Evaluation of *Bacillus* biocontrol potential through the lens of secondary metabolite diversity in genomic, enzymatic, and chemical space

by

Md Jahangir Alam

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2022

Keywords: *Bacillus*, biological agents, plant pathogens, biosynthetic gene clusters, secondary metabolites, cytochromes P450

Approved by

Douglas Goodwin, Chair, Professor, Chemistry and Biochemistry, Auburn University
Holly Ellis, Professor, Biochemistry and Molecular Biology, East Carolina University
Steven Mansoorabadi, Associate Professor, Chemistry and Biochemistry, Auburn University
Angela Calderón, Associate Professor, Drug Discovery and Development, Auburn University

Abstract

Global food security is under threat by plant-pathogenic oomycetes and fungi. The overuse and misuse of pesticides against such pathogens triggers fungicide resistance and long-term environmental contamination. This warrants exploration of sustainable and eco-friendly alternative strategies to chemical pesticides. Biological agents over synthetic pesticides have been considered an attractive alternative for plant and crop protection. In particular, bacteria inhabiting plant rhizospheres are well-suited as biopesticides and biofertilizers due to their ability to suppress root-associated plant pathogens and enhance plant growth through the production of various secondary metabolites, phytohormones, and lytic enzymes. To date, many bacteria-derived secondary metabolites have been shown to control plant pathogens and promote plant growth by direct or indirect mechanisms. Such secondary metabolites are produced by a tightly linked set of enzymes encoded by biosynthetic gene clusters (BGCs). *Bacillus* species, prolific producers of such metabolites, are commonly referred to as plant growth-promoting rhizobacteria (PGPR) for their ability to spur plant growth and exert disease biocontrol. The secondary metabolites imparting these properties are structurally and chemically diverse and include lipopeptides, bacteriocins, polyketides, siderophores, and terpenes. Previous studies have demonstrated that secondary metabolites within these classes directly contribute to *Bacillus* biological activity against plant pathogens. Accordingly, secondary metabolites with growth-promotional and antimicrobial potential continue to be of great interest for controlling plant diseases, promoting plant health, and developing drugs against infectious and chronic diseases.

For more than a century, numerous antibiotics and antibiotic scaffolds have been discovered in bacteria, fungi, and plants. *Bacillus* species living in the soil and plant rhizosphere

are a potential source of novel and industrially relevant antimicrobial compounds. The full potential of such natural products is only recently becoming known due to advances in genome sequencing, bioinformatics algorithms, and analytical instrumentation such as HPLC, LC-MS, and NMR. These advances have facilitated the identification of novel enzymes with primary and accessory roles in natural product biosynthesis, including their catalytic mechanisms and the regulatory systems associated with their activity. Studies have revealed a great diversity of biosynthetic accessory enzymes in soil-borne bacteria including *Streptomyces*, *Pseudomonas*, and *Bacillus*. In particular, cytochromes P450 are often found as accessory enzymes in natural product biosynthetic pathways. To date, more than 20 P450s have been identified in the biosynthesis of different classes of natural products, including macrolides (EryF and EryK), polyenes (PimD and CYP105P1), glycopeptides (OxyB, OxyC, and OxyD), polyketides (AurH, CalO2, and CYP158), alkaloids (StaP), and fatty acids ($P450_{BioI}$ and CYP74). These P450s have been reported to catalyze broad diversity of reactions, including hydroxylation, epoxidation, heterocyclization, and ring-coupling. Although these P450s have been characterized structurally and functionally, their physiological roles are far less understood. They are anticipated to be involved in either functionalization of structurally diverse natural products to establish/maintain a competing advantage or in the metabolic breakdown of antibiotic compounds to reduce cellular toxicity. Thus, investigations of accessory P450s for physiological, catalytic, and structural properties have continued to be of great interest. Previously, analysis of 128 *Bacillus* genomes identified 112 accessory P450 enzymes in the biosynthetic of various classes of secondary metabolites. However, little is known about the structural, catalytic, and physiological roles of these enzymes.

The ***long-term goal*** of this research is to exploit *Bacillus* species as probiotics and biocontrol agents and develop their antimicrobial secondary metabolites as potential pesticides/fungicides. A deeper understanding of their biocontrol potential, through the lens of their secondary metabolite diversity in genomic, enzymatic, and chemical space, is a major step in that direction. Toward achieving these ends, the ***objective*** of this research is to identify *Bacillus* secondary metabolites, evaluate *Bacillus* biological activity against plant pathogens, and elucidate the functional roles of associated cytochrome P450 enzymes. In this study, a total of 288 novel PGPR *Bacillus* strains were evaluated as commercially viable biological agents against root-associated plant pathogens. A comprehensive discovery framework incorporating genome mining, antibiosis, chemical, and phenotypic screening was established to evaluate *Bacillus* biocontrol potential. In addition, 1,562 *Bacillus* genome sequences were analyzed to identify BGC-affiliated P450s that may be involved in functional modification of secondary metabolites. The P450s were further studied *in silico* for structural and functional features and a representative CYP102A2 predicted to be associated with plantazolicin biosynthesis was further investigated *in vitro* towards elucidating its catalytic role.

Our antibiosis screening of 288 *Bacillus* strains against a plant-pathogenic oomycete, *Phytophthora nicotianae* showed 59 strains with strong antibiosis activity, whereas 41 and 188 strains were weak and non-inhibitors, respectively. Importantly, fifty-six out of 59 strong-inhibitory strains were distributed within only five species: *B. velezensis, B. subtilis*, *B. pumilus*, *B. safensis*, and *B. altitudinis*. The high concentration of anti-*P. nicotianae* activity among five *Bacillus* species suggested these species may carry common factors responsible for this antibiosis activity. In order to investigate the breadth of antibiosis activity, the most promising 59 *P.*

nicotianae-inhibitory *Bacillus* strains were further evaluated against three root-associated plant-pathogenic fungi: *Fusarium oxysporum*, *F. graminearum*, and *Rhizoctonia solani*. All strains from *B. velezensis* and all but one *B. subtilis* strain exhibited strong antibiosis activity against all three fungi; however, strains from *B. pumilus*, *B. safensis,* and *B. altitudinis* exhibited no or weak inhibition against these fungi. Accounting for the breadth of inhibition, *B. velezensis* and *B. subtilis* were classified as "*generalists*" while *B. pumilus*, *B. safensis,* and *B. altitudinis* were classified as specialists for their relatively narrow inhibitory properties.

BGC analysis of all 288 *Bacillus* strains showed that highly conserved antimicrobial peptide and polyketide-producing BGCs were distributed among the *generalists* and *specialists* but virtually absent among non-inhibitory species. In addition, *generalists* invariably carried 2 or 3 lipopeptide BGCs that were predicted to produce fengycin and surfactin or iturin, fengycin, and surfactin, respectively. Consistent with BGC predictions, chemical analyses of extracts from the culture media of representative strains showed that iturin/bacillomycin L, fengycin, and surfactin were produced by *B. velezensis* while fengycin and surfactin were produced by *B. subtilis*. As a contrast, only surfactin among these three was produced by *B. pumilus*, *B. safensis*, and *B. altitudinis*. Evaluation of purified bacillomycin L (an iturin), fengycin, and surfactin for antibiosis activity against *P. nicotianae* and *F. oxysporum* in plate-based and 96-well plate microtiter-based assays showed each compound exerted antibiosis activity against the target pathogens.

Eighteen strains from the most promising antifungal/anti-oomycete species, *Bacillus velezensis,* were further compared for their abilities to exert intraspecies inhibition, mount intraspecies resistance, and in addition, to show antifungal activity, biofilm formation, and extent of antimicrobial secondary metabolite production. Interestingly, a wide range of intraspecies

inhibition and resistance responses were observed. The AB01 and JJ951 strains showed the most robust antibacterial activity against other *B. velezensis* strains, while AB01, JJ951, AP46, and JJ747 showed the greatest resistance to inhibition by other strains. Phylogenetic analyses based on 16S rRNA sequence from all 18 strains indicated that conserved regulatory/resistance factors may be responsible for these observed intraspecies interactions. Further, evaluation of antibiosis activity of each *B. velezensis* strain against *F. oxysporum*, *F. graminearum*, and *R. solani* showed a variable inhibitory response depending on the fungus being evaluated. Accounting for overall antifungal activity, AP215 exhibited more robust inhibition than any of the other strains evaluated, but substantial antifungal inhibition was also observed for AB01, JJ1284, AP52, JM204, AP81, AP202, JM199, and JJ747. Comparison of the production of three polyketides (bacillaene, difficidin, and macrolactin W) and three lipopeptides (iturin/bacillomycin L, fengycin, and surfactin) showed that AB01 and JJ951 produced larger quantities of these antimicrobial compounds than any other strains.

From mammals to bacteria, cytochromes P450 typically play one (or both) of two roles in metabolism, the derivatization of metabolites to specialize their function (e.g., steroid hormone biosynthesis) or to facilitate detoxification/excretion of xenobiotic compounds. Both functions could play an important role in secondary metabolite biosynthesis and resistance to toxicity exerted by the same. In order to investigate the diversity of antimicrobial secondary metabolite scaffolds, how that influences the variation in their bioactivity, and to what extent this is mediated by cytochromes P450, a comprehensive analysis was carried out using 1,562 *Bacillus* genomes. From a total of 5,051 cytochrome P450 genes, we identified 614 integrated within biosynthetic gene clusters (BGCs) as “accessory genes”. The most common BGC-affiliated P450 families were

CYP113, CYP134, CYP109, CYP107, and CYP102 and these were associated with the secondary metabolite BGCs of difficidin, cyclodipeptide, bacillibactin, bacillaene, and plantazolicin, respectively. Interestingly, amino acid sequence conservation of enzymes within each P450 family showed CYP113 and CYP134 were the most highly conserved, whereas CYP107 and CYP109 were the least conserved. This observation indicated that phylogenetically-related amino acid sequence conservation of each P450 family may be linked with substrate specificity with respective biosynthetic pathway. High-resolution homology models of a representative enzyme from each BGC-affiliated P450 family showed distinct structural features at the active site that may be related to substrate specificity with the respective biosynthetic pathway. Molecular docking simulations of each P450 with candidate substrates showed favorable binding of CYP113, CYP134, CYP109, CYP107, and CYP102 with the putative substrates difficidin, cyclodipeptide, dihydroxybenzoate, bacillaene, and plantazolicin, respectively.

