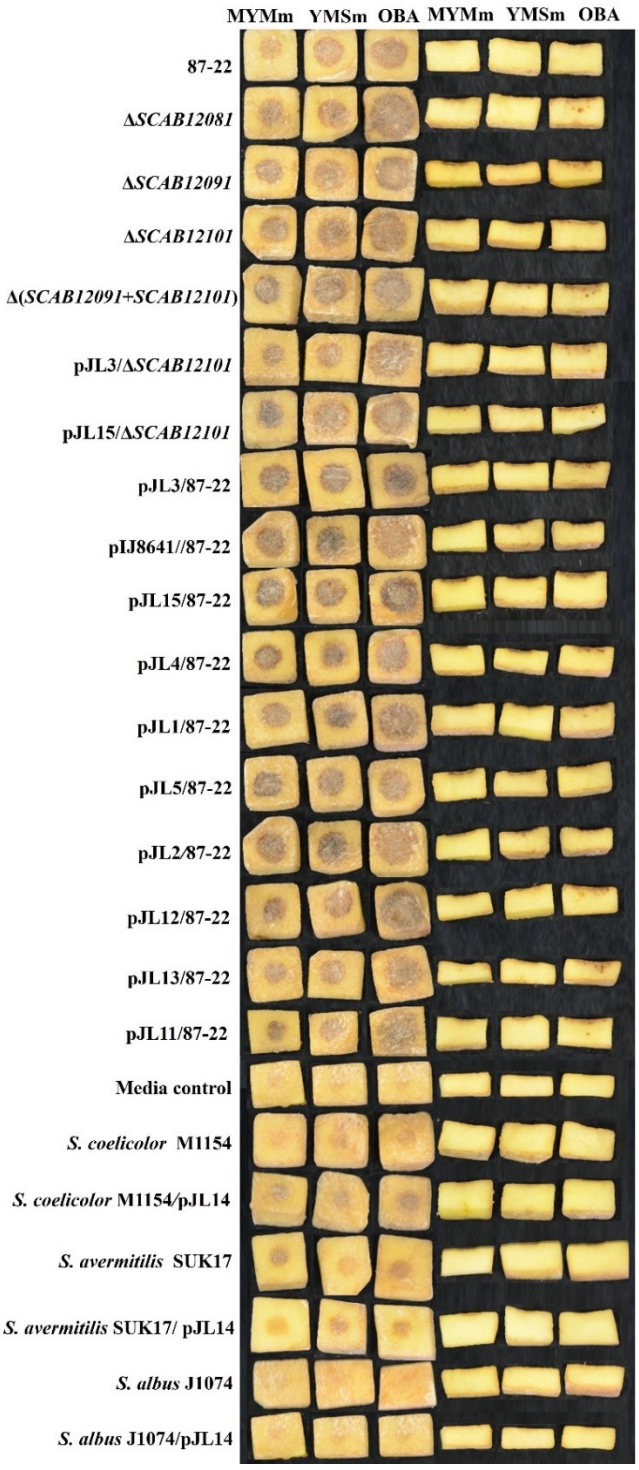


Figure 3.29 Potato tuber slice bioassay for detecting phytotoxic activity in ethyl acetate extracts of the engineered *Streptomyces* strains cultivated on MYMm, YMSm and OBA.

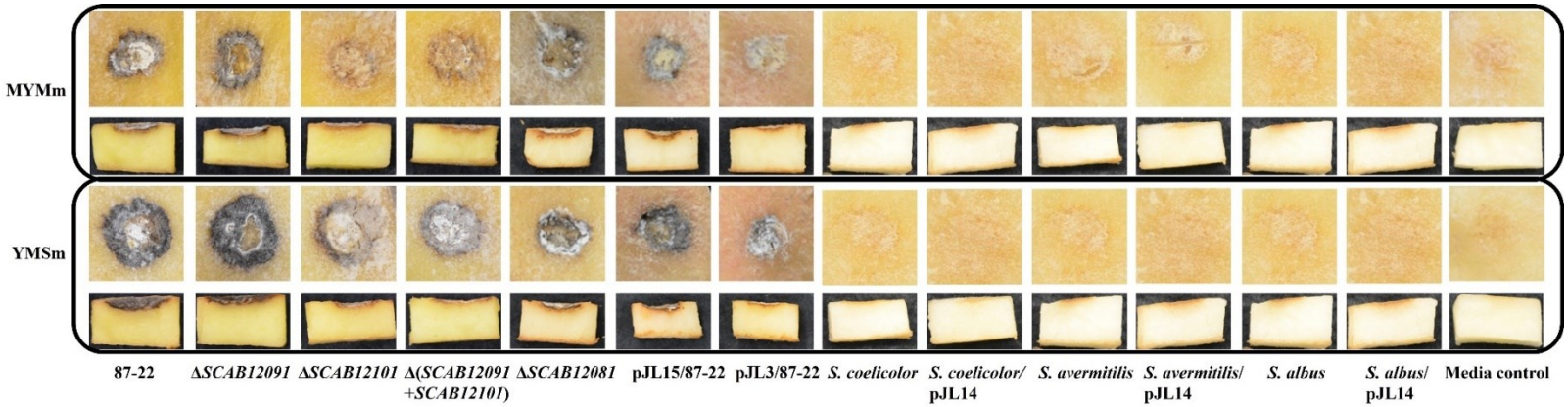

MYMm
YMSm
87-22
ΔSCAB12091
ΔSCAB12101
Δ(SCAB12091 +SCAB12101)
ΔSCAB12081
pJL15/87-22
pJL3/87-22
S. coelicolor
S. coelicolor/ pJL14
S. avermitilis
S. avermitilis/ pJL14
S. albus
S. albus/ pJL14
Media control

Figure 3.30 Potato tuber slice bioassay to assess the virulence phenotype of the *S. scabiei* deletion mutants Δ*SCAB12091*, Δ*SCAB12101* and Δ(*SCAB12091*+*SCAB12101*), the *S. scabiei SCAB12111* overexpression strain (87-22/pJL3), and of the heterologous expression strains *S. avermitilis* SUK17/pJL14, *S. coelicolor* M1154/pJL14 and *S. albus* J1074/pJL14. The strains were cultured on MYMm and YMSm prior to inoculation of the potato tuber slices. *S. scabiei* 87-22 and 87-22/pJL15 (empty vector control) were used as positive controls. The corresponding cosmid-deficient strains were used as negative controls.

3.3.2.2 Bioassays for detecting antimicrobial activity

To test whether or not the product(s) of the phenazine-like BGC displays antimicrobial activity, three different methods were employed: the agar plug assay method, disk diffusion assay method, and the soft agar overlay assay method. No activity against any of the indicator organisms was detected with the agar plug and disk diffusion assays for any of the engineered strains (data not shown). However, when using the soft agar overlay method, antifungal activity was detected from the WT *S. scabiei* 87-22 and all other *S. scabiei* engineered strains cultured on YMSm (Figure 3.31). In addition, WT *S. scabiei* harbouring the *SCAB12111* overexpression plasmid (87-22/pJL3) and the Δ*SCAB12101* mutant both showed activity against *B. subtilis*, while *S. avermitilis* SUK17/pJL14 showed activity against *Staphylococcus epidermidis* (Figure 3.31). However, the activities were not reproducible when the assays were repeated several (> 3) times. Specifically, the antibacterial activities were detected the first time the assay was performed, but with repeated assays, the activities were detected in some instances but not in others.

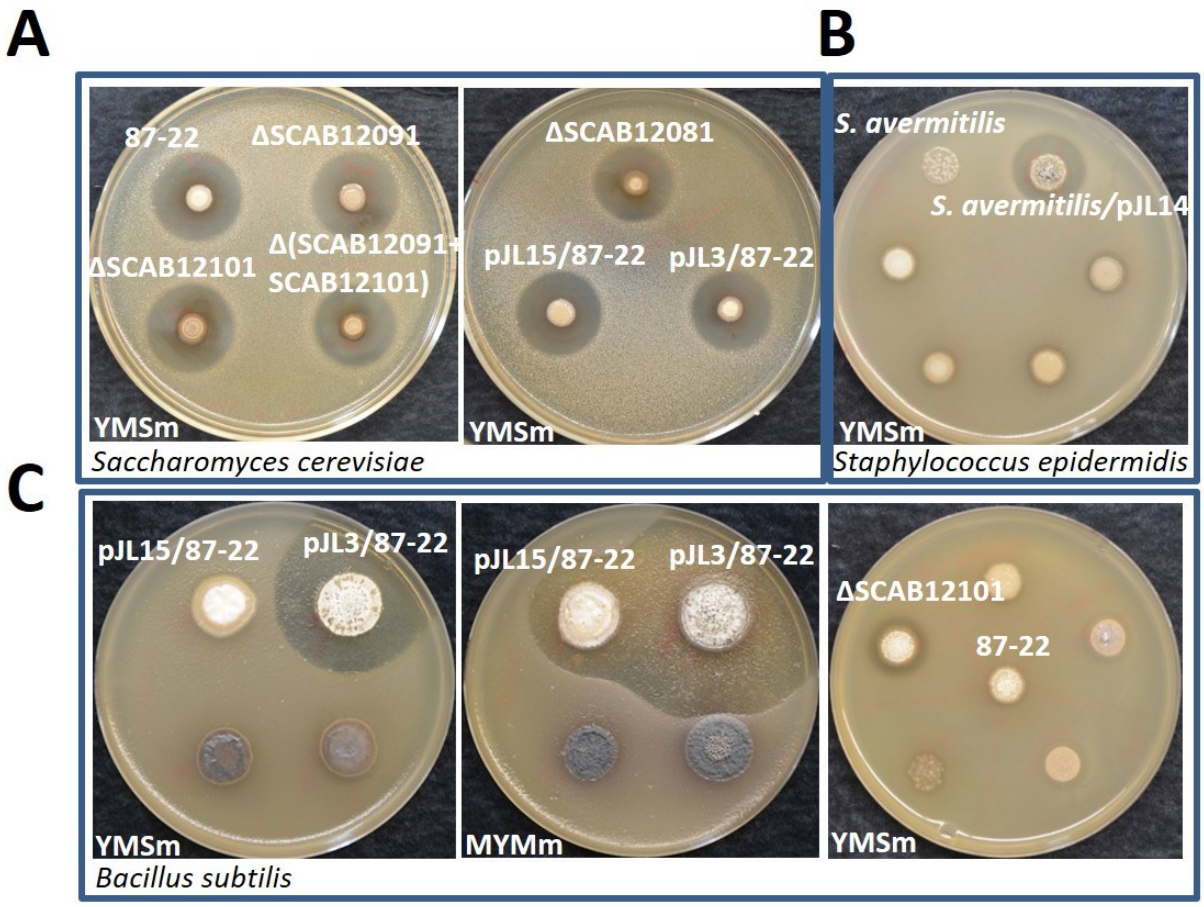


Figure 3.31 Soft agar overlay assay to assess the bioactivity of the *Streptomyces* engineered strains. *Streptomyces* spores was cultured on YMSm/MYMm at 28 °C for 4 days and then overlayed with soft nutrient agar containing *S. cerevisiae* (**A**), *S. epidermidis* (**B**) or *B. subtilis* (**C**). Photographs were taken 24 hrs following addition of the overlay.

3.3.3 Morphological analysis of the engineered *Streptomyces* strains

To investigate whether or not the phenazine-like BGC affects the morphological phenotype of the engineered *Streptomyces* strains, the strains along with their corresponding controls were cultured for 7 days on four different agar media: MYMm, YMSm, OBA and SFM (Figure 3.32 – Figure 3.34). No noticeable differences in morphology were observed for most of the engineered strains cultured on these four media. The exceptions were Δ*SCAB12091*, Δ*SCAB12101*,

Δ(*SCAB12091*+*SCAB12101*), Δ*SCAB12101*/pJL15, Δ*SCAB12101*/pJL3, and 87-22/pIJ8641, which all showed delays in growth on SFM. The morphological differences between 87-22/pJL15 and 87-22/pIJ8641, both of which were used as vector control strains, suggest that overexpression of the *egfp* gene in the pIJ8641 vector may have some effects on morphological development in *S. scabiei*. It is noteworthy that a light reddish color was observed on the backside of the MYMm, YMSm and OBA for the WT *S. scabiei* strain overexpressing *SCAB12111* from the SP44 promoter (Figure 3.33), and on the backside of OBA for *S. albus* strain harbouring Cosmid 2194 (Figure 3.34). Whether this observed pigment is related to the phenazine-like BGC requires further investigation.

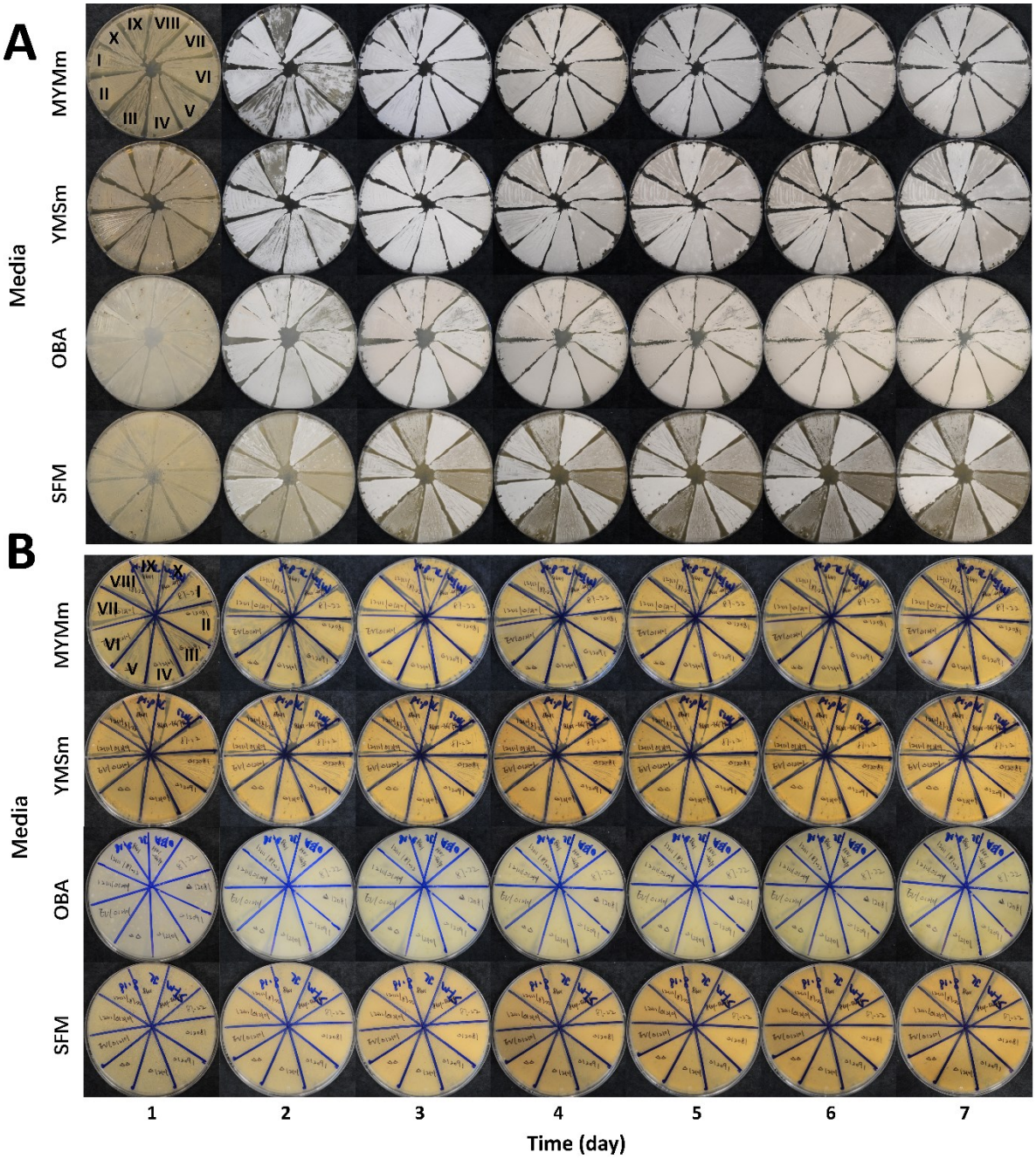


Figure 3.32 Growth of the engineered *S. scabiei* strains and their control strains on MYMm, YMSm, OBA and SFM agar. I, WT 87-22; II, Δ*SCAB12081*; III, Δ*SCAB12091*; IV, Δ*SCAB12101*; V, Δ(*SCAB12091*+*SCAB12101*); VI, Δ*SCAB12101*/pJL15; VII, Δ*SCAB12101*/pJL3; VIII, 87-22/pJL3; IX, 87-22/pIJ8641; X, 87-22/pJL15. **A:** The front side of the plates. **B:** The backside of the plates. Photos were taken every 24 hours for 7 days.

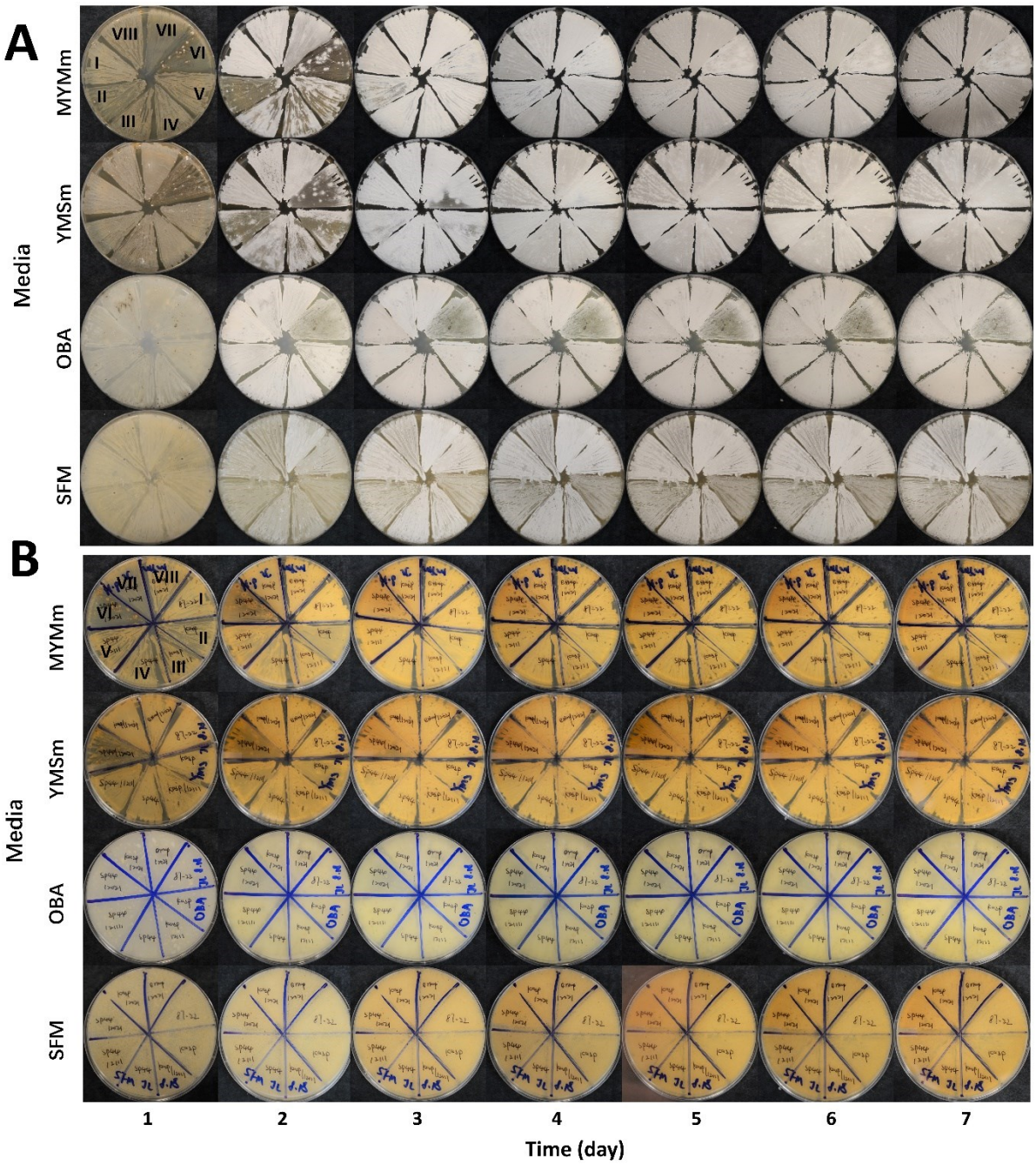


Figure 3.33 Growth of *Streptomyces* engineered strains and their control strains on MYMm,YMSm, OBA and SFM agar. I, 87-22; II, pIJ8641/*kasO*p*/87-22; III, pIJ8641/*kasO*p*/*SCAB12111*/87-22; IV, pIJ8641/SP44/87-22; V, pIJ8641/SP44/*SCAB12111*/87-22; VI, SP44/*SCAB12021*/87-22; VII, *kasO*p*/*SCAB12021*/87-22; VIII, ermEp*/*SCAB12021*/87-22. **A:** The front side of the plates. **B:** The backside of the plates.

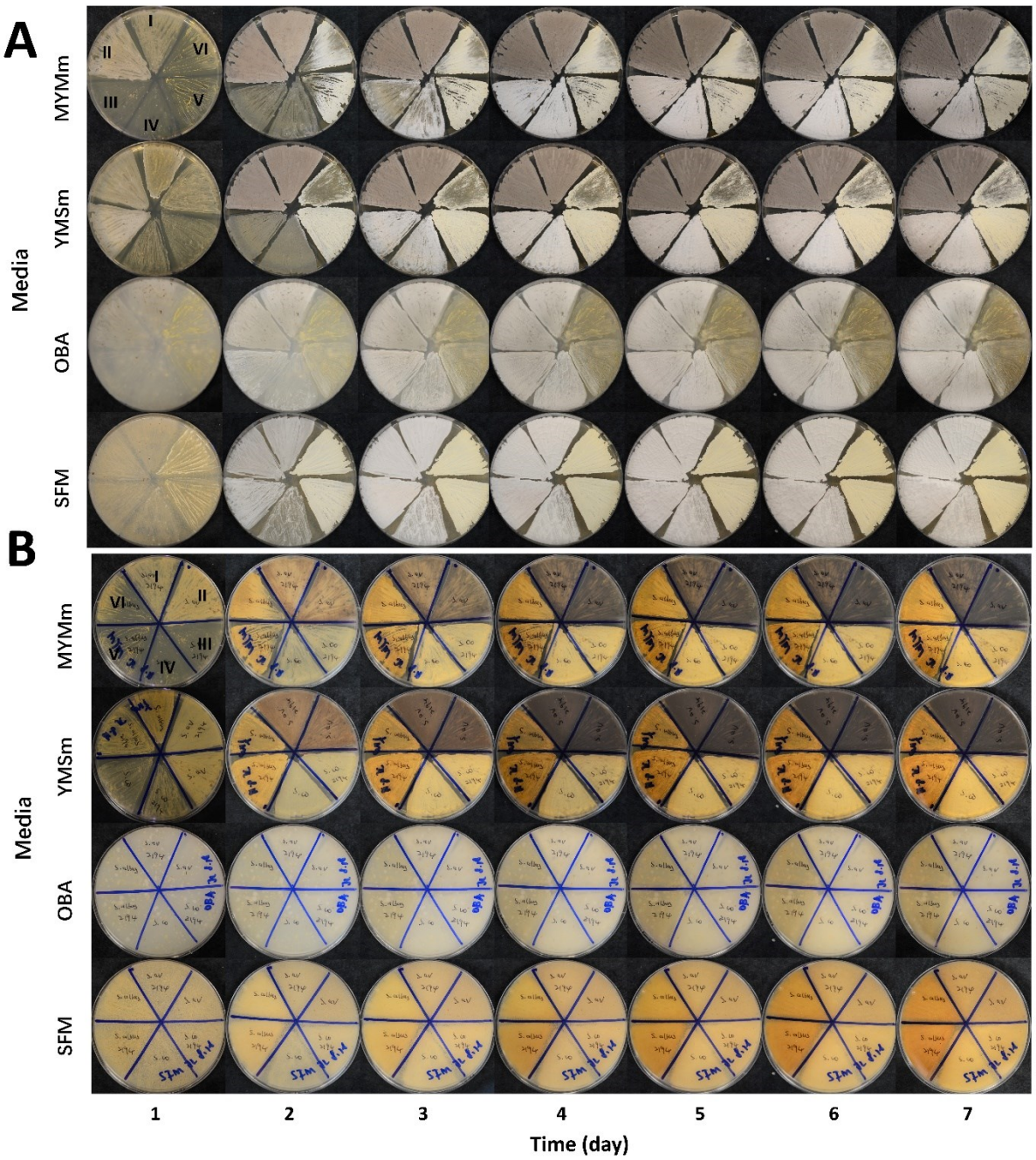


Figure 3.34 Growth of *Streptomyces* engineered strains and their control strains on MYMm,YMSm, OBA and SFM agar. I, *S. avermitilis* SUK17/cosmid 2194; II, *S. avermitilis* SUK17; III, *S. coelicolor* M1154/cosmid 2194; IV, *S. coelicolor* M1154; V, *S. albus* J1074/cosmid 2194; VI, *S. albus* J1074. **A:** The front side of the plates. **B:** The backside of the plates.

3.4 Metabolomics analysis of WT *S. scabiei* and the Δ*SCAB12101* mutant

Metabolomic analysis can identify and quantify all the metabolites that are produced in a cellular system. To investigate the metabolite potential of *S. scabiei,* and to potentially identify the product(s) of the phenazine-like BGC, a comparative metabolomics analysis of *S. scabiei* 87-22 and the Δ*SCAB12101* mutant was conducted. The Δ*SCAB12101* mutant was chosen because the gene expression analysis showed that deletion of *SCAB12101* not only induced the expression of the core phenazine biosynthetic genes*,* but also influences the expression of other genes within the phenazine-like BGC (Figure 3.27).

3.4.1 Metabolomics analysis of WT *S. scabiei*

To begin, a detailed analysis of the metabolites that can be produced by WT *S. scabiei* 87-22 was conducted. The bacterium was cultured on three different agar media, YMSm, MYMm and OBA, which have been reported to support the production of specialized metabolites (Ikeda et al. 1987; Johnson et al. 2007; Bignell et al. 2010b; Kitani et al. 2011; Komatsu et al. 2013; Hegemann et al. 2015; Fyans et al. 2016; Daniel-Ivad et al. 2017; Gehrke et al. 2019). The plate cultures were extracted with ethyl acetate, and the recovered metabolites were subjected to untargeted LC-MS^2 in both positive and negative ionization mode in order to detect as many compounds as possible. The resulting spectral data were analyzed using the Feature-Based Molecular Networking (FBMN) workflow within the GNPS platform to generate molecular networks along with quantitative results for statistical analysis within the networks (Nothias et al. 2020). In addition, the recently described BioDendro workflow was employed, which enables hierarchical clustering of MS^2 spectra and presents the results as a tree (Rawlinson et al. 2020). After molecular networking and background and media subtraction, a total of 6,260 metabolites were detected (Figure 3.35A and B). The majority of these molecules were detected under all three culturing conditions, though

distinct sets of metabolites were also found in each extract, with the OBA extract containing the highest number of unique metabolites (Figure 3.35B). The 15 most intense ions (by peak area) detected in the extracts (Figure 3.35C; Table 3.6) were annotated by performing a GNPS library search (Wang et al. 2016) or by using NAP (da Silva et al. 2018), SIRIUS (Böcker et al. 2009), MetWork (Beauxis and Genta-Jouve 2019) or the CFM-ID 3.0 web server (Allen et al. 2014). Of these compounds, CFA-Ile, pyochelin, concanamycin A and thaxtomin A are known metabolites produced by *S. scabiei* (Li et al. 2019b), while the other compounds have not been previously reported to be produced by this organism.

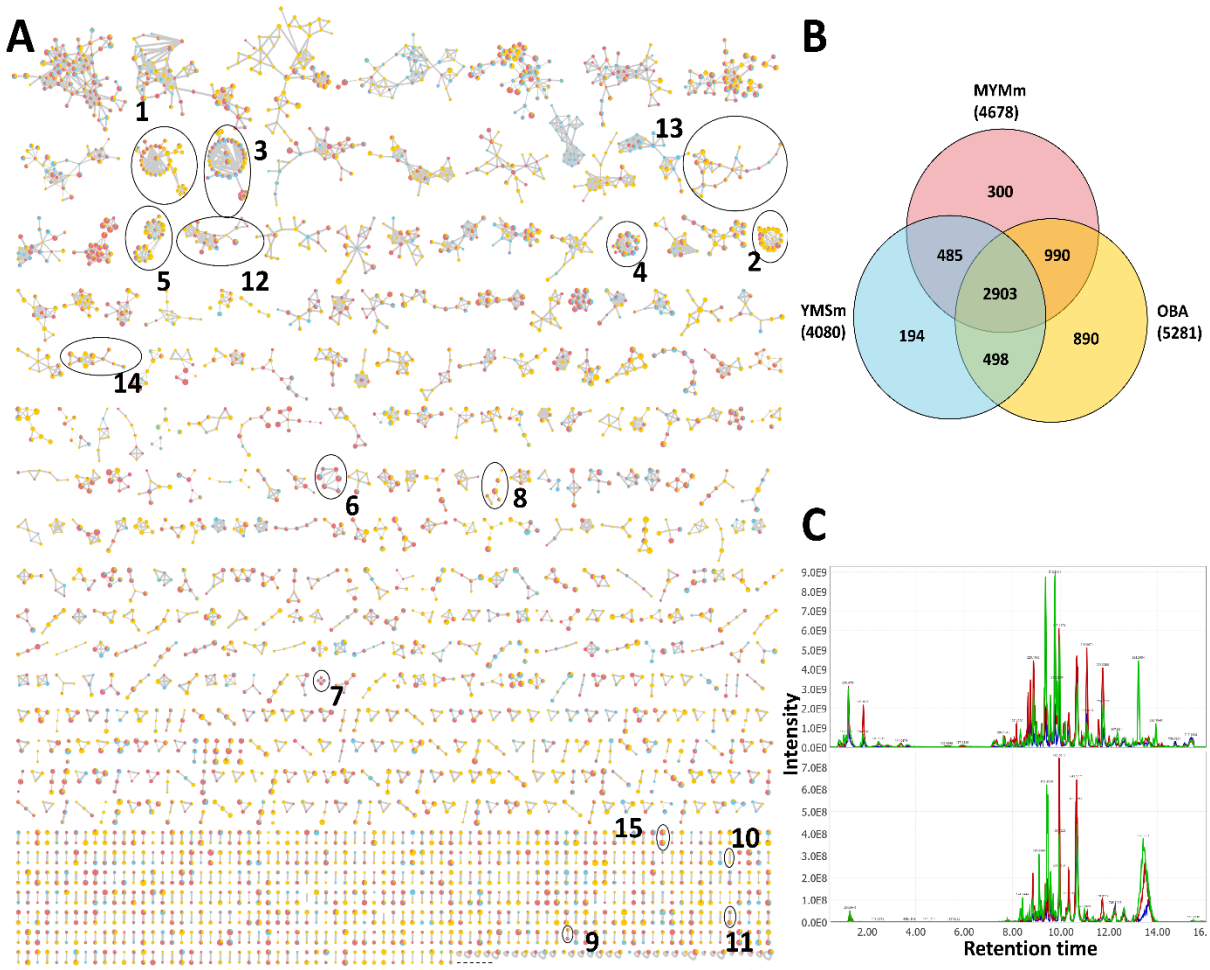


Figure 3.35 Metabolomics analysis of *S. scabiei* 87-22. **A:** FBMN of *S. scabiei* metabolites extracted from MYMm, YMSm and OBA and analyzed by untargeted LC-MS^2 in both positive and negative ionization modes. Each node represents one fragmentation spectrum from a detected compound, and node size represents the summed intensity (peak area) of the ion from all samples. Edge thickness indicates the relative similarity of MS^2 data between nodes. The pie charts indicate the relative abundance of each compound in the different extracts: MYM-red, YMS-blue, OBA-yellow. The networks containing annotated compounds circled in thick black: 1, thaxtomin A in negative ionization mode; 2, thaxtomin A in positive ionization mode; 3, CFA-Ile in negative ionization mode; 4, CFA-Ile in positive ionization mode; 5, concanamycin A in positive ionization mode; 6, bottromycin A2 in positive ionization mode; 7, bottromycin A2 in negative ionization mode; 8, deferrioxamine E in positive and negative ionization mode; 9, pyochelin in positive ionization mode; 10, indole acetic acid in positive ionization mode; 11, ectoine in positive ionization mode; 12, cyclo(L-Val-L-Pro) in positive ionization mode; 13, 211A decahydroquinoline cis and mairine B in positive ionization mode; 14, andrachcinidine in positive ionization mode; 15, aerugine in positive ionization mode. The dots at the bottom of the figure indicate that other networks were detected but are not shown. **B:** Venn diagram displaying node counts according to distribution among the *S. scabiei* culture extracts: MYM-red, YMS-blue,

OBA-yellow. **C:** Base peak chromatogram from LC-MS^2 analysis of *S. scabiei* culture extracts in positive (upper) and negative (lower) mode: MYM-red: YMS-blue, OBA-green.

Table 3.6 The 15 most intense ions (by peak area) detected in the *S. scabiei* 87-22 culture extracts.

m/z **$[M+H]^+$**	***m/z*** **$[M-H]^-$**	**Retention Time (min)**	**Formula†**	**Compound Name‡**	**Annotated Using**
212.2006	210.1860	9.7489	$C_{13}H_{25}NO$	(2*S*,4a*S*,5*S*,6*R*,8a*R*)-5-methyl-2-propyl-decahydroquinolin-6-ol (211A Decahydroquinoline cis)	NAP
322.2004	320.1865	15.6427	$C_{18}H_{27}NO_4$	CFA-Ile	CFM-ID
224.0365	222.0219	11.72	$C_{10}H_9NO_3S$	2-(1,3-benzoxazol-2-ylthio)propanoic acid	NAP
197.1153	195.1007	9.9636	$C_{10}H_{16}N_2O_2$	Cyclo(L-Val-L-Pro)	GNPS library
228.1957	226.1811	8.8787	$C_{13}H_{25}NO_2$	1-[(2*S*,6*R*)-6-[(2*S*)-2-hydroxypentyl]piperidin-2-yl]propan-2-one (andrachcinidine)	NAP
228.196	226.1814	10.079	$C_{13}H_{25}NO_2$	1-[(2*S*,6*R*)-6-[(2*S*)-2-hydroxypentyl]piperidin-2-yl]propan-2-one (andrachcinidine)	NAP
325.0676	323.0530	10.9854	$C_{14}H_{16}N_2O_3S_2$	Pyochelin	GNPS library
308.1888	306.1706	10.3184	$C_{18}H_{29}NO_3$	Methyl-substituted CFA-Ile	CFM-ID
184.1675	182.1529	9.2423	$C_{11}H_{21}NO$	(4*R*,4a*R*,7*R*,7a*R*)-2,4,7-trimethyl-hexahydro-1H-cyclopenta[c]pyridin-7a-ol (mairine B)	NAP
210.0596	208.0450	9.8113	$C_{10}H_{11}NO_2S$	2-[(4*R*)-4-(hydroxymethyl)-4,5-dihydro-1,3-thiazol-2-yl]phenol (aerugine)	NAP
421.1512	419.1366	9.6884	$C_{22}H_{20}N_4O_5$	Thaxtomin A derivative	MetWork
213.2039	211.1893	9.8261	*	*	
699.4439	697.4293	13.6265	$C_{40}H_{60}NO_9$	Steroids and steroid derivatives	NAP
888.5071	**	12.6819	$C_{46}H_{75}NO_{14}$	Concanamycin A	GNPS library
439.1598	437.1452	9.4529	$C_{22}H_{22}N_4O_6$	Thaxtomin A	GNPS library

† Predicted using SIRIUS

‡ Annotated using the GNPS library, NAP, MetWork, or CFM-ID 3.0 web server

* No annotation/structure prediction

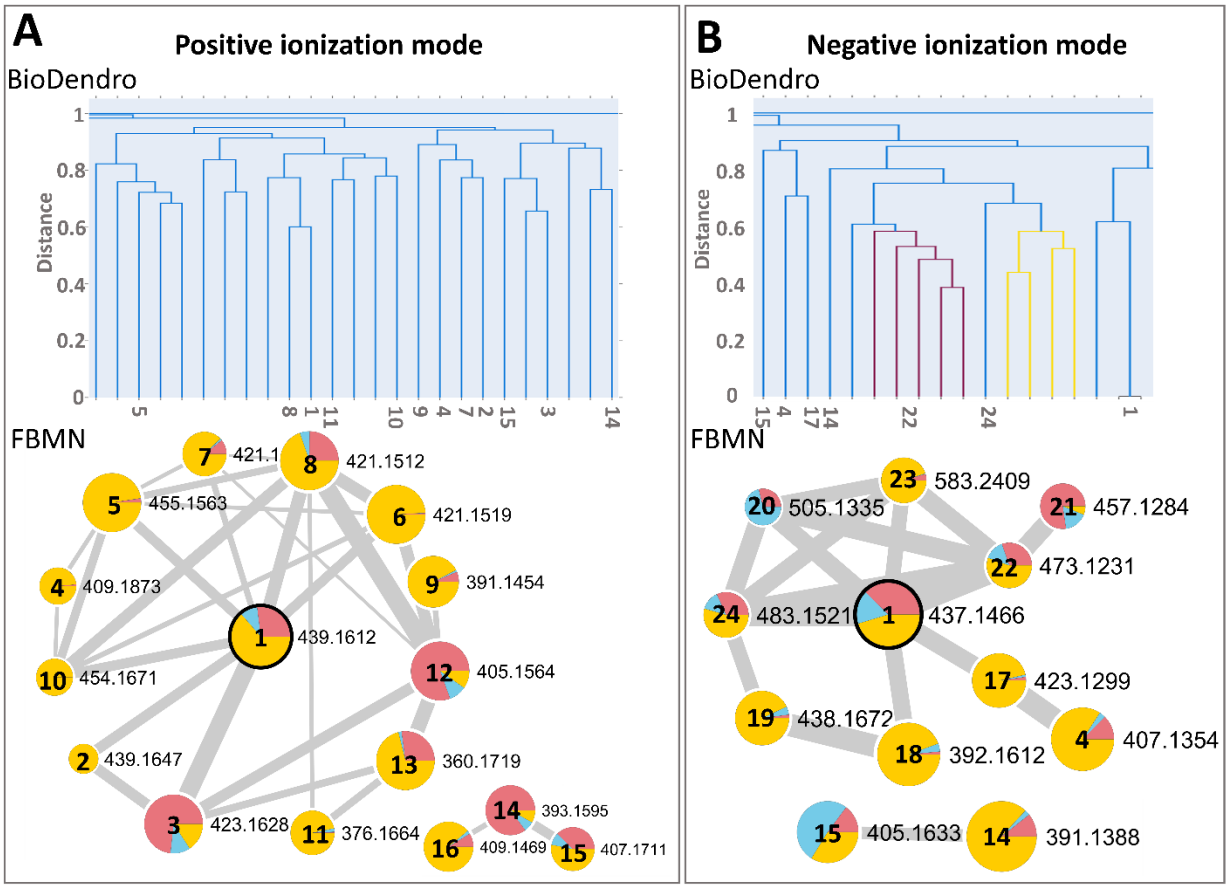
** Not detected

The following sections describe the known metabolites detected in the WT *S. scabiei* 87-22 metabolome along with molecules that were not previously known to be produced by this organism.