To facilitate elucidation of a functional role for a CYP102 in plantazolicin biosynthesis, a representative CYP102A2 ($P450_{BM3}$) protein from *Bacillus amyloliquefaciens* FZB42 (*Ba*CYP102A2) (NCBI accession: WP_012117030) was synthesized, cloned into a pET21 vector, and used to transform *E. coli* (BL21-[DE3]) for expression. The enzyme was purified using Ni-NTA affinity, anion-exchange, and size-exclusion chromatographies. The identity of full-length *Ba*CYP102A2 enzyme was confirmed by the presence of a major band at ~120 kDa in SDS-PAGE gel, a heme-based Soret band at 418.5 nm in UV-vis spectra, along with the type-I spectral shift upon substrate binding and sigmoidal and hyperbolic kinetic responses of NADPH oxidation by *Ba*CYP102A2 in the presence of the archetypal substrates sodium dodecyl sulfate (SDS) and oleic acid (OA), respectively. Fitting of substrate titration curves and steady-state kinetic responses of

*Ba*CYP102A2 using the substrates SDS, myristic acid (MA), and palmitic acid (PA) produced sigmoidal responses while OA produced a hyperbolic response.

In summary, the research of this dissertation has revealed that *Bacillus* antibiosis against *P. nicotianae* is highly conserved within five *Bacillus* species (*B. velezensis*, *B. subtilis*, *B. pumilus*, *B. safensis*, and *B. altitudinis*). *B. velezensis* and *B. subtilis* exhibited broad antibiosis activity against *P. nicotianae* as well as multiple fungal pathogens. Comparison of genomic, antibiosis, and chemical profiles showed that the production of phylogenetically conversed lipopeptides correlates with these observed antibiotic abilities. Further quantitative evaluation of 18 *B. velezensis* strains showed wide range of intraspecies inhibition activity, resistance to such inhibition, and antifungal activity that may be linked with the expression of regulatory/resistance factors and/or production of antimicrobial compounds. Strains AB01 and JJ951 exhibited robust intraspecies inhibition activity, and strains AB01, JJ951, AP46, and JJ747 exhibited the greatest resistance to inhibition by other *B. velezensis* strains. Strain AP215 exhibited the most robust antifungal activity. Evaluation of phylogeny and amino acid sequence conservation of 614 BGC-affiliated *Bacillus* P450s showed enzymes from each P450 family may be linked with specific secondary metabolite biosynthesis. *In silico* structural and substrate-binding evaluation of a representative enzyme from five BGC-affiliated P450 families showed distinct structural features that matched with favorable substrate binding associated with the respective secondary metabolite biosynthesis. Finally, a representative *Ba*CYP102A2 enzyme predicted to be involved in plantazolicin biosynthesis was expressed, purified, and further evaluated for substrate binding titration and steady-state kinetic responses with typical CYP102 substrates.

Acknowledgments

"If I have seen further, it is by standing on the shoulders of giants."— Isaac Newton

It's been an honor and great privilege to work with you, Dr. Goodwin. I've learned from each of your actions (and inactions) throughout my stay at Auburn. You have been decisive, precise, clear, bold, and patient in each step to guide me in the right direction. I am truly indebted to your contribution to making me a scientist. You have been the light in the darkness, steadfast in turmoil, and a force in the face of obstacles. I came to Auburn not only for a degree but also to learn communication, leadership, and collaboration. You provided me a great platform to learn and collaborate more than what I could ever ask for. You taught me how to be a good man, husband, father, son, and responsible human being. I couldn't ask anything better. You've become a mentor and a role model for rest of my life. Thank you.

My sincere thanks to my committee members, Dr. Mansoorabadi, Dr. Ellis, and Dr. Calderon. Dr. Mansoorabadi, thanks for your guidance and suggestions in many aspects of my research. Also thank you for providing opportunities to meet with guest lecturers and fellow graduate students to discuss science and prospective careers. You always have been a cheerful person in our biochemistry seminar, coming up with critical questions and suggestions. Thanks for your support throughout my stay at Auburn. Dr. Ellis, thanks for your help and advice in my research as well as my career. You never hesitated to give me suggestions, and you were always ready to take any questions and concerns related to my research as well as personal matters. You were very helpful and approachable with questions and suggestions. Dr. Calderon, thanks for your

critical questions, suggestions, and feedback on my research. I have immensely benefitted from your guidance and your mass spectrometry class.

Dr. Liles, it's been a great opportunity to work with you. You helped me to become a good scientist, critical thinker, storyteller, and leader. I learned a great deal from each of our meetings. You have been a force of nature in lifting me up and showing great interest in my work. In each of our meetings and e-mail correspondence, I felt that my effort and research were worthwhile. You are an exceptional scientist, mentor, and leader who always encourage students to become independent scientists. Thank you for providing the necessary training and opportunity to work in your lab. Dr. Noel, you have been a great person to work with and talk to. You have always been open to any question and happy to provide the resources I needed. Also, thanks for giving me access to your lab. Thank you for being very supportive throughout my stay at Auburn. Many thanks to Dr. Kloepper for his effort in collecting *Bacillus* strains and allowing us to research on them. Also, thanks for your huge contribution to the field of plant-growth promotion and rhizobacteria.

I am grateful forever to my parents who brought me into this world and raise me to be a strong, independent, and kind human being. I am truly grateful to my wife, Maliha. Thank you for your patience and sacrifices. I wouldn't be able to come to this stage without your constant support and encouragement. Areebah, my dear daughter, you are a blessing in my life. Watching you grow up inspired me to keep working and pushing through good and bad days. You have been my inspiration and my reason to work as hard as I could.

I would like to thank my lab mates, Hui Xu, Jessica Krewall, Tarfi Aziz, Callie Barton, Rejaul Islam, and Chidozie Ugochukwu. Hui, thanks for sharing your ideas on doing research and writing the dissertation. Jessica, thanks for teaching so many lab techniques, being always supportive, and helping me whenever I needed it. You've been a great trainer on instrumentation, a great teacher of scientific concepts, and a role model. Tarfi, thanks for the many discussions we had on science and career. You helped me to keep my feet on the ground and taught me many lab techniques and preparation for the future. Callie, thanks for being so helpful and providing support in our lab. You have been the most reliable person to talk about many aspects of our lab. Rejaul, thank you for training me on how to work in a biochemistry lab. You've been a tremendous help throughout my stay at Auburn. You've made my research easier. You trained me to work on protein expression, purification, and characterization. Chidozie, thanks for your time in discussing so many topics, including research, science, scientist, career, and mentorship. I learned immensely about so many aspects of your life and experiences. You have been my go-to buddy to share my thoughts and get immediate feedback on almost anything. Rene Fuanta, thanks for the frequent visits and helpful discussions about my research and career choices. You have been a role model throughout my stay at Auburn.

I would like to thank the three undergrad students who worked with me over the past two years. Raegan, I am not sure how much you've learned from me, but I learned a lot from you. I am impressed with your persistence, eagerness to keep trying, and not to give up. Ben, thanks for showing up, and keep trying on the task that I gave you. You taught me how not to give up and how to show up even though things are not working. You are a great person to work with. Nina,

it’s been a great opportunity to work with you. I learned a lot from you about how to speak convincingly, how to present, how to multitask, and in fact, how to sell.

I would like to thank the Bangladeshi student community at Auburn University for supporting me in many aspects of my personal and academic life. Many thanks to Shamim Iqbal, Nirob Saha, and all current Bangladeshi students in the Chemistry and Biochemistry department for helping me throughout my stay at Auburn.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

PGPR	**P**lant **G**rowth-**P**romoting **R**hizobacteria
BGC	**B**iosynthetic **G**ene **C**luster
NRPS	**N**on-**R**ibosomal **P**eptide **S**ynthetases
PKS	**P**olyketide **S**ynthases
PKS-NRPS	**P**olyketide **S**ynthases-**N**on-**R**ibosomal **P**eptide **S**ynthetases hybrid
RiPP	**R**ibosomally synthesized and **P**ost-translationally modified **P**eptide
CDPS	Cyclodipeptide **S**ynthases
antiSMASH	**anti**biotics & **S**econdary **M**etabolite **A**nalysis **S**hell
BiG-SCAPE	**B**iosynthetic **G**ene **S**imilarity **C**lustering and **P**rospecting **E**ngine
P. nicotianae	***P****hytophthora nicotianae*
F. oxysporum	***F****usarium oxysporum*
F. graminearum	***F****usarium graminearum*
R. solani	***R****hizoctonia solani*
ISR	**I**nduced **S**ystemic **R**esistance
EPA	**E**nvironmental **P**rotection **A**gency
VOC	**V**olatile **O**rganic **C**ompound
HPLC	**H**igh-**P**ressure **L**iquid **C**hromatography
MS	**M**ass **S**pectrometry
ESI	**E**lectrospray **I**onization