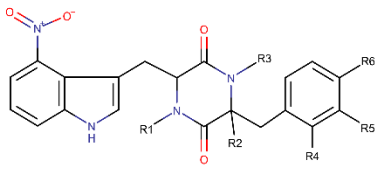
3.4.1.1 Thaxtomins

Two thaxtomin networks were annotated from the metabolome of WT *S. scabiei*, with one in positive ionization mode (15 compounds) and the other in negative ionization mode (9 compounds) (Figure 3.35A; Figure 3.36A and B; Table S1). The results suggest that OBA is the best medium for supporting the production of most compounds in the networks, though production was detected in the other two media in some instances. As expected, thaxtomin A (**1**; *m/z* 439.1612, $[M+H]^+$; *m/z* 437.1466, $[M-H]^-$) was the predominant thaxtomin analogue and was annotated using the GNPS library (Figure 3.36). The annotation was confirmed by comparing the key MS^2 fragments with those reported for thaxtomin A (Winn et al. 2018) (Table S1). Based on the precursor ion mass, MS^2 fragments and retention time (Table S1), **2** was predicted to be an isomer of thaxtomin A. Two thaxtomin A isomers, *p*-isomer and *o*-isomer, have been reported to be produced in minor amounts by *S. scabiei* (King et al. 2001), and by using CFM-ID 3.0, it was predicted that the *p*-isomer is the best candidate match for **2**. Compound **3** (*m/z* 423.1628, $[M+H]^+$) was predicted to be thaxtomin B based on its precursor ion mass and MS^2 fragmentation pattern (Table S1) (Jiang et al. 2018b). Based on the mass, comparative chromatographic data and CFM-ID prediction, **4** (*m/z* 409.1873, $[M+H]^+$) was predicted to be a monooxygenated analogue of thaxtomin C, **5** (*m/z* 455.1563, $[M+H]^+$) was predicted to be a monooxygenated derivative of thaxtomin A, and **17** (*m/z* 423.1299, $[M-H]^-$) was predicted to be a 15-de-*N*-methyl analogue of thaxtomin A. The three thaxtomin analogues mentioned above have all been previously reported

to be isolated from *S. scabiei* (King and Lawrence 1996; King and Calhoun 2009). Using MetWork, **6** (*m/z* 421.1519, $[M+H]^+$), **7** (*m/z* 421.1477, $[M+H]^+$), and **8** (*m/z* 421.1512, $[M+H]^+$), were putatively identified as C-14 dehydrated analogues of thaxtomin A, of which **8** was among the 15 most intense ions detected in the culture extracts in the current study (Table 3.6). Dehydrated analogues of thaxtomin A have only been reported as biotransformation products of *Aspergillus niger* (Lazarovits et al. 2004) and have not been detected in *S. scabiei* before. Notably, thaxtomin C (**14**; *m/z* 393.1595, $[M+H]^+$; *m/z* 391.1388, $[M-H]^-$) and thaxtomin D (**15**; *m/z* 407.1711, $[M+H]^+$; *m/z* 405.1633, $[M-H]^-$) were not readily annotated using molecular networking, whereas they could be annotated in the hierarchical MS^2 spectra trees generated by BioDendro (Figure 3.36A and B). When crosschecked with the FBMN network, thaxtomin C and D were found to form a separate network from the main thaxtomin network in both the positive and negative ionization mode (Figure 3.36A and B). On the other hand, other related ions from the network were not clustered in the BioDendro tree, which illustrates how different similarity and visualization methods are complementary to explore spectral similarity in metabolomics experiments. Thaxtomin C is the major form of thaxtomin produced by the sweet potato pathogen *Streptomyces ipomoeae* (King et al. 1994), and trace amounts of thaxtomin C and thaxtomin D have also been reported from *S. scabiei* (King and Calhoun 2009; Jiang et al. 2018a), but their production has never been reported from *S. scabiei* cultured on MYMm and YMSm. Other compounds in these networks have mass fragmentation pattern similar to thaxtomins (Table S1), but the predicted structures based on NAP, SIRIUS, MetWork, or CFM-ID are unrelated to the thaxtomins and thus might be new derivatives.



C



Compound	R1	R2	R3	R4	R5	R6
1 (Thaxtomin A)	Me	OH	Me	H	OH	H
2 (Thaxtomin A *p*-isomer)	Me	OH	Me	H	H	OH
3 (Thaxtomin B)	Me	OH	Me	H	H	H
4 (Hydroxy thaxtomin C)	Me	OH	H	H	H	H
5 (Hydroxy thaxtomin A)	Me	OH	Me	H	OH	OH
14 (Thaxtomin C)	Me	H	H	H	H	H
15 (Thaxtomin D)	Me	H	Me	H	H	H
17 (15-de-*N*-methyl-thaxtomin A)	Me	OH	H	H	OH	H

D

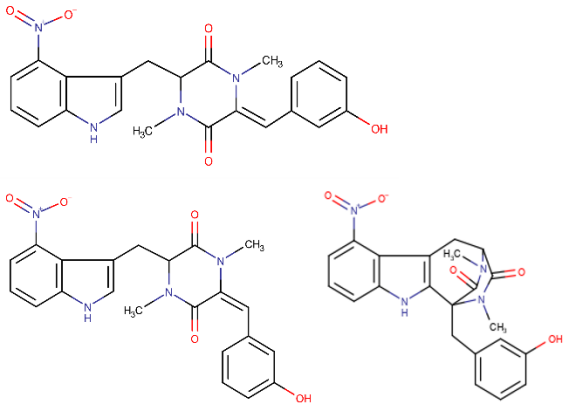


Figure 3.36 Visualization of fragmentation spectra for thaxtomin metabolites from *S. scabiei* 87-22. Similarity between MS^2 spectra was explored using BioDendro (top) and FBMN (bottom) in positive ionization mode (**A**) and in negative ionization mode (**B**). In FBMN, edges are created if the cosine score is > 0.7 and there are at least four matched fragment ions. In the BioDendro trees, all links connecting nodes with Jaccard distances ≥ 0.6 are indicated in blue, and those with distances < 0.6 are indicated in other colours. The numbers indicated in the trees correspond to the node numbers shown in the FBMN networks. Features in the FBMN networks are as described in

the legend for Figure 2. Nodes that have MS^2 matches to the GNPS library are outlined in black. (**C**) The chemical structures of the metabolites annotated in the thaxtomin networks. (**D**) The chemical structures of the dehydrated analogues of thaxtomin A that have been described before (Lazarovits et al. 2004).

3.4.1.2 Coronafacoyl Phytotoxins

The metabolomics analysis of WT *S. scabiei* using FBMN allowed the annotation of two putative coronafacoyl phytotoxin networks, one in positive ionization mode and one in negative ionization mode (Figure 3.37A and B), and they contained 10 and 14 compounds, respectively (Table S2). Most of the compounds from both networks could be detected in all three of the culture extracts, although they were generally more abundant in the MYMm and YMSm.

Coronafacoyl phytotoxins possess a readily recognized MS^2 fragmentation pattern, which includes peaks at *m/z* 191 and 163 (Mitchell 1984). The mass spectra for all of the compounds in the two networks displayed the characteristic fragments *m/z* 191 and 163 except for **27**, **29**, **42** and **43**, which had prominent fragments corresponding to *m/z* 177 and *m/z* 149 (Figure 3.38; Table S2). Based on the precursor ion mass, MS^2 fragments and CFM-ID prediction, **25** was annotated as CFA-Ile (*m/z* 322.2004, $[M+H]^+$; *m/z* 320.1863, $[M-H]^-$) (Figure 3.37C; Figure 3.38) and was the main coronafacoyl analogue as expected (Bignell et al. 2018). Compound **35** (*m/z* 320.1865, $[M-H]^-$) was annotated as an isomer of CFA-Ile on the basis of the precursor ion mass, retention time and mass fragmentation pattern (Table S2), and the only other reported isomer for CFA-Ile is *N*-coronafacoyl-L-*allo*-isoleucine (CFA-aIle) produced by *Pseudomonas savastanoi* (Bignell et al. 2018). The mass fragmentation patterns of **27**, **29**, **42** and **43** suggested that these metabolites differ from the other coronafacoyl derivatives in that they contain a methyl group at position C-7 of the bicyclic hydrindane ring instead of an ethyl group (Figure 3.38). Based on the mass fragmentation pattern and CFM-ID prediction, **27** was annotated as the methyl-substituted derivative of CFA-Ile

(Figure 3.37C; Figure 3.38) and was among the most intense ions detected in the extracts (Table S2). Compounds **28** (*m/z* 308.1856, $[M+H]^+$) and **29** (*m/z* 308.1856, $[M+H]^+$) both have the same mass as **27**, but their fragmentation patterns matched that of CFA-Ile. As *N*-coronafacoyl-valine (CFA-Val) is a known coronafacoyl phytotoxin (Bignell et al. 2018), and it was previously shown that *S. scabiei* likely produces this compound (Fyans et al. 2015), compounds **28** and **29** were annotated as isomers of CFA-Val. Using MetWork, **37** and **38** were predicted to be decarboxylated derivatives of CFA-Ile and CFA-Val, respectively. Using BioDendro, four additional coronafacoyl-related compounds were detected, **46** (*m/z* 322.1978, $[M+H]^+$), **47** (*m/z* 308.1895, $[M+H]^+$), **48** (*m/z* 645.107, $[M-H]^-$) and **49** (*m/z* 322.2010, $[M-H]^-$), with mass fragmentation patterns similar to that of CFA-Ile. Based on their precursor ion mass and fragmentation pattern, **46** and **47** were annotated as isomers of CFA-Ile and CFA-Val, respectively. The remaining compounds in these two networks have mass fragmentation patterns similar to coronafacoyl phytotoxins, but the predicted structures based on NAP, SIRIUS, MetWork, or CFM-ID are completely unrelated to the coronafacoyl phytotoxins, suggesting that they may be novel coronafacoyl analogues.

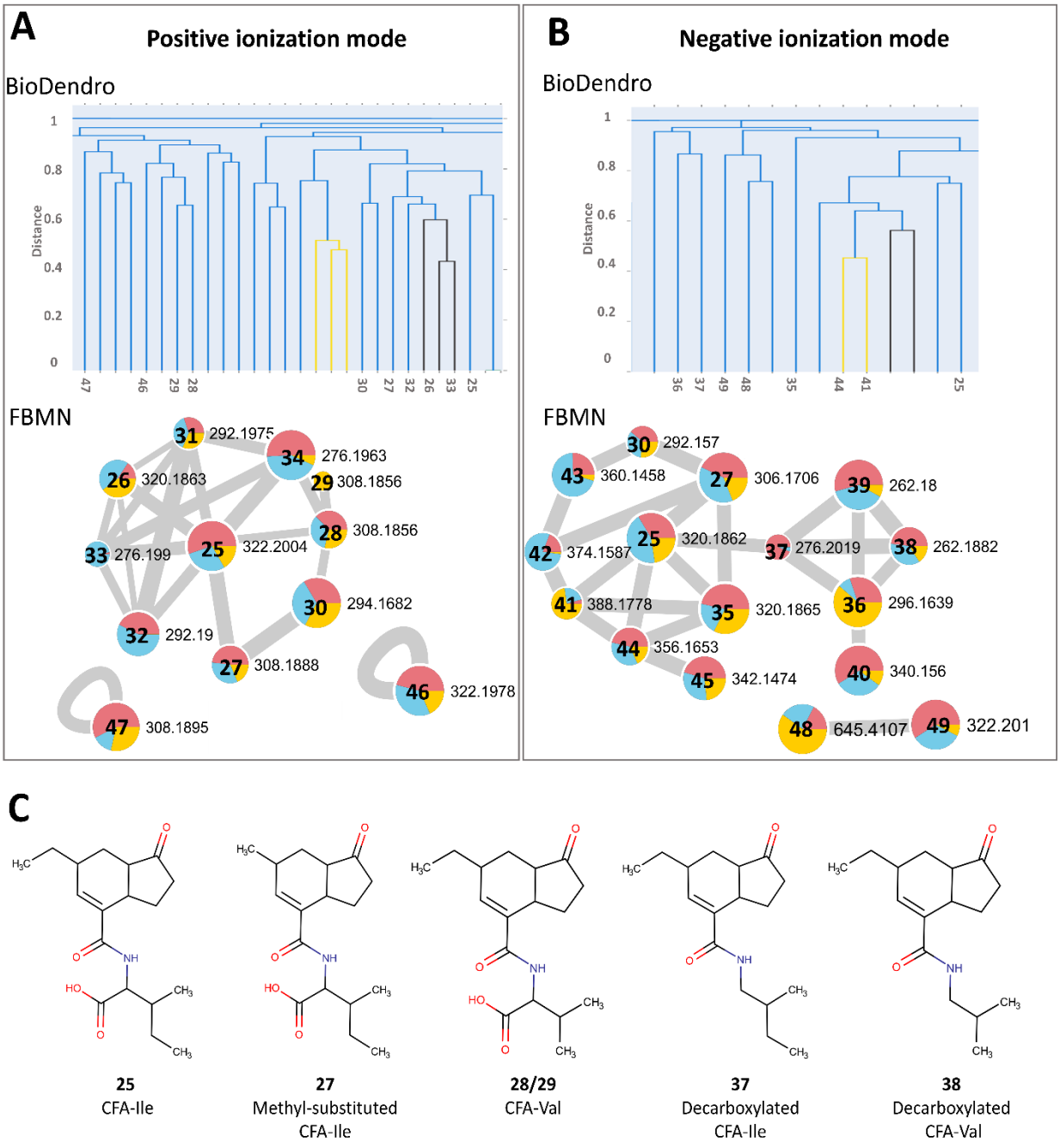


Figure 3.37 Visualization of fragmentation spectra for the putative coronafacoyl phytotoxins from *S. scabiei* 87-22. Similarity between MS^2 spectra was explored using BioDendro (top) and FBMN (bottom) in positive ionization mode (**A**) and in negative ionization mode (**B**). The features of the BioDendro trees and the FBMN networks are as described in the legends for Figures 3.35 and 3.36. (**C**) The chemical structures of metabolites annotated in the coronafacoyl phytotoxin networks.

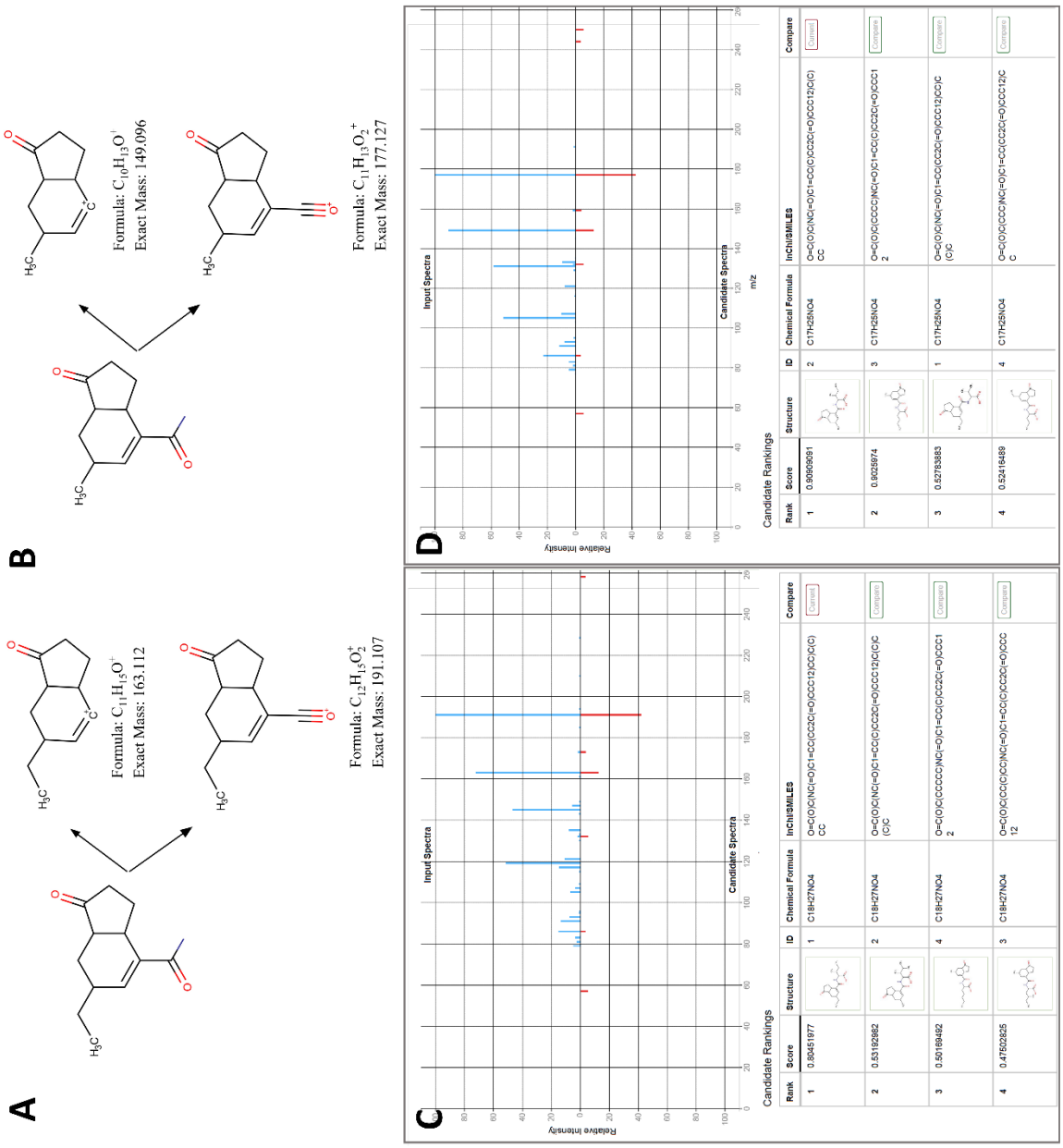
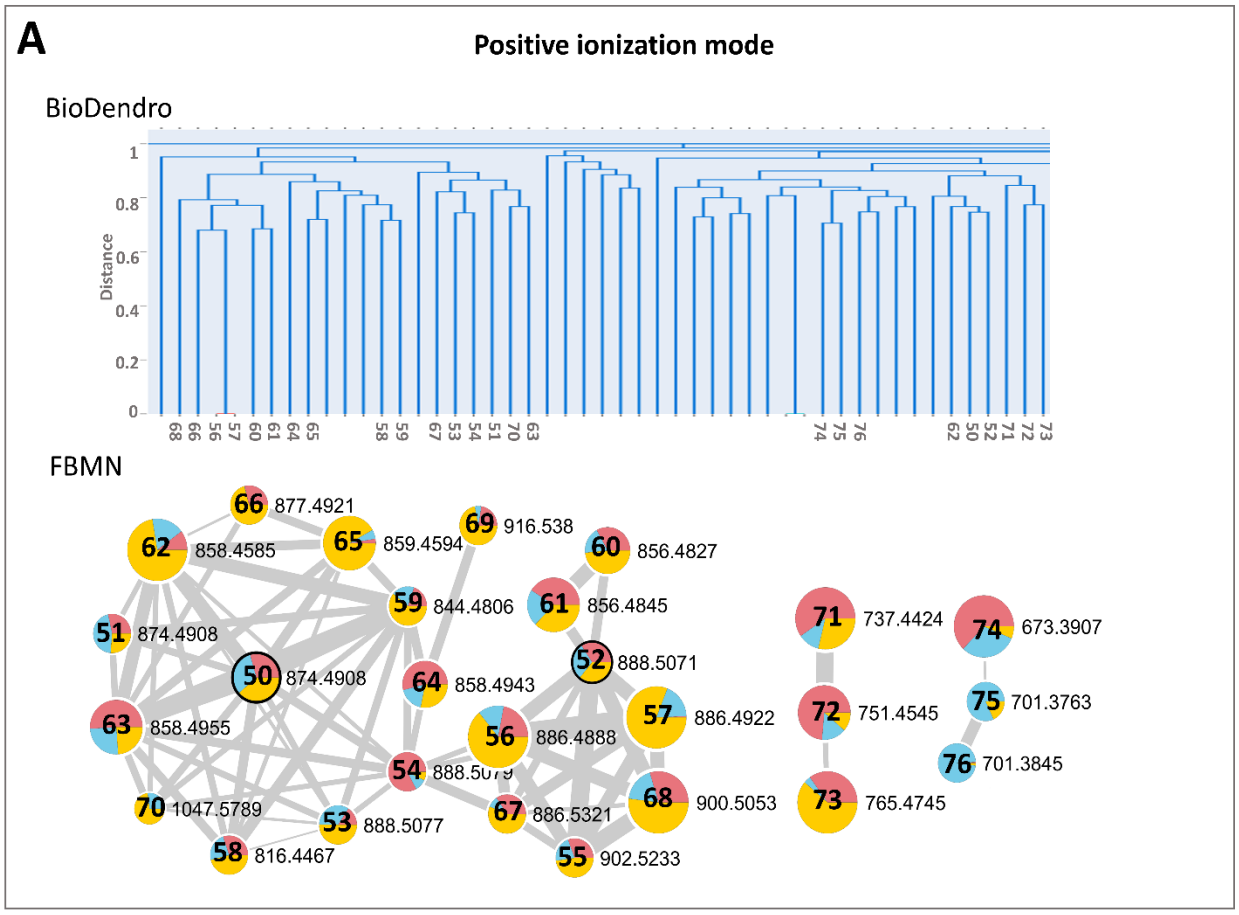


Figure 3.38 Mass fragmentation pattern of coronafacoyl phytotoxins. (**A**) The fragmentation products shown represent the main products generated from the coronafacic acid polyketide moiety that is conserved in all coronafacoyl phytotoxins. (**B**) Predicted mass fragmentation pattern of the methyl-substituted coronafacoyl phytotoxin derivatives. (**C**) CFM-ID prediction result for compound **25**. The input spectrum shown is for **25**, and the predicted candidate spectrum is for CFA-Ile, which was ranked as the best candidate for **25**. (**D**) CFM-ID prediction result for compound **27**. The input spectrum shown is for **27,** and the predicted candidate spectrum is for the methyl-substituted analogue of CFA-Ile, which was ranked as the best candidate for **27**.

3.4.1.3 Concanamycins

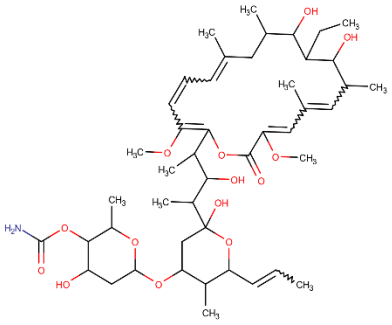
The production of the phytotoxic metabolites concanamycin A and B by *S. scabiei* in oat-based media has been reported before (Natsume et al. 1996, 1998, 2001; Fyans et al. 2016), whereas production in MYMm and YMSm has not been previously investigated. A concanamycin MS^2 network consisting of 21 compounds was annotated in the metabolome of *S. scabiei* 87-22 in positive ionization mode (Figure 3.39; Table S3), whereas no concanamycin derivatives were detected in negative ionization mode. OBA was the best medium for supporting production of most of the concanamycins, though **50** and **52** were evenly distributed across all three media tested, where **54**, **63**, and **64** were found to be most abundant in the MYMm extract.

Compound **50** (*m/z* 874.4908, $[M+Na]^+$) was annotated as concanamycin B by spectral matching with the GNPS library (Figure 3.39B), and **51** was manually annotated to be an isomer of concanamycin B based on the precursor ion mass, retention time and MS^2 fragments (Table S3). Compound **52** (*m/z* 888.5071, $[M+Na]^+$) was annotated as concanamycin A using the GNPS library and inspection of its MS^2 fragmentation pattern (Haydock et al. 2005), and **53** and **54** were manually annotated as isomers of concanamycin A based on the precursor ion mass, retention time and MS^2 fragments (Table S3). The only previously reported isomer for concanamycin A was O-methyl-concanamycin B (Kinashi et al. 1984), which could be either **53** or **54** (Figure 3.39B). Using MetWork, **55** (*m/z* 902.5233, $[M+Na]^+$) was predicted to be O-methyl-concanamycin A, and **56** (*m/z* 886.4888, $[M+Na]^+$) and **57** (*m/z* 886.4922, $[M+Na]^+$) were predicted to be oxidized analogues of concanamycin A based on the expected molecular formula differences (Figure 3.39B). Seven additional previously unreported concanamycin-related compounds (**71** to **76**) were detected using BioDendro (Figure 3.39A; Table S3).



B

Compound	R1	R2
50 (Concanamycin B)	CH_3	H
52 (Concanamycin A)	CH_2CH_3	H
53/54 (O-Methyl-concanamycin B)	CH_3	CH_3
55 (O-Methyl-concanamycin A)	CH_2CH_3	CH_3



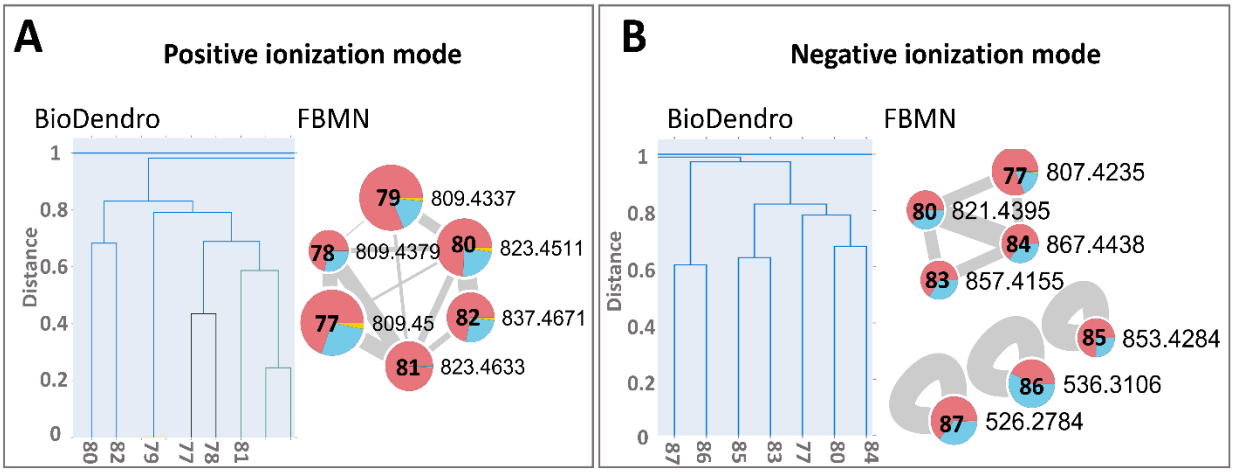
56/57
Oxidized concanamycin A

Figure 3.39 Visualization of fragmentation spectra for the concanamycin metabolites from *S. scabiei* 87-22. (**A**) Similarity between MS^2 spectra was explored using BioDendro (top) and FBMN (bottom) in positive ionization mode. The features of the BioDendro tree and the FBMN network are as described in the legends for Figures 3.35 and 3.36. (**B**) The chemical structure of metabolites annotated in the concanamycin network.

3.4.1.4 Bottromycins

Two putative bottromycin networks were annotated in the WT *S. scabiei* metabolome, one each in positive and negative ionization mode and containing six and four molecules, respectively. The masses and fragmentation patterns of the metabolites are in good agreement with those of reported bottromycin molecules (Crone et al. 2016) (Figure 3.40A and B; Table S4). MYMm was found to be the best medium for bottromycin production followed by YMSm, whereas very little production was detected when *S. scabiei* was cultured on OBA.

Based on their precursor ion masses, MS^2 fragmentation patterns and retention times, **77** (*m/z* 809.4500, $[M+H]^+$; *m/z* 807.4235, $[M-H]^-$) was annotated as bottromycin B2, and **78** was annotated as bottromycin B1 (Figure 3.40A and B; Table S4). In addition, using NAP, it was predicted that **80** (*m/z* 823.4511, $[M+H]^+$; *m/z* 821.4395, $[M-H]^-$) is bottromycin A2; **79** (*m/z* 809.4377, $[M+H]^+$) is bottromycin A2 acid based on shared MS^2 fragments (141.1, 169.1, 268.2, 301.1, 363.2, 476.3, 639.4); **81** is an isomer of bottromycin A2 and **82** (*m/z* 837.4670, $[M+H]^+$) is bottromycin C2. Analysis of the hierarchical MS^2 tree created by BioDendro revealed the presence of three additional putative bottromycin-related compounds: **85** (*m/z* 853.4284, $[M+H]^+$), **86** (*m/z* 526.2784, $[M+H]^+$) and **87** (*m/z* 536.3106, $[M+H]^+$), where the mass and fragmentation pattern of **85** matched that of carboxylated O-desmethyl bottromycins A2 (Table S4) (Crone et al. 2016). Other compounds in these networks have mass fragmentation pattern similar to bottromycins, but the predicted structures based on NAP, SIRIUS, MetWork, or CFM-ID are completely unrelated to the bottromycins.



C

Compound	R1	R2	R3	R4
77 (Bottromycin B2)	H	H	Me	Me
79 (Bottromycin A2 acid)	Me	H	Me	OH
80 (Bottromycin A2)	Me	H	Me	Me
82 (Bottromycin C2)	Me	Me	Me	Me

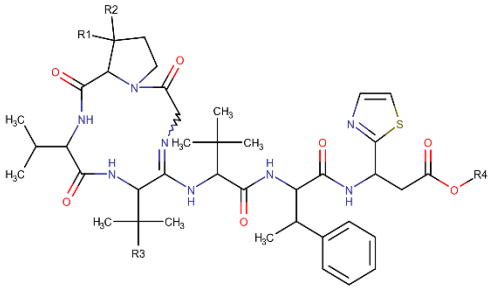


Figure 3.40 Visualization of fragmentation spectra for the putative bottromycins from *S. scabiei* 87-22. Similarity between MS^2 spectra was explored using BioDendro (left) and FBMN (right) in positive ionization mode (**A**) and in negative ionization mode (**B**). The features of the BioDendro trees and the FBMN networks are as described in the legends for Figures 3.35 and 3.36. (**C**) The chemical structures of metabolites annotated in the bottromycin networks.

3.4.1.5 Siderophores

The metabolomics analysis conducted here allowed for the annotation of one pyochelin network in positive ionization mode and two desferrioxamine networks in positive and negative ionization modes (Figure 3.41A; Table S5). However, no scabichelin metabolites were detected in any of the samples in either positive or negative ionization mode. MYMm and OBA supported desferrioxamine and pyochelin production, while only low levels of the metabolites were detected in YMSm extracts. Spectral matching with the GNPS library enabled the annotation of **88** (*m/z*

601.3532, $[M+H]^+$; *m/z* 599.3431, $[M-H]^-$), **92** (*m/z* 325.0607, $[M+H]^+$) and **94** (*m/z* 583.3459, $[M-H]^-$) as desferrioxamine E, pyochelin and dehydroxynocardamine (a derivative of desferrioxamine), respectively, which was further strengthened by comparison with published spectra (Seipke et al. 2011; Senges et al. 2018). The use of NAP allowed the putative annotation of **89** (*m/z* 401.2392, $[M+H]^+$) as bisucaberin (da Silva et al. 2018), which is part of a family of dihydroxamate siderophores originally isolated from the marine bacterium *Alteromonas haloplanktis* (Fujita et al. 2012; Senges et al. 2018).

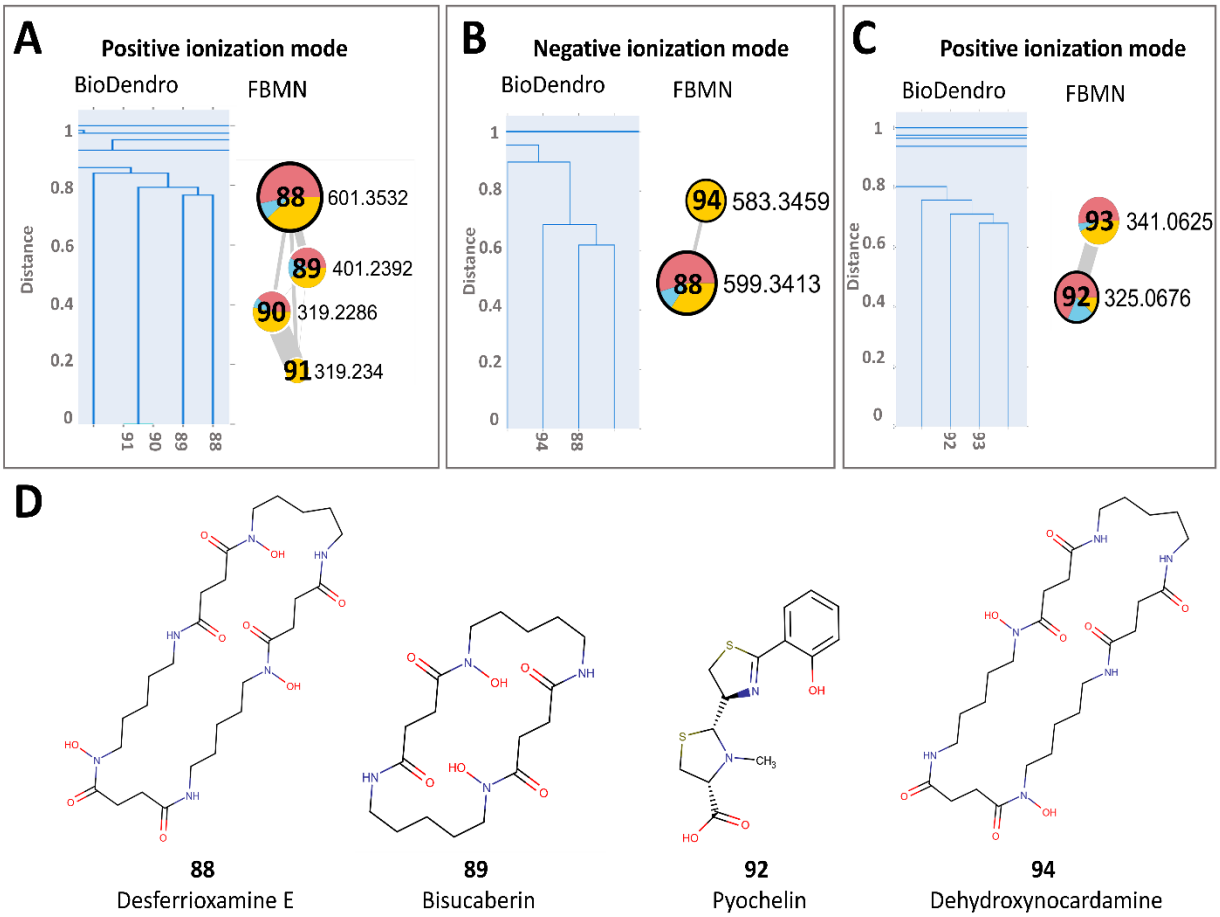


Figure 3.41 Visualization of the similarity of fragmentation spectra for the siderophore metabolites from *S. scabiei* 87-22. Spectral dendrogram and network of desferrioxamine E in positive ionization mode (**A**) and in negative ionization mode (**B**), and the dendrogram and network of pyochelin in positive ionization mode (**C**). The dendrograms were created by BioDendro and the networks by FBMN, and the features of each are as described in the legends for Figures 3.35 and 3.36. (**D**) The putative structures of metabolites annotated in the siderophore networks.

3.4.1.6 Other Compounds

In the current study, a compound corresponding to indole-3-acetic acid (IAA) (*m/z* 176.0800, $[M + H]^+$) (Figure 3.42) was detected in the WT *S. scabiei* OBA extract but not in the other extracts (Figure 3.35A). Also, ectoine (*m/z* 143.0700, $[M + H]^+$) (Figure 3.42) was detected in MYMm, YMSm and OBA extracts, with levels being highest in MYMm. Interestingly, several putative compounds not previously known to be produced by *S. scabiei* 87-22 were also identified from the list of the 15 most intense ions detected. These included cyclo(L-Val-L-Pro), aerugine, decahydroquinoline, andrachcinidine and mairine B (Figure 3.42, Table 3.6).

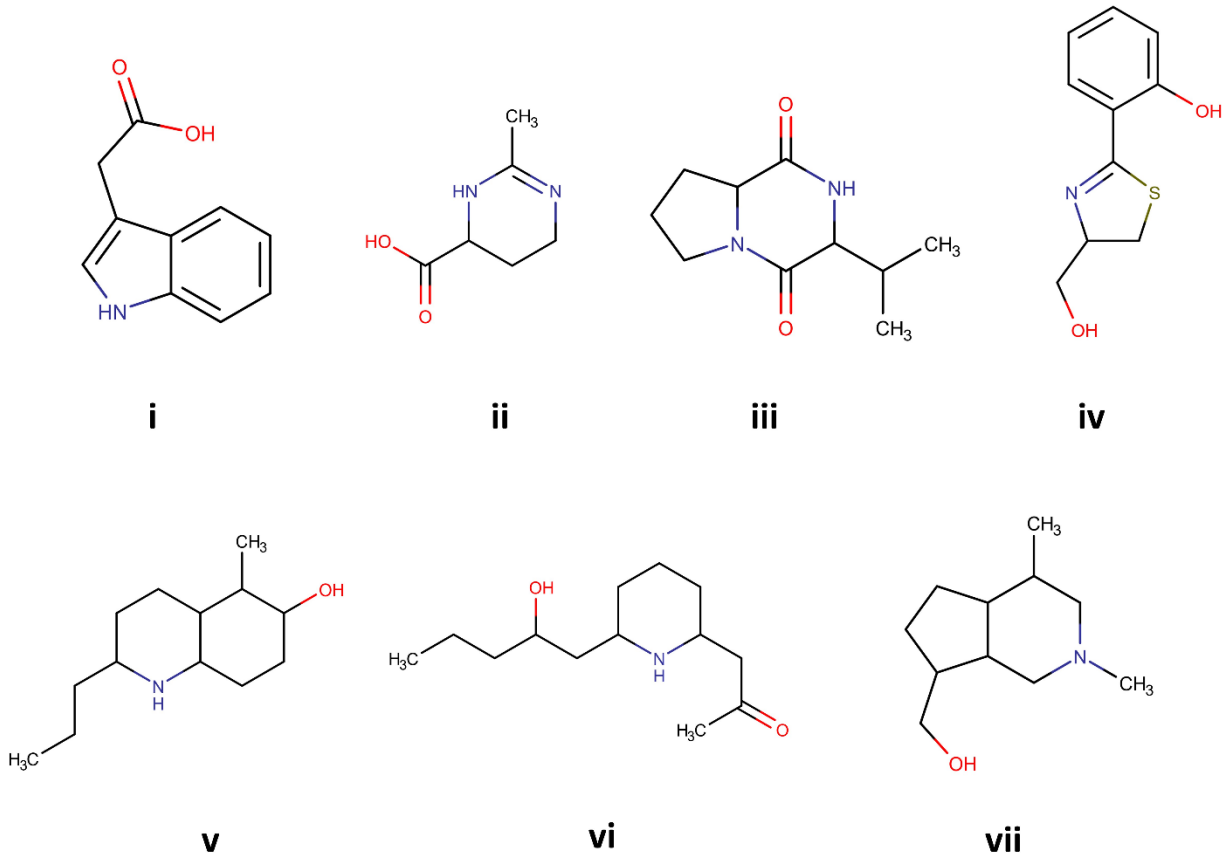


Figure 3.42 Chemical structures of select metabolites identified/predicted from the metabolome of *S. scabiei*. (i) Indole-3-acetic acid (IAA); (ii) Ectoine; (iii) Cyclo(L-Val-L-Pro); (iv) Aerugine; (v) 211A decahydroquinoline cis; (vi) Andrachcinidine; (vii) Mairine B.

3.4.2 Comparative metabolomics analysis of WT *S. scabiei* and the Δ*SCAB12101* mutant

To compare the metabolic profiles of WT *S. scabiei* 87-22 and the Δ*SCAB12101* mutant, the two strains were cultured on MYMm, YMSm and OBA, and the metabolites were extracted and subjected to LC-MS^2 as described in Section 3.4.1. The resulting MS^2 data were analyzed using the GNPS classical molecular networking. After molecular networking and background and media subtraction, a total of 2687 and 3023 metabolites were detected in positive ionization mode in *S. scabiei* 87-22 and Δ*SCAB12101*, respectively, (Figure 3.43; Table S6), while 1591 and 1883 metabolites were detected in negative ionization mode in 87-22 and Δ*SCAB12101*, respectively (Figure 3.43; Table S7).

Using the MolNetEnhancer workflow, chemical classification of molecular families could be retrieved. As shown in Figure 3.43, Table S6 and Table S7, 52.8% of the metabolites detected in positive ionization mode could be annotated and classified into 58 chemical classes, of which a higher number of metabolites were observed in the Δ*SCAB12101* extracts across 31 classes. In negative ionization mode, 61.5% of the metabolites could be annotated and classified into 55 chemical classes, of which a higher number of metabolites were observed in the Δ*SCAB12101* extracts across 26 classes. The predominant chemical classes detected were carboxylic acids and derivatives, prenol lipids, benzene and substituted derivatives, and organooxygen compounds. Other chemical classes included flavonoids, coumarins, indoles and diazanaphthalenes. In the class of diazanaphthalenes in positive ionization mode, five direct parents were found, including phenazines and derivatives, quinazolinamines, quinazolines, quinoxalines and stilbenes (Table 3.7). Moreover, another eight phenazines and derivatives were also found in the class of carboxylic acids and derivatives (Table 3.8).

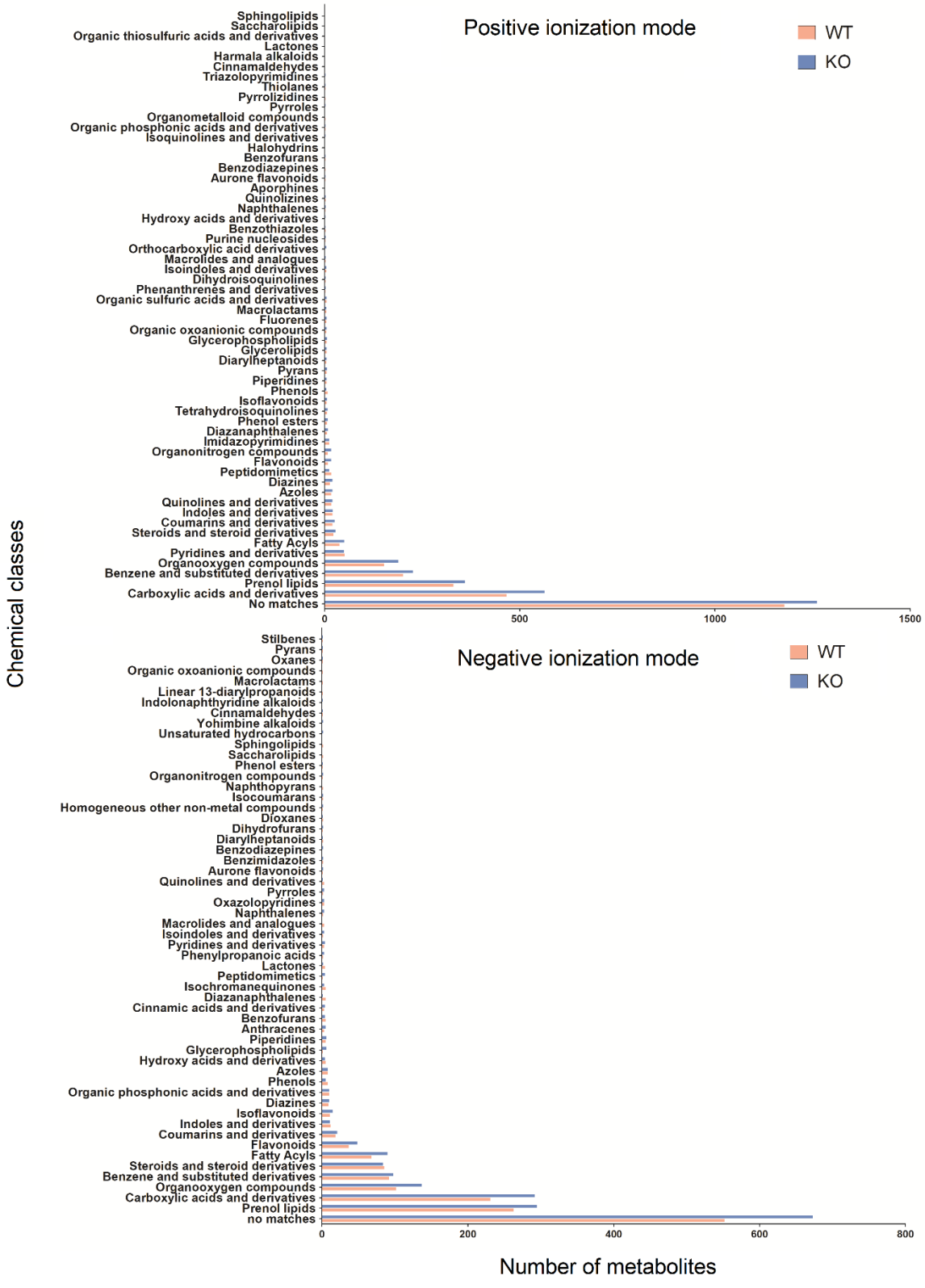


Figure 3.43 Chemical family analysis of metabolites detected in the *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO) extracts in positive and negative ionization mode.

Table 3.7 Diazanaphthalenes class analysis of metabolites detected in the *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO) extracts in positive ionization mode. Chemical classes highlighted in bold are those where the number of molecules detected in the KO extracts was greater than the number detected in the WT extracts.

Direct parent	**Only in WT**	**Only in KO**	**Both in WT/KO**	**WT**	**KO**	**Total**
Phenazines and derivatives	1	2	1	2	3	4
Quinazolinamines	0	2	0	0	2	2
Quinazolines	0	2	1	1	3	3
Quinoxalines	0	0	1	1	1	1
Stilbenes	2	0	0	2	0	2

Table 3.8 Carboxylic acids and derivatives class analysis of metabolites detected in the *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO) extracts in positive ionization mode. Chemical classes highlighted in bold are those where the number of molecules detected in the KO extracts was greater than the number detected in the WT extracts.

Direct parent	**Only in WT**	**Only in KO**	**Both in WT/KO**	**WT**	**KO**	**Total**
Phenazines and derivatives	1	0	2	3	2	3
Dipeptides	5	21	50	55	71	76
Alpha amino acids	7	11	30	37	41	48
Alpha amino acids and derivatives	8	16	22	30	38	46
Piperidines	3	11	31	34	42	45
N-acyl-alpha amino acids	8	14	13	21	27	35
Triterpenoids	3	12	15	18	27	30
Cysteine and derivatives	4	5	18	22	23	27
Coumarins and derivatives	2	9	14	16	23	25
Imidolactams	2	1	9	11	10	12
Medium-chain keto acids and derivatives	3	2	5	8	7	10
Pyrazines	0	4	6	6	10	10
Benzene and substituted derivatives	3	0	6	9	6	9
Delta amino acids and derivatives	0	8	1	1	9	9
Isothioureas	0	0	9	9	9	9
Macrolides and analogues	3	2	4	7	6	9
Oligopeptides	1	7	1	2	8	9
Tyrosine and derivatives	0	2	7	7	9	9
Angular furanocoumarins	1	4	3	4	7	8
Azepines	1	2	5	6	7	8
Glycosylamines	1	2	5	6	7	8
Indenes and isoindenes	2	1	5	7	6	8

Isoindoles	4	2	2	6	4	8
L-alpha-amino acids	3	2	3	6	5	8
Nitrobenzenes	0	2	6	6	8	8
Histidine and derivatives	0	2	5	5	7	7
Long-chain fatty acids	1	3	3	4	6	7
Phenylalanine and derivatives	1	1	5	6	6	7
Flavonoid-7-O-glycosides	1	0	5	6	5	6
Organic phosphonic acids	0	1	5	5	6	6
Acylaminobenzoic acid and derivatives	1	1	3	4	4	5
Alpha amino acid amides	2	1	2	4	3	5
Aspartic acid and derivatives	0	0	5	5	5	5
Indole-3-acetic acid derivatives	0	0	5	5	5	5
Pyrimidones	0	4	1	1	5	5
1,2-aminoalcohols	1	1	2	3	3	4
2'-Hydroxychalcones	1	1	2	3	3	4
Benzylisoquinolines	1	3	0	1	3	4
Hippuric acids	2	1	1	3	2	4
Isoleucine and derivatives	0	2	2	2	4	4
Macrolactams	0	4	0	0	4	4
Tetrahydroisoquinolines	0	1	3	3	4	4
1-hydroxy-2-unsubstituted benzenoids	0	0	3	3	3	3
6-alkylaminopurines	1	0	2	3	2	3
Arginine and derivatives	1	1	1	2	2	3
Aromatic monoterpenoids	0	2	1	1	3	3
Beta amino acids and derivatives	1	0	2	3	2	3

Eudesmanolides, secoeudesmanolides, and derivatives	0	3	0	0	3	3
Fatty alcohol esters	1	0	2	3	2	3
Ketoximes	0	2	1	1	3	3
Peptides	0	2	1	1	3	3
1,4-benzodiazepines	0	0	2	2	2	2
2-heteroaryl carboxamides	0	0	2	2	2	2
6-aminopurines	0	1	1	1	2	2
Acetamides	1	1	0	1	1	2
Aldoximes	0	1	1	1	2	2
Alpha-halocarboxylic acids	0	1	1	1	2	2
Benzoic acids	1	0	1	2	1	2
Benzoic acids and derivatives	1	1	0	1	1	2
Benzothiazines	0	0	2	2	2	2
Benzothiazoles	1	0	1	2	1	2
Benzoxazines	1	0	1	2	1	2
Carboxylic acid esters	0	1	1	1	2	2
Cyclic carboximidic acids	0	2	0	0	2	2
Cyclohexenones	2	0	0	2	0	2
Fatty amides	0	0	2	2	2	2
Hexoses	2	0	0	2	0	2
Hydroxy fatty acids	0	1	1	1	2	2
Ketals	0	2	0	0	2	2
m-Xylenes	0	0	2	2	2	2
Phenylpiperidines	0	2	0	0	2	2
Phenylpropanoic acids	0	0	2	2	2	2
Piperidinones	0	1	1	1	2	2

Proline and derivatives	2	0	0	2	0	2
Pyrrolizidines	2	0	0	2	0	2
Stilbenes	0	1	1	1	2	2
Styrenes	0	0	2	2	2	2
Trifluoromethylbenzenes	0	0	2	2	2	2
Tropane alkaloids	0	0	2	2	2	2
Asparagine and derivatives	0	0	1	1	1	1
Leucine and derivatives	0	0	1	1	1	1
N-phenylureas	0	1	0	0	1	1
Phenylpyrazoles	1	0	0	1	0	1
Pyranopyridines	1	0	0	1	0	1
Sesquiterpenoids	0	0	1	1	1	1
Tricarboxylic acids and derivatives	1	0	0	1	0	1

3.4.2.1 Phenazines

Three networks in positive ionization mode were classified as phenazines and derivatives using MolNetEnhancer (Figure 3.44; Table S8) (Ernst et al. 2019). Crosschecked with NAP, **95** (*m/z* 271.068, $[M + H]^+$) was annotated as 5,10-dihydrophenazine-1,6-dicarboxylic acid (DHPDC), which was only present in the extract of Δ*SCAB12101* cultured on OBA. Compound **96** (*m/z* 227.081, $[M + H]^+$) was annotated as 5,10-dihydrophenazine-1-carboxylic acid (DHPCA) in the extracts from both the WT strain cultured on OBA and the deletion strain cultured on OBA and YMSm. Notably, the intensity of **96** was relatively greater in the deletion mutant extracts than in the WT extracts (Figure 3.44). However, the NAP Consensus top-ranked candidates for **98, 99** and **101** are not phenazine derivatives, though MolNetEnhancer classified them as phenazines. When

the candidate lists generated from NAP for **98, 99** and **101** were checked, it was found that phenazine derivatives received the third highest score, which suggests that the candidate with the highest scoring match in NAP may not be always the best candidate. Thus, **98** (*m/z* 214.159, $[M + H]^+$) was annotated as phenazine-1,9-diol, which got the third highest score in NAP Consensus ranking. Similarly, **99 (***m/z* 255.076 $[M + H]^+$**)** was annotated as 9-methoxyphenazine-1-carboxylic acid, and **101 (***m/z m/z* 229.061 $[M + H]^+$**)** was annotated as phenazine-2,3,7-triol. Based on the precursor ion mass and MS^2 fragments, **100 (***m/z* 269.092 $[M + H]^+$) was annotated as 9-methoxy-phenazine-1-carboxylic acid methyl ester. Compound **97** (*m/z* 214.159, $[M + H]^+$) has mass fragmentation pattern similar to phenazines; however, the predicted structures using NAP or MetWork are all unrelated to phenazines, suggesting that it may be a novel phenazine derivative.

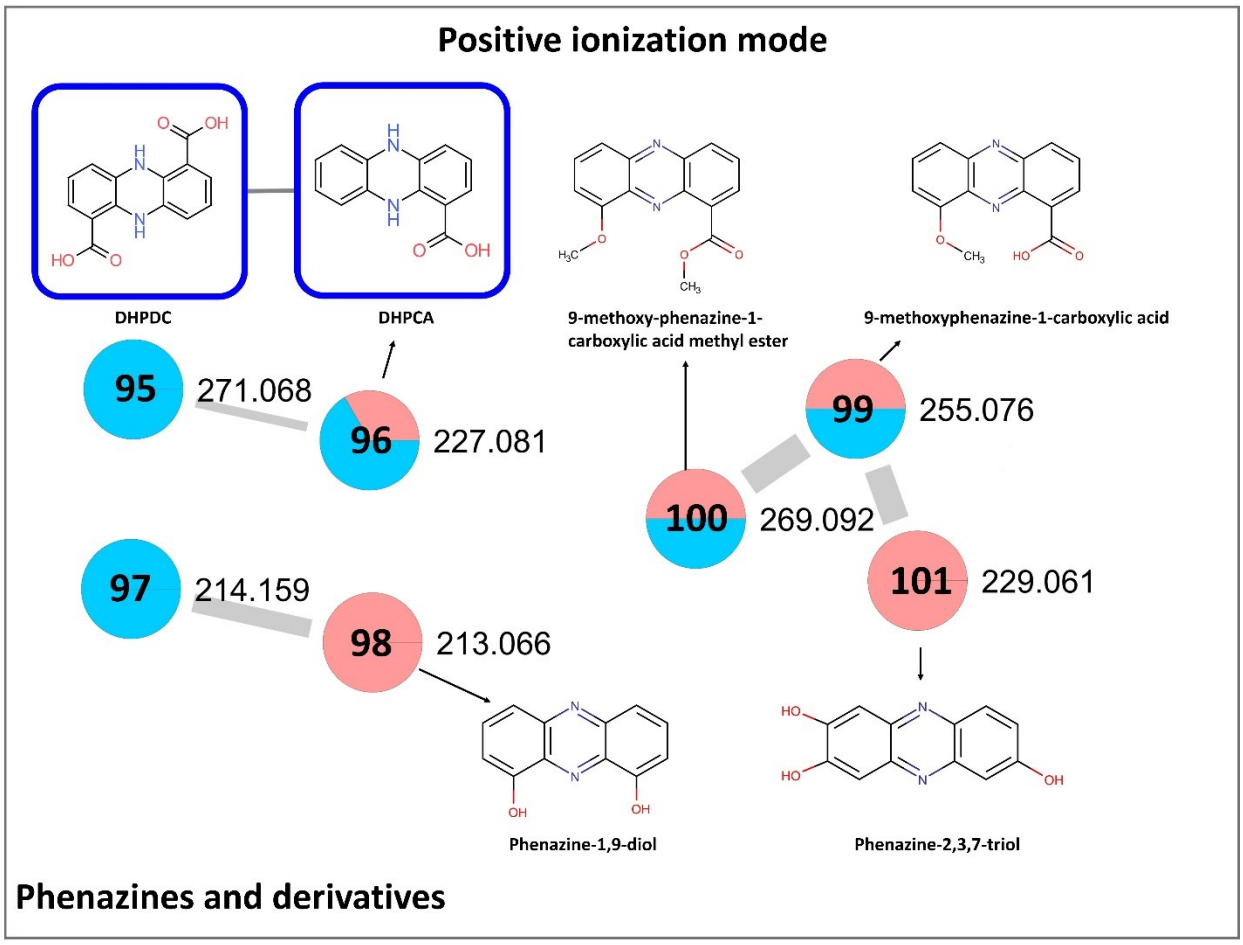


Figure 3.44 Mass spectral networks of phenazines and derivatives annotated in the *S. scabiei* 87-22 and Δ*SCAB12101* extracts in positive ionization mode. In the classical molecular network, each node represents one fragmentation spectrum from a detected compound, and node size represents the summed intensity (peak area) of the ion from all samples. Edges are created if the cosine score is > 0.6 and there are at least three matched fragment ions. Edge thickness indicates the relative similarity of MS^2 data between nodes. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* - blue. The NAP Consensus top ranked candidates were boxed in thick blue.

In addition to using molecular networking to predict the product of the phenazine-like BGC, the differences between the networks for known molecules produced by *S. scabiei* were also investigated, and the results are presented in the following sections.

3.4.2.2 Thaxtomins

Using classical molecular networking, three thaxtomin networks were annotated from the metabolomes of WT *S. scabiei* and Δ*SCAB12101*, with one in positive ionization mode and the other two in negative ionization mode (Figure 3.46). Besides the compounds previously detected from the metabolome of *S. scabiei* 87-22 (see Section 3.4.1), six additional compounds were detected that were only present in Δ*SCAB12101* OBA extract (Figure 3.46; Table S9). Based on the precursor ion mass, MS^2 fragments and retention time, **102** (*m/z* 407.17 $[M + H]^+$) was predicted to be an isomer of thaxtomin D, and **107** (*m/z* 407.136 $[M - H]^-$) was annotated as an isomer of hydroxy thaxtomin C. The remaining four compounds have mass fragmentation patterns similar to thaxtomin, but the predicted structures based on NAP are completely unrelated to the thaxtomins.

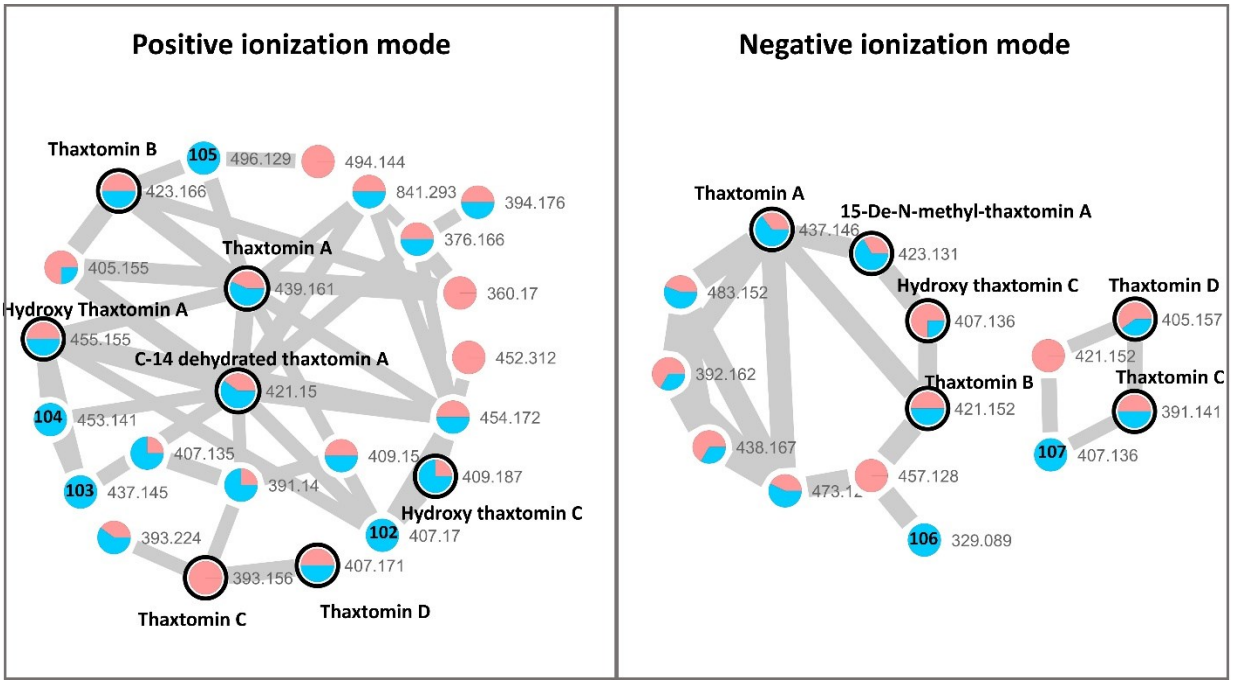


Figure 3.45 Thaxtomin mass spectral network from *S. scabiei* 87-22 and Δ*SCAB12101* in positive negative ionization mode. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* - blue. Nodes that have MS^2 matches to the GNPS libraries are outlined in black. Features in the molecular network are as described in the legend for Figure 3.44.

3.4.2.3 Coronafacoyl phytotoxins

Two coronafacoyl phytotoxin networks were annotated from the metabolomes of WT *S. scabiei* and Δ*SCAB12101* using GNPS classical molecular networking (Figure 3.47). Besides the compounds previously detected from *S. scabiei* 87-22 (Section 3.4.1), three additional compounds were detected, which were only present in the Δ*SCAB12101* MYMm and YMSm extracts (Figure 3.47; Table S10). NAP annotated **108** (*m/z* 306.169 $[M + H]^+$) as coronatine (COR); however, the *S. scabiei* 87-22 genome does not have the key biosynthetic genes required for COR production, suggesting that the NAP prediction for this compound is inaccurate. Compound **108** has a similar MS^2 fragmentation pattern as the methyl-substituted CFA-Ile, which suggests that this compound may be a methyl-substituted coronafacoyl phytotoxin derivative. Compounds **109** (*m/z* 320.541

[M - H]$^-$) and **110** (*m/z* 356.163 [M - H]$^-$) have mass fragmentation patterns similar to coronafacoyl phytotoxins (Figure 3.47; Table S10); however, the predicted structures based on NAP are completely unrelated to the coronafacoyl phytotoxins.

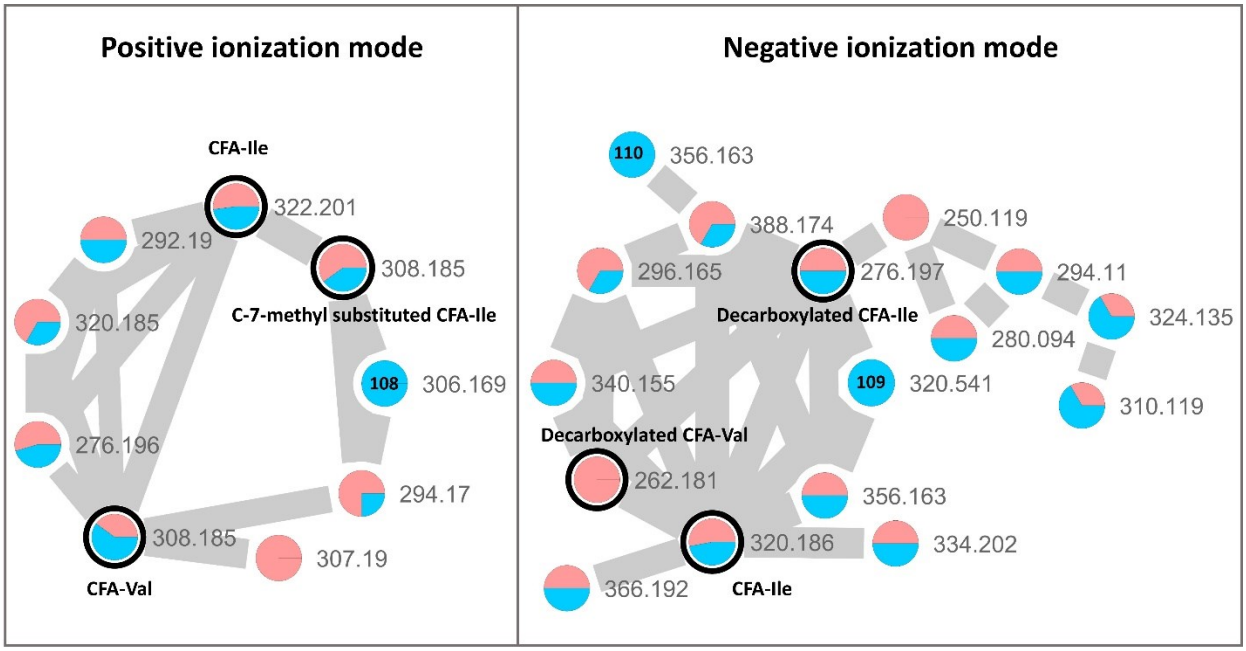


Figure 3.46 Coronafacoyl phytotoxins mass spectral network from *S. scabiei* 87-22 and Δ*SCAB12101* in positive negative ionization mode. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* – blue. Nodes that have MS^2 matches to the GNPS libraries are outlined in black. Features in the molecular networks are as described in the legend for Figure 3.44.

3.4.2.4 Concanamycins

A concanamycin network in positive ionization mode was detected using GNPS classical molecular networking from the metabolomes of WT *S. scabiei* and Δ*SCAB12101* (Figure 3.48). The results show that the O-methyl-concanamycin B detected in the WT extracts was absent from the Δ*SCAB12101* mutant extracts (Figure 3.48). Other metabolites in the network were also absent from the mutant extracts, though their identities are unknown. One compound (**111,** *m/z* 873.508 $[M + Na]^+$) was detected only in the Δ*SCAB12101* OBA extract (Figure 3.48). Based on the NAP prediction, **111** was annotated as avermectin A1b; however, the *S. scabiei* 87-22 genome does not harbour known avermectin biosynthetic genes, suggesting that the NAP prediction for this compound is inaccurate. The key MS^2 fragments (196.1, 378.2, 379.2, 485.2, 485.3, 486.2, 486.3, 873.5) of **111** are similar to those of the concanamycins, suggesting that it may be a novel concanamycin analogue.

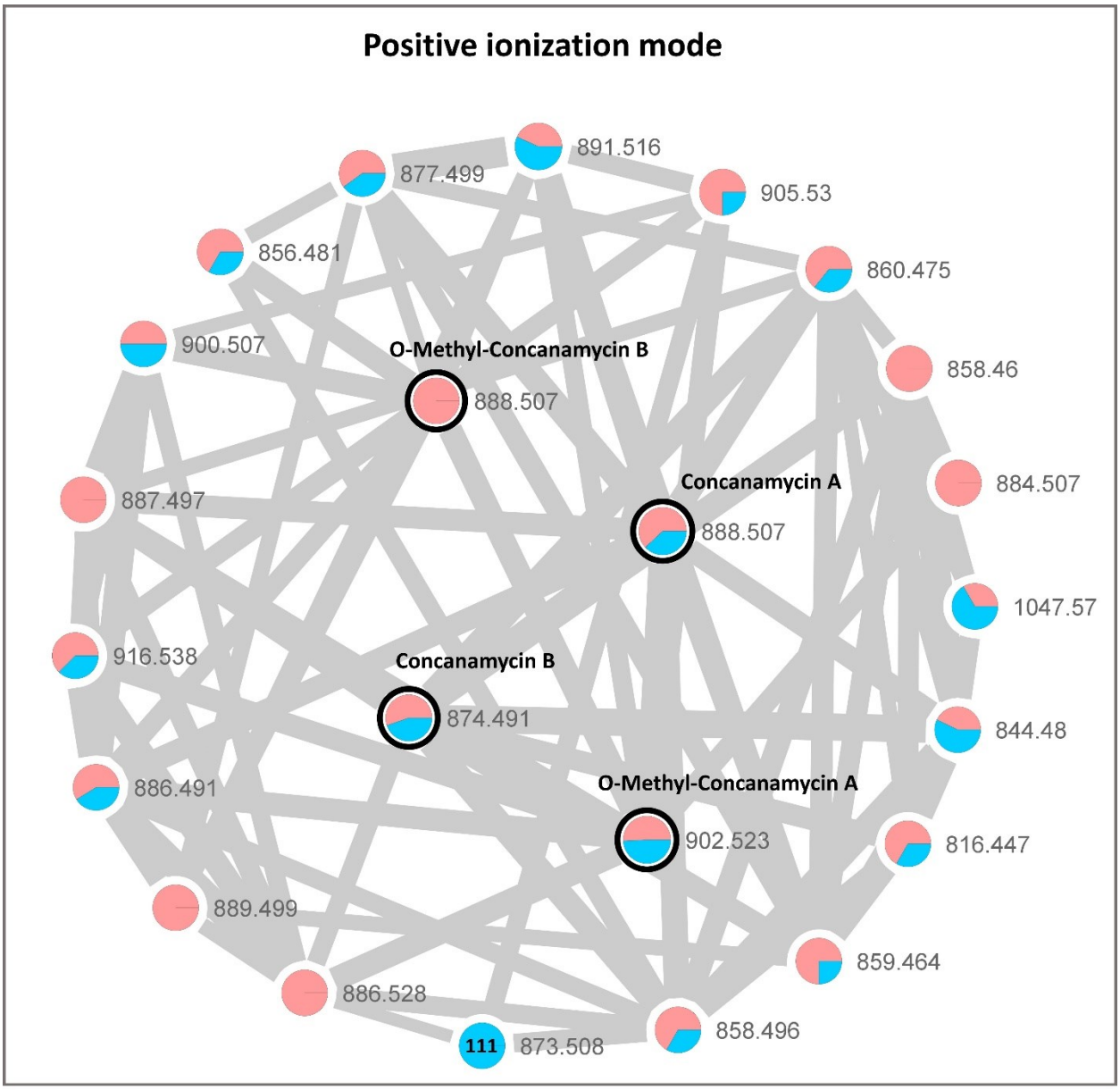


Figure 3.47 Concanamycin mass spectral network from *S. scabiei* 87-22 and Δ*SCAB12101* in positive negative ionization mode. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* – blue. Nodes that have MS^2 matches to the GNPS libraries are outlined in black. Features in the molecular network are as described in the legend for Figure 3.44.

3.4.2.5 Other compounds

Using GNPS classical molecular networking, two bottromycin networks were annotated from the metabolomes of *S. scabiei* 87-22 and Δ*SCAB12101* in positive and negative ionization mode (Figure 3.49). All of the compounds detected previously from the metabolome of *S. scabiei* 87-22 (Section 3.4.1) were also detected in the Δ*SCAB12101* mutant extracts. Overall, the analysis suggests that Δ*SCAB12101* generally produces fewer bottromycin-related compounds than WT *S. scabiei* 87-22 (Figure 3.49).