LC	**L**iquid **C**hromatography
DAD	**D**iode **A**rray **D**etector
TIC	**T**otal **I**on **C**urrent
EIC	**E**xtracted **I**on **C**hromatogram
NMR	**N**uclear **M**agnetic **R**esonance
SDS	**S**odium **d**odecyl **s**ulfate
MA	**M**yristic **A**cid
PA	**P**almitic **A**cid
OA	**O**leic **A**cid
AMP	**A**mpicillin
CAM	**C**hloramphenicol
IPTG	**I**sopropyl-β-d-thiogalactopyranoside
PMSF	**P**henylmethylsulphonyl fluoride

Chapter 1: Literature review

1.1. Introduction

Global food production must increase to feed the increasing populations in the future (1). In order to enhance crop yield, improve food quality, and control plant diseases, there is an urgent need for effective and sustainable fertilizers and pesticides. Along with the abiotic stresses (i.e., excessive heat, drought, and cold), the biotic stresses imposed by plant pathogens such as bacteria, fungi, oomycetes, and viruses need to be controlled effectively (2, 3). In particular, plant-pathogenic fungi and oomycetes cause substantial economic loss worldwide (4). In addition to known plant pathogens, emerging pathogens are considered a rising threat to global food security (4). Use of chemical pesticides to combat plant pathogens, a century-old disease control strategy, is no longer working effectively. Since the 1960s, the efficacy of multi-site and single-site pesticides has been declining rapidly due to the loss of plant pathogen sensitivity to these agents (5). In particular, pesticide resistance has been detected in plant pathogens against the most effective single-site pesticides, including benzimidazoles, phenylamines, dicarboximides, and quinone outside inhibitors (QoIs) (6, 7). The lack of novel modes of action accompanied by declining innovations in pesticide discovery warrants exploration of effective and alternative strategies for sustainable pathogen control.

Bacteria living in the plant-rhizosphere develop various chemical interactions with plants; these include beneficial, neutral, and detrimental interactions. Some bacteria that inhabit soil and plant-rhizosphere promote plant growth directly or indirectly. For such beneficial roles, they are known as plant growth-promoting rhizobacteria (PGPR) (8, 9). PGPR serve as biofertilizers through one or more of multiple mechanisms, including nitrogen fixation, phosphate

solubilization, iron acquisition, and modulation of phytohormones such as indole acetic acid (IAA), cytokinin, gibberellin, and ethylene (10, 11). In addition, they act as biocontrol agents by producing antibiotics and lytic enzymes, competing against pathogenic microorganisms for nutrients, and triggering induced systemic resistance (ISR) against plant pathogens (9). An enormous effort initiated by public research institutions and agrochemical companies has yielded several commercial bioproducts as biofertilizers and biocontrol agents in the past three decades. Some prominent PGPR strains registered with the U.S. Environmental Protection Agency (EPA) include *Agrobacterium radiobacter* K84, *Aspergillus flavus* AF36, *Bacillus licheniformis* SB3086, *Bacillus pumilus* GB34, *Pseudomonas chlororaphis* 63-28, and *Streptomyces griseoviridis* K61 (12).

Bacteria from the genus *Bacillus* are rod-shaped, Gram-positive, and endospore-forming microbes that inhabit soil, water, plant-rhizosphere, and extreme environments (13, 14). They are well known as a predominant PGPR due to their inherent ability in plant growth and disease biocontrol, facilitated by their well-adapted physiology in the plant-rhizosphere. Past studies have demonstrated that strains from *Bacillus subtilis*, *Bacillus velezensis*, *Bacillus pumilus*, and *Bacillus thuringiensis* have excellent abilities in plant growth promotion and disease suppression (15). This ability is primarily attributed to their role in phosphate solubilization, nitrogen fixation, siderophore production, phytohormone modulation, and production of volatile organic compounds (VOCs) (15, 16). To date, more than two dozen antimicrobial and industrially-relevant compounds with diverse physicochemical properties have been identified in the *Bacillus* species (15). The most common classes of such natural products are cyclic lipopeptides, macrolides, bacteriocins, and terpenes (17). These are widely known as secondary or specialized metabolites since they are

nonobligatory for growth and reproduction; however, they are advantageous in specific ecological niches. Some of the *Bacillus*-based commercial biological agents distributed worldwide are Avogreen (*B. subtilis*), Ballad Plus i Sonata (*B. pumilus*), RhizoVital 42 (*B. amyloliquefeciens*), EcoGuard TM (*B. licheniformis*), and Botrybel (*B. velezensis*) (15).

Since the early 20th century, bioassay-based chemical fermentation has been the primary strategy for natural product discovery (18, 19). However, this process became inefficient over time due to a high rate of compound rediscovery coupled with a lack of novelty of bioactive compounds. This caused a rapid decline in the rate of novel antibiotic discovery in the 1970s (19, 20). Fortuitously, the genomic revolution, accompanying the advancement of computation and bioinformatics algorithms in the early 21st century has promised to restore the progress in natural product discovery (21, 22). An increasing amount of publicly accessible genome sequences, fueled by low sequencing cost, has allowed virtually anyone to predict the biosynthetic pathways of natural products encoded by organisms of interest. Today, hundreds of bioinformatics tools, databases, and specialized discovery pipelines are publicly available, enabling investigators to identify, analyze, compare, and dereplicate biosynthetic gene clusters (BGCs) that may produce antibiotic natural products (23–25). In parallel, significant advances have taken place in analytical instrumentations for high-throughput and reliable chemical screening and /or unequivocal chemical identification. These include NMR, MS, and HPLC. In addition, several integrated and streamlined genomic (PRISM4, GNP) and chemoinformatic (NPClassifier, NC-MFP, MAP4) pipelines have been developed in this area for a more specialized discovery pipeline (26–30). This allows reliable predictions of natural product structure from genome sequences as well as analysis of a large amount of chemical data produced by analytical instrumentation (26, 31). Over the past

two decades, several multi-faceted technologies have also been developed to discover untapped microbial biosynthetic potential at a large scale, including functional genomics, synthetic genomics, and metagenomics.

1.2. Plant pathogens

Continued and sustainable improvement in crop yield and food quality is crucial to ensure global food security into the foreseeable future (4). Since the agricultural revolution, plants and crops have always been under threat from various abiotic and biotic stresses (3). Major abiotic stresses include drought, heat, salinity, and cold cause crop failure and food scarcity. Biotic stresses such as insects, weeds, and plant-pathogenic organisms are equally, if not more, destructive to crop yield and food quality. Plant-pathogenic microorganisms are found across multiple kingdoms, including bacteria, viruses, fungi, oomycetes, and nematodes (2). In agriculture, each pathogen or a combination of pathogens, brings unique challenges to detection, treatment, and sustainable protection of plants and crops. In particular, some of the planet's most prolific and destructive plant-pathogens are fungi and oomycetes and account for enormous losses in crop yield on an annual basis (32). In 2008, the projected economic loss from the crop failure due to invasive fungal pathogen was estimated to be $21 billion per year in the United States (33).

1.2.1. Plant-pathogenic fungi and oomycetes

Fungi and oomycetes are the most notorious eukaryotic plant pathogens (32). They cause substantial economic loss in agriculture and lead to hunger, starvation, and famines by triggering total crop failure. These two pathogens have been a threat to plants and crops throughout the history of agriculture. Among many, two major famines that were caused by these two pathogens are the great Irish potato famine in 1845 and the great Bengal famine in 1943 (34, 35). The Irish

potato famine was originally triggered by an oomycete, *Phytophthora infestans*, the cause of potato late blight. The Bengal famine was triggered by a fungus, *Cochliobolus miyabeanus*, the causative agent of brown spot disease of rice and maize. Based on their scientific/economic impact, Dean *et al.* recently ranked fungal pathogens from a survey of the international community (36). According to the survey, the top 10 fungi are, in order from first to tenth, *Magnaporthe oryzae*, *Botrytis cinerea*, *Puccinia spp., Fusarium graminearum*, *F. oxysporum*, *Blumeria graminis*, *Mycosphaerella graminicola*, *Colletotrichum spp.*, *Ustilago maydis*, and *Melampsora lini*. Another similar survey was conducted by Kamoun *et al.* to rank the oomycete pathogens based on their scientific and economic impact (37). In order from first to tenth, the top ten are *Phytophthora infestans*, *Hyaloperonospora arabidopsidis*, *P. ramorum*, *P. sojae*, *P. capsica*, *Plasmopara viticola*, *P. cinnamomi*, *P. parasitica*, *Pythium ultimum*, and *Albugo candida*.

Major crops that are affected by fungal pathogens include rice, wheat, corn, and many vegetables (38). *Magnaporthe oryzae* is the most destructive disease of rice that causes rice blast, a threat to approximately one-half of the world's population (39). *Botrytis cinerea*, known as grey mold, is the second most significant fungal species in terms of pathogenic destruction, that can infect more than 500 plant genera and 150 botanical families (40). *Fusarium graminearum* is a highly destructive fungal pathogen that causes enormous economic loss by infecting many cereal species (e.g., wheat, barley, oats, and corn) (41). *Fusarium oxysporum* is a ubiquitous soil-borne fungal pathogen that causes a wide range of plant diseases (42, 43). *Rhizoctonia solani* causes a variety of diseases in soybean, sugar beet, rice, turfgrass, and potatoes (44–46).

Fungal-like pathogens called oomycetes (also known as water molds) are a eukaryotic organisms that causes enormous economic loss in agriculture (32). They resemble fungi in some

aspects of phenotype including mycelial growth and mode of nutrition; however, they are phylogenetically distantly related (Fig. 1.1A) (47).