The metabolomics analysis also identified one pyochelin network in positive ionization mode and two desferrioxamine E networks in positive and negative ionization mode (Figure 3.50). The results suggest that Δ*SCAB12101* produces more desferrioxamine-related compounds and fewer pyochelin-related compounds than WT *S. scabiei* 87-22. Besides the compounds previously detected from *S. scabiei* 87-22 (Section 3.4.1), three additional desferrioxamine E-related compounds were identified from the combined analysis of the *S. scabiei* 87-22 and Δ*SCAB12101* extracts (Figure 3.50; Table S11). Compound **112** (*m/z* 617.351 $[M + H]^+$) was annotated as 6-hydoxylated analogue of nocardamine based on its precursor ion mass, MS^2 fragments, and MetWork prediction. Based on the mass fragmentation pattern and NAP/MetWork prediction, **113** and **114** are suggested to be novel desferrioxamine analogues (Figure 3.50; Table S11).

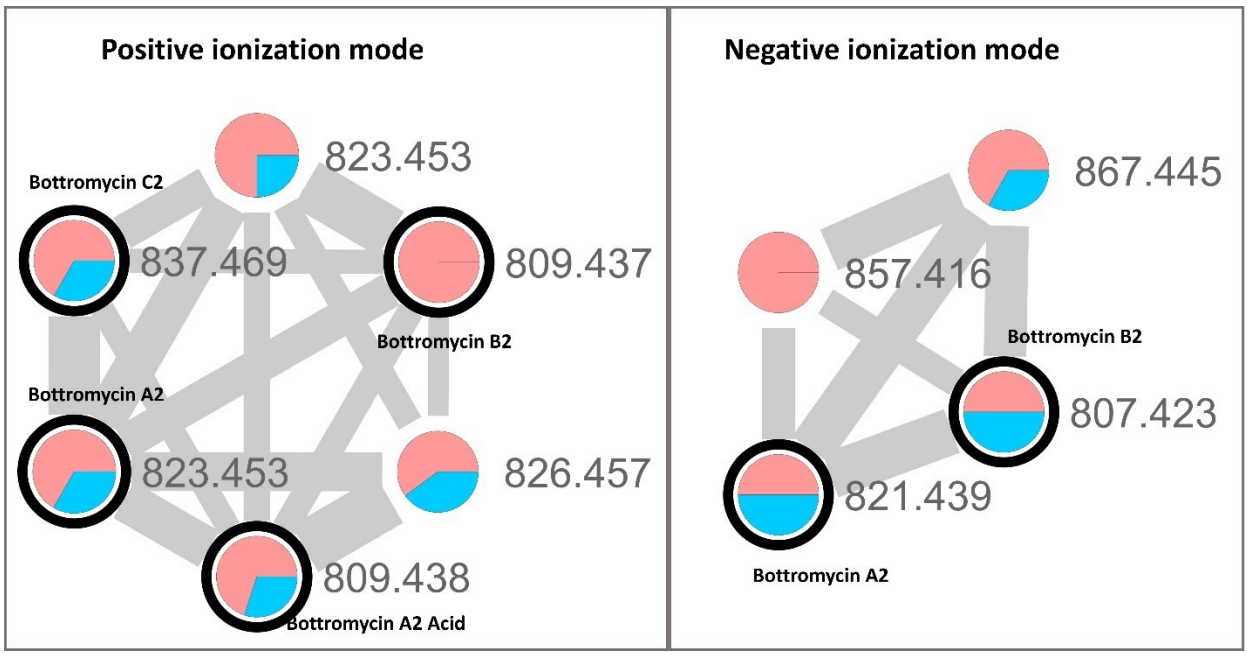


Figure 3.48 Bottromycins mass spectral network from *S. scabiei* 87-22 and Δ*SCAB12101* in positive and negative ionization mode. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* - blue. Nodes that have MS^2 matches to the GNPS libraries are outlined in black. Features in the molecular network are as described in the legend for Figure 3.44.

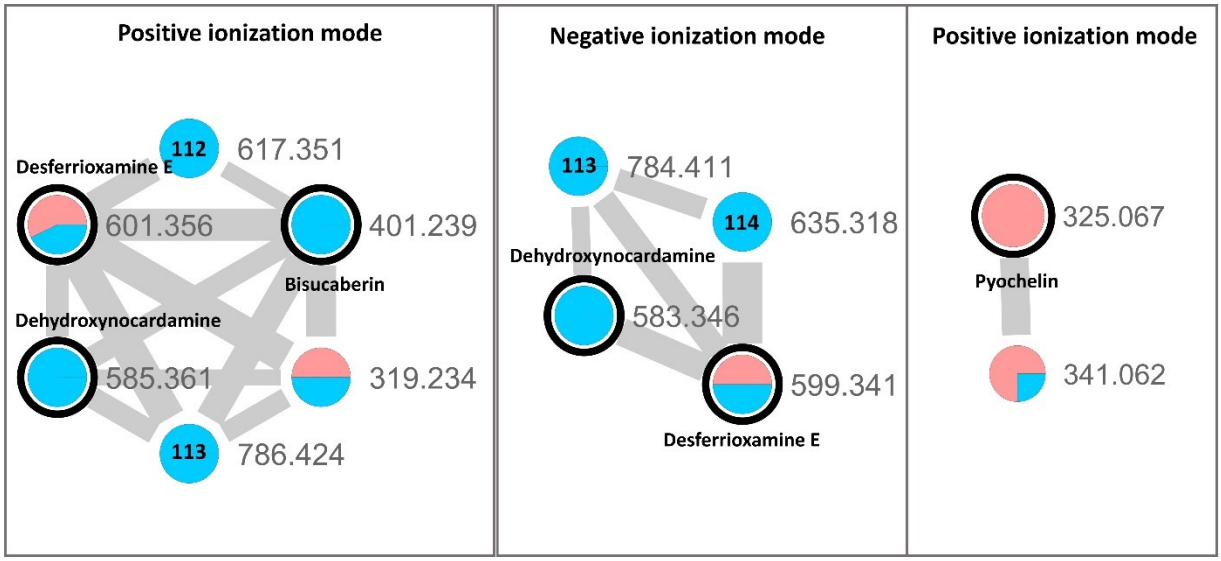


Figure 3.49 Siderophores mass spectral network from *S. scabiei* 87-22 and Δ*SCAB12101* in positive and negative ionization mode. The pie charts indicate the relative abundance of each compound in the two strains: *S. scabiei* 87-22 – red; Δ*SCAB12101* - blue. Nodes that have MS^2 matches to the GNPS libraries are outlined in black. Features in the molecular network are as described in the legend for Figure 3.44

CHAPTER 4: DISCUSSION

The main objective of this thesis was to investigate an orphan specialized metabolite BGC that is present on the chromosome of *S. scabiei* 87-22. The first approach taken was to use different bioinformatics tools to predict the boundaries of this BGC and the nature of the molecule(s) that is produced. As this unknown BGC is not expressed or is expressed at very low levels in *S. scabiei* 87-22 under known laboratory conditions, different engineered strains were constructed to activate the BGC, either in the native host or in different heterologous hosts. Bioassays were conducted to examine the bioactivity and pathogenicity of the engineered strains, and growth studies were conducted to examine the effects of the BGC on morphological development. Expression of genes within the BGC was analyzed in a subset of engineered strains using semi-quantitative RT-PCR. Finally, metabolomics analysis was performed on a promising engineered strain in order to detect the product(s) of the cryptic BGC. The main findings of this thesis and their significance will be discussed in the following sections.

4.1 Bioinformatics analysis of the orphan phenazine-like BGC

The orphan phenazine-like BGC was initially annotated more than ten years ago, and the annotation was based on the presence of genes encoding proteins that are similar to known phenazine biosynthetic enzymes (Yaxley 2009). However, the boundaries of the BGC, the function of each gene, and the conservation of the BGC within the genome sequences of other *Streptomyces* species remained unclear.

4.1.1 Genomic analysis of specialized metabolite BGCs in *S. scabiei* using antiSMASH and DeepBGC

To obtain an updated annotation of the specialized metabolite BGCs present within the *S. scabiei* 87-22 genome, two genome mining tools, antiSMASH 5.0 (Blin et al. 2019b) and DeepBGC 1.0 (Hannigan et al. 2019), were both employed. AntiSMASH 5.0 uses a machine learning technique called Hidden Markov Models (HMM) and a human-defined rules-based approach to identify BGCs for specialized metabolites (Blin et al. 2019b). In contrast, DeepBGC excels at detecting novel BGC classes by employing a deep learning technique called Recurrent Neural Networks (RNNs) and Pfam domains to detect BGCs in genome sequences (Hannigan et al. 2019). In this study, 24 BGCs were detected by both DeepBGC and AntiSMASH in the *S. scabiei* 87-22 genome (Table 3.1 and 3.2). DeepBGC detected 112 more BGCs than AntiSMASH; however, there were ten BGCs that were only detected by AntiSMASH and not by DeepBGC. AntiSMASH can predict the closest known compound for a BGC, while DeepBGC can not (Kautsar et al. 2019). For example, both AntiSMASH and DeepBGC were able to detect a genomic region containing the thaxtomin BGC (Table 3.1 and 3.2), but only AntiSMASH was able to identify it as the one involved in thaxtomin biosynthesis (50% similarity to the thaxtomin BGC in the MiBIG database). For an unknown compound with known activity, DeepBGC will be much more helpful for finding the corresponding BGC (Hannigan et al. 2019). Overall, AntiSMASH and DeepBGC are complementary to each other as they each have unique advantages in predicting BGCs. The combined application of AntiSMASH and DeepBGC allowed for comparisons and a better understanding of the biosynthetic potential of *S. scabiei* 87-22. However, neither antiSMASH nor DeepBGC classified the cryptic BGC as a phenazine: antiSMASH predicted that the cryptic BGC is a butyrolactone BGC from *SCAB12041* to *SCAB12121* (region 8 in

antiSMASH), and DeepBGC (Region 23 in DeepBGC) predicted it as “Other” BGC from *SCAB12081* to *SCAB12101* (Table 3.1; Table 3.2). So, to determine the boundaries of the BGC and the conservation of the BGC within the genome sequences of other *Streptomyces* species, other bioinformatics tools, such as large-scale, pan-genomic mining tools and whole genome alignment tools, need to be employed.

4.1.2 Large-scale network analysis, classification, and phylogenetic analysis of BGCs

A disadvantage of both AntiSMASH and DeepBGC is that they can only be used to detect BGCs on a single-genome basis. With the increased availability of sequenced bacterial genomes, genome mining for BGCs has expanded from a single-genome basis to large-scale, pan-genomic mining. Large-scale clustering of these BGCs allows researchers to compare and classify BGCs into GCFs, and to investigate the evolution of the gene clusters within and across the GCFs in the different organisms. In this study, two new computational workflows, BiG-SCAPE and CORASON (Navarro-Muñoz et al. 2020), which are tightly integrated with antiSMASH and MIBiG, were recruited to explore the conservation and evolution of the *S. scabiei* 87-22 BGCs (predicted by antiSMASH) in the genomes of other species in the *Streptomycetaceae* family. A sequence similarity network consisting of 5890 BGCs as unique nodes from 188 genome sequences (187 from *Streptomycetaceae* family and one from *Pseudomonas aeruginosa* PAO1) was generated using BIG-SCAPE. BIG-SCAPE classified these 5890 BGCs into eight types, with most of the BGCs classified into the "Others" type (Figure 3.4), and it grouped the BGCs into 431 GCFs (≥ three BGCs). Out of the 34 BGCs identified in *S. scabiei* 87-22 by antiSMASH, 26 were associated with other BGCs (Figure 3.5). BGCs involved in the biosynthesis of ectoine, desferrioxamine E, hopene and spore pigment form six large networks (>100 BGCs) with BGCs from other *Streptomyces*, indicating that these regions are well-conserved across the

Streptomycetaceae family. In contrast, region 8 (the phenazine-like BGC) formed a small network with *S. bottropensis* region 023, *S. stelliscabiei* region 001, and *Streptomyces* sp. 1222.2 region 004, which suggests that the phenazine-like BGC is only conserved in these four *Streptomyces* species with ≥75% similarity (Figure 3.6). To study the evolutionary relationship of the *S. scabiei* 87-22 region 008 (the phenazine-like BGC) within and across different GCFs, a multi-locus phylogenetic tree was constructed using CORASON. Region 008 formed a distinct clade with the same regions identified in the similarity network by BiG-SCAPE, and these in turn formed a clade with two other regions from *Kitasatospora* species (Figure 3.7). Thus, the *S. scabiei* phenazine-like BGC appears to be most closely related to BGCs identified in only a small number of *Streptomyces* and *Kitasatospora* species.

Phylogenetic analysis of individual genes within the *S. scabiei* BGC (*SCAB12021, SCAB12031, SCAB12041, SCAB12051, SCAB12061, SCAB12071, SCAB12081, SCAB12091, SCAB12101*, and *SCAB12111*) demonstrated that the genes form a distinct clade with those from *S. stelliscabiei*, *Streptomyces* sp. 1222.2, *S. bottropensis*, *Kitasatospora* sp. Root107 and *Kitasatospora* sp. MMS16–BH015, which supports the CORASON results suggesting that the phenazine–like BGC from *S. scabiei* has a very close relationship with the BGCs from the above species. In addition, the phylogenetic analysis suggested that there is a close relationship with genes from *S. ipomoeae, Saccharothrix* sp. ST 888 and *Streptacidiphilus pinicola.* Interestingly, *S. scabiei, S. stelliscabiei, S. bottropensis* and *S. ipomoeae* have all been reported as plant pathogenic species (Bukhalid et al. 2002; Wanner 2009; Bignell et al. 2010a; Guan et al. 2012), which suggests that the conserved phenazine-like BGC may play a role in pathogenesis. *Kitasatospora* sp. Root107 may also be a plant-associated bacterium as it was isolated from the

root of *Arabidopsis thaliana* (BioSample: SAMN04155693). However, there is little information about *Streptomyces* sp.1222.2 and *Saccharothrix* sp. ST–888.

4.1.3 Whole-genome alignment using Mauve

Comparative genomics represents a new approach to identify new types of BGCs. For example, the Eustáquio group identified the diazaquinomycin BGC in *Streptomyces* sp. F001 and in *Micromonospora* sp. B006 using progressive MAUVE alignments (Braesel et al. 2019), and the Machida group identified the kojic acid and oxylipin BGCs from *Aspergillus* species using comparative genomics (Itaru et al. 2014). In this study, whole-genome alignment by the progressive Mauve algorithm was employed to predict the boundaries of the phenazine-like BGC (Darling et al. 2010). Alignment of *S. scabiei* 87-22 chromosome with that of *S. stelliscabiei*, *Streptomyces* sp. 1222.2 and *S. bottropensis* indicated that the region from *SCAB11861* to *SCAB12211* is well conserved and intact in these four species. Based on this, the phenazine–like BGC in *S. scabiei* is predicted to be 42.9 kb in size and consists of 34 genes (Figure 3.9; Table 3.4), which is much larger than initially thought (Yaxley 2009). A comparison of the phenazine–like BGC from the above four species revealed that these genes are of high similarity and the gene arrangements of these phenazine–like biosynthetic genes are all identical, though some of the genomes are not completely closed. The phenazine–like BGC is located in the arm region of the chromosome of *S. scabiei* and *Streptomyces* sp. 1222.2 (Figure 3.8), which suggests that this BGC may be involved in secondary/specialized metabolism instead of primary metabolism (Bentley et al. 2002). The location of the BGC in *S. stelliscabiei* and *S. bottropensis,* however, is unclear as the two genomes may not have been properly assembled (Figure 3.8).

4.1.4 Predicted functions for genes in the putative phenazine–like BGC from *S. scabiei*

The phenazine-like BGC contains three main phenazine biosynthetic genes, *phzC* (*SCAB12021*), *phzD* (*SCAB12041*) and *phzE* (*SCAB12051*). While no *phzF* (*SCAB43981*) and *phzG* (*SCAB51331*) homologues were identified within the BGC, homologues were found in the core region of the *S. scabiei* 87-22 chromosome. *phzF* (*SCAB43981*) is predicted to be located in a T1PKS/NRPS BGC (Region#18 predicted by antiSMASH; Table 3.1), and no homologous BGCs were found in the MIBiG database. *phzG* (*SCAB51331*) is predicted to be located in an unknown BGC predicted by DeepBGC (Region#69; Table 3.2). It has been reported that *phzC-G* are essential in phenazine biosynthesis, whereas *phzA and phzB* are important but non-essential biosynthetic proteins (McDonald et al. 2001; Guo et al. 2017). The functions of PhzC, PhzD and PhzE are well established by experiments; however, the functions of PhzA, PhzB, PhzF and PhzG still remain hypothetical. In the bioinformatics analysis conducted here, SCAB12031 is homologous to a protein from *S. kebangsaanensis* that was previously predicted to be a PhzA homologue, but protein alignment analysis suggests that SCAB12031 lacks key residues required for PhzA activity.

SCAB12111 encodes a protein that is similar to TylR, a positive activator of tylosin production in *S. fradiae* (Bate et al. 1999, 2006; Stratigopoulos et al. 2004). Based on this and on the location of *SCAB12111* within the phenazine-like BGC, it was hypothesized that *SCAB12111* functions as a positive activator of genes within the BGC. Analysis of the phenazine-like BGC also revealed the presence of two genes encoding GBL receptor homologues, *SCAB12091* and *SCAB12101*, and one GBL biosynthesis gene *SCAB12081*. GBLs are quorum sensing communication signals employed by *Streptomyces* to regulate specialized metabolite production and morphological development (Nishida et al. 2007). The GBL regulatory system usually consists

of a GBL synthase that produces one or more GBLs, and a cognate GBL receptor. In some instances, there are also GBL receptor homologues, called pseudo GBL receptors, which function as transcriptional repressors but do not bind GBL molecules. For example, the virginiamycin, actinorhodin, jadomycin, and tylosin gene clusters of *S. virginiae*, *S. coelicolor*, *S. venezuelae* and *S. fradiae,* respectively, harbour both GBL receptors (BarA, ScbR, JadR3, and TylP) and pseudo GBL receptors (BarB, ScbR2, JadR2, and TylQ) (Xu and Yang 2019). Notably, the two predicted GBL receptors within the *S. scabiei* phenazine-like BGC do not seem to have resulted from gene duplication, as the phylogenetic analysis revealed that *SCAB12091* has a closer relationship with genuine GBL receptors, while *SCAB12101* is more closely related to other pseudo GBL receptors. Thus, *SCAB12091* is likely a genuine GBL receptor and *SCAB12101* a pseudo GBL receptor (Figure 3.21) (Xu and Yang 2019). Notably, the *SCAB12091* gene is located next to the GBL synthase gene *SCAB12081*, which is consistent with reports that the GBL synthase gene is usually situated next to the genuine GBL receptor gene (Takano 2006). The bioinformatic analysis also revealed the presence of a gene, *SCAB11971,* encoding a homologue of AfsR, which is a conserved pleiotropic regulator in *Streptomyces* spp. AfsR-like proteins are known to be associated with specialized metabolite BGCs and function as transcriptional activators (Horinouchi 2003), and thus it is predicted that *SCAB11971* may encode an activator of the phenazine-like BGC. Other genes in the phenazine–like BGC may be responsible for the transport (*SCAB12001, SCAB12151, SCAB12171,* and *SCAB11991*) or modification of the final product of the BGC.

4.2 Gene expression, morphology, pathogenicity, and bioactivity analysis of the engineered *Streptomyces* strains

In order to activate the expression of the cryptic phenazine-like BGC, the two putative negative repressor genes *SCAB12091* (GBL receptor) and *SCAB12101* (pseudo GBL receptor)

were deleted individually and in combination, and the putative activator gene *SCAB12111* was overexpressed in the WT strain as well as in the *SCAB12101* deletion strain. In addition, strong, constitutive promoters were introduced upstream of the phenazine biosynthetic genes in *S. scabiei*, and some of the genes from the phenazine-like BGC were moved into three different heterologous hosts. A *SCAB12081* (GBL synthase) deletion strain, which was initially constructed as part of a collaboration with another research group, was also included in this analysis. The engineered strains were cultured on different media, and the expression of the phenazine-like BGC in some strains was analyzed by semi-quantitative RT-PCR. The morphology and bioactivity of the engineered strains were also analyzed.

4.2.1 Culture medium composition affects the gene expression pattern

In this study, the WT strain along with selected engineered strains were cultured on five different media, MYMm, SFM, YMSm, ISP-4, and SA agar, to test which medium is suitable for the expression of the cryptic phenazine-like BGC. MYMm is a modified version of MYM, a rich medium used for assessing the production of specialized metabolites and other natural products by *Streptomyces* spp. (Hegemann et al. 2015; Daniel-Ivad et al. 2017; Gehrke et al. 2019). YMSm is a modified version of YMS, usually used for cultivation and antibiotic production in *S. avermitilis* (Ikeda et al. 1987; Kitani et al. 2011; Komatsu et al. 2013). SFM is a complex medium used for routine cultivation and antibiotic production in *Streptomyces* species (Kieser et al. 2000). ISP-4 is used for cultivation and characterization of *Streptomyces* as per the International *Streptomyces* Project (Shirling and Gottlieb 1966). SA is a semi-synthetic medium used for detecting specialized metabolite production by *S. clavuligerus* (Ferguson et al. 2016; AbuSara et al. 2019). Semi-quantitative RT-PCR results showed that the expression of the core phenazine biosynthetic genes could be detected when two of the engineered strains [Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*)]

were cultured on MYMm and YMSm, which suggests that these two media support the activation of the phenazine-like BGC. Therefore, MYMm and YMSm were selected in subsequent analyses for testing gene expression in the engineered strains.

4.2.2 The pseudo GBL receptor *SCAB12101* is a negative repressor of genes in the phenazine-like BGC

Results presented in this study show that deletion of the predicted GBL receptor gene *SCAB12091* did not activate the expression of the core phenazine genes, while deletion of the pseudo GBL receptor gene *SCAB12101* did (Figure 3.26 and 3.27). This suggests that *SCAB12101* is a negative repressor of the cryptic phenazine-like BGC, whereas *SCAB12091* may function as a transcriptional activator rather than a repressor (Figure 4.1). It is not uncommon for a GBL receptor and pseudo-GBL receptor to exhibit opposing effects on the regulation of a BGC in *Streptomyces* species. For example, in *S. filipinensis,* the GBL receptor SfbR activates filipin production, whereas the pseudo GBL receptor SfbR2 repress its production (Barreales et al. 2020). Similarly in *S. pristinaespiralis,* the GBL receptor SpbR is an activator of pristinamycin biosynthesis, whereas the pseudo GBL receptors PapR3 and PapR5 are negative repressors (Mast et al. 2015). In *S. aureofaciens*, the GBL receptor SagR activates auricin biosynthesis while the pseudo GBL receptor Aur1R represses its biosynthesis (Novakova et al. 2010; Wang et al. 2011). Furthermore, the GBL receptor JadR3 in *S. venezuelae* functions as an activator, and the pseudo GBL receptor JadR2 functions as both an activator and a repressor (Xu et al. 2010; Zou et al. 2014). Further studies will be needed to determine what role *SCAB12091* plays in the regulation of the phenazine-like BGC. RT-PCR also indicated that the predicted pseudo-GBL receptor gene *SCAB12101* not only contributes to the regulation of expression of the core phenazine biosynthetic genes within the phenazine-like BGC, but also influences the expression of other genes within the BGC (Figure

3.27). Possibly, this may be due to the ability of pseudo GBL receptors to directly bind to multiple promoter regions and regulate their biosynthesis (Li et al. 2015; Xu and Yang 2019).

4.2.3 *SCAB12101* may regulate the phenazine-like BGC via *SCAB12111*

The results of the gene expression analysis also indicated that overexpression of predicted regulatory gene *SCAB12111* in WT *S. scabiei* induces the expression of the core phenazine genes (Figure 3.27). Notably, expression of *SCAB12111* itself was induced in the Δ*SCAB12101* mutant (Figure 3.27), and overexpression of *SCAB12111* within the Δ*SCAB12101* mutant led to higher levels of phenazine biosynthetic gene expression compared to the Δ*SCAB12101* mutant (Figure 3.27). Together, these results suggest that *SCAB12111* is an activator of the core phenazine biosynthetic genes, and that the pseudo GBL receptor *SCAB12101* represses the biosynthetic genes by directly inhibiting the transcription of *SCAB12111* (Figure 4.1). Notably, some potential binding sites for *SCAB12101* were identified upstream of *SCAB12111* using the EMBOSS palindrome searching tool (Figure 3.16A), supporting the idea that *SCAB12101* may regulate the phenazine-like BGC via *SCAB12111,* though further studies are needed to confirm this. Xu et al. also demonstrated that the pseudo GBL receptor ScbR2 in *S. coelicolor* represses the expression of coelimycin P1 BGC by directly inhibiting the expression of the cluster-situated activator KasO (Xu et al. 2010; Gomez-Escribano et al. 2012b). Additionally, the pseudo GBL receptor JadR2 from *S. venezuelae* inhibits the expression of the jadomycin BGC via repression of the JadR1 activator (Xu et al. 2010).

4.2.4 GBL biosynthesis may negatively regulate expression of the core phenazine biosynthetic genes

Generally, GBLs produced by GBL synthase homologues are required for specialized

metabolite production. A reduced production of specialized metabolites is anticipated when deleting the GBL synthase gene, whereas an opposite or no effect is anticipated when deleting the corresponding GBL receptor gene(s) (Takano 2006; Zou et al. 2014). However, improvements in specialized metabolite production have also been reported in GBL synthase gene deletion strains. For example, in *S. bingchenggensis*, deletion of the GBL synthase gene *sbbA* increased the production of milbemycin, whereas deletion of the GBL receptor gene *sbbR* decreased production (He et al. 2018). Similarly in *S. coelicolor* and *S. lividans,* inactivation of the GBL synthase gene *scbA* resulted in an increased production of both actinorhodin and undecylprodigiosin (Butler et al. 2003; Takano et al. 2001; Li et al. 2015). In this study, the gene expression analysis indicated that the deletion of the predicted GBL synthase gene *SCAB12081* induced the expression of the core phenazine genes in *S. scabiei*. This may indicate that the putative GBL(s) produced by *SCAB12081* negatively affects gene expression by serving as a ligand for the GBL receptor *SCAB12091* and preventing it from binding to its target promoter(s), which might include the *SCAB12111* promoter. By deleting *SCAB12081*, this would presumably abolish GBL production, thereby possibly enabling *SCAB12091* to bind to its target promoter(s) and activate gene expression (Figure 4.1).

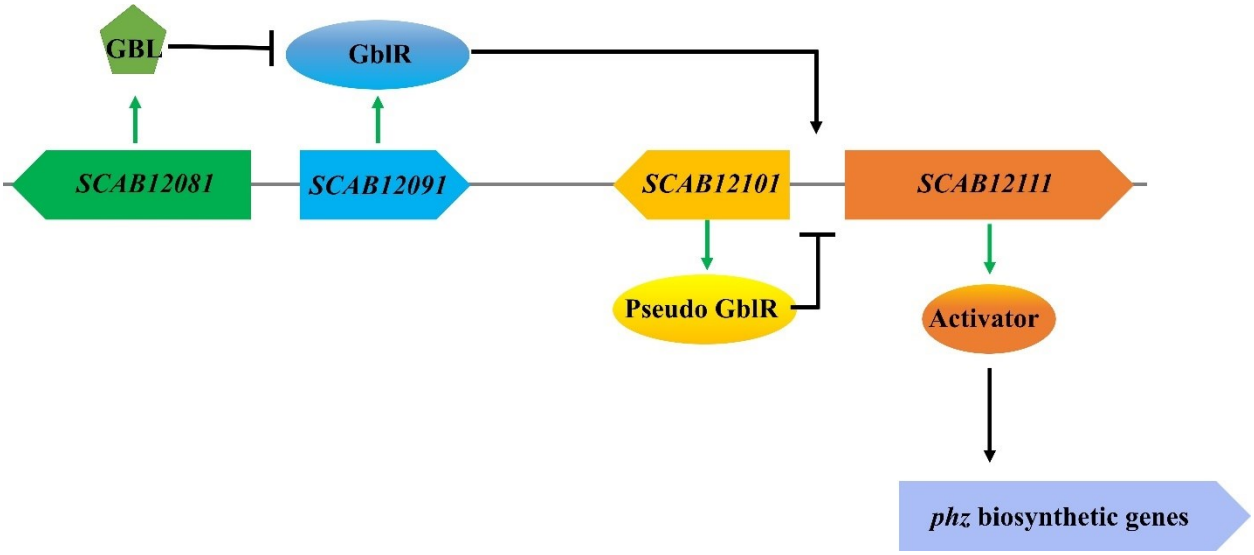


Figure 4.1 Proposed GBL regulatory model in *S. scabiei*. Regulatory interactions are shown by black arrowhead (activation) or black lines ending with a bar (repression). Lines in green represent transcription and translation of the biosynthetic genes.

4.2.5 No correlation was observed between the expression of the phenazine-like BGC and the pathogenicity of *S. scabiei*

To determine whether or not the product of the phenazine-like BGC exhibits phytotoxic activity like other specialized metabolites produced by *S. scabiei,* potato tuber tissue bioassays were performed using ethyl acetate extracts from different *Streptomyces* engineered strains cultured on MYMm, YMSm and OBA. The results indicated that the extracts prepared from all of the *Streptomyces* engineered strains were able to cause slightly pitted necrosis of the potato tissue, with the exception of the heterologous expression strains. The most severe symptoms were observed with the OBA culture extracts. OBA is a plant-based medium that is known to support the production of phytotoxins such as thaxtomins (Johnson et al. 2007), CFA-Ile (Bignell et al. 2010b), and concanamycins (Fyans et al. 2016). Compared to the WT *S. scabiei* 87-22 extracts, the extracts from many of the engineered strains [Δ*SCAB12081,* Δ*SCAB12091,* Δ*SCAB12101*,

Δ(*SCAB12091+SCAB12101)*, pJL3/87-22], displayed more severe effects on the potato tuber tissue (Figure 3.29). However, there was no correlation between the severity of necrosis/pitting observed and the expression profile of the phenazine-like BGC in the corresponding strain. The severity of tissue damage may correlate with the amounts of other known phytotoxins (thaxtomins, concanamycins, coronafacoyl phytotoxins) produced by the different strains. It has been shown that pure thaxtomin A can cause necrosis on excised potato tuber tissue (Loria et al. 2006), and CFA-Ile induces tuber tissue hypertrophy (Bignell et al. 2018); furthermore, Natsume and colleagues showed that pure concanamycin A induced slightly sunken lesions on potato tuber tissue (Natsume et al. 2017). Further studies are needed to examine the relative amounts of these different phytotoxins produced by the engineered strains.

To assess the virulence phenotype of some of the different S*treptomyces* engineered strains, potato tuber tissue bioassays were performed using agar plugs from well-sporulated culture plates. The results showed that the WT *S. scabiei* readily colonized the surface of the potato tuber tissue and caused severe deep-pitted necrosis of the potato tuber tissue (Figure 3.30). Deletion of *SCAB12091* and *SCAB12081* did not affect the pathogenicity of *S. scabiei*, but deletion of *SCAB12101* appeared to reduce the pathogenicity of *S. scabiei*, and similar findings were observed for the Δ(*SCAB12091+SCAB12101*) mutant and *SCAB12111* overexpression strain (87-22/pJL3). As *SCAB12101* is proposed to regulate the phenazine-like BGC via the *SCAB12111* activator, the reduced pathogenicity of the Δ*SCAB12101*, Δ(*SCAB12091+SCAB12101*), and 87-22/pJL3 strains may suggest that *SCAB12111* is not only a CSR, but a pleiotropic regulator that controls the expression of multiple pathways, thereby affecting the production of multiple specialized metabolites. It is not uncommon that CSRs can directly control the expression of genes in other BGCs. For example, GdmRIII is not only a cluster-situated positive activator in the biosynthesis

of geldanamycin in *Streptomyces autolyticus* CGMCC0516, but also a negative repressor in the biosynthesis of elaiophylin (Jiang et al. 2017). Similarly, JadR1 is not only a cluster-situated positive activator in the biosynthesis of jadomycin in *S. venezuelae,* but also a negative reprssor in the expression of biosynthetic genes within the chloramphenicol BGC (Xu et al. 2010).

4.2.6 The product of the phenazine-like BGC may have antibacterial activity

Methods like the agar plug assay, disk diffusion assay and the soft agar overlay assay are commonly used to determine the antimicrobial activity of actinobacteria. We did not detect any antimicrobial activity when using the agar plug assay method and the disk diffusion assay method, which may be due to the concentration of secreted metabolites being too low to have inhibitory effects. However, when using the soft agar overlay method, antifungal activity was readily detected from the WT *S. scabiei* 87-22 and all other *S. scabiei* engineered strains cultured on YMSm (Figure 3.31). Concanamycins are biologically active against fungi and are most likely responsible for the observed antifungal activity, though this would need to be confirmed (Bignell et al. 2014). The *SCAB12111* overexpression strain (87-22/pJL3) and the Δ*SCAB12101* mutant showed activity against the Gram-positive bacterium *B. subtilis*, which may potentially relate to the production of phenazines, as phenazine natural products exhibit broad-spectrum activity against various organisms including bacteria and fungi (Laursen and Nielsen 2004; Mavrodi et al. 2006). In addition, *S. avermitilis* SUK17/pJL14 showed activity against *S. epidermidis,* another Gram-positive bacterium. However, the observed activities were not reproducible when the soft agar overlay assays were repeated multiple times. It has been reported that zones of inhibition in the soft agar overlay assay can result from nutrient deprivation, pH effects or antibiotic production (Nkanga and Hagedorn 1978). Thus, the modified agar overlay method developed by Nkanga and

Hagedorn could be used to eliminate the nutrient deprivation and pH effects to improve the reproducibility of the results (Nkanga and Hagedorn 1978).

4.2.7 The phenazine-like BGC is not involved in morphological development

Morphological development in *Streptomyces* spp. is under the control of various small molecules. For example, A-factor and its derivatives, cyclic di-GMP, and cyclic-AMP, have also been reported to be involved in morphological differentiation of *Streptomyces* (Horinouchi 2007; Tschowri et al. 2014). To investigate whether or not the expression of the phenazine-like BGC affects the morphological phenotype of the engineered *Streptomyces* strains, these strains along with their controls were cultured on four different agar media (MYMm, YMSm, OBA and SFM) for 7 days (Figure 3.32 – Figure 3.34). However, morphological analysis showed that no noticeable difference was observed in the morphology of these engineered strains cultured on MYMm, YMSm, and OBA, except that delays in growth were observed in Δ*SCAB12091*, Δ*SCAB12101*, Δ(*SCAB12091*+*SCAB12101*), Δ*SCAB12101*/pJL15, Δ*SCAB12101*/pJL3, and 87-22/pIJ8641 cultured on SFM, which suggests that the medium composition of SFM may have effects on the the primary metabolism of the above strains. However, the phenazine-like BGC does not seem to be involved in morphological development.

4.3 Metabolomics analysis

The metabolomics analysis described in this study was conducted for two purposes: (1) to explore the metabolic potential of WT *S. scabiei,* and (2) to try and identify the product(s) of the phenazine-like BGC. For the latter, the Δ*SCAB12101* mutant was chosen for comparative metabolic profiling since the gene expression results from this study suggested that this mutant

may produce detectable levels of the metabolite(s) of interest. The main findings of the analysis and their significance are discussed below.

4.3.1 Detection of new analogues of known metabolites as well as molecules not known to be produced by *S. scabiei*

To characterize the metabolic potential of WT *S. scabiei*, the bacterium was cultured on three different agar growth media: YMSm, MYMm, and OBA. This follows the OSMAC principle, where a single strain has the potential to produce different molecules under different environmental conditions (Romano et al. 2018). As different compounds have different ionization preferences, the extracts prepared from the plate cultures were subjected to untargeted LC-MS^2 in both positive and negative ionization mode to detect as many compounds as possible. The resulting spectral data were analyzed using the FBMN workflow within the GNPS platform (Nothias et al. 2020) and BioDendro workflow (Rawlinson et al. 2020). These two metabolomics tools use different and complementary similarity and visualization methods to explore spectral similarity in metabolomics experiments. For instance, FBMN uses a cosine score to measure the spectral similarity between two fragmentation spectra and presents the results as molecular networks along with isomer discrimination and quantitative results for statistical analysis within the networks (Nothias et al. 2020). In contrast, BioDendro uses Bray-Curtis or Jaccard to calculate the pairwise distances between all spectra and enables hierarchical clustering of MS^2 spectra and presents the results as a tree (Rawlinson et al. 2020).