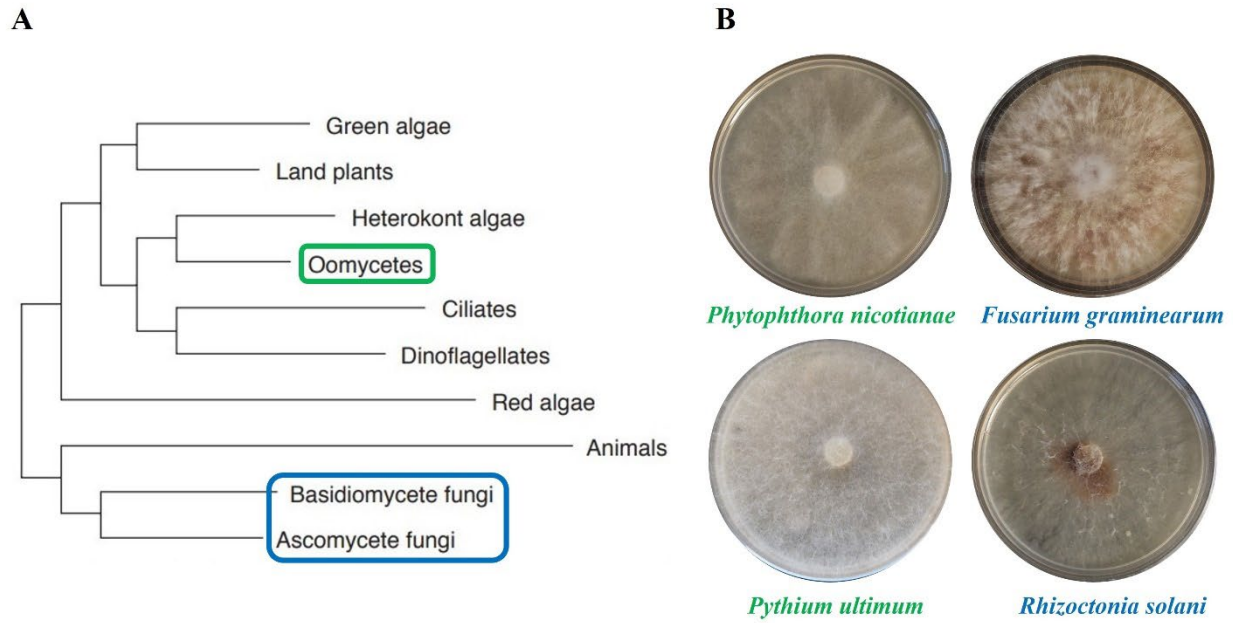


Figure 1.1: Phylogeny and growth morphology of oomycetes and fungi. (A) The phylogenetic tree shows evolutionary relationships between the major eukaryotic groups; reproduced from (48) and adapted from (49) (B) Growth morphology of plant-pathogenic oomycetes (*Phytophthora nicotianae* and *Pythium ultimum*) and fungi (*Fusarium graminearum*, and *Rhizoctonia solani*).

Plant-pathogenic oomycetes cause root rot in a wide range of hosts, including soybean, tomato, pea, sugar beet, and tomato (50). The two most economically important oomycetes genera in terms of plant pathogenicity are *Phytophthora* and *Pythium* (51). Diseases caused by *Phytophthora* include late blight of potato and tomato, and root/stem rots of many plant species (52). *Pythium* causes a wide range of diseases, including root rots, blight, and seed damping-off of many types of plants (53). The *Phytophthora infestans* is a very destructive plant pathogen that causes potato late blight (54). The *Phytophthora sojae* is known as a root rot pathogen that causes destructive diseases to soybean and various agriculturally and ornamentally important plants (55).

Phytophthora capsici is known as a highly dynamic and destructive pathogen of vegetables, including cucurbits, pepper, tomato, eggplants, snap, and lima beans (56). *Pythium ultimum* causes damping-off and root rot diseases of diverse plant species, including corn, soybean, potato, wheat, and ornamental plants (37).

Fungi and oomycetes are highly similar in their growth morphology, but they are phylogenetically distantly related (Fig. 1.1A) (32). They are fundamentally different in their physiology, biochemistry, and genetics. For instance, fungi are generally haploid, whereas oomycetes are diploid; fungal hyphae are septate, whereas oomycete hyphae are non-septate; fungal cell walls consist of chitin, whereas oomycete cell walls mainly consist of 1,3-/1-6- or 1-4-β glucans (cellulose) (57, 58). Even with these differences, both organisms are remarkably similar in their mode of action to attack plants and defend toxic compounds produced by plants. Both groups of organisms utilize an extensive toolbox of virulence factors, including cell-wall-degrading enzymes (CWDEs) and elicitors (pathogen-derived molecules) (59–61). The primary roles of these factors include loosening and otherwise compromising the plant cell wall, to facilitate successful penetration and acquisition of nutrients, countering/nullifying plant defense compounds/mechanisms, and disrupting/silencing the defense-related genes of the host plant (62, 63). In addition, they have an incredible ability to rapidly evolve their pathogenicity factors to counter the plant defense or hostile environment created by plant (64, 65).

1.2.2. Existing treatments against plant-pathogenic fungi and oomycetes

Crop-destroying pathogens, including fungi and oomycetes, account for nearly 30% of combined pre and post-harvesting agricultural loss (66). The five most heavily-produced foods worldwide (rice, wheat, sugarcane, maize, and soybean) are all under threat from these pathogens.

For more than a century, fungicides (including oomycides) have been used against plant-pathogenic fungi and oomycetes for crop protection (67).

Since the early 19th century, inorganic chemicals/materials such as saltwater, copper, sulfur, and phosphorous have been used as fungicides as a plant protection strategy (68). The use of organic fungicides began in the 1940s (6). To develop highly efficient, less toxic compounds for protecting plants and improving crop yield, the agrochemical industry has engaged further research on organic fungicides. Their efforts have led to the commercialization of several organic fungicides between the 1940s and 1970s with a multi-site mode of action. Some of these broad-spectrum fungicide families are dithiocarbamates, phthalimides, and chlorothalonil (6). In the 1960s, new fungicides with a site-specific modes of action were developed to boost crop protection (6). Prominent single-site fungicides are methyl benzimidazole, carbamates, demethylation inhibitors, the quinone outside inhibitors, and succinate dehydrogenase inhibitors (67). Currently, fungicides with more than forty-five different modes of action are commercially available for controlling plant pathogenic fungi and oomycetes (6).

1.2.3. Pesticide resistance mechanism and consequences

The multi-target fungicides and inorganic chemicals have remained effective and less susceptible to resistance against plant-pathogenic fungi and oomycetes. This is mainly because the resistance development is slow and unfavorable against the non-systematic fungicides with multiple modes of action. However, the site-specific fungicides are highly susceptible to resistance development due to their single modes of action, such as disrupting a particular cellular process or binding to a specific protein in an important cellular process (67). The confirmed case of resistance against site-specific fungicides was first documented in the 1960s, a few years after the use of site-

specific fungicides (67). Since then, fungicide resistance among fungi and oomycetes has remained a challenge to food and crop protection.

Major factors contributing to the development of resistance are the chemistry and mode of action of the fungicidal compound, the biology and reproductive ability of the target pathogen, and the frequency of fungicide application (69). The most common resistance mechanism of site-specific fungicides is the development of a single mutation in the target protein (altered target site) that reduce the sensitivity of the target protein to the fungicide in question (70, 5). Other mechanisms include (a) synthesis of an alternative enzyme that substitutes for the target enzyme, (b) overproduction of the fungicide target, (c) an active efflux or reduced uptake of the fungicide, and (d) a metabolic breakdown of the fungicide (5). Fungicide resistance against the non-specific, multi-site fungicides is less common; however, their appearance has been reported as well. One of the main mechanisms of the resistance development involves the overexpression of drug efflux transporters, including the ABC transporter AtrB and the MFS transporter MfsM2 (69). Resistance to demethylation inhibitors (DMIs) was observed due to the mutation in the 14α demethylase (CYP51) gene and the overexpression of the CYP51 gene (5). Further, resistance to benzimidazole was reported to be correlated with point mutations at the benzimidazole-binding site of the target protein β-tubulin (71).

1.2.4. The urgency of alternative and sustainable strategies

Fungicide resistance is considered a rising threat pose by plant-pathogenic fungi and oomycetes. A recent case study showed that *Botrytis cinerea,* a major pathogen of cultivated fruits, vegetables, and ornamental flowers is a "high risk" organism for rapid resistance development (72). This fungus poses a high resistance risk to major fungicides, including benzimidazoles,

dicarboximides, anilinopyrimidines, strobilurins, and SDHIs (boscalid). Hahn *et al.* reported that there are no equivalent alternative chemical protection against this highly destructive pathogen (69). Most commonly applied fungicides for crop protection inhibit either sterol biosynthesis, respiration, methionine biosynthesis, tubulin function, or signal transduction. Fungicide resistance to all these modes of action has already been reported. It has been speculated that the lack of alternative modes of action may seriously confront the durability of anti-resistance strategies (6). For a sustainable fungal control system, a better understanding of the biochemistry of the development of resistance and pathogenicity would provide sufficient resources to identify novel targets and develop effective strategies to exploit them. Along with finding novel modes of action, preventative measures such as risk assessment and resistance management with integrated disease management should be considered. Another protective measure is the early diagnosis in molecular detection of gene expression associated with fungicide resistance using PCR-based detection technologies (5).

The biological agents are defined as the disease-suppressive microorganisms that can improve plant health and suppress plant pathogens (73). To date, biocontrol is one of the poorly understood areas of plant-microbe interactions since it involves interactions among the plant, the pathogen, the biocontrol agent, and the microbial community. Despite the complexity and lack of understanding, the rate of implementation of commercial biocontrol products is increasing rapidly. As of 2017, biological control represented about 5% of the total agrochemical pesticide sales, and this figure is expected to increase by 10-15% per year (74).

1.3. PGPR as promising biofertilizers and biopesticides

Biological agents have been considered promising alternatives to synthetic fungicides against plant-pathogenic fungi and oomycetes. Bacteria inhabiting plant rhizospheres have been shown to suppress plant pathogens and promote plant growth (8, 9), earning them the title plant growth-promoting rhizobacteria (PGPR) (8). In recent years, an increasing interest in the biological control of plant pathogens has been observed due to the urgency of environmentally friendly alternatives to chemical pesticides (75). Biological agents have been considered more sustainable than synthetic fungicides and are anticipated to be more specific to target pathogens as compared to the chemical fungicides (76, 77). In addition to the efficacy in protection, many strains from the genera *Bacillus*, *Pseudomonas,* and *Agrobacterium* play an important role in plant growth (15). Moreover, these PGPR can produce heat and desiccation-tolerant endospores, maintain high cell viability, and have a prolonged shelf-life (78).

It has been shown that PGPR can increase crop yield by as high as 57%, depending on the type of crop (78). The mechanisms to increase plant growth fall into two main categories: direct and indirect (Fig. 1.2). For direct growth promotion, they can serve as sources of growth factors, nutrients, and/or hormones. Examples of the direct growth promotion mechanisms include nitrogen fixation, phosphate solubilization, iron acquisition by bacterially-generated siderophores, and regulation of hormones (indole acetic acid, cytokinin, and gibberellin) (78–83). On the other hand, indirect plant growth promotion occurs by suppressing the growth or activity of pathogenic microorganisms inhabiting plant rhizosphere through the production of antibiotics, and lytic enzymes, competing for nutrients, and triggering induced systemic resistance (ISR) against pathogens (9, 84–86). In the future, PGPR are expected to be used widely as alternatives to

chemical pesticides in agriculture, horticulture, silviculture, and environmental cleanup strategies (86).

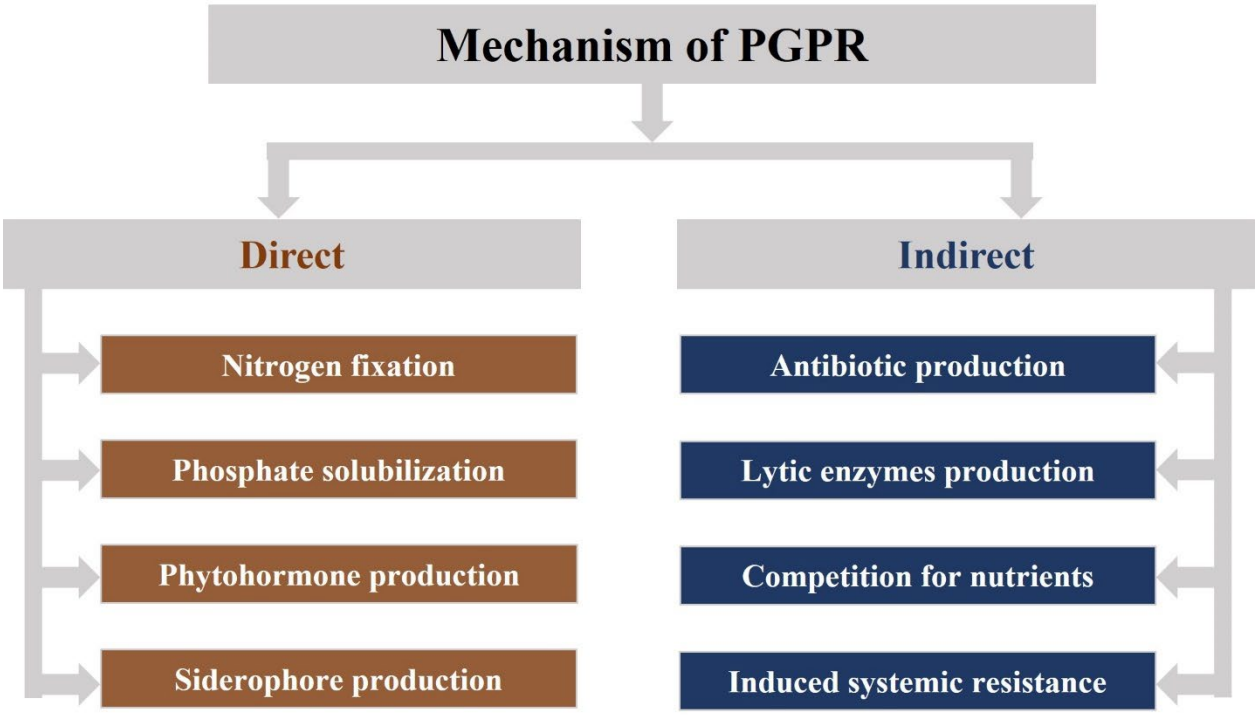


Figure 1.2: Direct and indirect modes of plant growth promotion by PGPR (9, 86, 87).

1.3.1. PGPR as biofertilizers

PGPR from the genera *Sinorhizobia*, *Bradyrhizobia*, *Mesorhizobia*, *Azotobacter*, *Azospirillum*, *Burkholderia*, and *Bacillus* fix atmospheric nitrogen in soil and make it available to plants. By utilizing nitrogenase enzyme (*nifH*) PGPR reduce atmospheric nitrogen (N_2) to ammonia (NH_3) in the first step of nitrogen fixation process. Although phosphorus is abundant in the plant rhizosphere, it is largely inaccessible due to its poor solubility (10). PGPR from the genera *Azospirillum*, *Bacillus*, *Rhizobium*, and *Pseudomonas* produce organic acids such as gluconic and citric acids as well as phosphatases that solubilize inorganic phosphate and make it accessible to plants. PGPR also play important roles in modulating phytohormone homeostasis that facilitate

plant growth and development (88). It has been confirmed that PGPR can produce plant growth-promoting phytohormones such as auxin, cytokinins, ethylene, gibberellins, abscisic acid, salicylic acid, and jasmonic acid (86). *Aeromonas punctata*, *Serratia macrcescens*, and *Azospirillum brasilense* have been shown to stimulate and induce morphological changes of plant root through the production of auxin (89). *Burkholderia* and *Phyllobacterium* have been shown to influence plant ethylene homeostasis (90, 91). *Bacillus subtilis*, *Pseudomonas fluorescens*, and *Burkholderia phytofirmans* synthesize the phytohormone cytokinin and influence the level of plant cytokinin concentration. *Pseudomonas, Bacillus, Stenotrophomonas, Serratia*, and *Arthrobacter* have been shown to influence plant growth by producing volatile organic compounds, including 2,3-butanediol, acetoin, 2-pentylfuran, E-11, 13-tetradecadien-1-ol, 2-butanone, and 2-methyl-*n*-1-tridecene (92). In addition, some PGPR directly produce indoleacetic acid and ethylene and/or stimulate plants to do the same. These are two of the most important biostimulants that affect plant growth, cell division, extension, differentiation, photosynthesis, pigment formation, biosynthesis of various metabolites, and nodule formation (93, 94).

In the plant rhizosphere, PGPR take up otherwise insoluble ferric iron (Fe^{3+}) through the action of various siderophores, and subsequently supply it to the plants in a soluble and bioavailable form (95). This can contribute to a deficiency of iron in the rhizosphere for other competing and/or pathogenic microbiomes on one hand and benefits plants with iron uptake on the other. Indeed, iron is a limiting factor for growth in nearly all biological contexts. Siderophores are invariably constructed to orient multiple functional groups for the near ideal coordination of Fe. In chemical terms, siderophores rely on four types of functional moieties for metal coordination: catecholates, phenolates, hydroxamates, and carboxylates (96). Siderophores forms

tight (association constant ~10^{20} to 10^{30} M^{-1}) and stable complexes with ferric iron which are then transported into the cytosol (97). In addition to iron scavenging, siderophores can have alternative functions, including non-iron metal transport, toxic metal sequestration, signaling, protection from oxidative stress, and antibiotic activity (96). In light of the importance of iron acquisition for growth and survival, it should come as no surprise that more than 500 siderophores of bacterial origin have been identified (88). Among PGPR, *Bacillus velezensis*, *Bacillus amyloliquefaciens*, and *Pseudomonas fluorescens* are but a handful of examples of bacteria whose siderophore production and excretion produces a net benefit to the plants with which they associate (98). It has been shown that *Pseudomonas* can exert plant-growth-promotion ability by depriving native microflora of iron, and at the same time, stimulate increases in crop yield of 144% (95).

1.3.2. PGPR as biopesticides

Several PGPR have been reported to be antagonistic against insects, fungi, bacteria, viruses, weeds, and/or nematodes. Their biocontrol abilities are mainly mediated by the production of antibiotics, lytic enzymes, siderophores, competition for nutrients, niche exclusion, and through the triggering of a plant's induced systemic resistance (ISR) (9). *Pseudomonas*, *Bacillus*, *Azospirillum*, *Rhizobium*, and *Serratia* produce one or more of the following antibiotics: phenazine-1-carboxyclic acid, 2,4-diacetyl phloroglucinol, oomycin, pyoluteorin, pyrrolnitrin, kanosamine, zwittermicin-A, and pantocin (99). Some PGPR produce lytic enzymes such as chitinases, cellulases, β-1,3-glucanases, proteases, and lipases to disrupt the cell wall, cell membrane and associated proteins of many pathogenic organisms (100). These lytic enzyme-producing PGPR strains have shown bioactivity against a range of phytopathogens, including *B. cinerea*, *Sclerotium rolfsii*, *F. oxysporum*, *Phytophthora spp.*, *R. solani*, and *Pythium. ultimum*.