The analyses conducted in this study revealed that the metabolic profile of *S. scabiei* varies among the three media tested, with most metabolites being observed in the plant-based medium OBA. Metabolites were annotated by comparing masses and fragmentation patterns with the GNPS open-access library (Wang et al. 2016) or by using Network Annotation Propagation (NAP)

(da Silva et al. 2018), SIRIUS (Böcker et al. 2009), MetWork (Beauxis and Genta-Jouve 2019), and Competitive Fragmentation Modeling for Metabolite Identification 3.0 (CFM-ID 3.0) (Allen et al. 2014). Using this approach, new analogues of known metabolites as well as molecules that were not previously known to be produced by *S. scabiei* were putatively annotated.

Eleven thaxtomin analogues have been previously identified from different pathogenic *Streptomyces* spp., and these vary in the presence or absence of hydroxyl and *N*-methyl groups on the thaxtomin backbone (King and Calhoun 2009). Using the FBMN and BioDendro, multiple thaxtomin analogues were annotated from the metabolome of WT *S. scabiei* 87-22, including thaxtomin A, thaxtomin B, thaxtomin C, thaxtomin D, *p*-isomer of thaxtomin A, hydroxy thaxtomin A, hydroxy thaxtomin C, and 15-de-*N*-methyl analogue of thaxtomin A. OBA is the best medium for supporting the production of most compounds in the networks, consistent with the presence of cellobiose and other cello-oligosaccharides in oat-based media, which are known inducers of thaxtomin production in *S. scabiei* and other *Streptomyces* species (Wach et al. 2007; Johnson et al. 2007). In contrast, the production of thaxtomins in MYMm and YMSm has not been described before, and neither contain cello-oligosaccharides or the other known thaxtomin inducer, suberin (Lerat et al. 2010). Thaxtomin C is the major form of thaxtomin produced by the sweet potato pathogen *Streptomyces ipomoeae* (King et al. 1994), and trace amounts of thaxtomin C and thaxtomin D have also been reported from *S. scabiei* (King and Calhoun 2009; Jiang et al. 2018a), but their production has never been reported from *S. scabiei* cultured on MYMm and YMSm. The other thaxtomin analogues mentioned above have all been previously reported to be isolated from *S. scabiei* (King and Lawrence 1996; King and Calhoun 2009). In addition, three C-14 dehydrated analogues of thaxtomin A were also putatively annotated, which have only been reported as

biotransformation products of *Aspergillus niger* (Lazarovits et al. 2004) and have not been detected in *S. scabiei* before.

CFA-Ile has been shown to be the main coronafacoyl phytotoxin produced by *S. scabiei*, though other minor compounds that are likely coronafacoyl derivatives have also been detected in culture extracts (Bown et al. 2017; Bignell et al. 2018). In this study, multiple coronafacoyl phytotoxins were putatively annotated from the metabolomics analysis of *S. scabiei* 87-22, including CFA-Ile, methyl-substituted derivative of CFA-Ile, CFA-Val, decarboxylated derivatives of CFA-Ile, and decarboxylated derivatives of CFA-Val. Most of the compounds from both networks could be detected in all three culture extracts from MYMm, YMSm and OBA. While production of coronafacoyl phytotoxins in oat-based media has been described before (Fyans et al. 2015), the current study is the first to detect production of these molecules in YMSm and MYMm. *N*-coronafacoyl-valine (CFA-Val) is a known coronafacoyl phytotoxin (Bignell et al. 2018), and it has been previously shown that *S. scabiei* likely produces this compound (Fyans et al. 2015). The decarboxylated derivatives of CFA-Ile and CFA-Val have never been reported previously. The production of methyl-substituted coronafacoyl derivatives was previously proposed based on our studies of the biosynthesis of CFA-Ile in *S. scabiei* (Bown et al. 2017), but this is the first time that such molecules have been detected in culture extracts of WT *S. scabiei.* It is notable that the production of methyl-substituted coronafacoyl derivatives has not been observed in other coronafacoyl phytotoxin-producing bacteria, and would presumably result from the incorporation of methylmalonyl-CoA instead of ethylmalonyl-CoA during synthesis of the coronafacic acid polyketide moiety (Bown et al. 2017).

The metabolomics analysis putatively annotated four concanamycin analogues, including concanamycin A, concanamycin B, O-methyl-concanamycin A, and O-methyl-concanamycin B.

OBA was the best medium for supporting production of most of the concanamycins, which is consistent with the reports that oat-based media can support the production of concanamycin A and B in *S. scabiei* and other *Streptomyces* spp. (Natsume et al. 1996, 1998, 2001; Fyans et al. 2016), while production in MYMm and YMSm has not been previously reported. The analysis also enabled the annotation of four bottromycin analogues, including bottromycin A2, B2, C2, and A2 acid, which have all been previously reported to be produced by *S. scabiei* (Gomez-Escribano et al. 2012a; Crone et al. 2016). MYMm was the best medium for bottromycin production in this study, and this medium is similar to the GYM medium that is typically used for bottromycin fermentations (Vior et al. 2020).

Several siderophore metabolites were annotated in the metabolome of the *S. scabiei* WT culture extracts, including desferrioxamine E, pyochelin and dehydroxynocardamine, and bisucaberin. The production of desferrioxamine and pyochelin has previously been confirmed in this organism (Seipke et al. 2011; Kodani et al. 2013). Bisucaberin is part of a family of dihydroxamate siderophores originally isolated from the marine bacterium *Alteromonas haloplanktis* (Fujita et al. 2012; Senges et al. 2018). Bisucaberin has also been reported in some *Streptomyces* species and identical building blocks are used for the biosynthesis of bisucaberins and desferrioxamines (Barona-Gómez et al. 2004; Kadi et al. 2008), suggesting that the two metabolites are likely synthesized by gene products from the same BGC.

Other compounds, such as IAA and ectoine, have also been annotated in the current study. IAA is the major active form of auxins, which are plant hormones responsible for cell division, differentiation, root architecture formation, apical dominance and senescence (Morel and Castro-sowinski 2013). Production of IAA has been reported in many plant pathogenic and plant growth-promoting microorganisms, including *S. scabiei*, and homologues of IAA biosynthetic genes were

previously reported in the genome of *S. scabiei* 87-22 (Bignell et al. 2010a). Ectoine is a water-soluble organic osmolyte, which is widely produced by *Streptomyces*, helping them to cope with extreme osmotic stress (Sadeghi et al. 2014).

Several compounds, such as cyclo(L-Val-L-Pro), aerugine, decahydroquinoline, andrachcinidine, and mairine B, were predicted to be present in the WT culture extracts and were not previously known to be produced by *S. scabiei* (Figure 3.52, Table 3.6). The cyclodipeptide cyclo(L-Val-L-Pro) and related metabolites are mainly formed by NRPSs or cyclodipeptide synthases, and they exhibit a variety of biological activities, including plant growth promotion (Ortiz-Castro et al. 2011), cell-to-cell communication (Belin et al. 2012), and have antimicrobial and anticancer properties (Mishra et al. 2017). The production of cyclodipeptides has been demonstrated in other *Streptomyces* spp. (Gosse et al. 2019); however, their corresponding BGCs have not been identified. Genomic analysis of *S. scabiei* revealed the presence of many short or incomplete NRPS BGCs, though none are predicted to utilize Val or Pro as substrates. In addition, cyclodipeptide synthase encoding genes were not identified in the *S. scabiei* genome based on homology searches. Thus, it is currently unclear which BGC is involved in the biosynthesis of cyclo(L-Val-L-Pro) in *S. scabiei*. Aerugine is a siderophore that has been reported to be produced by *Pseudomonas* and *Streptomyces* spp. (Inahashi et al. 2017), and is proposed to be derived from the hydrolytic cleavage and subsequent reduction of pyochelin (Inahashi et al. 2017). Therefore, aerugine and pyochelin, which was also detected in our study, are likely synthesized using the same BGC. The decahydroquinolines are lipophilic alkaloids that have important pharmacological activities and have been reported in extracts from the skin of neotropical poison frogs (Tokuyama et al. 1987). Andrachcinidine is a 2,6-disubstituted piperidine alkaloid that has been isolated from the small perennial plant *Andrachne aspera* Spreng and may function as a chemical defense agent

(Apichaisataienchote et al. 2006). Mairine B is a skytanthine-type monoterpenoid alkaloid that has been isolated from the plant *Incarvillea mairei* (Xing et al. 2010). The production of decahydroquinoline, andrachcinidine and mairine B by *Streptomyces* species has not been previously reported. Therefore, further studies will be required to characterize these metabolites and to identify the BGCs responsible for their production in *S. scabiei.*

4.3.2 Comparative metabolomics analysis of WT *S. scabiei* and the Δ*SCAB12101* mutant leads to the putative identification of phenazine derivatives

To putatively identify the product(s) of the phenazine-like BGC, a comparative metabolomics analysis was conducted. For this purpose, classical molecular networking was employed to generate networks using positive and negative ionization mode data in GNPS (Wang et al. 2016). Although FBMN allows for isomer discrimination and can approximate quantitative interpretation of the molecular network (Nothias et al. 2020), it is less sensitive than classical networking, and FBMN requires pre-processing of the data and has more user-defined parameters to set, and thus it is more complicated than classical GNPS. Thus, to rapidly discover novel compounds, classical molecular networking is sufficient. After molecular networking, background and media subtraction, 336 metabolites were detected in the Δ*SCAB12101* extracts in positive ionization mode that were absent from the WT extracts. In negative ionization mode, the Δ*SCAB12101*extracts contained 292 metabolites what were absent from the WT extracts. Possibly, this may be due to cross-regulation of multiple BGCs by the pseudo GBL receptor, such that deletion of *SCAB12101* activates production of multiple specialized metabolites (Xu et al. 2010; Wang et al. 2011; Mingyar et al. 2014; Li et al. 2015; Zhu et al. 2016; Xu and Yang 2019). Further investigations are required to determine whether this is the case.

In this study, the MolNetEnhancer was employed to discover the chemical families that we are interested in, for example, the phenazines. MolNetEnhancer combines the outputs from classical molecular networking, MS2LDA (substructural annotation), NAP (in silico annotation), and DEREPLICATOR (peptidic natural products annotation), and can classify the chemical families and subfamilies through ClassyFire, which provides a more comprehensive chemical overview of metabolomics data (Van Der Hooft et al. 2016; Mohimani et al. 2017; Ernst et al. 2019). Using the MolNetEnhancer workflow, phenazine derivatives in the class of diazanaphthalenes and carboxylic acids were detected in the mutant extracts. In addition, flavonoids, coumarins, and indoles were detected, which could be attributed to the cross-pathway regulation by the pseudo GBL receptor.

The metabolomics analysis of Δ*SCAB12101* and WT *S. scabiei* putatively annotated six phenazine derivatives, including DHPDC, DHPCA, phenazine-1,9-diol, 9-methoxyphenazine-1-carboxylic acid, 9-methoxy-phenazine-1-carboxylic acid methyl ester and phenazine-2,3,7-triol. The phenazine biosynthetic pathway is derived from the shikimic acid pathway, and their biosynthesis has been well elucidated; however, there are still gaps in understanding the last few steps (Laursen and Nielsen 2004; Blankenfeldt and Parsons 2014; Guttenberger et al. 2017; Guo et al. 2020). PhzC catalyzes the first step of the shikimate pathway, while PhzA/B, PhzD, PhzE, PhzF, and PhzG catalyze the transformation of chorismic acid into DHPDC and DHPCA. Both DHPDC and DHPCA are central intermediates in the biosynthesis of strain-specific phenazines (Figure 4.2) (Laursen and Nielsen 2004; Blankenfeldt and Parsons 2014; Guttenberger et al. 2017; Guo et al. 2020). The other annotated phenazine derivatives could be novel phenazine derivatives, and their natural production and biosynthesis pathways have not been reported before. Further experiments are needed to elucidate their structure.

DHPDC and the unknown phenazine derivative **97** were only present in the extracts of Δ*SCAB12101* cultured on OBA and YMSm, respectively. The results also suggested that Δ*SCAB12101* produces more DHPCA than WT strain, while the amount of the 9-methoxyphenazine-1-carboxylic acid and 9-methoxy-phenazine-1-carboxylic acid methyl ester ions in the deletion mutant extracts was almost the same in the WT extracts. Interestingly, phenazine-1,9-diol and phenazine-2,3,7-triol were only present in the WT extracts, which suggests that there could be other phenazine BGCs in the genome of *S. scabiei* 87-22. BlastP analysis was conducted using the amino acid sequence of PhzC (SCAB12021), PhzD (SCAB12041), PhzE (SCAB12051), PhzF (SCAB43891), and PhzG (SCAB51331). It was found that there are two homologues of PhzC (SCAB12021) encoded in the *S. scabiei* genome: SCAB67661 (41.3% identity, 54.5% similarity) and SCAB1981 (42.7% identity, 53.0% similarity), and they all have the same conserved domain DAHP_synth_2. PhzE (SCAB12051) also has two homologues encoded in the genome: SCAB67631 (51.3% identity, 63.1% similarity) and SCAB1971 (24.8% identity, 35.6% similarity), and they all have the same conserved domains, a Chorismate_bind domain and a GATase1 domain. In contrast, no homologue of PhzD (SCAB12041) was identified. *SCAB67661* and *SCAB67631* are separate by one gene, *SCAB67651*; however, they are not located in any BGC predicted by antiSMASH or DeepBGC. *SCAB1981* and *SCAB1971* are neighbours, and they are found to be located in an unknown BGC (region#2 predicted by antiSMASH and DeepBGC) (Table 3.1 and 3.2). AntiSMASH predicted that this unknown BGC is a 35 kb betalactone, which has 8% similarity to the esmeraldin (phenazine derivative) BGC; DeepBGC predicted that this unknown BGC is 139 kb NRP, which has antibacterial activities (Table 3.1 and 3.2). The Blast search also showed that PhzF (SCAB43891) has one homologue, SCAB76381 (30.1% identity, 35.6% similarity), and PhzG (SCAB51331) has one homologue, SCAB51781

(39.1% identity, 50.7% similarity). So, it is possible that there are other phenazine BGCs located in the genome of *S. scabiei. P. aeruginosa* have been reported to have two copies of a phenazine BGC, and each copy is sufficient for the phenazine production. In addition, the expression of the two copies of the operon is differentially regulated (Mavrodi et al. 2001). Therefore, the two compounds, phenazine-1,9-diol and phenazine-2,3,7-triol, which were only present in the WT extracts, may be produced by another phenazine BGC that is present in the genome of *S. scabiei* 87-22.

Besides the compounds previously detected from the metabolome of *S. scabiei* 87-22, additional thaxtomins, coronafacoyl phytotoxins, concanamycins, and siderophores, which were only present in Δ*SCAB12101*, were also putatively annotated. The results also suggest that the mutant generally produces less bottromycin-related compounds, more desferrioxamine-related compounds and less pyochelin-related compounds compared with the WT strain. This may indicate that *SCAB12101* has pleiotropic effects on the biosynthesis of other metabolites in *S. scabiei*, though further studies are required to confirm this.

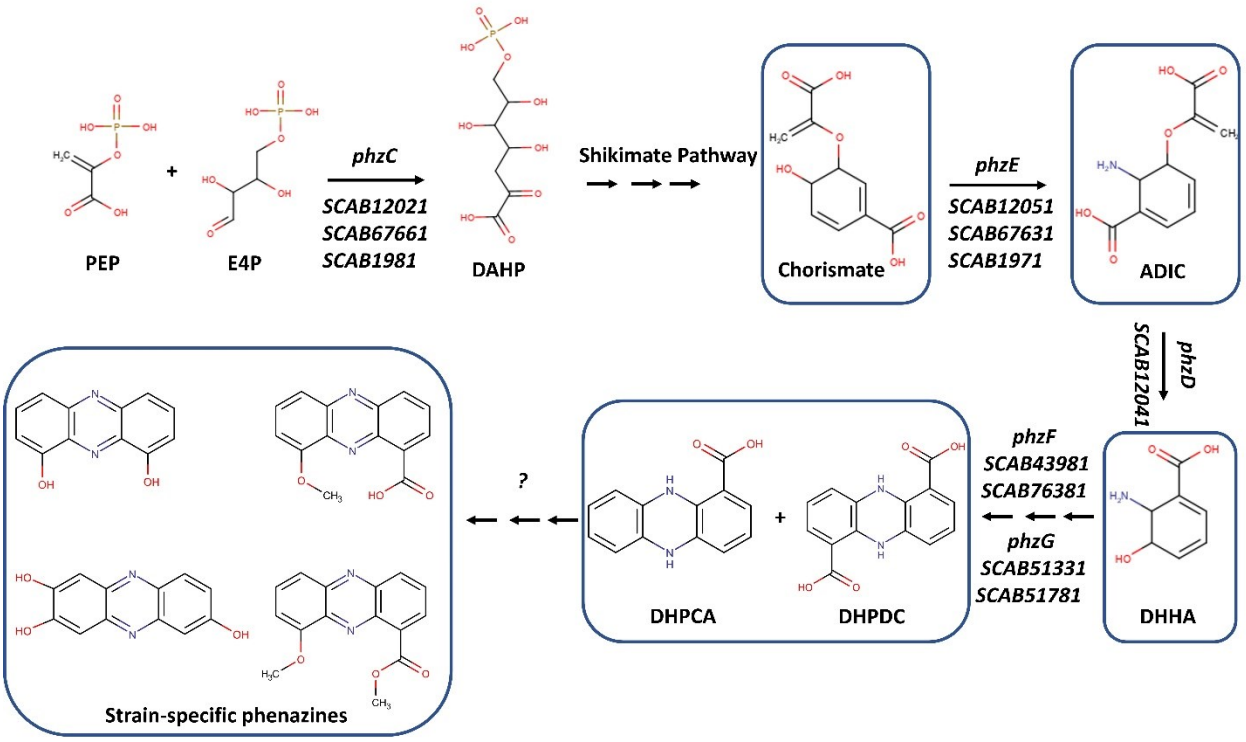


Figure 4.2 Hypothetical biosynthetic pathway for the biosynthesis of phenazine derivatives in *S. scabiei.* Molecules that were detected in the metabolomics analysis of *S. scabiei* 87-22 and Δ*SCAB12101* are boxed. Chorismate (*m/z* 179.034 $[M-H]^-$) was annotated by GNPS library; ADIC (*m/z* 226.074 $[M+H]^+$; *m/z* 224.056 $[M-H]^-$) was annotated by NAP; DHHA (*m/z* 156.066 $[M+H]^+$; *m/z* 154.05 $[M-H]^-$) was annotated by NAP; the annotations for DHPCA, DHPDC and other phenazine derivatives were described in Section 3.4.2.1.

4.4 Conclusions and future directions

4.4.1 Conclusions

Overall, this study provides novel insights into a cryptic and silent phenazine-like BGC that is present in the genome of the potato CS pathogen *S. scabiei* 87-22. Using different bioinformatics tools, the boundaries of the BGC were predicted, and it is proposed that the BGC is much larger than initially thought. Molecular and culture-based strategies were successfully employed to activate expression of the BGC, and this also enabled a better understanding of the

role of predicted regulatory genes in controlling the expression of the BGC. Metabolomics analysis conducted in this study suggest that the BGC is involved in the production of phenazine-related metabolites. This analysis additionally provided new insights into the overall specialized metabolic potential of *S. scabiei* and enabled the prediction of several novel compounds that were not previously known to be produced by this organism.

4.4.2 Future directions

Several important questions remain to be addressed regarding the phenazine-like BGC. In particular, the identity of the molecules produced by the BGC require further characterization to confirm that they are phenazine-related metabolites. Given that the *S. scabiei* Δ*SCAB12101*/*SCAB12111* overexpression strain constructed in this study was shown to exhibit very high expression of the core phenazine biosynthetic genes, this strain could be used for conducting large-scale purification of the molecules from culture extracts for structural characterization and bioactivity (antimicrobial, phytotoxic) studies. If the resulting molecules are shown to exhibit phytotoxic activity, then the role of the molecules in the pathogenicity of *S. scabiei* can be assessed by constructing a biosynthetic mutant in a strain that does not produce the principal pathogenicity determinant thaxtomin A. This will enable the detection of any subtle changes in the pathogenic phenotype of *S. scabiei* resulting from mutation of the phenazine-like BGC. Furthermore, confocal microscopy using GFP-labeled WT and BGC mutant strains of *S. scabiei* could be performed to determine if the resulting metabolites play a role in plant tissue colonization during infection. Similar work was used to show that the virulence-associated Nec1 protein from *S. turgidiscabies* is important for radish root colonization by the pathogen (Joshi et al. 2007a). GFP-based reporter plasmids could additionally be used as described before (Bignell

et al. 2010b) to determine whether genes within the phenazine-like BGC are expressed when the pathogen is colonizing plant tissues.

The phenazine-like BGC is conserved in the genomes of other actinobacteria, including two *Streptomyces* spp. that are pathogenic to plants. Although the role of the resulting metabolite(s) in the pathogenicity of *S. scabiei* remains unclear, evidence is presented in this thesis suggesting that the molecule(s) may exhibit antibacterial activity, and thus may have a function in inter-microbial interactions as described for other phenazine-producing organisms (Mazzola et al. 1992; Guttenberger et al. 2017; Biessy and Filion 2018). If the pure molecules are shown in bioactivity assays to exhibit antimicrobial activity, then the role of the molecules in the competition of *S. scabiei* in the soil environment can be assessed by culturing the phenazine producers (*S. scabiei* engineered strain) and non-producers (*S. scabiei* phenazine biosynthetic mutant) in pasteurized soil (reduced microbial competition) for 2-7 days, respectively. Then the populations of the *S. scabiei* strains could be estimated by serial dilution of multiple soil samples (Mazzola et al. 1992).

Future work could also be conducted to verify the boundaries proposed in this study for the phenazine-BGC. *S. scabiei* deletion mutants of genes near the proposed boundaries could be generated to determine whether such genes are required for metabolite biosynthesis. In addition, the regulation of the BGC requires further investigation. Based on the results of this study, it is hypothesized that *SCAB12091* is a genuine GBL receptor that activates expression of the phenazine-like BGC, and so overexpression of the gene could be conducted to determine whether this enhances expression of genes in the BGC. Also, the nature of the GBL(s) produced by the predicted GBL synthase *SCAB12081*, and the impact of the GBL(s) on *SCAB12091* activity could be investigated further. The target gene(s) and binding site(s) of *SCAB12091*, *SCAB12101* and *SCAB12111*, all of which are predicted to be DNA binding proteins, could be elucidated by

overexpressing and purifying the proteins from *E. coli* and conducting electrophoretic mobility shift and DNA footprinting assays. Putative binding sites for *SCAB12101* and/or *SCAB12091* were identified in the current study and could be used as a starting point for such investigations.

Finally, the metabolomics analysis conducted in the current study suggests that *S. scabiei* is capable of producing novel derivatives of known metabolites as well as molecules that were not previously known to be produced by this organism. Further research could aim to characterize the structure and bioactivity of the novel metabolites and to identify the associated BGC in cases where this is currently unknown. Furthermore, the function(s) of these novel molecules for the producing organism would be an interesting avenue of research to pursue.

REFERENCES

AbuSara NF, Piercey BM, Moore MA, et al (2019) Comparative genomics and metabolomics analyses of clavulanic acid-producing *Streptomyces* species provides insight into specialized metabolism. Front Microbiol 10:2550. https://doi.org/10.3389/fmicb.2019.02550

Ahmadi MK, Pfeifer BA (2016) Recent progress in therapeutic natural product biosynthesis using *Escherichia coli*. Curr Opin Biotechnol 42:7–12. https://doi.org/10.1016/j.copbio.2016.02.010

Ahuja EG, Janning P, Mentel M, et al (2008) PhzA/B catalyzes the formation of the tricycle in phenazine biosynthesis. J Am Chem Soc 130:17053–17061. https://doi.org/10.1021/ja806325k

Allen F, Pon A, Wilson M, et al (2014) CFM-ID: A web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res 42:94–99. https://doi.org/10.1093/nar/gku436

Andryukov B, Mikhailov V, Besednova N (2019) The biotechnological potential of secondary metabolites from marine bacteria. J Mar Sci Eng 7:176. https://doi.org/10.3390/jmse7060176

Apichaisataienchote B, Korpraditskul V, Fotso S, Laatsch H (2006) Aerugine, an antibiotic from *Streptomyces fradiae* strain SU-1. Kasetsart J (Nat Sci) 40:335–340

Bai C, Zhang Y, Zhao X, et al (2015) Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. Proc Natl Acad Sci U S A 112:12181–12186. https://doi.org/10.1073/pnas.1511027112

Balagurunathan R, Radhakrishnan M, Shanmugasundaram T, et al (2020) Protocols in actinobacterial research. Springer US, New York, NY

Baltz RH (2016) Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. J Ind Microbiol Biotechnol 43:343–370. https://doi.org/10.1007/s10295-015-1682-x

Baltz RH (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol 37:759–772. https://doi.org/10.1007/s10295-010-0730-9

Barona-Gómez F, Wong U, Giannakopulos AE, et al (2004) Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J Am Chem Soc 126:16282–16283. https://doi.org/10.1021/ja045774k

Barreales EG, Payero TD, Jambrina E, Aparicio JF (2020) The *Streptomyces filipinensis* gamma-butyrolactone system reveals novel clues for understanding the control of secondary metabolism. Appl Environ Microbiol 86:e00443-20. https://doi.org/10.1128/AEM.00443-20

Barry SM, Kers JA, Johnson EG, et al (2012) Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. Nat Chem Biol 8:814–816. https://doi.org/10.1038/nchembio.1048

Bate N, Bignell DRD, Cundliffe E (2006) Regulation of tylosin biosynthesis involving "SARP-helper" activity. Mol Microbiol 62:148–156. https://doi.org/10.1111/j.1365-2958.2006.05338.x

Bate N, Butler R, Gandecha AR, Cundliffe E (1999) Multiple regulatory genes in the tylosin

biosynthetic cluster of *Streptomyces fradiae*. Chem Biol 6:617–624

Beauxis Y, Genta-Jouve G (2019) MetWork: A web server for natural products anticipation. Bioinformatics 35:1795–1796. https://doi.org/10.1093/bioinformatics/bty864

Belin P, Moutiez M, Lautru S, et al (2012) The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. Nat Prod Rep 29:961–979. https://doi.org/10.1039/c2np20010d

Bender CL, Alarcón-Chaidez F, Gross DC, et al (1999) *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol Mol Biol Rev 63:266–292

Bentley SDD, Chater KFF, Cerdeño-Tárraga AM, et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a

Bérdy J (2005) Bioactive microbial metabolites. J Antibiot (Tokyo) 58:1–26. https://doi.org/10.1038/ja.2005.1

Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol 8:208–215. https://doi.org/10.1016/j.mib.2005.02.016

Bicz J (2013) Investigations of siderophore and tetronic acid biosynthesis in *Streptomyces scabies* 87.22. University of Warwick

Biessy A, Filion M (2018) Phenazines in plant-beneficial *Pseudomonas* spp.: biosynthesis, regulation, function and genomics. Environ Microbiol 20:3905–3917. https://doi.org/10.1111/1462-2920.14395

Bignell DRD, Cheng Z, Bown L (2018) The coronafacoyl phytotoxins: structure, biosynthesis, regulation and biological activities. Antonie Van Leeuwenhoek 111:649–666. https://doi.org/10.1007/s10482-017-1009-1

Bignell DRD, Fyans JK, Cheng Z (2014) Phytotoxins produced by plant pathogenic *Streptomyces* species. J Appl Microbiol 116:223–235. https://doi.org/10.1111/jam.12369

Bignell DRD, Huguet-Tapia JC, Joshi M V., et al (2010a) What does it take to be a plant pathogen: Genomic insights from *Streptomyces* species. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 98:179–194. https://doi.org/10.1007/s10482-010-9429-1

Bignell DRD, Seipke RF, Huguet-Tapia JC, et al (2010b) *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant–microbe interactions. Mol Plant-Microbe Interact 23:161–175. https://doi.org/10.1094/MPMI-23-2-0161

Blankenfeldt W, Kuzin AP, Skarina T, et al (2004) Structure and function of the phenazine biosynthetic protein PhzF from *Pseudomonas fluorescens*. Proc Natl Acad Sci U S A 101:16431–16436. https://doi.org/10.1073/pnas.0407371101

Blankenfeldt W, Parsons JF (2014) The structural biology of phenazine biosynthesis. Curr Opin Struct Biol 29:26–33. https://doi.org/10.1016/j.sbi.2014.08.013

Blin K, Pascal Andreu V, de los Santos ELC, et al (2019a) The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 47:D625–D630. https://doi.org/10.1093/nar/gky1060

Blin K, Shaw S, Steinke K, et al (2019b) antiSMASH 5.0: updates to the secondary

metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310

Böcker S, Letzel MC, Lipták Z, Pervukhin A (2009) SIRIUS: Decomposing isotope patterns for metabolite identification. Bioinformatics 25:218–224. https://doi.org/10.1093/bioinformatics/btn603

Bode HB, Bethe B, Höfs R, Zeeck A (2002) Big effects from small changes: Possible ways to explore nature's chemical diversity. ChemBioChem 3:619–627. https://doi.org/10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9

Bown L, Altowairish MS, Fyans JK, Bignell DRDD (2016) Production of the *Streptomyces scabies* coronafacoyl phytotoxins involves a novel biosynthetic pathway with an F420-dependent oxidoreductase and a short-chain dehydrogenase/reductase. Mol Microbiol 101:122–135. https://doi.org/10.1111/mmi.13378

Bown L, Li Y, Berrué F, et al (2017) Coronafacoyl phytotoxin biosynthesis and evolution in the common scab pathogen *Streptomyces scabiei*. Appl Environ Microbiol 83:1–15. https://doi.org/10.1128/AEM.01169-17

Braesel J, Lee JH, Arnould B, et al (2019) Diazaquinomycin biosynthetic gene clusters from marine and freshwater actinomycetes. J Nat Prod 82:937–946. https://doi.org/10.1021/acs.jnatprod.8b01028

Bukhalid RA, Takeuchi T, Labeda D, Loria R (2002) Horizontal transfer of the plant virulence gene, *nec1*, and flanking sequences among genetically distinct *Streptomyces* strains in the diastatochromogenes cluster. Appl Environ Microbiol 68:738–744. https://doi.org/10.1128/AEM.68.2.738-744.2002

Chalupowicz L, Barash I, Schwartz M, et al (2006) Comparative anatomy of gall development on *Gypsophila paniculata* induced by bacteria with different mechanisms of pathogenicity. Planta 224:429–437. https://doi.org/10.1007/s00425-006-0229-9

Chambers MC, Maclean B, Burke R, et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920. https://doi.org/10.1038/nbt.2377

Chan YA, Podevels AM, Kevany BM, Thomas MG (2009) Biosynthesis of polyketide synthase extender units. Nat Prod Rep 26:90–114. https://doi.org/10.1039/b801658p

Chen CW, Huang CH, Lee HH, et al (2002) Once the circle has been broken: Dynamics and evolution of *Streptomyces* chromosomes. Trends Genet 18:522–529. https://doi.org/10.1016/S0168-9525(02)02752-X

Cheng Z (2018) Regulation of coronafacoyl phytotoxin production in the potato common scab pathogen *Streptomyces scabies*. Memorial University of Newfoundland

Cheng Z, Bown L, Piercey B, Bignell DRD (2019) Positive and negative regulation of the virulence-associated coronafacoyl phytotoxin in the potato common scab pathogen *Streptomyces scabies*. Mol Plant-Microbe Interact 32:1348-1359. https://doi.org/10.1094/MPMI-03-19-0070-R

Cheng Z, Bown L, Tahlan K, Bignell DRD (2015) Regulation of coronafacoyl phytotoxin production by the PAS-LuxR family regulator CfaR in the common scab pathogen *Streptomyces scabies*. PLoS One 10:e0122450. https://doi.org/10.1371/journal.pone.0122450

Chin-A-Woeng TFC, Thomas-Oates JE, Lugtenberg BJJ, Bloemberg G V (2001) Introduction of the *phzH* gene of *Pseudomonas chlororaphis* pcl1391 extends the range of

biocontrol ability of phenazine-1-carboxylic acid-producing *Pseudomonas* spp. strains. Mol Plant-Microbe Interact 14:1006–1015. https://doi.org/10.1094/mpmi.2001.14.8.1006

Cimermancic P, Medema MH, Claesen J, et al (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421. https://doi.org/10.1016/j.cell.2014.06.034

Crone WJK, Leeper FJ, Truman AW (2012) Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. Chem Sci 3:3516–3521. https://doi.org/10.1039/c2sc21190d

Crone WJK, Vior NM, Santos-Aberturas J, et al (2016) Dissecting bottromycin biosynthesis using comparative untargeted metabolomics. Angew Chemie Int Ed 55:9639–9643. https://doi.org/10.1002/anie.201604304

Cruz-Morales P, Kopp JF, Martínez-Guerrero C, et al (2016) Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. Genome Biol Evol 8:1906–1916. https://doi.org/10.1093/gbe/evw125

Cummings M, Breitling R, Takano E (2014) Steps towards the synthetic biology of polyketide biosynthesis. FEMS Microbiol Lett 351:116–125. https://doi.org/10.1111/1574-6968.12365

Cuthbertson L, Nodwell JR (2013) The TetR family of regulators. Microbiol Mol Biol Rev 77:440–475. https://doi.org/10.1128/mmbr.00018-13

da Silva RR, Wang M, Nothias LF, et al (2018) Propagating annotations of molecular networks using in silico fragmentation. PLoS Comput Biol 14:1–26.

https://doi.org/10.1371/journal.pcbi.1006089

Daniel-Ivad M, Hameed N, Tan S, et al (2017) An engineered allele of *afsQ1* facilitates the discovery and investigation of cryptic natural products. ACS Chem Biol 12:628–634. https://doi.org/10.1021/acschembio.6b01002

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147

Dees MW, Wanner LA (2012) In search of better management of potato common scab. Potato Res 55:249–268. https://doi.org/10.1007/s11540-012-9206-9

Dejong CA, Chen GM, Li H, et al (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat Chem Biol 12:1007–1014. https://doi.org/10.1038/nchembio.2188

Delaney SM, Mavrodi D V., Bonsall RF, Thomashow LS (2001) *phzO*, a gene for biosynthesis of 2-hydroxylated phenazine compounds in *Pseudomonas aureofaciens* 30-84. J Bacteriol 183:318–327. https://doi.org/10.1128/JB.183.1.318-327.2001

Desriac F, Jégou C, Balnois E, et al (2013) Antimicrobial peptides from marine proteobacteria. Mar Drugs 11:3632–3660. https://doi.org/10.3390/md11103632

Elliot MA, Buttner MJ, Nodwell JR (2007) Myxobacteria. ASM Press, Washington, DC, USA

English AL, Boufridi A, Quinn RJ, Kurtböke DI (2017) Evaluation of fermentation conditions triggering increased antibacterial activity from a near-shore marine intertidal

environment-associated *Streptomyces* species. Synth Syst Biotechnol 2:28–38. https://doi.org/10.1016/j.synbio.2016.09.005

Ernst M, Kang K Bin, Caraballo-Rodríguez AM, et al (2019) Molnetenhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. Metabolites 9:144. https://doi.org/10.3390/metabo9070144

Ferguson NL, Peña-Castillo L, Moore MA, et al (2016) Proteomics analysis of global regulatory cascades involved in clavulanic acid production and morphological development in *Streptomyces clavuligerus*. J Ind Microbiol Biotechnol 43:537–555. https://doi.org/10.1007/s10295-016-1733-y

Foulston LC, Bibb MJ (2010) Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. Proc Natl Acad Sci U S A 107:13461–13466. https://doi.org/10.1073/pnas.1008285107

Francenia Santos-Sánchez N, Salas-Coronado R, Hernández-Carlos B, Villanueva-Cañongo C (2019) Shikimic acid pathway in biosynthesis of phenolic compounds. In: Plant physiological aspects of phenolic compounds. IntechOpen, pp 1–15

Fu J, Teucher M, Anastassiadis K, et al (2010) A recombineering pipeline to make conditional targeting constructs. In: Methods in enzymology, 2nd edn. Elsevier Inc., pp 125–144

Fujita MJ, Kimura N, Yokose H, Otsuka M (2012) Heterologous production of bisucaberin using a biosynthetic gene cluster cloned from a deep sea metagenome. Mol Biosyst 8:482–485. https://doi.org/10.1039/c1mb05431g

Fyans JK, Altowairish MS, Li Y, Bignell DRD (2015) Characterization of the coronatine-like phytotoxins produced by the common scab pathogen *Streptomyces scabies*. Mol plant-microbe Interact 28:443–54. https://doi.org/10.1094/MPMI-09-14-0255-R

Fyans JK, Bown L, Bignell DRD (2016) Isolation and characterization of plant-pathogenic *Streptomyces* species associated with common scab-infected potato tubers in Newfoundland. Phytopathology 106:123–131. https://doi.org/10.1094/PHYTO-05-15-0125-R

Gehrke EJ, Zhang X, Pimentel-Elardo SM, et al (2019) Silencing cryptic specialized metabolism in *Streptomyces* by the nucleoid-associated protein Lsr2. Elife 8:1–28. https://doi.org/10.7554/eLife.47691