Chitinase of *B. thuringiensis* has been demonstrated to exhibit insecticidal activity through the hydrolysis of chitin (101). Production of chitinase by *Pseudomonas* strains has been shown to be correlated with antifungal activity (102). PGPR-derived cellulase has been reported to enhance root colonization through the increase of nodulation. Cellulases have also been reported to contribute to cytolytic effects in *Phytophthora parasitica* (103). Some PGPR exert antibiosis activity against plant pathogens by producing a variety of volatile organic compounds (VOCs). Examples of prominent fungicidal VOCs include S-methyl methanethiosulfonate, 1,3,5-trichloro-2-methoxy benzene, dimethyl sulfide, *S*-methyl thioacetate, and methyl thiocyanate (104, 105).

1.3.3. Commercial PGPR bioproducts for agricultural applications

The first commercial biological agent against plant pathogens is *Agrobacterium radiobacter*, developed in 1979 in the U.S. (12). Since then, the most successful bioproducts against plant diseases have been developed based on the PGPR *Bacillus* species, including *B. thuringiensis* and *B. subtilis* (106, 107). *Bacillus*-based bioproducts are most effective against *Fusarium* and *Rhizoctonia* and could be used in cotton, peanut, soybean, corn, vegetables, and small grain crops. Besides *Bacillus spp.*, other PGPR used as active ingredients in commercial bioproducts are *Agrobacterium*, *Azospirillum*, *Burkholderia*, *Pseudomonas*, and *Streptomyces* (12). Some of the most important commercial biological agents, associated bioproducts, and intended crops are presented in Table 1.1.

Table 1.1: Prominent commercial PGPR-based biofertilizers and biopesticides (9, 12, 15, 108)

PGPR Strains	Name of bioproducts	Intended crop
Agrobacterium radiobacter	Diegall, Galltrol-A, Nogall, Norbac 84 C	Fruit, nut, ornamental nursery stock, and trees
Ampelomyces Quisqualis	AQ10 BioFungicide	Fruit, vegetable, and ornamental crops
Aspergillus flavus	Aspergillus flavus AF36, Afla-guard	Cotton, Peanut
Azospirillum brasilense	Azo-Green	Turf and forage crops
Bacillus licheniformis	Ecoguard; Novozymes Biofungicide Green Relief	Ornamental plants and ornamental turf
Bacillus pumilus	GB34 Concentrate Biological Fungicide	Soybean
Bacillus amyloliquefaciens	Quantum 4000	Broccoli, cabbage, cantaloupe, cauliflower, celery, cucumber, lettuce, ornamentals, peppers, tomato, and watermelon
Bacillus subtilis	Epic, HiStick N/T, Kodiak, Rhizo-Plus, Serenade, Subtilex, Taegro	Barley, beans, cotton, legumes peanut, pea, rice, and soybean
Bacillus thuringiensis	Biobit, Dipel, Delfin, Certan, Acrobe, Skeetal, Vectobc, Trident, Novodor, Foil	Vegetables, forestry, apiculture, mosquito control
Burlkholderia cepacian	Blue Circle, Deny, Intercep	Alfalfa, barley, beans, clover, cotton, maize, peas, sorghum, vegetables, and wheat
Candida oleophila	Aspire	Postharvest diseases
Gliocladium catenulatum	Primastop	Soilborne pathogens
Gliocladium virens	Soilgard	Ornamentals, vegetables, cotton
Pseudomonas aureofaciens	Bio-Ject Spot-Less	Golf course turf
Pseudomonas chlororaphis	AtEze	Vegetables and ornamentals in greenhouses
Pseudomonas syringae	Bio-save10	Citrus and pome fruit
Pseudomonas fluorescens	BlightBan A506, Conquer, Victus, MVP, M-trak	Almond, apple, cherry, mushroom, peach, pear, potato, strawberry, and tomato
Pseudozyma flocculosa	Sporodex L	Roses and cucumbers in greenhouses
Streptomyces griseovirdis	Mycostop	Field, ornamental, and vegetable crops
Trichoderma harzianum	Binab T, Root Shield; Plant Shield, Trichodex	Wounds in the ornamental, shade, and forest trees, Greenhouses, nurseries, turf, home gardens, planting boxes, and outdoor soil, Most food crops,

In order to commercialize PGPR for plant growth promotion (biofertilizer) and disease control (biopesticide), it is important to understand how they modulate plant defense mechanisms, deploy biocontrol actions, and control plant pathogens (109). Major criteria of successful commercial PGPR are (a) demand for an economical and viable market, (b) consistent and broad-spectrum action, (c) safety and stability, (d) longer shelf life, (e) low capital costs, and (f) easy availability (108). A comprehensive strategy for evaluating PGPR as a commercially viable biological agent include, but not limited to (i) understanding the effects of environmental parameters, (ii) mechanisms of action and ecology, (iii) field testing, (iv) production of biocontrol agents, (v) formulation, (vi) delivery, and (vii) registration (12). Further, a proposed model for biocontrol research and development program includes (a) collecting diverse microorganisms, (b) screening for bioactivity, (c) identifying biocontrol mechanism, (d) collecting new strains with conserved biocontrol mechanism, and (e) utilizing new strains to expand biocontrol program (73).

1.4. *Bacillus* as promising alternatives to synthetic fungicides

Bacteria from the genus *Bacillus* are Gram-positive, rod-shaped, endospore-forming, catalase-positive, and aerobic microorganisms of the phylum *Firmicutes* (110). They are ubiquitous in nature and can inhabit soil, water, and the plant rhizosphere. They can also survive in extreme environments such as high pH (7.0 to 9.0), high temperature, and high salt concentrations. Most *Bacillus* species are non-pathogenic, with only two exceptions: *Bacillus anthracis*, the causative agent of anthrax, and *Bacillus cereus*, a causative agent of foodborne illness (111, 112). *B. subtilis* is the model *Bacillus* species and is one of the most studied prokaryotes in laboratories (113). The whole genome of *Bacillus* is approximately 4.2 Mb, comprising nearly 4,100 protein-coding genes (114). Many *Bacillus* species produce industrially

relevant enzymes, including the protein barnase (a ribonuclease), α-amylase, protease subtilisin, and the restriction enzyme BamH1 (115–117). Several *Bacillus* species, including *B. amyloliquefaciens, B. licheniformis*, *B. pumilus,* and *B. subtilis* have been used as active ingredients in commercial biofungicides (118). *B. thuringiensis* has been considered a highly specific, safe and effective bioinsecticide (119). The former group has been used against broad range of pathogens/diseases, including, rust, powdery mildew, downy mildews, blast, *Rhizoctonia*, *Phytophthora*, *Fusarium*, *Pythium*, *Aspergillus*, and *Botryodiplodia*. The latter was the first commercial biopesticide developed based on *B. thuringiensis* to control lepidopteran, dipteran, and coleopteran pests (120). Its pesticidal ability is mainly mediated by the production of proteins δ-endotoxins or Cry proteins (121).

1.4.1. Economically important *Bacillus* phenotypes

Bacillus species can produce endospore, a dormant, tough, and non-productive structure that enables them to lie dormant in unfavorable conditions for extended periods (122). *Bacillus* endospores are highly resistant to ultraviolet (UV) radiation, abiotic stresses, extreme heat, and chemicals such as peroxide and hypochlorite, which make them useful in agriculture, biotechnology, and medicine (122). In addition, *Bacillus* species can produce biofilm, which allows them to cluster into non-biological surfaces as a protected mode of growth and defense in hostile environments (123). Such biofilm is a dynamic and structurally complex system with characteristics of a multicellular organism and a complete ecosystem that has applications in medicine, agrochemical, food, and aquaculture industry (124). *Bacillus* species produce a higher quantity of useful enzymes such as amylase, xylanase, lichenase, lipase, cellulase, and pectinase, which are routinely used in industry to produce important chemicals, including riboflavin,

menaquinone-7, inositol, or N-acetylglucosamine (125–127). Importantly, some *Bacillus* species produce various classes of antibiotics, siderophore, and phytohormones that could potentially be used in agriculture, medicine, and biotechnological applications (127, 128).

1.4.2. Common *Bacillus* species with economic importance

Bacillus species are widely considered for agricultural and biotechnological applications due to their potential use as biofertilizers, biopesticides, and producers of industrially-relevant enzymes and biofilms (15, 107, 126, 129). Several PGPR *Bacillus* species are used as active ingredients in commercial biofertilizers and biopesticides, including *B. subtilis* (Kodiak, Subtilex, VAULT, SERENADE, RhizoPlus, Taegro, POMEX), *B. pumilus* (Yield Shield, SONATA), *B. amyloliquefaciens* (BioYield, RhizoVital 42), *B. licheniformis* (EcoGuard TM), and *B. velezensis* (Botrybel) (15, 107). *Bacillus* species with the ability to produce higher quantities of industrially-relevant enzymes are *B. subtilis*, *B. amyloliquefaciens*, *B. licheniformis,* and *B. pumilus* (116, 126, 127). The most commonly known *Bacillus* species for robust biofilm and spore formation is *B. subtilis* (130). Other *Bacillus* species known for biofilm production are *B. velezensis* and *B. amyloliquefaciens* (131, 132).

1.4.3. *Bacillus* as plant growth-promoting agents

It has been shown that a commercial *Bacillus* bioproduct developed from *B. subtilis* could increase crop yield by 40% (133). *Bacillus* species are predominant plant growth-promoting bacteria (8, 134). Their endospore-forming ability allows them to survive for very long periods of time under highly unfavorable conditions. *Bacillus* species produce exopolysaccharides and siderophores at the time of water scarcity and salinity in order to prevent movement of toxic ions and adjust the ionic balance and water transport in plant tissues. They have an innate ability to

colonize plant roots and interact symbiotically or synergistically with plants in the plant rhizosphere. *Bacillus* take up nutrients from the roots and, in exchange, provide various antibiotics, biostimulants, and phytohormones for growth as well as protection against abiotic and biotic stresses. They stimulate plant immunity against stresses by altering stress-responsive genes, proteins, phytohormones. *Bacillus* species regulate intracellular phytohormone metabolism and increase plant stress tolerance by synthesizing indole-3-acetic acid, gibberellic acid, and 1-aminocyclopropane-1-carboxylate (ACC) deaminase (15). Further, *Bacillus* species provide nitrogen, phosphorus, and iron to plants as a soluble form by utilizing their own enzymes and producing organic acids and siderophores (135). In addition, they prime plants by creating hostile environments and produce less toxic compounds that trigger induce systemic resistance (ISR) (136). *Bacillus* species with ISR ability include *B. amyloliquefaciens*, *B. subtilis*, *B. pasteurii*, *B. cereus*, *B. pumilus*, *B. mycoides*, and *B. sphaericus* (136).

1.4.4. *Bacillus* as biocontrol agents

Bacillus species produce several classes of structurally diverse secondary (or specialized) metabolites that could be used as biocontrol agents, fungicides, drugs, and biosurfactants (17). On average, *Bacillus* species devote 4 to 8% of their total genomes to the biosynthesis of secondary metabolites that are not directly associated with their reproduction, growth, or survival (17, 137). The most important class of bioactive secondary metabolites produced by *Bacillus* species are ribosomal and non-ribosomal peptides (including lipopeptides), polyketides, bacteriocins, and siderophores (15). The non-ribosomal peptides, especially lipopeptides, have broad-spectrum antagonistic activity against plant-pathogenic bacteria, viruses, fungi, and oomycetes. The most commonly observed *Bacillus* cyclic lipopeptides are surfactin, iturin, and fengycin. Each of them

purportedly kill target pathogens by disrupting cell membranes though the precise mechanisms by which each does so are not identical. (75).

Surfactin, a broad-spectrum antimicrobial, causes cell membrane disruption by repetitive hydrophobic and hydrophilic integrations through a process of membrane destabilization (138). The antagonistic activity of iturin is attributed to the formation of ion-conducting pores as opposed to large-scale membrane disruption or solubilization (139). Fengycin exerts antagonistic activity by interacting with the cell membrane, disrupting membrane structure, and increasing permeability (140). *Bacillus* species also produce bacteriocin, a class of ribosomally-synthesized post-translationally modified peptides (RiPP) (141). Bacteriocins are small, heat-stable, amphiphilic peptides that cause cell damage to target pathogens. Generally, they exhibit narrow-spectrum antibacterial activity against closely related bacteria (15). Other peptides with antagonistic activity produced by *Bacillus* species are gageostatin, dihydroisocoumarins, bacilysin, and rhizocticins (17). Some common polyketides produced by *Bacillus* species are bacillaene, difficidin, and macrolactin. They are known as strong-inhibitor of protein biosynthesis, although their exact mode of action is far less understood (17). Important secondary metabolites with antimicrobial activity produced by *Bacillus* species are shown in Table 1.2.

1.4.5. *Bacillus* secondary metabolites: chemical and antimicrobial properties

Bacillus-derived peptides and polyketides are the most extensively investigated secondary metabolites due to their broad spectrum of antimicrobial properties (17). Compounds from these two broad classes exhibit antibiosis activity against fungi, oomycetes, bacteria, and/or viruses. Non-ribosomal peptides have exhibited a broad antibiosis activity that includes multiple groups of bacteria (142); however, they are also among a more limited group of compounds that show potent antifungal and/or anti-oomycete activity. The polyketides, discussed further below, also display rather broad antibiosis activity, but they are typically most effective against bacteria rather than filamentous eukaryotes like fungi and oomycetes.

A major class of non-ribosomal peptides are the lipopeptides. The term "lipopeptide" accurately describes the two main components of these compounds. One is a core peptide composed of seven to ten amino acids, many of which are canonical proteogenic L-amino acids; however, owing to their non-ribosomal synthetic origin, these core peptides also often include D-amino acids and other non-proteogenic species (e.g., L-ornithine). The core peptide is also invariably cyclized (Fig. 1.3). The second main component is a fatty acid side chain. The most obvious variation in the fatty acid is chain length, ranging from 14C to 20C or more. Variability also includes β-OH derivatives, unsaturated chains, and branched (or *iso)* fatty acid derivatives.

The most commonly observed non-ribosomal peptides produced by *Bacillus* species are iturin, fengycin, and surfactin (Fig. 1.3). Iturin contains seven amino acid residues in its core moiety and a fatty acid side chain (143). The core peptide of all iturins starts with an ^{L}Asn-^{D}Tyr-^{D}Asn. The

final four amino acids vary, giving rise to the most commonly observed iturin derivatives iturin, bacillomycin D, bacillomycin F, bacillomycin L, mycosubtilin, and mojavensin. Each of these iturinic compounds contains a single fatty acid whose chain length varies from C14-C18. Iturinic compounds are widely known for strong antifungal/anti-oomycete activity against eukaryotic plant pathogens (143). Some of them also display antibacterial activity (144).

Table 1.2: Prominent antimicrobial secondary metabolites produced by *Bacillus* species (15)

Secondary metabolite	**BGC class[1]**	**Bioactivity(s)**
Bacillaene	PKS-NRPS	Antibacterial
Bacillibactin	NRPS	Iron acquisition
Bacilysin	Other	Antifungal, antibacterial
Difficidin	PKS	Antibacterial
Fengycin/Plipastatin	NRPS	Antifungal
Gageostatin	NRPS	Antifungal
Iturin/Bacillomycin	NRPS	Antifungal
Lichenysin	NRPS	Hemolytic, cytotoxic
Locillomycin	NRPS	Antibacterial
Macrolactin	PKS	Antifungal, antibacterial, cytotoxic
Mycosubtilin	PKS-NRPS	Antifungal
Plantazolicin	RiPP	Antibacterial
Rhizocticin	Other	Antifungal
Subtilin	RiPP	Antibacterial
Surfactin	NRPS	Hemolytic, cytotoxic
Zwittermicin	PKS-NRPS	Antifungal, antibacterial

[1]BGC class: BGC classifications were based on the structural and chemical properties of secondary metabolites, where non-ribosomal peptide synthetases (NRPS), polyketides (PKS), a hybrid of PKS and NRPS (PKS-NRPS), ribosomally-synthesized and post-translationally modified peptides (RiPP), terpenes, and BGCs outside these five classes (other).

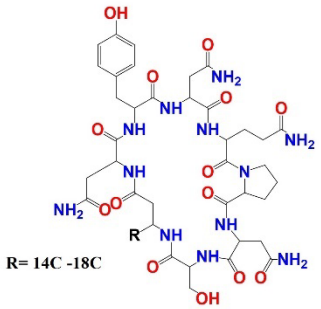
Fengycin has a core peptide which consists of six proteogenic and four non-proteogenic amino acids, including an invariant ^{D}ornithine at position two (145). Taking fengycin A as the

archetype for the group, the sequence of the core peptide is ^{L}Glu-^{D}Orn-^{D}Tyr-^{D}Thr-^{L}Glu-^{D}Ala-^{L}Pro-^{L}Gln-^{L}Tyr-^{L}Ile. A striking feature of fengycin is a lactone ring established between the carboxyl group of amino acid 10 (Ile in fengycin A) and the phenol oxygen of ^{D}Tyr which is invariant among fengycins at position 3. The most common fengycin core peptides vary at positions six and ten, where aliphatic side chains Ala, Val, Leu, and Ile are with variation in the core peptide amino acids produced by *Bacillus* species are represented across the range of fengycin compounds. For example, fengycin B is in all other ways identical to fengycin A, except that a ^{D}Val appears in place of ^{D}Ala at position six. Other less commonly observed fengycin derivatives are C, D, S, X, and Y (146). Each of these derivatives also contains a variable-length β-hydroxy fatty acid side chain (14C to 21C or more). Fengycins are widely known for strong antibiosis activity against filamentous fungi and oomycetes (147–149).

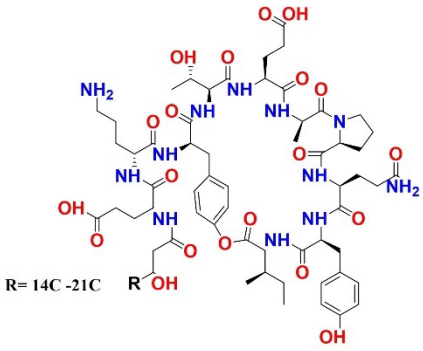
Surfactin is a heptapeptide which consists of five proteogenic amino acids and two non-proteogenic amino acids. The sequence of the core peptide is ^{L}Glu-^{L}Leu-^{D}Leu-^{L}Val-^{L}Asp-^{D}Leu-^{L}Leu (150). Three surfactin derivatives have been described in literature: surfactin, pumilacidin, and lichenysin (151). These are broadly referred as "surfactin-like" lipopeptides. Compared to surfactin, pumilacidin has Leu in position four and an Ile/Val in position seven. Lichenysin differs from surfactin by a change in the first amino acid, where Gln appears instead of a Glu. Each of these derivatives also contain a single fatty side acid chain whose length varies from 12C to 19C. Surfactin-like compounds are widely known as "biosurfactants" due to their strong amphiphilic properties (152). They exhibit broad antibiosis activity against both eukaryotic and prokaryotic organisms, including bacteria, fungi, and viruses (153–156).