Goering AW, McClure RA, Doroghazi JR, et al (2016) Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. ACS Cent Sci 2:99–108. https://doi.org/10.1021/acscentsci.5b00331

Gokulan K, Khare S, Cerniglia C (2014) Metabolic pathways: Production of secondary metabolites of bacteria. In: Encyclopedia of Food Microbiology. Elsevier, pp 561–569

Gomez-Escribano JP, Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4:207–215. https://doi.org/10.1111/j.1751-7915.2010.00219.x

Gomez-Escribano JP, Song L, Bibb MJ, Challis GL (2012a) Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. Chem Sci 3:3522–3525. https://doi.org/10.1039/C2SC21183A

Gomez-Escribano JP, Song L, Fox DJ, et al (2012b) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. Chem Sci 3:2716. https://doi.org/10.1039/c2sc20410j

Gosse JT, Ghosh S, Sproule A, et al (2019) Whole genome sequencing and metabolomic study of cave *Streptomyces* isolates ICC1 and ICC4. Front Microbiol 10:1–12. https://doi.org/10.3389/fmicb.2019.01020

Grininger M (2020) The role of the iterative modules in polyketide synthase evolution. Proc Natl Acad Sci 117:8680–8682. https://doi.org/10.1073/pnas.2004190117

Gross H (2007) Strategies to unravel the function of orphan biosynthesis pathways: Recent examples and future prospects. Appl Microbiol Biotechnol 75:267–277. https://doi.org/10.1007/s00253-007-0900-5

Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep 26:1408–1446. https://doi.org/10.1039/b817075b

Guan D, Grau BL, Clark CA, et al (2012) Evidence that thaxtomin C is a pathogenicity determinant of *Streptomyces ipomoeae* , the causative agent of *Streptomyces* soil rot disease of sweet potato. Mol Plant-Microbe Interact 25:393–401. https://doi.org/10.1094/MPMI-03-11-0073

Guerinot M Lou (1994) Microbial iron transport. Annu Rev Microbiol 48:743–772. https://doi.org/10.1146/annurev.mi.48.100194.003523

Guo F, Xiang S, Li L, et al (2015) Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. Metab Eng 28:134–142.

https://doi.org/10.1016/j.ymben.2014.12.006

Guo S, Wang Y, Bilal M, et al (2020) Microbial synthesis of antibacterial phenazine-1,6-dicarboxylic acid and the role of PhzG in *Pseudomonas chlororaphis* GP72AN. J Agric Food Chem 68:2373–2380. https://doi.org/10.1021/acs.jafc.9b07657

Guo S, Wang Y, Dai B, et al (2017) PhzA, the shunt switch of phenazine-1,6-dicarboxylic acid biosynthesis in *Pseudomonas chlororaphis* HT66. Appl Microbiol Biotechnol 101:7165–7175. https://doi.org/10.1007/s00253-017-8474-3

Gust B, Challis GL, Fowler K, et al (2003a) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100:1541–1546. https://doi.org/10.1073/pnas.0337542100

Gust B, Chandra G, Jakimowicz D, et al (2004) λ Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. In: Advances in applied microbiology. pp 107–128

Gust B, O'Rourke S, Bird N, et al (2003b) Recombineering in *Streptomyces coelicolor*. In: FEMS Online Protoc. http://www.jic.ac.uk/science/molmicro/summerschool2007/material/Recombineering.pdf

Guttenberger N, Blankenfeldt W, Breinbauer R (2017) Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. Bioorganic Med Chem 25:6149–6166. https://doi.org/10.1016/j.bmc.2017.01.002

Guzmán-Chávez F, Zwahlen RD, Bovenberg RAL, Driessen AJM (2018) Engineering of the filamentous fungus Penicillium chrysogenum as cell factory for natural products. Front Microbiol 9:2768. https://doi.org/10.3389/fmicb.2018.02768

Hannigan GD, Prihoda D, Palicka A, et al (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res 47:e110–e110. https://doi.org/10.1093/nar/gkz654

Hanson JR (2003) Natural products: the secondary metabolites, 1st editio. The Royal Society of Chemistry

Harir M, Bendif H, Bellahcene M, Fortas and Rebecca Pogni Z (2018) *Streptomyces* Secondary Metabolites. In: Basic Biology and Applications of Actinobacteria. IntechOpen

Haydock SF, Appleyard AN, Mironenko T, et al (2005) Organization of the biosynthetic gene cluster for the macrolide concanamycin A in *Streptomyces neyagawaensis* ATCC 27449. Microbiology 151:3161–3169. https://doi.org/10.1099/mic.0.28194-0

He H, Ye L, Li C, et al (2018) SbbR/SbbA, an important ArpA/AfsA-like system, regulates milbemycin production in *Streptomyces bingchenggensis*. Front Microbiol 9:1–17. https://doi.org/10.3389/fmicb.2018.01064

Healy FG, Krasnoff SB, Wach M, et al (2002) Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by *Streptomyces acidiscabies*. J Bacteriol 184:2019–2029. https://doi.org/10.1128/JB.184.7.2019-2029.2002

Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2015) Lasso peptides: An intriguing class of bacterial natural products. Acc Chem Res 48:1909–1919. https://doi.org/10.1021/acs.accounts.5b00156

Herrmann KM (1995) The shikimate pathway as an entry to aromatic secondary metabolism. Plant Physiol 107:7–12. https://doi.org/10.1104/pp.107.1.7

Hertweck C (2009) The biosynthetic logic of polyketide diversity. Angew Chemie - Int Ed 48:4688–4716. https://doi.org/10.1002/anie.200806121

Hibbing ME, Fuqua C, Parsek MR, Peterson SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8:15–25. https://doi.org/10.1038/nrmicro2259

Hiltunen LH, Ojanperä T, Kortemaa H, et al (2009) Interactions and biocontrol of pathogenic *Streptomyces* strains co-occurring in potato scab lesions. J Appl Microbiol 106:199–212. https://doi.org/10.1111/j.1365-2672.2008.03992.x

Hodgson DA (2000) Primary metabolism and its control in streptomycetes: A most unusual group of bacteria. In: Advances in Microbial Physiology. pp 47–238

Horinouchi S (2007) Mining and polishing of the treasure trove in the bacterial genus *Streptomyces*. Biosci Biotechnol Biochem 71:283–299. https://doi.org/10.1271/bbb.60627

Horinouchi S (2003) AfsR as an integrator of signals that are sensed by multiple serine/threonine kinases in Streptomyces coelicolor A3(2). J Ind Microbiol Biotechnol 30:462–467. https://doi.org/10.1007/s10295-003-0063-z

Hou B, Lin Y, Wu H, et al (2018) The novel transcriptional regulator LmbU promotes lincomycin biosynthesis through regulating expression of its target genes in *Streptomyces lincolnensis*. J Bacteriol 200:1–17. https://doi.org/10.1128/JB.00447-17

Huo L, Rachid S, Stadler M, et al (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. Chem Biol 19:1278–1287. https://doi.org/10.1016/j.chembiol.2012.08.013

Iavicoli A, Boutet E, Buchala A, Métraux J (2003) Induced systemic resistance in arabidopsis thaliana in response to root inoculation with *Pseudomonas fluorescens* CHA0. Mol Plant-Microbe Interact 16:851–858. https://doi.org/10.1094/MPMI.2003.16.10.851

Ibrahim A, Yang L, Johnston C, et al (2012) Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. Proc Natl Acad Sci U S A 109:19196–19201. https://doi.org/10.1073/pnas.1206376109

Ikeda H, Ishikawa J, Hanamoto A, et al (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531. https://doi.org/10.1038/nbt820

Ikeda H, Kotaki H, Omura S (1987) Genetic studies of avermectin biosythesis in *Streptomyces avermitilis*. J Bacteriol 169:5615–5621. https://doi.org/10.1128/jb.169.12.5615-5621.1987

Inahashi Y, Zhou S, Bibb MJ, et al (2017) Watasemycin biosynthesis in *Streptomyces venezuelae*: thiazoline C-methylation by a type B radical-SAM methylase homologue. Chem Sci 8:2823–2831. https://doi.org/10.1039/c6sc03533g

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23–28. https://doi.org/10.1016/0378-1119(90)90336-P

Itaru T, Myco U, Hideaki K, et al (2014) Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: Application to sequenced genomes of Aspergillus and ten other filamentous fungal species. DNA Res 21:447–457. https://doi.org/10.1093/dnares/dsu010

Jiang G, Zhang Y, Powell MM, et al (2018a) High-yield production of herbicidal thaxtomins and thaxtomin analogs in a nonpathogenic *Streptomyces* strain. Appl Environ Microbiol 84:1–12. https://doi.org/10.1128/AEM.00164-18

Jiang G, Zuo R, Zhang Y, et al (2018b) One-pot biocombinatorial synthesis of herbicidal thaxtomins. ACS Catal 8:10761–10768. https://doi.org/10.1021/acscatal.8b03317

Jiang M, Yin M, Wu S, et al (2017) GdmRIII, a TetR family transcriptional regulator, controls geldanamycin and elaiophylin biosynthesis in *Streptomyces autolyticus* CGMCC0516. Sci Rep 7:4803. https://doi.org/10.1038/s41598-017-05073-x

Jiang M, Zhang H (2016) Engineering the shikimate pathway for biosynthesis of molecules with pharmaceutical activities in *E. coli*. Curr Opin Biotechnol 42:1–6. https://doi.org/10.1016/j.copbio.2016.01.016

Johnson EG, Joshi M V., Gibson DM, Loria R (2007) Cello-oligosaccharides released from host plants induce pathogenicity in scab-causing *Streptomyces* species. Physiol Mol Plant Pathol 71:18–25. https://doi.org/10.1016/j.pmpp.2007.09.003

Johnson EG, Krasnoff SB, Bignell DRD, et al (2009) 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. Mol Microbiol 73:409–418. https://doi.org/10.1111/j.1365-2958.2009.06780.x

Johnston CW, Skinnider MA, Wyatt MA, et al (2015) An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. Nat Commun 6:8421. https://doi.org/10.1038/ncomms9421

Joshi M, Rong X, Moll S, et al (2007a) *Streptomyces turgidiscabies* secretes a novel

virulence protein, Nec1, which facilitates infection. Mol plant-microbe Interact 20:599–608. https://doi.org/10.1094/MPMI-20-6-0599

Joshi M V., Bignell DRD, Johnson EG, et al (2007b) The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and virulence in *Streptomyces scabies*. Mol Microbiol 66:633–642. https://doi.org/10.1111/j.1365-2958.2007.05942.x

Kadi N, Song L, Challis GL (2008) Bisucaberin biosynthesis: An adenylating domain of the BibC multi-enzyme catalyzes cyclodimerization of N-hydroxy-N-succinylcadaverine. Chem Commun 5119–5121. https://doi.org/10.1039/b813029a

Kautsar SA, Blin K, Shaw S, et al (2019) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458. https://doi.org/10.1093/nar/gkz882

Keatinge-Clay AT (2016) The structural relationship between iterative and modular PKSs. Cell Chem Biol 23:540–542. https://doi.org/10.1016/j.chembiol.2016.05.005

Kieser T, Bibb M, Buttner M, Chater K (2000) Practical *Streptomyces* genetics. John Innes Foundation, Norwich, England

Kinashi H (2011) Giant linear plasmids in *Streptomyces*: A treasure trove of antibiotic biosynthetic clusters. J Antibiot (Tokyo) 64:19–25. https://doi.org/10.1038/ja.2010.146

Kinashi H, Someno K, Sakaguchi K (1984) Isolation and characterization of concanamycins A, B and C. J Antibiot (Tokyo) 37:1333–1343. https://doi.org/10.7164/antibiotics.37.1333

King RR, Calhoun LA (2009) The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. Phytochemistry 70:833–841.

https://doi.org/10.1016/j.phytochem.2009.04.013

King RR, Harold Lawrence C, Gray JA (2001) Herbicidal properties of the thaxtomin group of phytotoxins. J Agric Food Chem 49:2298–2301. https://doi.org/10.1021/jf0012998

King RR, Lawrence CH (1996) Characterization of new thaxtomin A analogues generated in vitro by *Streptomyces scabies*. J Agric Food Chem 44:1108–1110. https://doi.org/10.1021/jf950243o

King RR, Lawrence CH, Calhoun LA, Ristaino JB (1994) Isolation and characterization of thaxtomin-type phytotoxins associated with *Streptomyces ipomoeae*. J Agric Food Chem 42:1791–1794. https://doi.org/10.1021/jf00044a042

King RR, Lawrence CH, Clark MC, Calhoun LA (1989) Isolation and characterization of phytotoxins associated with *Streptomyces scabies*. J Chem Soc Chem Commun 13:849–850. https://doi.org/10.1039/c39890000849

Kinoshita H, Ipposhi H, Okamoto S, et al (1997) Butyrolactone autoregulator receptor protein (BarA) as a transcriptional regulator in *Streptomyces virginiae*. J Bacteriol 179:6986–6993. https://doi.org/10.1128/jb.179.22.6986-6993.1997

Kitani S, Miyamoto KT, Takamatsu S, et al (2011) Avenolide, a *Streptomyces* hormone controlling antibiotic production in *Streptomyces avermitilis*. Proc Natl Acad Sci U S A 108:16410–5. https://doi.org/10.1073/pnas.1113908108

Kobayashi Y, Ichioka M, Hirose T, et al (2010) Bottromycin derivatives: Efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. Bioorganic Med Chem Lett 20:6116–6120. https://doi.org/10.1016/j.bmcl.2010.08.037

Kodani S, Bicz J, Song L, et al (2013) Structure and biosynthesis of scabichelin, a novel tris-hydroxamate siderophore produced by the plant pathogen *Streptomyces scabies* 87.22. Org Biomol Chem 11:4686–4694. https://doi.org/10.1039/c3ob40536b

Komatsu M, Komatsu K, Koiwai H, et al (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol 2:384–396. https://doi.org/10.1021/sb3001003

Kramer J, Özkaya Ö, Kümmerli R (2020) Bacterial siderophores in community and host interactions. Nat Rev Microbiol 18:152–163. https://doi.org/10.1038/s41579-019-0284-4

Lambrecht M, Okon Y, Vande Broek A, Vanderleyden J (2000) Indole-3-acetic acid: a reciprocal signalling molecule in bacteria–plant interactions. Trends Microbiol 8:298–300. https://doi.org/10.1016/S0966-842X(00)01732-7

Laponogov I, Sadawi N, Galea D, et al (2018) ChemDistiller: An engine for metabolite annotation in mass spectrometry. Bioinformatics 34:2096–2102. https://doi.org/10.1093/bioinformatics/bty080

Laursen JB, Nielsen J (2004) Phenazine natural products: biosynthesis, synthetic analogues, and biological activity. Chem Rev 104:1663–1685. https://doi.org/10.1021/cr020473j

Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: Rates of change and exchange. J Mol Evol 44:383–397. https://doi.org/10.1007/PL00006158

Lazarovits G, Hill J, King RR, Calhoun LA (2004) Biotransformation of the *Streptomyces scabies* phytotoxin thaxtomin A by the fungus *Aspergillus niger*. Can J Microbiol 50:121–126. https://doi.org/10.1139/w03-110

Lemos ML, Dopazo CP, Toranzo AE, Barja JL (1991) Competitive dominance of antibiotic-producing marine bacteria in mixed cultures. J Appl Bacteriol 71:228–232. https://doi.org/10.1111/j.1365-2672.1991.tb04452.x

Lerat S, Simao-Beaunoir A-M, Wu R, et al (2010) Involvement of the plant polymer suberin and the disaccharide cellobiose in triggering thaxtomin A biosynthesis, a phytotoxin produced by the pathogenic agent *Streptomyces scabies*. Phytopathology 100:91–96. https://doi.org/10.1094/PHYTO-100-1-0091

Li Q-A, Mavrodi D V., Thomashow LS, et al (2011) Ligand binding induces an ammonia channel in 2-amino-2-desoxyisochorismate (ADIC) synthase PhzE. J Biol Chem 286:18213–18221. https://doi.org/10.1074/jbc.M110.183418

Li X, Wang J, Li S, et al (2015) ScbR- and ScbR2-mediated signal transduction networks coordinate complex physiological responses in *Streptomyces coelicolor*. Sci Rep 5:14831. https://doi.org/10.1038/srep14831

Li Y, Liu J, Adekunle D, et al (2019a) TxtH is a key component of the thaxtomin biosynthetic machinery in the potato common scab pathogen *Streptomyces scabies*. Mol Plant Pathol 20:1379–1393. https://doi.org/10.1111/mpp.12843

Li Y, Liu J, Díaz-Cruz G, et al (2019b) Virulence mechanisms of plant-pathogenic *Streptomyces* species: an updated review. Microbiology 165:1025–1040. https://doi.org/10.1099/mic.0.000818

Lin Y, Kieser HM, Hopwood DA, Chen CW (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol Microbiol 10:923–933. https://doi.org/10.1111/j.1365-2958.1993.tb00964.x

Loria R (1995) Differential production of thaxtomins by pathogenic *Streptomyces* species in vitro. Phytopathology 85:537. https://doi.org/10.1094/Phyto-85-537

Loria R, Bignell DRD, Moll S, et al (2008) Thaxtomin biosynthesis: The path to plant pathogenicity in the genus *Streptomyces*. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 94:3–10. https://doi.org/10.1007/s10482-008-9240-4

Loria R, Bukhalid RA, Fry BA, King RRR (1997) Plant pathogenicity in the genus *Streptomyces*. Plant Dis 81:836–846. https://doi.org/10.1094/PDIS.1997.81.8.836

Loria R, Kers J, Joshi M (2006) Evolution of plant pathogenicity in *Streptomyces*. Annu Rev Phytopathol 44:469–487. https://doi.org/10.1146/annurev.phyto.44.032905.091147

Luo M, Fadeev E a, Groves JT (2005) Mycobactin-mediated iron acquisition within macrophages. Nat Chem Biol 1:149–153. https://doi.org/10.1038/nchembio717

MacNeil DJ, Gewain KM, Ruby CL, et al (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111:61–68. https://doi.org/10.1016/0378-1119(92)90603-M

Martínez-Núñez MA, López VEL y (2016) Nonribosomal peptides synthetases and their applications in industry. Sustain Chem Process 4:1–8. https://doi.org/10.1186/s40508-016-0057-6

Mast Y, Guezguez J, Handel F, Schinko E (2015) A complex signaling cascade governs pristinamycin biosynthesis in *Streptomyces pristinaespiralis*. Appl Environ Microbiol 81:6621–6636. https://doi.org/10.1128/AEM.00728-15

Mavrodi DV, Ksenzenko VN, Chatuev BM, et al (1997) Structural and functional

organization of *Pseudomonas fluorescens* genes encoding enzymes of phenazine-1-carboxylic acid biosynthesis. Mol Biol (Mosk) 31:74–80

Mavrodi D V., Ksenzenko VN, Bonsall RF, et al (1998) A seven-gene locus for synthesis of phenazine-1-carboxylic acid by *Pseudomonas fluorescens* 2-79. J Bacteriol 180:2541–2548

Mavrodi D V, Blankenfeldt W, Thomashow LS (2006) Phenazine compounds in fluorescent *Pseudomonas* spp. biosynthesis and regulation. Annu Rev Phytopathol 44:417–445. https://doi.org/10.1146/

Mavrodi D V, Bonsall RF, Delaney SM, et al (2001) Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from *Pseudomonas aeruginosa* PAO1. J Bacteriol 183:6454–6465. https://doi.org/10.1128/JB.183.21.6454

Mazzola M, Cook RJ, Thomashow LS, et al (1992) Contribution of phenazine antibiotic biosynthesis to the ecological competence of fluorescent pseudomonads in soil habitats. Appl Environ Microbiol 58:2616–2624. https://doi.org/10.1128/AEM.58.8.2616-2624.1992

McCormick JR, Flärdh K (2012) Signals and regulators that govern *Streptomyces* development. FEMS Microbiol Rev 36:206–231. https://doi.org/10.1111/j.1574-6976.2011.00317.x

McDonald M, Mavrodi D V., Thomashow LS, Floss HG (2001) Phenazine biosynthesis in *Pseudomonas fluorescens*: Branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. J Am Chem Soc 123:9459–9460. https://doi.org/10.1021/ja011243+

Miethke M, Marahiel MA (2007) Siderophore-based iron acquisition and pathogen control.

Microbiol Mol Biol Rev 71:413–451. https://doi.org/10.1128/MMBR.00012-07

Mingyar E, Feckova L, Novakova R, et al (2014) A γ-butyrolactone autoregulator-receptor system involved in the regulation of auricin production in *Streptomyces aureofaciens* CCM 3239. Appl Microbiol Biotechnol 99:309–325. https://doi.org/10.1007/s00253-014-6057-0

Mishra AK, Choi J, Choi SJ, Baek KH (2017) Cyclodipeptides: An overview of their biosynthesis and biological activity. Molecules 22:1–13. https://doi.org/10.3390/molecules22101796

Mitchell RE (1984) A naturally-occurring structural analogue of the phytotoxin coronatine. Phytochemistry 23:791–793. https://doi.org/10.1016/S0031-9422(00)85028-1

Mohimani H, Gurevich A, Mikheenko A, et al (2017) Dereplication of peptidic natural products through database search of mass spectra. Nat Chem Biol 13:30–37. https://doi.org/10.1038/nchembio.2219

Mohimani H, Kersten RD, Liu WT, et al (2014a) Automated genome mining of ribosomal peptide natural products. ACS Chem Biol 9:1545–1551. https://doi.org/10.1021/cb500199h

Mohimani H, Liu W-T, Mylne JS, et al (2011) Cycloquest: Identification of cyclopeptides via database search of their mass spectra against genome databases. J Proteome Res 10:5575–5575. https://doi.org/10.1021/pr201019q

Mohimani H, Liu WT, Kersten RD, et al (2014b) NRPquest: Coupling mass spectrometry and genome mining for nonribosomal peptide discovery. J Nat Prod 77:1902–1909. https://doi.org/10.1021/np500370c

Moore BS, Hertweck C (2002) Biosynthesis and attachment of novel bacterial polyketide

synthase starter units. Nat Prod Rep 19:70–99. https://doi.org/10.1039/b003939j

Moore BS, Hopke JN (2001) Discovery of a new bacterial polyketide biosynthetic pathway. ChemBioChem 2:35–38. https://doi.org/10.1002/1439-7633(20010105)2:1<35::AID-CBIC35>3.3.CO;2-T

Morel MA, Castro-sowinski S (2013) The complex molecular signaling network in microbe–plant interaction. In: Arora NK (ed) Plant microbe symbiosis: fundamentals and advances. Springer India, New Delhi, pp 169–199

Mosunova O, Navarro-Muñoz JC, Collemare J (2020) The biosynthesis of fungal secondary metabolites: from fundamentals to biotechnological applications. In: Reference Module in Life Sciences. Elsevier, pp 1–19

Myronovskyi M, Luzhetskyy A (2016) Native and engineered promoters in natural product discovery. Nat Prod Rep 33:1006–1019. https://doi.org/10.1039/C6NP00002A

Natsume M, Komiya M, Koyanagi F, et al (2005) Phytotoxin produced by *Streptomyces* sp. causing potato russet scab in Japan. J Gen Plant Pathol 71:364–369. https://doi.org/10.1007/s10327-005-0211-6

Natsume M, Ryu R, Abe H (1996) Production of phytotoxins, concanamycins A and B by *Streptomyces* spp. causing potato scab. Japanese J Phytopathol 62:411–413. https://doi.org/10.3186/jjphytopath.62.411

Natsume M, Taki M, Tashiro N, Abe H (2001) Phytotoxin production and aerial mycelium formation by *Streptomyces scabies* and *S. acidiscabies* in vitro. J Gen Plant Pathol 67:299–302

Natsume M, Tashiro N, Doi A, et al (2017) Effects of concanamycins produced by *Streptomyces scabies* on lesion type of common scab of potato. J Gen Plant Pathol 83:78–82. https://doi.org/10.1007/s10327-017-0696-9

Natsume M, Yamada A, Tashiro N, Abe H (1998) Differential production of the phytotoxins thaxtomin A and concanamycins A and B by potato common scab-causing *Streptomyces* spp. Japanese J Phytopathol 64:202–204. https://doi.org/10.3186/jjphytopath.64.202

Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, et al (2020) A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68. https://doi.org/10.1038/s41589-019-0400-9

Neilands JB (1995) Siderophores: Structure and function of microbial iron transport compounds. J Biol Chem 270:26723–26726. https://doi.org/10.1074/jbc.270.45.26723

Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. J Nat Prod 70:461–477. https://doi.org/10.1021/np068054v

Nguyen CT, Dhakal D, Pham VTT, et al (2020) Recent advances in strategies for activation and discovery/Characterization of cryptic biosynthetic gene clusters in *Streptomyces*. Microorganisms 8:616. https://doi.org/10.3390/microorganisms8040616

Nguyen DD, Wu C-H, Moree WJ, et al (2013) MS/MS networking guided analysis of molecule and gene cluster families. Proc Natl Acad Sci 110:E2611–E2620. https://doi.org/10.1073/pnas.1303471110

Nigam PS, Singh A (2014) Metabolic Pathways: Production of Secondary Metabolites - Fungi, Second Edi. Elsevier

Nishida H, Ohnishi Y, Beppu T, Horinouchi S (2007) Evolution of γ-butyrolactone synthases and receptors in *Streptomyces*. Environ Microbiol 9:1986–1994. https://doi.org/10.1111/j.1462-2920.2007.01314.x

Nishimura K, Hosaka T, Tokuyama S, et al (2007) Mutations in rsmG, encoding a 16S rRNA methyltransferase, result in low-level streptomycin resistance and antibiotic overproduction in *Streptomyces coelicolor* A3(2). J Bacteriol 189:3876–3883. https://doi.org/10.1128/JB.01776-06

Niu G, Chater KF, Tian Y, et al (2016) Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp. FEMS Microbiol Rev. https://doi.org/10.1093/femsre/fuw012

Nivina A, Yuet KP, Hsu J, Khosla C (2019) Evolution and diversity of assembly-line polyketide synthases. Chem Rev 119:12524–12547. https://doi.org/10.1021/acs.chemrev.9b00525

Nkanga EJ, Hagedorn C (1978) Detection of antibiotic producing *Streptomyces* inhabiting forest soils. Antimicrob Agents Chemother 14:51–59. https://doi.org/10.1128/AAC.14.1.51

Nodwell JR (2014) Are you talking to me? A possible role for γ-butyrolactones in interspecies signalling. Mol Microbiol 94:483–485. https://doi.org/10.1111/mmi.12787

Nothias L-F, Petras D, Schmid R, et al (2020) Feature-based molecular networking in the GNPS analysis environment. Nat Methods 17:905–908. https://doi.org/10.1038/s41592-020-0933-6

Novakova R, Kutas P, Feckova L, Kormanec J (2010) The role of the TetR-family

transcriptional regulator Aur1R in negative regulation of the auricin gene cluster in *Streptomyces aureofaciens* CCM 3239. Microbiology 156:2374–2383. https://doi.org/10.1099/mic.0.037895-0

Ochi K, Okamoto S, Tozawa Y, et al (2004) Ribosome engineering and secondary metabolite production. Adv Appl Microbiol 56:155–184. https://doi.org/10.1016/S0065-2164(04)56005-7

Ortíz-Castro R, Contreras-Cornejo HA, Macías-Rodríguez L, López-Bucio J (2009) The role of microbial signals in plant growth and development. Plant Signal Behav 4:701–712. https://doi.org/10.4161/psb.4.8.9047

Ortiz-castro R, Díaz-pérez C, Martínez-trujillo M, et al (2011) Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. Proc Natl Acad Sci U S A 108:7253–7258. https://doi.org/10.1073/pnas.1006740108

Otaka T, Kaji A (1976) Mode of action of bottromycin A2. Release of aminoacyl- or peptidyl-tRNA from ribosomes. J Biol Chem 251:2299–2306

Otaka T, Kaji A (1983) Mode of action of bottromycin A2: Effect of bottromycin A2 on polysomes. FEBS Lett 153:53–59. https://doi.org/10.1016/0014-5793(83)80118-5

Otaka T, Kaji A (1981) Mode of action of bottromycin A2: Effect on peptide bond formation. FEBS Lett 123:173–176. https://doi.org/10.1016/0014-5793(83)80118-5

Parsons JF, Calabrese K, Eisenstein E, Ladner JE (2003) Structure and mechanism of *Pseudomonas aeruginosa* PhzD, an isochorismatase from the phenazine biosynthetic pathway. Biochemistry 42:5684–5693. https://doi.org/10.1021/bi027385d

Parsons JF, Calabrese K, Eisenstein E, Ladner JE (2004) Structure of the phenazine biosynthesis enzyme PhzG. Acta Crystallogr Sect D Biol Crystallogr 60:2110–2113. https://doi.org/10.1107/S0907444904022474

Parsons JF, Greenhagen BT, Shi K, et al (2007) Structural and functional analysis of the pyocyanin biosynthetic protein PhzM from *Pseudomonas aeruginosa*. Biochemistry 46:1821–1828. https://doi.org/10.1021/bi6024403

Pierson III LS, Thomashow LS (1992) Cloning and Heterologous Expression of the Phenazine Biosynthetic Locus from *Pseudomonas aureofaciens* 30-84. Mol Plant-Microbe Interact 5:330–339. https://doi.org/10.1094/MPMI-5-330

Pierson LS, Gaffney T, Lam S, Gong F (1995) Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium *Pseudomonas aureofaciens* 30â€“84. FEMS Microbiol Lett 134:299–307. https://doi.org/10.1111/j.1574-6968.1995.tb07954.x

Pissowotzki K, Mansouri K, Piepersberg W (1991) Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes strELMB2N. MGG Mol Gen Genet 231:113–123. https://doi.org/10.1007/BF00293829

Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11:395. https://doi.org/10.1186/1471-2105-11-395

Ptashne M (2014) The chemistry of regulation of genes and other things. J Biol Chem 289:5417–5435. https://doi.org/10.1074/jbc.X114.547323

Rangaswamy V, Jiralerspong S, Parry R, Bender CL (1998a) Biosynthesis of the *Pseudomonas* polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. Proc Natl Acad Sci 95:15469–15474. https://doi.org/10.1073/pnas.95.26.15469

Rangaswamy V, Mitchell R, Ullrich M, Bender C (1998b) Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. J Bacteriol 180:3330–3338. https://doi.org/10.1128/JB.180.13.3330-3338.1998

Rawlinson C, Jones D, Rakshit S, et al (2020) Hierarchical clustering of MS/MS spectra from the firefly metabolome identifies new lucibufagin compounds. Sci Rep 10:6043. https://doi.org/10.1038/s41598-020-63036-1

Reen FJ, Romano S, Dobson ADW, O'Gara F (2015) The sound of silence: Activating silent biosynthetic gene clusters in marine microorganisms. Mar Drugs 13:4754–4783. https://doi.org/10.3390/md13084754

Remali J, Sarmin N 'Izzah M, Ng CL, et al (2017) Genomic characterization of a new endophytic *Streptomyces kebangsaanensis* identifies biosynthetic pathway gene clusters for novel phenazine antibiotic production. PeerJ 5:e3738. https://doi.org/10.7717/peerj.3738

Ren H, Wang B, Zhao H (2017) Breaking the silence: new strategies for discovering novel natural products. Curr Opin Biotechnol 48:21–27. https://doi.org/10.1016/j.copbio.2017.02.008

Robert-seilaniantz A, Navarro L, Bari R, Jones JDG (2007) Pathological hormone imbalances. Curr Opin Plant Biol 10:372–379. https://doi.org/10.1016/j.pbi.2007.06.003

Romano S, Jackson SA, Patry S, Dobson ADW (2018) Extending the "one strain many compounds" (OSMAC) principle to marine microorganisms. Mar Drugs 16:1–29. https://doi.org/10.3390/md16070244

Ruiz-Barba JL, Cathcart DP, Warner PJ, Jimenez-Diaz R (1994) Use of *Lactobacillus plantarum* LPCO10, a bacteriocin producer, as a starter culture in Spanish-style green olive fermentations. Appl Environ Microbiol 60:2059–2064. https://doi.org/10.1128/aem.60.6.2059-2064.1994

Sadeghi A, Karimi E, Dahaji PA, et al (2012) Plant growth promoting activity of an auxin and siderophore producing isolate of *Streptomyces* under saline soil conditions. World J Microbiol Biotechnol 28:1503–1509. https://doi.org/10.1007/s11274-011-0952-7

Sadeghi A, Soltani BM, Nekouei MK, et al (2014) Diversity of the ectoines biosynthesis genes in the salt tolerant *Streptomyces* and evidence for inductive effect of ectoines on their accumulation. Microbiol Res 169:699–708. https://doi.org/10.1016/j.micres.2014.02.005

Sambrook JF, Russell DW (2001) Molecular Cloning: A Laboratory Manual, Third. Cold Spring Harbor Laboratory Press, NY

Schlatter D, Fubuh A, Xiao K, et al (2009) Resource amendments influence density and competitive phenotypes of *Streptomyces* in soil. Microb Ecol 57:413–420. https://doi.org/10.1007/s00248-008-9433-4

Schofield LR, Anderson BF, Patchett ML, et al (2005) Substrate ambiguity and crystal structure of *Pyrococcus furiosus* 3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase: An ancestral 3-Deoxyald-2-ulosonate-phosphate synthase? Biochemistry 44:11950–11962. https://doi.org/10.1021/bi050577z

Seipke RF, Kaltenpoth M, Hutchings MI (2012) *Streptomyces* as symbionts: An emerging and widespread theme? FEMS Microbiol Rev 36:862–876. https://doi.org/10.1111/j.1574-6976.2011.00313.x

Seipke RF, Song L, Bicz J, et al (2011) The plant pathogen *Streptomyces scabies* 87-22 has a functional pyochelin biosynthetic pathway that is regulated by TetR- and AfsR-family proteins. Microbiology 157:2681–2693. https://doi.org/10.1099/mic.0.047977-0

Seki-Asano M, Okazaki T, Yamagishi M, et al (1994) Isolation and characterization of new 18-membered macrolides FD-891 and FD-892. J Antibiot (Tokyo) 47:1226–1233

Sélem-Mojica N, Aguilar C, Gutiérrez-García K, et al (2019) EvoMining reveals the origin and fate of natural product biosynthetic enzymes. Microb Genomics 5:1–16. https://doi.org/10.1099/mgen.0.000260

Senges CHR, Al-Dilaimi A, Marchbank DH, et al (2018) The secreted metabolome of *Streptomyces chartreusis* and implications for bacterial chemistry. Proc Natl Acad Sci U S A 115:2490–2495. https://doi.org/10.1073/pnas.1715713115

Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7:285–295. https://doi.org/10.1016/S1367-5931(03)00020-6

Shima J, Hesketh A, Okamoto S, et al (1996) Induction of actinorhodin production by *rpsL*

(encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). J Bacteriol 178:7276–7284. https://doi.org/10.1128/jb.178.24.7276-7284.1996

Shirling EB, Gottlieb D (1966) Methods for characterization of *Streptomyces* species. Int J Syst Bacteriol 16:313–340. https://doi.org/10.1099/00207713-16-3-313

Sidda JD, Corre C (2012) Gamma-butyrolactone and furan signaling systems in Streptomyces. In: Methods in enzymology, 1st edn. Elsevier Inc., pp 71–87

Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA (2017) PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res 45:W49–W54. https://doi.org/10.1093/nar/gkx320

Stratigopoulos G, Bate N, Cundliffe E (2004) Positive control of tylosin biosynthesis: Pivotal role of TylR. Mol Microbiol 54:1326–1334. https://doi.org/10.1111/j.1365-2958.2004.04347.x

Sun J, Kelemen GH, Fernández-Abalos JM, Bibb MJ (1999) Green fluorescent protein as a reporter for spatial and temporal gene expression in *Streptomyces coelicolor* A3(2). Microbiology 145:2221–2227. https://doi.org/10.1099/00221287-145-9-2221

Süssmuth RD, Mainz A (2017) Nonribosomal peptide synthesis—principles and prospects. Angew Chemie - Int Ed 56:3770–3821. https://doi.org/10.1002/anie.201609079

Takahashi Y, Nakashima T (2018) Actinomycetes, an inexhaustible source of naturally occurring antibiotics. Antibiotics 7:45. https://doi.org/10.3390/antibiotics7020045

Takano E (2006) γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic

production and differentiation. Curr Opin Microbiol 9:287–294. https://doi.org/10.1016/j.mib.2006.04.003

Takano E, Chakraburtty R, Nihira T, et al (2001) A complex role for the γ-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). Mol Microbiol 41:1015–1028. https://doi.org/10.1046/j.1365-2958.2001.02562.x

Thanapipatsiri A, Gomez-Escribano JP, Song L, et al (2016) Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. ChemBioChem 17:2189–2198. https://doi.org/10.1002/cbic.201600396

Thirumurugan D, Cholarajan A, Raja SSS, Vijayakumar R (2018) An introductory chapter: secondary metabolites. In: Secondary metabolites - sources and applications. InTech, pp 3–22

Tietz JI, Mitchell DA (2016) Using genomics for natural product structure elucidation. Curr Top Med Chem 16:1645–1694. https://doi.org/10.2174/1568026616666151012111439

Tokuyama T, Nishimori N, Shimada A, et al (1987) New classes of amidine, indolizidine and quinolizidine alkaloids from a poison-frog, *Dendrobates pumilio* (Dendrobatidae). Tetrahedron 43:643–652. https://doi.org/10.1016/S0040-4020(01)89998-1

Tschowri N, Schumacher MA, Schlimpert S, et al (2014) Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. Cell 158:1136–1147. https://doi.org/10.1016/j.cell.2014.07.022

Van Der Hooft JJJ, Wandy J, Barrett MP, et al (2016) Topic modeling for untargeted substructure exploration in metabolomics. Proc Natl Acad Sci U S A 113:13738–13743.

https://doi.org/10.1073/pnas.1608041113

Van Wezel GP, McDowall KJ (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. Nat Prod Rep 28:1311. https://doi.org/10.1039/c1np00003a

Ventura M, Canchaya C, Tauch A, et al (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 71:495–548. https://doi.org/10.1128/MMBR.00005-07

Vining LC (1990) Functions of secondary metabolites. Annu Rev Microbiol 44:395–427. https://doi.org/10.1146/annurev.mi.44.100190.002143

Vior NM, Cea-Torrescassana E, Eyles TH, et al (2020) Regulation of bottromycin biosynthesis involves an internal transcriptional start site and a cluster-situated modulator. Front Microbiol 11:1–16. https://doi.org/10.3389/fmicb.2020.00495

Wach MJ, Krasnoff SB, Loria R, Gibson DM (2007) Effect of carbohydrates on the production of thaxtomin A by *Streptomyces acidiscabies*. Arch Microbiol 188:81–88. https://doi.org/10.1007/s00203-007-0225-x

Waisvisz JM, van der Hoeven MG, van Peppen J, Zwennis WCM (1957) Bottromycin. I. A new sulfur-containing antibiotic. J Am Chem Soc 79:4520–4521. https://doi.org/10.1021/ja01573a072

Wang G, Inaoka T, Okamoto S, Ochi K (2009) A novel insertion mutation in *Streptomyces coelicolor* ribosomal S12 protein results in paromomycin resistance and antibiotic overproduction. Antimicrob Agents Chemother 53:1019–1026.

https://doi.org/10.1128/AAC.00388-08

Wang J, Wang W, Wang L, et al (2011) A novel role of ’pseudo’γ-butyrolactone receptors in controlling γ-butyrolactone biosynthesis in *Streptomyces*. Mol Microbiol 82:236–250. https://doi.org/10.1111/j.1365-2958.2011.07811.x

Wang L, Tahlan K, Kaziuk TL, et al (2004) Transcriptional and translational analysis of the ccaR gene from *Streptomyces clavuligerus*. Microbiology 150:4137–4145. https://doi.org/10.1099/mic.0.27245-0

Wang M, Carver JJ, Phelan V V., et al (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828–837. https://doi.org/10.1038/nbt.3597

Wang W, Li X, Wang J, et al (2013) An engineered strong promoter for Streptomycetes. Appl Environ Microbiol 79:4484–4492. https://doi.org/10.1128

Wanner LA (2009) A patchwork of *Streptomyces* species isolated from potato common scab lesions in North America. Am J Potato Res 86:247–264. https://doi.org/10.1007/s12230-009-9078-y

Wanner LA, Kirk WW, Qu XS (2014) Field efficacy of nonpathogenic *Streptomyces* species against potato common scab. J Appl Microbiol 116:123–133. https://doi.org/10.1111/jam.12336

Ward AC, Bora N (2006) Diversity and biogeography of marine actinobacteria. Curr Opin Microbiol 9:279–286. https://doi.org/10.1016/j.mib.2006.04.004

Weber T, Kim HU (2016) The secondary metabolite bioinformatics portal: Computational

tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol 1:69–79. https://doi.org/10.1016/j.synbio.2015.12.002

Weller DM, Mavrodi D V, van Pelt JA, et al (2012) Induced systemic resistance in Arabidopsis thaliana against *Pseudomonas syringae* pv. tomato by 2,4-diacetylphloroglucinol-producing *Pseudomonas fluorescens*. Phytopathology 102:403–412. https://doi.org/10.1094/PHYTO-08-11-0222

Willey JM, Gaskell AA (2011) Morphogenetic signaling molecules of the streptomycetes. Chem Rev 111:174–187. https://doi.org/10.1021/cr1000404

Winn M, Francis D, Micklefield J (2018) De novo biosynthesis of "non-natural" thaxtomin phytotoxins. Angew Chemie Int Ed 57:6830–6833. https://doi.org/10.1002/anie.201801525

Xie C, Deng J-J, Wang H-X (2015) Identification of AstG1, A LAL family regulator that positively controls ansatrienins production in *Streptomyces* sp. XZQH13. Curr Microbiol 70:859–864. https://doi.org/10.1007/s00284-015-0798-6

Xing A, Tian J, Liu C, et al (2010) Three new monoterpene alkaloids and a new caffeic acid ester from *Incarvillea mairei* var. multifoliolata. Helv Chim Acta 93:718–723. https://doi.org/10.1002/hlca.200900284

Xu G, Wang J, Wang L, et al (2010) "Pseudo" γ-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. J Biol Chem 285:27440–27448. https://doi.org/10.1074/jbc.M110.143081

Xu G, Yang S (2019) Regulatory and evolutionary roles of pseudo γ-butyrolactone receptors in antibiotic biosynthesis and resistance. Appl Microbiol Biotechnol 103:9373–9378.

https://doi.org/10.1007/s00253-019-10219-0

Xu J, Tozawa Y, Lai C, et al (2002) A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). Mol Genet Genomics 268:179–189. https://doi.org/10.1007/s00438-002-0730-1

Yamanaka K, Reynolds KA, Kersten RD, et al (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci U S A 111:1957–62. https://doi.org/10.1073/pnas.1319584111

Yang D, Park SY, Park YS, et al (2020) Metabolic engineering of *Escherichia coli* for natural product biosynthesis. Trends Biotechnol 38:745–765. https://doi.org/10.1016/j.tibtech.2019.11.007

Yaxley AM (2009) Study of the complete genome sequence of *Streptomyces scabies* (or scabiei) 87.22. University of Warwick

Zaburannyi N, Rabyk M, Ostash B, et al (2014) Insights into naturally minimised *Streptomyces albus* J1074 genome. BMC Genomics 15:97. https://doi.org/10.1186/1471-2164-15-97

Zarins-Tutt JS, Barberi TT, Gao H, et al (2016) Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. Nat Prod Rep 33:54–72. https://doi.org/10.1039/C5NP00111K

Zhang MM, Wang Y, Ang EL, Zhao H (2016) Engineering microbial hosts for production of bacterial natural products. Nat Prod Rep 33:963–987. https://doi.org/10.1039/c6np00017g

Zhang X, Hindra, Elliot MA (2019) Unlocking the trove of metabolic treasures: activating silent biosynthetic gene clusters in bacteria and fungi. Curr Opin Microbiol 51:9–15. https://doi.org/10.1016/j.mib.2019.03.003

Zhang Z, Du C, de Barsy F, et al (2020) Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. Sci Adv 6:. https://doi.org/10.1126/sciadv.aay5781

Zhou X, Wu H, Li Z, et al (2011) Over-expression of UDP-glucose pyrophosphorylase increases validamycin A but decreases validoxylamine A production in *Streptomyces hygroscopicus* var. jinggangensis 5008. Metab Eng 13:768–776. https://doi.org/10.1016/j.ymben.2011.10.001

Zhu J, Sun D, Liu W, et al (2016) AvaR2, a pseudo γ-butyrolactone receptor homologue from *Streptomyces avermitilis*, is a pleiotropic repressor of avermectin and avenolide biosynthesis and cell growth. Mol Microbiol 102:562–578. https://doi.org/10.1111/mmi.13479

Zou Z, Du D, Zhang Y, et al (2014) A γ-butyrolactone-sensing activator/repressor, JadR3, controls a regulatory mini-network for jadomycin biosynthesis. Mol Microbiol 94:490–505. https://doi.org/10.1111/mmi.12752

APPENDIX 1: RECIPES USED FOR BUFFERS, SOLUTIONS AND MEDIA

5× Tris-borate-EDTA (TBE) buffer

Tris base 54 g
Boric acid 27.5g
Disodium EDTA 4.6875 g
Adjust pH to 8.3, add distilled water to 1000 mL

10× DNA Loading Dye

Glycerol	3.9 mL
10% SDS	0.5 mL
0.5M EDTA	0.2 mL
Bromophenol Blue	25 mg
Xylene Cyanol	25 mg

Add distilled water to 10 mL, aliquot 1 mL solution and store at -20 °C.

TB Buffer

HEPES	0.119 g
$CaCl_2 \cdot 2H2O$	0.119 g
KCl	0.932 g

Adjust pH to 6.7 with KOH

$MnCl_2 \cdot 4H_2O$	0.544 g

Add distilled water to 50 mL, filter sterilized and store at 4°C

Luria-Bertani (LB) Broth/Agar

Luria-Bertani powder (Fisher Scientific, Canada) 25 g
Add distilled water to 1000 mL
For LB Agar, dispense 250 mL into 4 bottles containing 4.5 g of agar

Low Salt Luria-Bertani (LB) Broth/Agar

Tryptone 10 g
Yeast Extract 5 g
NaCl 2.5 g
Adjust pH to 7-7.2, add distilled water to 1000 mL
For low salt LB Agar, dispense 250 mL into 4 bottles containing 4.5 g of agar

Trypticase Soy Broth (TSB)

Trypticase soy broth powder (BD Biosciences, Canada) 30 g
Add distilled water to 1000 mL

2 × Yeast extract-Tryptone (YT)
Tryptone 16g
Yeast extract 10g
NaCl 5g
Add distilled water to 1000 mL

Modified (MYMm) Agar*
Maltose 4 g
Yeast Extract 4 g
Malt Extract Broth 10 g
Agar 20 g
Add distilled water to 1000 mL

* MYMm is the same as MYM, except that the malt extract in the latter media was replaced with Bacto Malt Extract Broth (BD Biosciences) (Hegemann et al. 2015; Daniel-Ivad et al. 2017; Gehrke et al. 2019)

Modified (YMSm) Agar*
Yeast Extract 4 g
Soluble Starch 4 g
Malt Extract Broth 10 g
10 mM $CoCl_2 \cdot 6H_2O$ 2100 µL
Adjust pH with NaOH to 7.2
Agar 20 g
Add distilled water to 1000 mL

* YMSm is the same as YMS, except that the malt extract in the latter media was replaced with Bacto Malt Extract Broth (BD Biosciences) (Li et al. 2019a)

Soy Flour Mannitol (SFM) Agar (Kieser et al. 2000)
Mannitol 20 g
Add distilled water to 1000 mL
Dispense 250 mL into 4 bottles containing 5 g of agar and 5 g of defatted soy flour

Oat Bran Agar (OBA, Johnson et al. 2007)*
Oatmeal 20 g
Adjust pH with NaOH to 7.2
Trace Element Solution 2 mL
Add distilled water to 1000 mL
Dispense 250 mL into 4 bottles containing 4.5 g of agar and autoclave for 20 min.
*The detailed steps for preparing OBA are as follows:

1. Weigh 20 g of oat bran and combine with 500 mL of water. Heat in a microwave at full power until it is boiling.
2. When it starts boiling, reduce power to 1 and boil for 15-20 minutes.

3. Place on stir plate and let it cool with stirring for 45-60 minutes. Speeding up cooling may result reduced and inconsistent thaxtomin production.
4. Adjust the pH of the broth to 7.2 with 1M NaOH
5. Add 2mL/L of Trace Element Solution.
6. Add water to bring the volume to 1 liter.
7. Dispense 250 mL into 4 bottles containing 4.5 g of agar and autoclave for 20 min.

Starch Asparagine Agar (SA)
Soluble Starch 10 g
L-Asparagine 2 g
MOPS 21 g
$MgSO_4 \cdot 7H_2O$ 0.6 g
K_2HPO_4 4.4 g
Add distilled water to 1000 mL
Dispense 250 mL into 4 bottles containing 5 g of agar and autoclave for 20 min, then add 1 mL of Trace Element Solution for SA

Trace Element Solution for SA
$FeSO_4 \cdot 7H_2O$ 100 mg
$MnCl_2 \cdot 4H_2O$ 100 mg
$ZnSO_4 \cdot 7H_2O$ 100 mg
$CaCl_2 \cdot 2H_2O$ 130 mg
Add distilled water to 100 mL
Filter sterilize and store at 4 °C

Potato Mash Agar (PMA)
Instant Mashed Potato Flakes (Idahoan, Canada) 12.5 g
Agar 5 g
Add distilled water to 250 mL, then autoclave

Super Optimal Broth (SOB)
Tryptone 20 g
Yeast Extract 5 g
NaCl 0.5 g
1 M KCl 2.5 mL
Adjust the pH with NaOH to 7.0
Add distilled water to 1000 mL
Autoclave, then add:
1 M $MgSO_4$ 1 mL
1 M $MgCl_2$ 1 mL

Super Optimal Broth with Catabolite Repression (SOC)
Add 20 mL of 1 M glucose to 1000 mL SOB

International *Streptomyces* Project Medium 4 (ISP-4)
ISP-4 powder (BD Biosciences) 8 g
Add distilled water to 1000 mL

X-Gal (2% w/v solution or 20 mg/mL)
X-Gal 200 mg
Dimethylformamide (DMF) 10 mL
Vortex to dissolve the X-Gal
Wrap the centrifuge tube in aluminum foil to protect from light and store at -20°C

Isopropyl β-D-thiogalactopyranoside (IPTG, 100 mM)
IPTG 0.238 g
Add distilled water to 10 mL
Filter sterilize and store at -20°C.

Nutrient Broth/Agar (NB/NA)
Nutrient broth (BD Biosciences, Canada) 28 g
Add distilled water to 1000 mL
For NA, dispense 250 mL into 4 bottles containing 4.5 g of agar

APPENDIX 2: SUPPLEMENTARY FIGURES

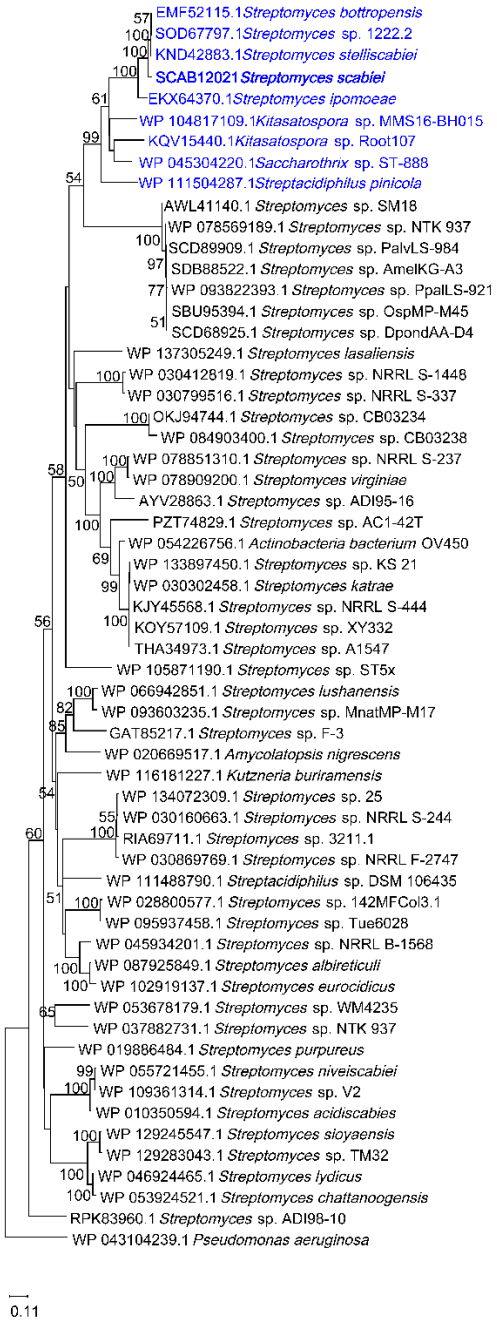


Figure S1 Phylogenetic relationships among homologues of SCAB12021 from the nr protein sequence database. Bootstrap values are shown for the branch points supported in ≥ 50% out of 1000 repetitions. The scale bar indicates the number of amino acid substitutions per site. The homologue from *Pseudomonas aeruginosa* was used as an outgroup. The clade containing SCAB12021 is indicated in blue font.

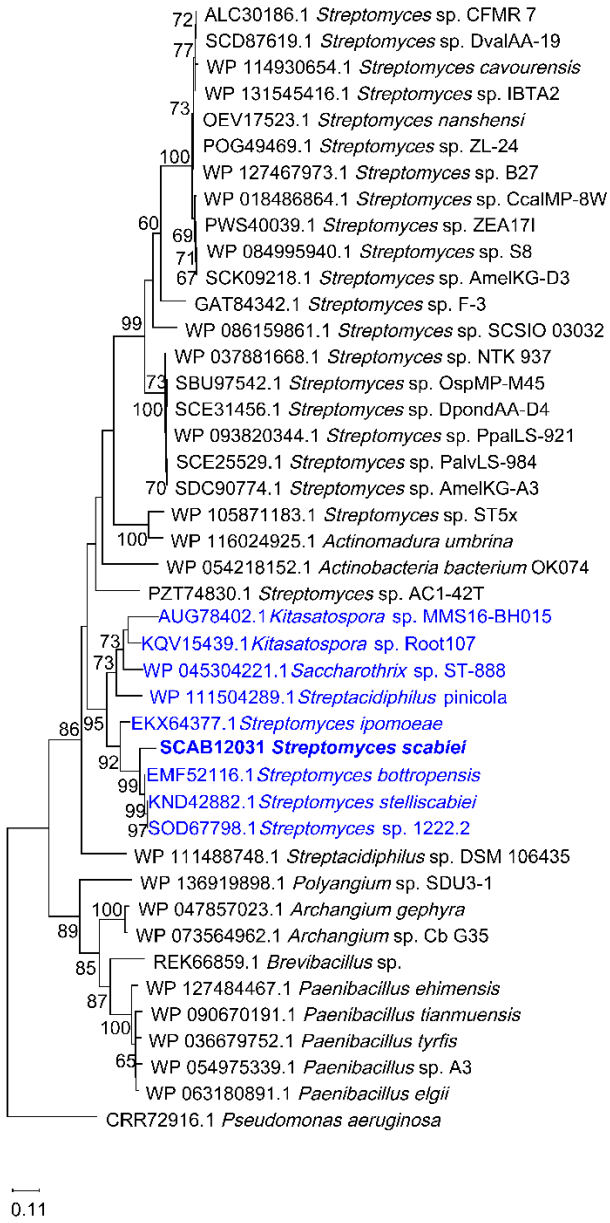


Figure S2 Phylogenetic relationships among homologues of SCAB12031 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The clade containing SCAB12031 is indicated in blue font.

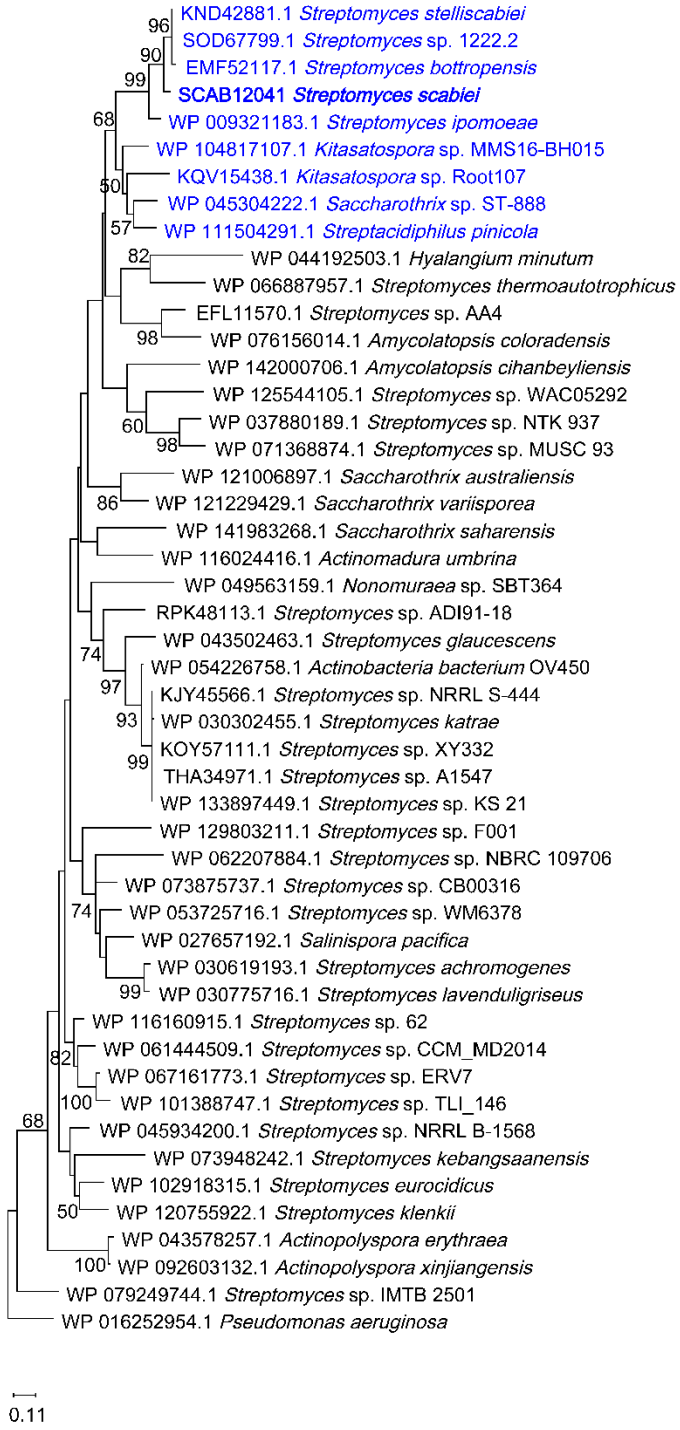


Figure S3 Phylogenetic relationships among homologues of SCAB12041 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The clade containing SCAB12041 is indicated in blue font.

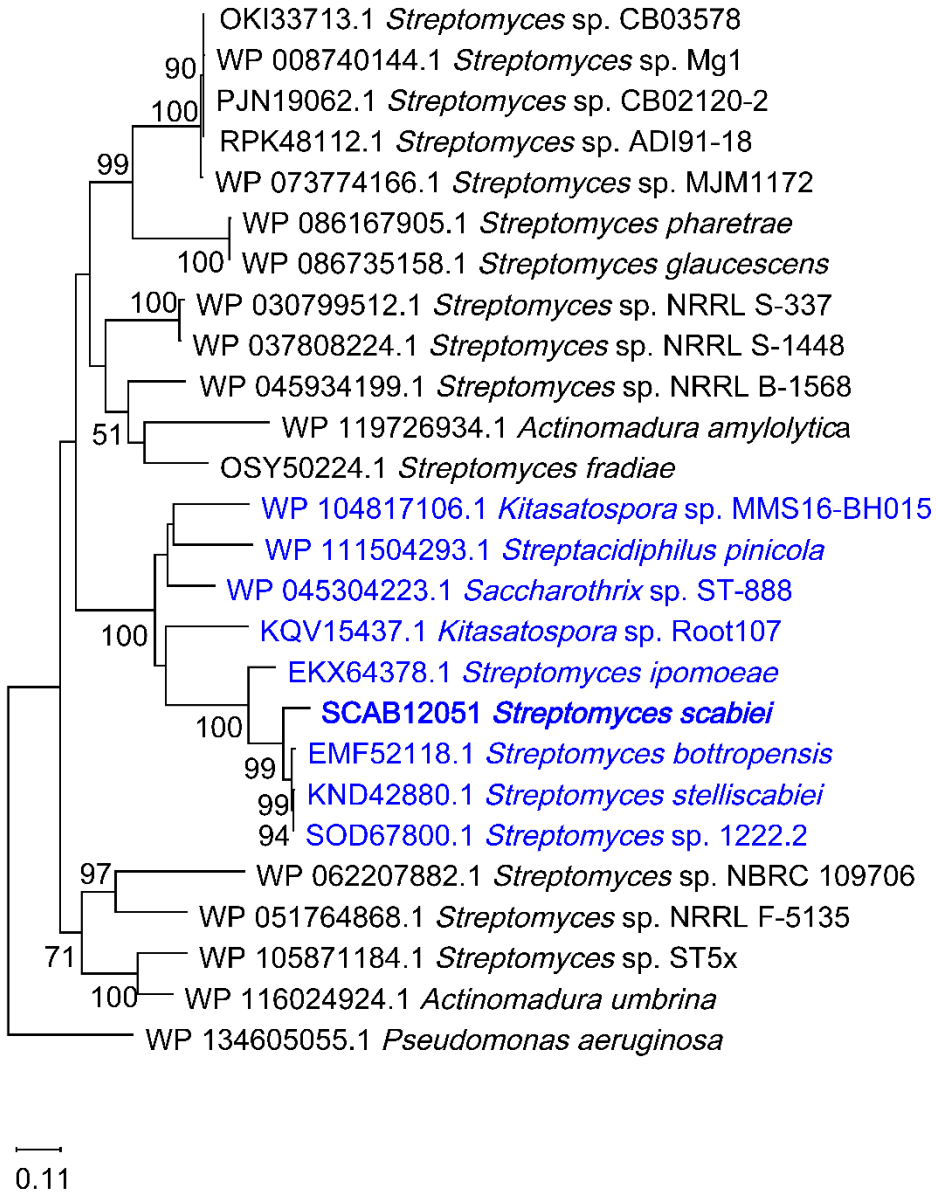


Figure S4 Phylogenetic relationships among homologues of SCAB12051 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The clade containing SCAB12051 is indicated in blue font.

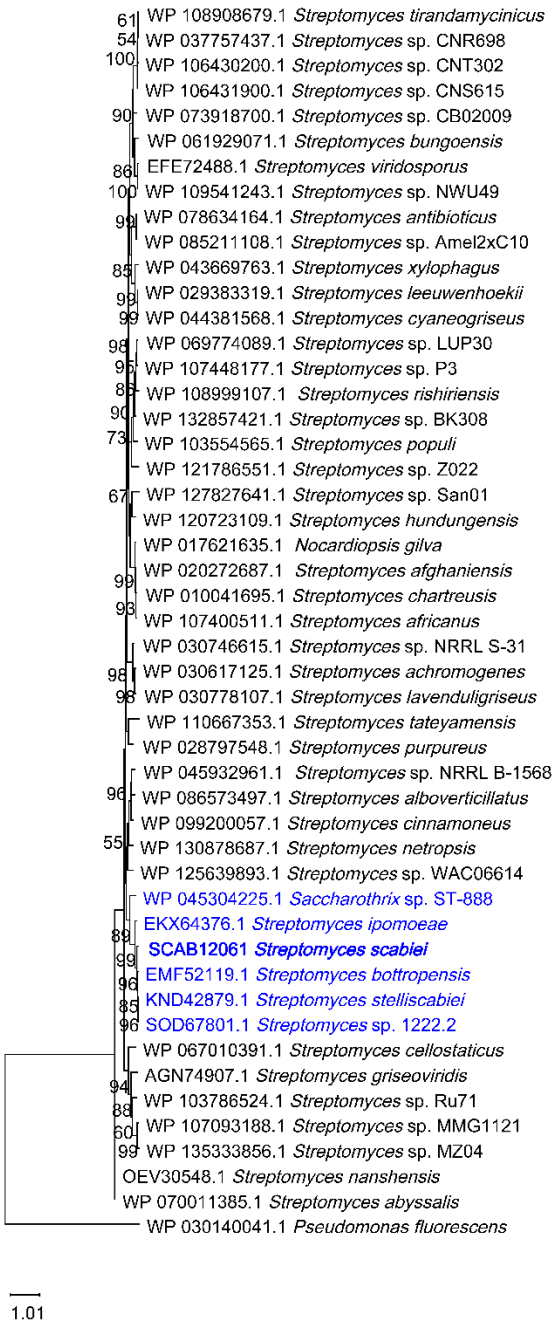


Figure S5 Phylogenetic relationships among homologues of SCAB12061 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The homologue from *Pseudomonas fluorescens* was used as an outgroup. The clade containing SCAB12061 is indicated in blue font.

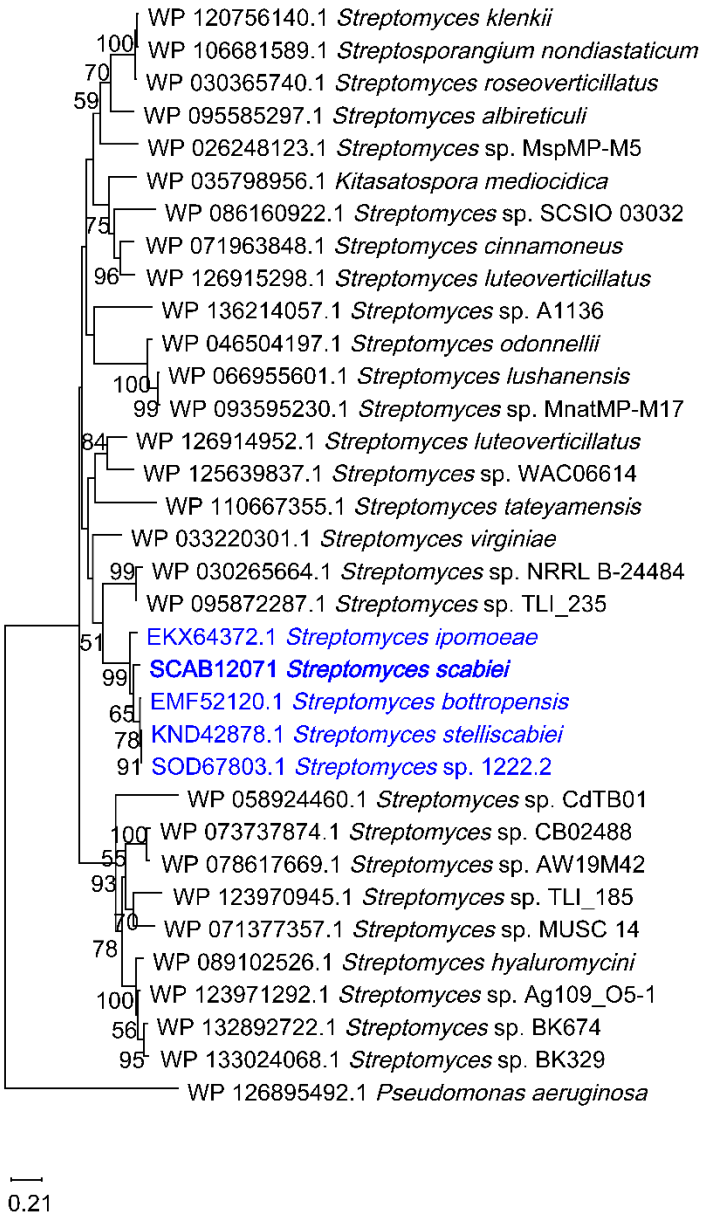


Figure S6 Phylogenetic relationships among homologues of SCAB12071 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The clade containing SCAB12071 is indicated in blue font.

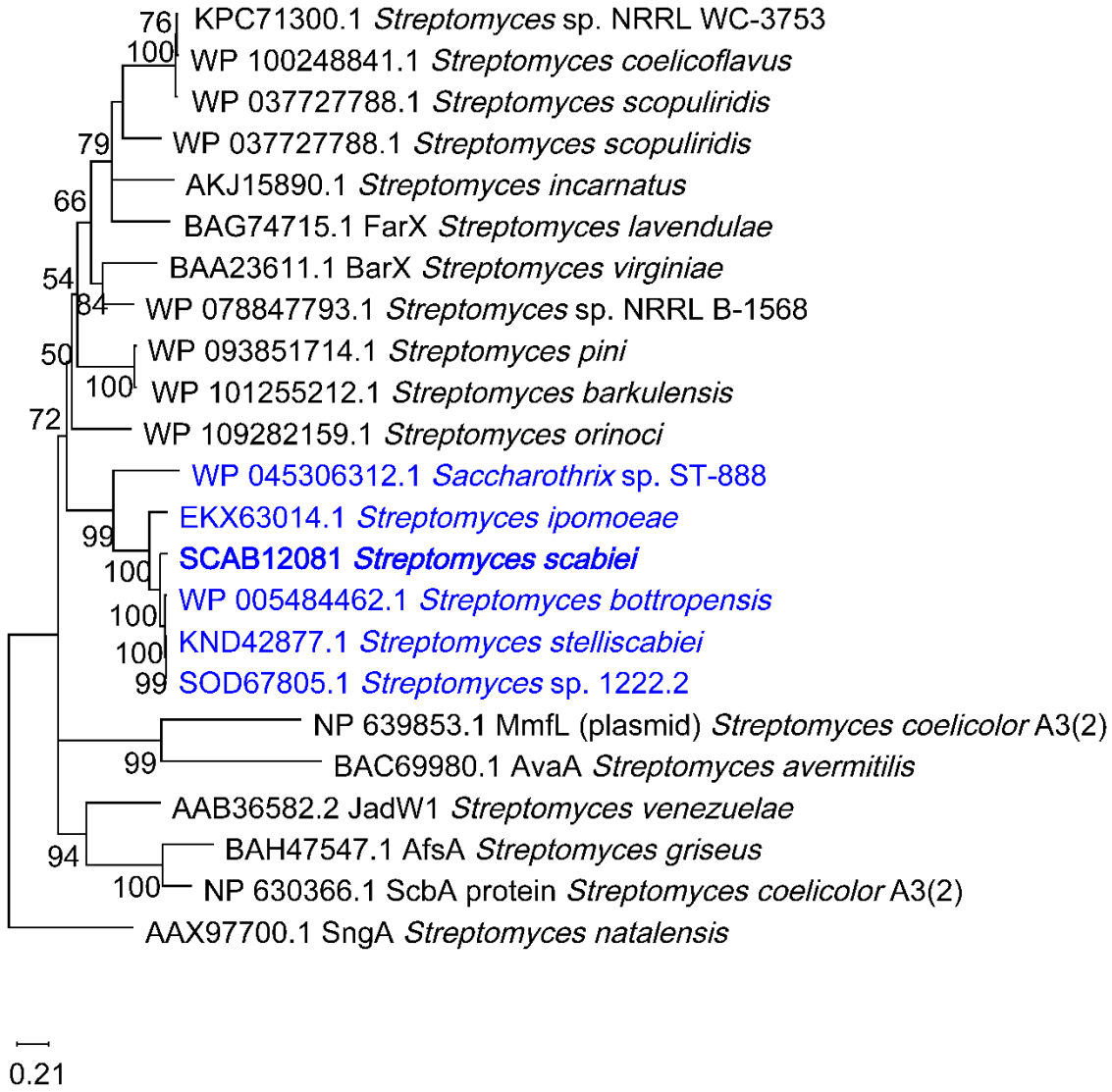


Figure S7 Phylogenetic relationships among homologues of SCAB12081 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The SngA homologue from *Streptomyces natalensis* was used as an outgroup. The clade containing SCAB12081 is indicated in blue font.

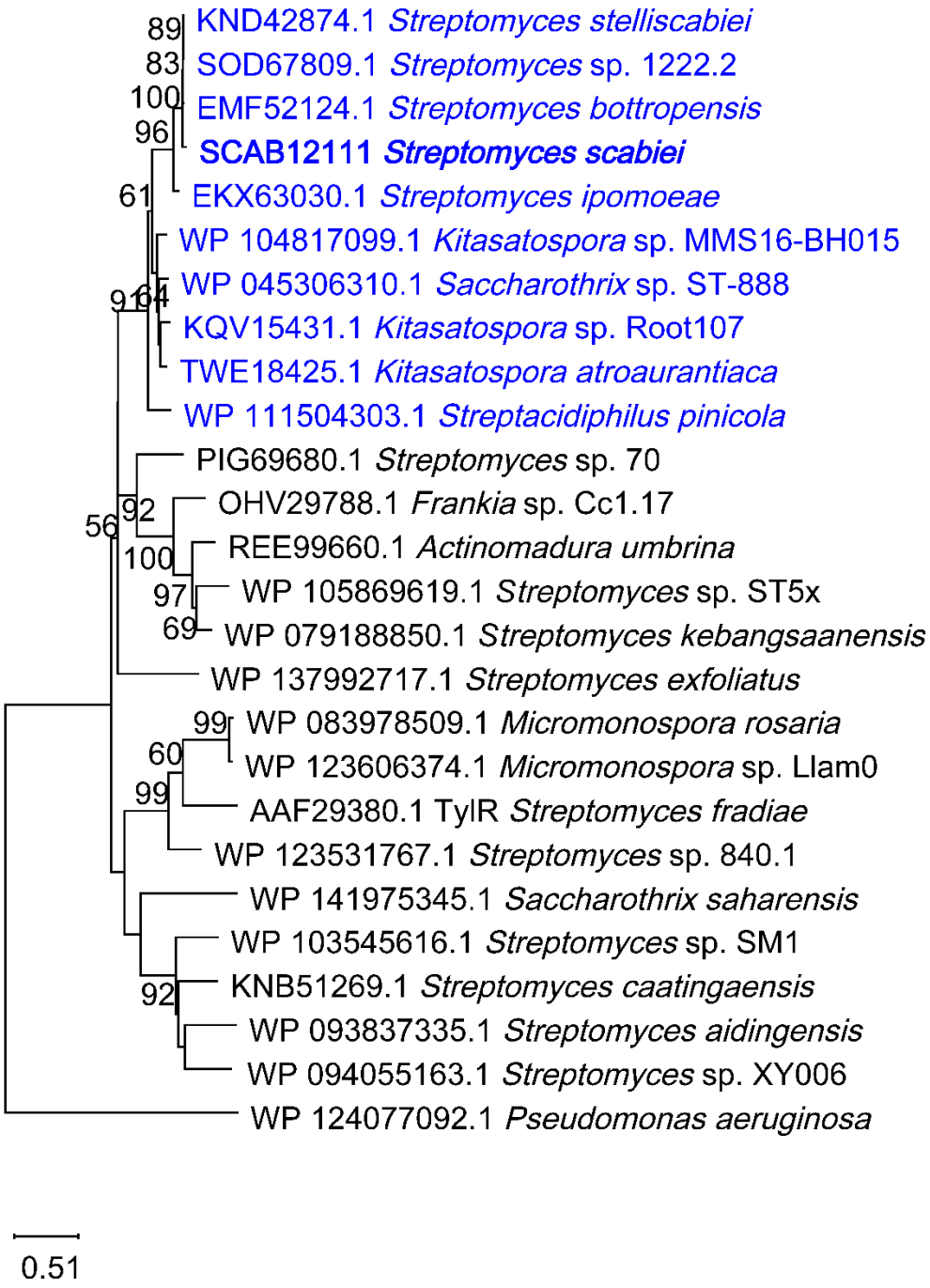
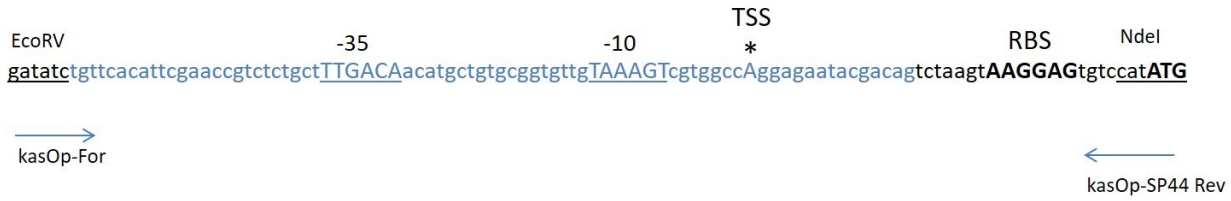


Figure S8 Phylogenetic relationships among homologues of SCAB12111 from the nr protein sequence database. Features in the phylogenetic tree are as described in the legend for Figure S1. The clade containing SCAB12111 is indicated in blue font.

*kas*Op* synthetic promoter:



SP44 synthetic promoter:

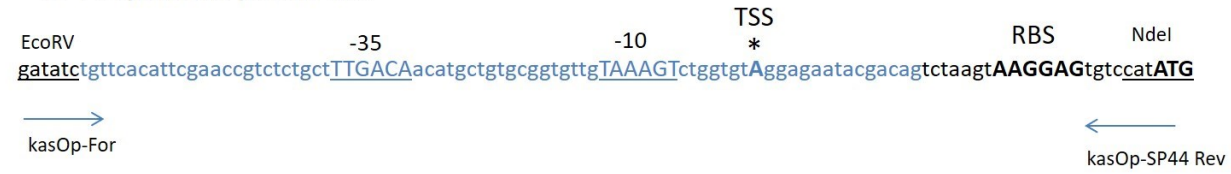


Figure S9 Structure of the SP44 and *kasO*p* promoter fragments used in this study. Two restriction sites, *Eco*RV and *Nde*I, were designed to clone the promoters into pIJ8641 and pIJ8668 plasmids, respectively.

APPENDIX 3: SUPPLEMENTARY TABLES

Table S1 MS data for compounds in the *S. scabiei* thaxtomin networks.

Compound	*m/z* $[M+H]^+$	Formula	Retention time	Key featured MS^2 fragments
1	439.1598	$C_{22}H_{22}N_4O_6$	9.4529	130.1, 159.1, 188.1, 219.1, 247.1, 362.1, 421.1, 439.1
2	439.1647	$C_{22}H_{22}N_4O_6$	15.759	131.1, 159.1, 188.1, 219.1, 247.1, 362.1, 421.1, 439.1
3	423.1628	$C_{22}H_{22}N_4O_5$	9.9188	130.1, 203.1, 231.1, 346.1, 405.1, 423.2
4	409.1873	$C_{21}H_{20}N_4O_5$	15.4933	145.1, 157.1, 216.1, 247.1, 259.1, 391.2 409.2
5	455.1563	$C_{22}H_{22}N_4O_7$	9.081	125.1, 160.1, 188.1, 219.1, 247.1, 419.1, 437.1, 455.2
6	421.1519	$C_{22}H_{20}N_4O_5$	15.811	130.1, 176.1, 188.1, 219.1, 247.1, 379.1, 421.1
7	421.1477	$C_{22}H_{20}N_4O_5$	10.685	130.1, 160.1, 188.1, 219.1, 247.1, 362.1, 421.1
8	421.1512	$C_{22}H_{20}N_4O_5$	9.6884	127.1.1, 140.0.1, 155.0, 229.1, 421.1
9	391.1454	$C_{21}H_{18}N_4O_4$	10.1804	139.1, 161.1, 175.1, 217.1, 391.1
10	454.1671	$C_{22}H_{20}N_4O_6$	8.2555	125.1, 145.1, 160.1, 188.1, 219.1, 247.1, 408.2, 436.2, 454.2
11	376.1664	$C_{22}H_{21}N_4O_3$	9.4168	130.1, 160.1, 188.1, 226.1, 247.1, 283.1, 317.1, 376.1
12	405.1564	$C_{22}H_{20}N_4O_4$	10.0567	134.1, 172.1, 203.1, 231.1, 405.1
13	360.1719	$C_{22}H_{21}N_3O_2$	9.9865	130.1, 161.1, 172.1, 203.1, 231.1, 247.1, 360.2
14	393.1595	$C_{21}H_{20}N_4O_4$	9.69	130.1, 134.1, 159.1, 175.1, 204.1, 231.1, 219.1, 309.1, 393.1
15	407.1711	$C_{22}H_{22}N_4O_4$	10.25	130.1, 134.1, 141.1, 159.1, 175.1, 203.1, 232.1, 361.2, 407.1

16	409.1469	$C_{21}H_{20}N_4O_5$	9.38	128.1, 130.1, 175.1, 218.1, 234.1, 235.1, 247.1, 409.1

Compound	***m/z*** **[M-H]$^-$**	**Formula**	**Retention time**	**Key featured MS^2 fragments**
1	437.1466	$C_{22}H_{22}N_4O_6$	9.5224	107.0, 127.1, 140.0, 155.0, 192.1, 245.1, 275.1, 329.1, 437.1
4	407.1354	$C_{21}H_{20}N_4O_5$	9.9821	87.1, 117.0, 141.0, 161.0, 215.1, 233.1 407.1
14	391.1388	$C_{21}H_{20}N_4O_4$	9.83	85.0, 113.0, 125.0, 156.0, 160.1, 169.0,174.1, 217.1, 391.1,
15	405.1633	$C_{22}H_{22}N_4O_4$	10.74	99.1, 112.0, 127.0, 131.0, 156.0, 172.1, 174.0, 174.1, 231.1, 405.2
17	423.1299	$C_{21}H_{20}N_4O_6$	9.2119	107.0, 126.0, 133.0, 141.0, 161.0, 192.1, 231.1, 249.1, 423.1
18	392.1612	$C_{22}H_{23}N_3O_4$	8.9232	107.0, 127.1, 140.0, 155.0, 192.1, 263.1, 284.1, 392.1
19	438.1672	$C_{24}H_{21}N_7O_2$	9.4797	107.0, 127.1, 140.0, 155.0, 192.1, 263.1, 284.1, 392.1, 438.2
20	505.1335	$C_{20}H_{22}N_6O_{10}$	9.3446	107.0, 127.0, 140.0, 155.0, 192.1, 245.1, 285.1, 329.1,437.1, 459.1, 505.1
21	457.1284	$C_{23}H_{18}N_6O_5$	9.9763	101.0, 127.1, 140.0, 155.0, 229.1, 247.1, 275.1, 421.1, 457.1
22	473.1231	$C_{23}H_{18}N_6O_6$	9.4428	107.0, 127.1, 140.0, 155.0, 192.1, 245.1, 275.1, 329.1, 437.1, 473.1
23	583.2409	$C_{27}H_{34}N_7O_8$	9.5195	107.0, 127.0, 140.0, 155.0, 192.1, 245.1, 275.1, 329.1,437.1, 583.1
24	483.1521	$C_{21}H_{22}N_7O_7$	9.4314	107.0, 133.0, 161.0,192.1, 245.0, 275.1, 329.1, 393.2, 437.1, 483.1

Table S2 MS data for compounds in the *S. scabiei* coronafacoyl phytotoxin networks.

Compound	*m/z* $[M+H]^+$	Retention time	Formula	Key featured MS^2 fragments
25	322.2004	15.6427	$C_{18}H_{27}NO_4$	119.1, 145.1, 163.1, 191.1, 276.1, 322.2
26	320.1863	10.1978	$C_{18}H_{25}NO_4$	119.1, 145.1, 163.1, 191.1, 302.2, 320.2
27	308.1888	10.3184	$C_{18}H_{29}NO_3$	105.1, 131.1, 149.1, 177.1, 262.2, 308.2
28	308.1856	15.534	$C_{18}H_{29}NO_3$	105.1, 119.1, 145.1, 163.1, 177.1, 191.1, 262.1, 308.2
29	308.1856	2.1334	$C_{18}H_{29}NO_3$	105.1, 119.1, 145.1, 163.1, 177.1, 191.1, 262.2, 308.2
30	294.1682	9.8598	$C_{17}H_{27}NO_3$	105.1, 131.1, 149.1, 177.1, 248.2, 294.2
31	292.1975	10.4771	$C_{17}H_{25}NO_3$	119.1, 145.1, 163.1, 191.1, 208.1, 292.2
32	292.1900	10.9579	$C_{17}H_{25}NO_3$	119.1, 145.1, 163.1, 191.1, 208.1, 292.2
33	276.1990	12.0761	$C_{17}H_{25}NO_2$	119.1, 145.1, 163.1, 191.1, 235.2, 276.2
34	276.1963	10.6999	$C_{17}H_{25}NO_2$	119.1, 145.1, 163.1, 191.1, 276.2
46	322.1978	10.6679	$C_{18}H_{27}NO_4$	119.1, 145.1, 147.1, 163.1, 164.1, 165.1, 191.1, 192.1, 193.1
47	308.1895	9.6214	$C_{18}H_{29}NO_3$	116.1, 163.1, 165.1, 193.1, 208.2, 234.1, 264.2, 308.2, 308.9
Compound	***m/z*** $[M-H]^-$	**Retention time**	**Formula**	**Key featured MS^2 fragments**
25	320.1862	10.6987	$C_{18}H_{27}NO_4$	107.1, 130.1, 163.1, 191.1, 276.2, 320.2
27	306.1706	10.2891	$C_{18}H_{29}NO_3$	107.1, 130.1, 149.1, 177.1, 262.2, 306.2
30	292.1570	9.7427	$C_{17}H_{27}NO_3$	116.1, 149.1, 177.1, 248.1, 292.2
35	320.1865	15.667	$C_{18}H_{27}NO_4$	107.1, 130.1, 163.1, 191.1, 276.2, 320.2
36	296.1639	11.388	$C_{19}H_{23}NO_2$	107.1, 149.1, 163.1, 184.1, 296.2
37	276.2019	10.0926	$C_{17}H_{27}NO_2$	107.1, 126.1, 163.1, 191.1, 276.2
38	262.1882	10.3717	$C_{16}H_{25}NO_2$	112.1, 134.1, 149.1, 163.1, 191.1, 218.1, 262.1
39	262.1800	10.3198	$C_{16}H_{25}NO_2$	112.1, 134.1, 149.1, 163.1, 191.1, 218.1, 262.1
40	340.1560	11.5821	$C_{20}H_{23}NO_4$	114.1, 163.1, 225.2, 296.2, 340.2
41	388.1778	10.6734	$C_{21}H_{27}NO_6$	107.1, 130.1, 163.1, 191.1, 276.2, 320.2,388.2

42	374.1587	10.1217	$C_{20}H_{25}NO_6$	107.1, 130.1, 149.1, 177.1, 262.2, 306.2, 374.2
43	360.1458	9.5999	$C_{19}H_{23}NO_6$	116.1, 149.1, 177.1, 248.2, 292.2, 360.1
44	356.1653	10.5861	$C_{21}H_{27}NO_4$	130.1, 163.1, 191.1, 276.2, 320.2, 356.2
45	342.1474	10.3147	$C_{20}H_{25}NO_4$	116.1, 163.1, 191.1, 262.2, 306.2
48	645.4107	10.4154	$C_{37}H_{54}N_6O_4$	101.0, 130.1, 163.1, 165.1, 193.1, 222.2, 278.2, 320.2, 321.2, 322.2, 645.3
49	322.2010	10.0492	$C_{18}H_{29}NO_4$	130.1, 147.1, 151.1, 163.1, 165.1, 193.1, 222.2, 234.2, 250.2, 276.2, 278.2, 322.2

Table S3 MS data for compounds in the *S. scabiei* concanamycin network.

Compound	*m/z* $[M+Na]^+$	Retention time	Formula	Key featured MS^2 fragments
50	874.4908	12.2078	$C_{45}H_{73}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 501.3, 701.4, 874.4
51	874.4908	11.4148	$C_{45}H_{73}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 501.3, 752.5, 874.4
52	888.5071	12.6819	$C_{46}H_{75}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 515.3, 715.4, 888.5
53	888.5077	14.4564	$C_{46}H_{75}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 501.3, 515.3, 888.5
54	888.5079	13.8185	$C_{46}H_{75}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 501.3, 515.3, 584.1, 888.5
55	902.5233	13.106	$C_{47}H_{77}NO_{14}$	151.1, 196.1, 237.1, 378.2, 410.2, 515.3, 902.5
56	886.4888	13.0704	$C_{46}H_{73}NO_{14}$	151.1, 196.1, 273.1, 378.2, 396.2, 501.3, 752.5, 888.5
57	886.4922	13.7441	$C_{46}H_{73}NO_{14}$	151.1, 196.1, 223.1, 273.1, 378.2, 396.2, 501.3, 515.3, 888.5
58	816.4467	12.3476	$C_{37}H_{67}N_3O_{15}$	184.1, 378.2, 396.2, 443.2, 816.4
59	844.4806	12.0611	$C_{44}H_{71}NO_{13}$	151.1, 196.1, 223.1, 378.2, 396.2, 471.3, 844.5
60	856.4827	12.2635	$C_{46}H_{67}N_5O_9$	196.1, 214.1, 378.2, 502.3, 667.4, 858.5
61	856.4845	13.0864	$C_{46}H_{67}N_5O_9$	196.1, 214.1, 378.2, 503.3, 667.4, 858.5
62	858.4585	11.9791	$C_{44}H_{69}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 485.3, 667.4, 858.5
63	858.4955	12.6547	$C_{44}H_{69}NO_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 485.3, 674.4, 858.5
64	858.4943	13.298	$C_{44}H_{69}NO_{14}$	196.1, 378.2, 471.3, 858.5
65	859.4594	11.8876	$C_{43}H_{68}N_2O_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 486.3, 859.5

66	877.4921	12.1672	$C_{47}H_{66}N_8O_7$	151.1, 196.1, 223.1, 397.2, 503.3, 877.4
67	886.5321	13.1887	$C_{41}H_{77}N_5O_{14}$	151.1, 196.1, 223.1, 378.2, 396.2, 501.3, 515.3, 888.5
68	900.5053	13.6831	$C_{47}H_{75}NO_{14}$	196.1, 273.1, 378.2, 514.3, 902.5
69	916.5380	13.3346	$C_{48}H_{79}NO_{14}$	196.1, 378.2,392.2, 410.2, 515.3, 520.3, 916.5
70	1047.5789	11.6424	$C_{56}H_{84}N_2O_{15}$	151.1, 196.1, 223.1, 378.2, 396.2, 411.2, 501.3, 674.4, 1047.6
71	737.4424	11.3834	$C_{36}H_{64}N_3O_{11}$	151.1, 398.2, 417.3, 443.3, 475.3, 537.3, 675.4, 701.4, 705.4, 719.4, 737.4
72	751.4545	12.0714	$C_{34}H_{68}N_2O_{14}$	151.1, 431.3, 489.3, 533.4, 551.3, 595.4, 627.4, 689.5, 715.4, 733.4, 751.5
73	765.4745	11.8169	$C_{38}H_{68}N_3O_{11}$	445.3, 503.3, 533.4, 595.4, 627.4, 645.4, 647.4, 715.4, 733.5, 765.5
74	673.3907	13.2486	$C_{34}H_{56}N_3O_9$	389.2, 405.2, 433.2, 447.3, 459.2, 463.3, 491.3, 629.4, 641.4, 645.4, 673.4
75	701.3763	14.8026	$C_{39}H_{54}N_2O_8$	205.1, 291.2, 415.2, 417.2, 475.3, 501.2, 519.3, 669.4, 683.4, 701.4
76	701.3845	14.8007	$C_{39}H_{54}N_2O_8$	182.1, 291.2, 415.2, 417.2, 475.3, 501.2, 519.3, 669.4, 683.4, 701.4

Table S4 MS data for compounds in the *S. scabiei* bottromycin networks.

Compound	*m/z* $[M+H]^+$	Retention time	Formula	Key featured MS^2 fragments
77	809.4500	9.3519	$C_{41}H_{60}N_8O_7S$	141.1, 169.1, 187.1, 268.2, 348.1, 349.2, 405.2, 457.1, 462.3, 623.4
78	809.4379	8.7965	$C_{41}H_{60}N_8O_7S$	141.1, 169.1, 187.1, 268.2, 348.1,349.2, 405.2, 457.1, 462.3, 623.4
79	809.4337	8.6803	$C_{41}H_{60}N_8O_7S$	141.1, 169.1, 268.2, 301.1, 363.2, 476.3, 639.4
80	823.4511	8.9427	$C_{42}H_{62}N_8O_7S$	86.1, 141.1, 169.1, 268.2, 313.1, 363.2, 391.2, 476.3, 637.4
81	823.4633	9.0455	$C_{42}H_{62}N_8O_7S$	134.1, 138.1, 141.1, 169.1,187.1, 268.2, 348.1. 363.2, 391.2, 412.2
82	837.4671	9.0237	$C_{43}H_{64}N_8O_7S$	98.1, 141.1, 155.1, 183.1, 282.1, 313.1, 350.2, 395.2, 490.3, 651.4
Compound	***m/z*** $[M-H]^-$	**Retention time**	**Formula**	**Key featured MS^2 fragments**
77	807.4235	8.5715	$C_{41}H_{60}N_8O_7S$	110.1, 139.1, 151.1, 226.2, 264.2, 332.1, 425.3, 652.4
80	821.4395	8.9382	$C_{42}H_{62}N_8O_7S$	98.0, 139.1, 151.1, 177.1, 226.2, 264.2, 425.3, 474.3
83	857.4155	8.8465	$C_{41}H_{62}N_8O_{10}S$	98.0, 139.1, 151.1, 177.1, 226.2, 264.2, 425.3, 474.3, 652.4
84	867.4438	9.0466	$C_{43}H_{64}N_8O_9S$	98.0, 139.1, 151.1, 177.1, 226.2, 264.2, 425.3, 474.3, 652.4
85	853.4284	8.6842	$C_{42}H_{62}N_8O_9S$	71.0, 151.1, 226.2, 264.2, 516.4, 807.4, 813.5
86	526.2784	9.7428	$C_{20}H_{41}N_6O_8S$	179.1, 196.1, 262.2, 389.2, 472.3, 490.3, 526.3
87	536.3106	9.2924	$C_{21}H_{43}N_8O_6S$	168.1, 185.1, 209.1, 279.1, 462.3, 463.3, 490.3

Table S5 MS data for compounds in the *S. scabiei* siderophore networks.

Compound	*m/z* $[M+H]^+$	Retention time	Formula	Key featured MS^2 fragments
88	601.3532	8.494	$C_{27}H_{48}N_6O_9$	84.1, 100.1, 102.1, 201.1, 283.1,401.2, 483.2, 583.3, 601.3
89	401.2392	8.4921	$C_{18}H_{32}N_4O_6$	84.1, 100.1, 102.1, 201.1, 243.1, 261.1, 283.1, 304.1, 356.2, 401.2
90	319.2286	7.5811	$C_{20}H_{30}O_3$	84.1, 100.1, 102.1, 201.1, 211.1, 241.1, 253.9, 301.1,319.2
91	319.2340	7.7256	$C_{20}H_{30}O_3$	84.1, 100.1, 102.1, 201.1, 211.1, 241.1, 253.9, 301.1,319.2
92	325.0676	10.9854	$C_{14}H_{16}N_2O_3S_2$	100.0, 128.1, 146.0, 172.0, 190.0, 206.0, 224.0, 281.1, 325.1
93	341.0625	9.9228	$C_{14}H_{16}N_2O_4S_2$	100.0, 128.1, 146.0, 172.0, 190.0, 206.0, 222.0, 295.1, 323.1
Compound	***m/z*** **$[M-H]^-$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**
88	599.3413	8.4597	$C_{27}H_{48}N_6O_9$	98.0, 136.1, 163.1, 181.1, 199.1, 217.1, 399.2, 599.3
94	583.3459	8.3615	$C_{27}H_{48}N_6O_8$	98.0, 136.1, 163.1, 181.1, 195.1, 237.1, 265.1, 365.2, 383.2, 583.3

Table S6 Chemical family analysis of metabolites detected in the *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO) extracts in positive ionization mode. Chemical classes highlighted in bold are those where the number of molecules detected in the KO extracts was greater than the number detected in the WT extracts.

Chemical Class	Only in WT	Only in KO	Both in WT&KO	WT	KO	Total
Total	993	1329	1694	2687	3023	4016
No matches	634	717	544	1178	1261	1895
Carboxylic acids and derivatives	98	196	368	466	564	662
Prenol lipids	72	101	258	330	359	431
Benzene and substituted derivatives	53	78	148	201	226	279
Organooxygen compounds	43	80	109	152	189	232
Pyridines and derivatives	14	11	38	52	49	63
Fatty Acyls	8	20	30	38	50	58
Steroids and steroid derivatives	7	13	15	22	28	35
Coumarins and derivatives	6	11	14	20	25	31
Indoles and derivatives	6	7	14	20	21	27
Quinolines and derivatives	6	9	11	17	20	26
Azoles	4	8	12	16	20	24
Diazines	0	7	13	13	20	20
Peptidomimetics	7	2	10	17	12	19
Flavonoids	0	7	9	9	16	16
Organonitrogen compounds	0	8	8	8	16	16
Imidazopyrimidines	3	3	9	12	12	15
Diazanaphthalenes	3	6	3	6	9	12
Phenol esters	2	4	4	6	8	10
Tetrahydroisoquinolines	3	4	3	6	7	10
Isoflavonoids	2	3	3	5	6	8

Phenols	3	0	4	7	4	7
Piperidines	2	2	3	5	5	7
Pyrans	1	2	4	5	6	7
Diarylheptanoids	1	2	3	4	5	6
Glycerolipids	1	1	4	5	5	6
Glycerophospholipids	0	1	5	5	6	6
Organic oxoanionic compounds	1	2	3	4	5	6
Fluorenes	0	1	4	4	5	5
Macrolactams	1	1	3	4	4	5
Organic sulfuric acids and derivatives	0	0	5	5	5	5
Phenanthrenes and derivatives	2	3	0	2	3	5
Dihydroisoquinolines	1	1	2	3	3	4
Isoindoles and derivatives	0	0	4	4	4	4
Macrolides and analogues	1	1	2	3	3	4
Orthocarboxylic acid derivatives	0	3	1	1	4	4
Purine nucleosides	1	2	1	2	3	4
Benzothiazoles	1	0	2	3	2	3
Hydroxy acids and derivatives	1	2	0	1	2	3
Naphthalenes	0	1	2	2	3	3
Quinolizines	0	0	3	3	3	3
Aporphines	1	1	0	1	1	2
Aurone flavonoids	0	0	2	2	2	2
Benzodiazepines	1	0	1	2	1	2
Benzofurans	0	0	2	2	2	2
Halohydrins	1	1	0	1	1	2
Isoquinolines and derivatives	0	1	1	1	2	2

Organic phosphonic acids and derivatives	0	0	2	2	2	2
Organometalloid compounds	1	0	1	2	1	2
Pyrroles	0	1	1	1	2	2
Pyrrolizidines	1	0	1	2	1	2
Thiolanes	0	0	2	2	2	2
Triazolopyrimidines	0	2	0	0	2	2
Cinnamaldehydes	0	0	1	1	1	1
Harmala alkaloids	0	1	0	0	1	1
Lactones	0	1	0	0	1	1
Organic thiosulfuric acids and derivatives	0	0	1	1	1	1
Saccharolipids	0	1	0	0	1	1
Sphingolipids	0	0	1	1	1	1

Table S7 Chemical family analysis of metabolites detected in the *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO) extracts in negative ionization mode. Chemical classes highlighted in bold are those where the number of molecules detected in the KO extracts was greater than the number detected in the WT extracts.

Chemical Class	**Only in WT**	**Only in KO**	**Both in WT&KO**	**WT**	**KO**	**Total**
Total	514	806	1077	1591	1883	2397
No matches	249	370	303	552	673	922
Prenol lipids	61	93	202	263	295	356
Carboxylic acids and derivatives	47	108	184	231	292	339
Organooxygen compounds	26	61	76	102	137	163
Benzene and substituted derivatives	30	36	62	92	98	128
Steroids and steroid derivatives	25	23	61	86	84	109
Fatty Acyls	14	36	54	68	90	104
Flavonoids	13	25	24	37	49	62
Coumarins and derivatives	7	9	12	19	21	28
Indoles and derivatives	4	3	8	12	11	15
Isoflavonoids	0	4	11	11	15	15
Diazines	0	1	9	9	10	10
Organic phosphonic acids and derivatives	0	0	10	10	10	10
Phenols	5	2	3	8	5	10
Azoles	1	1	7	8	8	9
Hydroxy acids and derivatives	3	2	2	5	4	7
Glycerophospholipids	0	5	1	1	6	6
Piperidines	0	1	5	5	6	6
Anthracenes	0	2	3	3	5	5
Benzofurans	1	0	4	5	4	5
Cinnamic acids and derivatives	1	2	2	3	4	5

Diazanaphthalenes	4	0	1	5	1	5
Isochromanequinones	2	0	3	5	3	5
Peptidomimetics	1	4	0	1	4	5
Lactones	2	0	2	4	2	4
Phenylpropanoic acids	1	2	1	2	3	4
Pyridines and derivatives	0	1	3	3	4	4
Isoindoles and derivatives	0	1	2	2	3	3
Macrolides and analogues	3	0	0	3	0	3
Naphthalenes	0	1	2	2	3	3
Oxazolopyridines	0	0	3	3	3	3
Pyrroles	0	1	2	2	3	3
Quinolines and derivatives	2	0	1	3	1	3
Aurone flavonoids	0	1	1	1	2	2
Benzimidazoles	0	0	2	2	2	2
Benzodiazepines	0	2	0	0	2	2
Diarylheptanoids	1	0	1	2	1	2
Dihydrofurans	0	1	1	1	2	2
Dioxanes	1	0	1	2	1	2
Homogeneous other non-metal compounds	0	1	1	1	2	2
Isocoumarans	0	0	2	2	2	2
Naphthopyrans	2	0	0	2	0	2
Organonitrogen compounds	0	1	1	1	2	2
Phenol esters	1	1	0	1	1	2
Saccharolipids	2	0	0	2	0	2
Sphingolipids	2	0	0	2	0	2
Unsaturated hydrocarbons	0	2	0	0	2	2
Yohimbine alkaloids	0	2	0	0	2	2

Cinnamaldehydes	0	0	1	1	1	1
Indolonaphthyridine alkaloids	0	1	0	0	1	1
Linear 1,3-diarylpropanoids	1	0	0	1	0	1
Macrolactams	1	0	0	1	0	1
Organic oxoanionic compounds	1	0	0	1	0	1
Oxanes	0	0	1	1	1	1
Pyrans	0		1	1	1	1
Stilbenes	0	0	1	1	1	1

Table S8 MS data for compounds in the phenazines networks of *S. scabiei* 87-22 (WT) and Δ*SCAB12101* (KO).

Compound	*m/z* $[M+H]^+$	Retention time	Formula	Key featured MS^2 fragments	Strain/media
95	271.068	434.332	$C_{14}H_{10}N_2O_4$	128.1, 166.1, 183.1, 225.1, 226.1, 253.1	KO/OBA
96	227.081	434.927	$C_{13}H_{10}N_2O_2$	70.1, 166.1, 181.1, 182.1, 183.1, 209.1	WT/OBA; KO/OBA,YMSm
97	214.159	500.551	$C_{12}H_9N_2O_2$	95.0, 105.1, 169.1, 185.1, 187.1, 215.2	KO/YMSm
98	213.066	501.803	$C_{12}H_8N_2O_2$	86.1, 95.1, 135.1, 169.1, 185.1, 187.1	WT/MYMm
99	255.076	631.949	$C_{14}H_{10}N_2O_3$	81.1, 93.1, 95.1, 107.1, 154.1, 181.1, 209.1, 227.1	WT/MYMm,YMSm;KO/MYMm,YMSm
100	269.092	662.921	$C_{15}H_{12}N_2O_3$	109.1, 135.1, 168.1, 181.1, 195.1, 196.1, 223.1, 241.1	WT/MYMm,YMSm; KO/MYMm,YMSm
101	229.061	828.889	$C_{12}H_8N_2O_3$	114.0, 117.0, 155.1, 156.0, 165.0, 183.1, 188.0, 201.1	WT/MYMm

Table S9 MS data of compounds that are only present in Δ*SCAB12101* in the thaxtomin networks.

Compound	***m/z*** **$[M+H]^+$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**	**Strain/ media**
102	407.17	565.826	$C_{22}H_{22}N_4O_4$	113.1, 164.0, 192.0, 205.0, 218.1, 233.0, 276.1, 304.1	KO/OBA
103	437.145	532.668	$C_{22}H_{20}N_4O_6$	107.0, 145.1, 148.1, 188.1, 219.1, 247.1, 259.1, 391.2	KO/OBA
104	453.141	557.512	$C_{22}H_{20}N_4O_7$	107.0, 113.0, 125.0, 146.1, 176.1, 233.1, 277.1, 407.1	KO/OBA
105	496.129	572.2	$C_{16}H_{23}N_4O_{14}$	148.1, 176.1, 188.1, 191.0, 217.1, 219.1, 245.1, 247.1	KO/OBA
Compound	***m/z*** **$[M-H]^-$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**	**Strain/ media**
106	329.089	754.62	$C_{15}H_{14}N_4O_5$	70.0, 98.0, 111.1, 112.0, 127.0, 140.0, 155.0, 329.1	KO/OBA
107	407.136	563.372	$C_{21}H_{20}N_4O_5$	107.0, 176.1, 188.1, 191.0, 219.1, 247.1, 259.1, 455.2	KO/OBA

Table S10 MS data for compounds in the coronafacoyl phytotoxin networks that are only present in the Δ*SCAB12101* extracts.

Compound	***m/z*** **$[M+H]^+$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**	**Strain/media**
108	306.169	607.775	$C_{17}H_{23}NO_4$	86.1, 93.1, 105.1, 107.1, 131.1, 149.1, 177.1, 195.1, 308.2	KO/YMSm
Compound	***m/z*** **$[M-H]^-$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**	**Strain/media**
109	320.541	643.0075	*	107.0, 126.1, 130.1, 161.1, 163.1, 191.1, 276.2, 320.2,	KO/MYMm, YMSm
110	356.163	938.998	$C_{21}H_{27}NO_4$	80.5, 120.8, 130.1, 137.4, 163.1, 181.2, 276.2, 320.2,	KO/YMSm

* No annotation/structure prediction

Table S11 MS data for compounds that are only present in the Δ*SCAB12101* extracts in the desferrioxamine E networks.

Compound	*m/z* $[M+H]^+$	Retention time	Formula	Key featured MS^2 fragments	Strain/media
112	617.351	506.098	C27H48 N6O10	84.1, 100.0, 100.1, 102.1, 181.1, 201.1, 219.1, 419.3	KO/ MYM
113	786.424	514.1205	C36H55 N11O9	84.1, 86.1, 100.0, 165.1, 166.1, 183.1, 201.1, 386.2	KO/ MYM
Compound	***m/z*** **$[M-H]^-$**	**Retention time**	**Formula**	**Key featured MS^2 fragments**	**Strain/media**
114	635.318	511.535	C28H44 N8O9	98.0, 136.1, 163.1, 181.1, 199.1, 217.1, 399.2, 599.3	KO/OBA
113	784.411	513.6455	C36H55 N11O9	98.0, 158.1, 178.1, 181.1, 183.1, 199.1, 237.1, 784.4	KO/ MYM

APPENDIX 4: SUPPLEMENTARY DATA

The LC-MS^2 data of WT *S. scabiei* 87-22 are available on the MassIVE public repository under the accession number MSV000085858 (https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=e08f94be407b4e9cabb24b23724657a4&view=advanced_view).

The jobs for WT *S. scabiei* 87-22 can be accessed via the following web links:

FBMN in mixed ionization mode:https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f0c337f3d4f44e9aa5acc3f2f2a84ad0.

FBMN in positive ionization mode:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=26d898d0fb114f9591dc6dea41686ecb

FBMN in negative ionization mode:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=3b14fdf65324481fa20c568edc0d0210#

NAP in positive ionization mode:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=d92ff1adc2964191b89766c1bb5a627d

NAP in negative ionization mode:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=fb9f5b648c9944bba278c00739c8cecd

MetWork projects:

https://metwork.pharmacie.parisdescartes.fr/projects

The LC-MS^2 data of Δ*SCAB12101* and WT *S. scabiei* 87-22 are available on the MassIVE public repository under the accession number MSV000086630 (https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=0d9e13631ef44cf8844984796f1190cd). The jobs for Δ*SCAB12101* and WT *S. scabiei* 87-22 can be accessed via the following web links:

Classical molecular networking in mixed ionization mode:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=afa06421667c4818958bef7672d5d57e

Classical molecular networking in positive ionization mode:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b4b4aa539e244d6a9a09e68054afa92b

Classical molecular networking in negative ionization mode:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=56ed5b2452ee4fac82d96cb1dc6053e8

NAP of Classic Molecular Networking V2 b4b4aa539e244d6a9a09e68054afa92b:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=05ca5d202e254081b16b93a1ee579f9e

NAP of Classic Molecular Networking V2 56ed5b2452ee4fac82d96cb1dc6053e8:

https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=f3145d6aeb9b4c82841df93e656520c9

DEREPLICATOR of Classic Molecular Networking V2 b4b4aa539e244d6a9a09e68054afa92b:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=3dddd57c58e54ce6b6966db8f7710c90

DEREPLICATOR of Classic Molecular Networking V2 56ed5b2452ee4fac82d96cb1dc6053e8:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7ee0825a603245cb875b97e67592ee30

Reanalysis of Classic Molecular Networking b4b4aa539e244d6a9a09e68054afa92b with MS2LDA MotifDB:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e9102828014b495899ec2989db9114a5

Reanalysis of Classic Molecular Networking 56ed5b2452ee4fac82d96cb1dc6053e8 with MS2LDA MotifDB:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=07037cf615ee448581f0263ba3fb7c7f

Molnetenhancer of Classic Molecular Networking b4b4aa539e244d6a9a09e68054afa92b:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2b9ce574beed479ea7d47ca638676beb

Molnetenhancer of Classic Molecular Networking 56ed5b2452ee4fac82d96cb1dc6053e8:

https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=5719d28b656b44a580d65f271d98afec