A

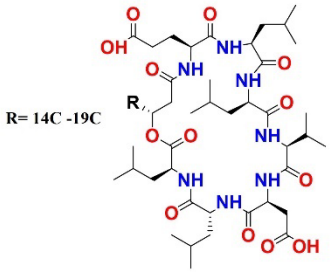
Iturin



Fengycin A



Surfactin



B

Iturin derivative	Core peptide sequence
Iturin	Asn-^{D}Tyr-^{D}Asn-Gln-Pro-Asn-Ser
Bacillomycin D	Asn-^{D}Tyr-^{D}Asn-Pro-Glu-Ser-Thr
Bacillomycin F	Asn-^{D}Tyr-^{D}Asn-Gln-Pro-Asn-Thr
Bacillomycin L	Asn-^{D}Tyr-^{D}Asn-Ser-Glu-Ser-Thr
Mycosubtilin	Asn-^{D}Tyr-^{D}Asn-Gln-Pro-Ser-Asn
Mojavensin	Asn-^{D}Tyr-^{D}Asn-Gln-Pro-Asn-Asn

Fengycin derivatives	Core peptide sequence
A	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Ala-Pro-Gln-Tyr-Ile
B	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Val-Pro-Gln-Tyr-Ile
C	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Val-Pro-Gln-Thr-Ile
D	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Val-Pro-Gln-Tyr-Val
S	Glu-^{D}Orn-^{D}Tyr-^{D}alloSer-Glu-^{D}Val-Pro-Gln-Tyr-Ile
X	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Ala-Pro-Leu-Tyr-Ile
Y	Glu-^{D}Orn-^{D}Tyr-^{D}alloThr-Glu-^{D}Val-Pro-Leu-Tyr-Ile

Surfactin derivatives	Core peptide sequence
Surfactin	Glu-Leu-^{D}Leu-Val-Asp-^{D}Leu-Leu
Pumilacidin	Glu-Leu-^{D}Leu-Leu-Asp-^{D}Leu-Val/Ile
Lichenysin	Gln-Leu-^{D}Leu-Val-Asp-^{D}Leu-Leu

Figure 1.3: Structures (A) and core peptide sequences (B) of iturins, fengycins, and surfactins. The structures of the quintessential lipopeptide in each group (iturin, fengycin A, and surfactin, respectively) are shown. The fatty acid side chain in each compound is denoted as "R," and the observed chain length range is indicated.

Three less-frequently observed *Bacillus* lipopeptides are rhizomide, locillomycin, and kannurin (Fig. 1.4). Rhizomide is a heptapeptide whose N-terminal ^{L}Leu residue is acetylated. Rhizomide is also distinct from the other lipopeptides in that it is not derivatized with a long-chain fatty acid (157). Three rhizomides (A-C) have been identified. Each rhizomide has the core peptide

sequence N-acetyl-^{L}Leu-^{L}Thr-^{D}Tyr-^{D}Ala-X-Y-^{L}Val, where X and Y are, respectively, ^{L}Ala ^{D}Ala (rhizomide A), ^{L}Ala ^{D}Ser (rhizomide B), or ^{L}Ser ^{D}Ala (rhizomide C). Rhizomide exhibits antibiosis activity against Gram-positive bacteria, including *B. subtilis and Staphylococcus aureus,* and the Gram-negative bacteria *Escherichia coli* (157, 158). Locillomycin contains a nonapeptide core moiety and a fatty acid side chain (159). Its core peptide sequence is ^{L}Thr-^{D}Gln-^{L}Asp-^{L}Gly-^{L}Asn-^{L}Asp-^{L}Gly-^{L}Tyr-^{L}Val. Only three locillomycin derivatives are known. All three of its derivatives have the common core peptide, but they are varied in the fatty acid side chain.

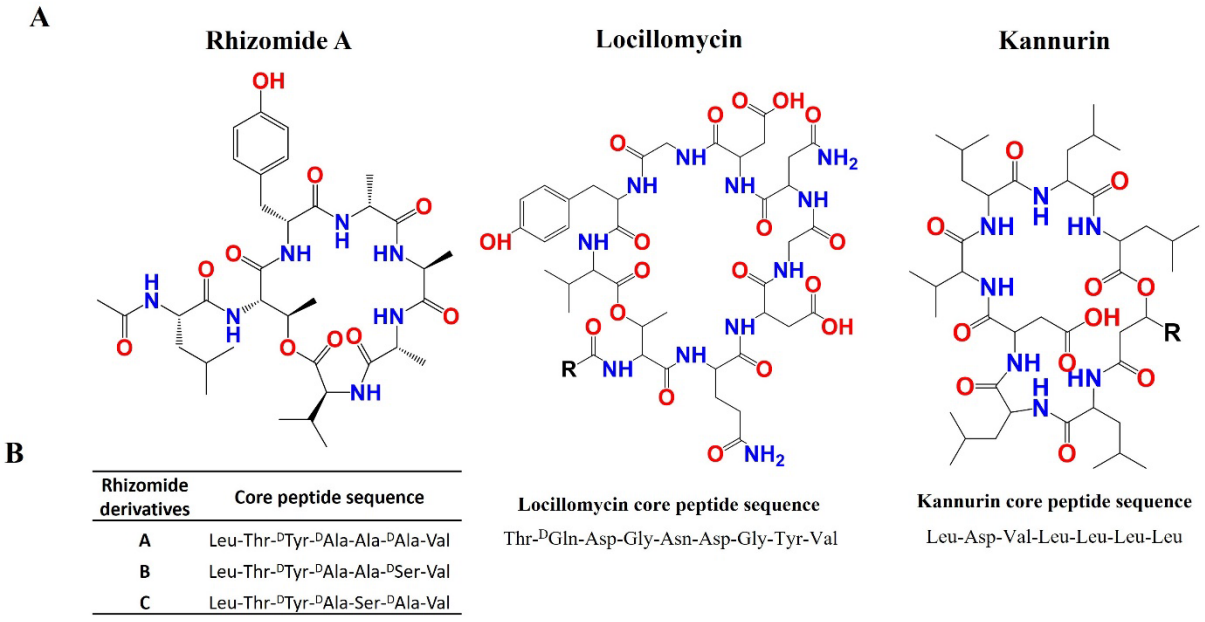


Figure 1.4: Structures (A) and observed core peptide sequences (B) of rhizomides, locillomycins, and kannurins. The structures of rhizomide A, locillomycin, and kannurin are presented as the quintessential representatives of each group. Locillomycins and kannurins vary only by the length of their fatty acid side chains (denoted by R).

Locillomycin exhibits a wide range of antibiotic activity against pathogenic microorganisms, including bacteria (*Staphylococcus aureus* and *Xanthomonas oryzae*) virus (porcine epidemic diarrhea virus, PEDV) (160, 161). Kannurin is a heptapeptide whose core peptide sequence is Leu-Asp-Val-Leu-Leu-Leu-Leu (162). It contains a single chain fatty acid. Three kannurin derivatives are known and they are varied in the fatty acid chain length. Kannurin exhibits antifungal activity against broad range of pathogens (163).

Bacillibactin, is also built on a non-ribosomal peptide foundation. Its core peptide structure is built on three repeating Gly-Thr units (Fig. 1.5). This core is derivatized with three 2,3-dihydroxybenzoate (DHB) groups (164). The function of bacillibactin contrasts starkly with the lipopeptides described above; it is one of the most commonly observed siderophores produced by *Bacillus* species. Typically, bacillibactin contributes to plant growth by sequestering sparingly soluble iron (Fe^{III}) and supplying it to plant as soluble form (Fe^{II}). However, it has been suggested that bacillibactin displays direct antibiosis activity against bacteria and plant-pathogenic fungi through iron scavenging and /or yet unknown intercellular antibiotic activity (165).

Plantazolicin is a linear peptide that contains a series of azole rings produced by post-translational modifications of a core peptide (166). It contrasts with all of the compounds described above as it is a ribosomally-derived structure (Fig. 1.5). It possesses narrow-spectrum antibiosis activity against Gram-positive bacteria, including *B. anthracis* and *B. cereus* (167).

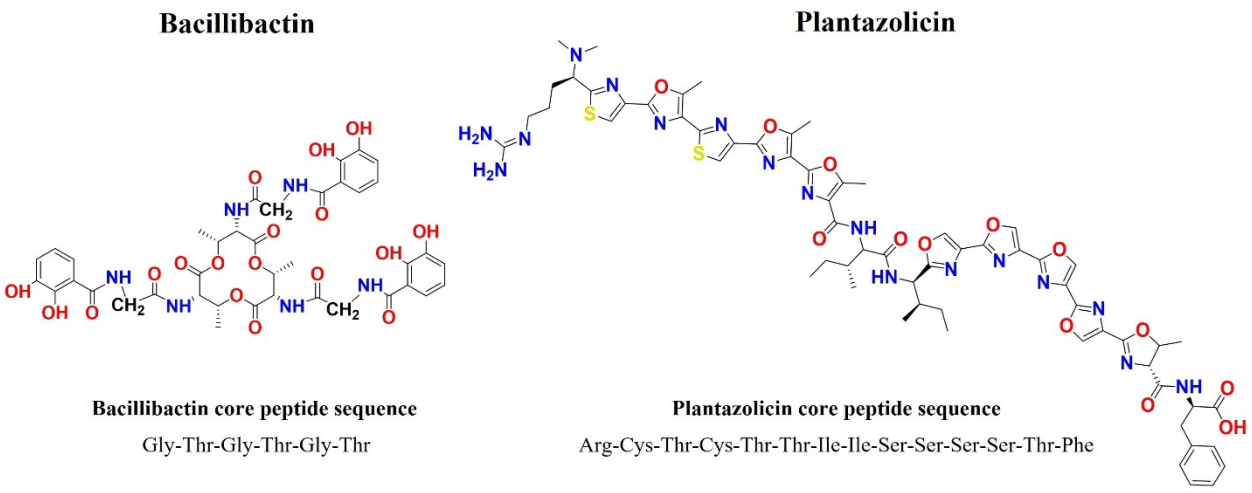


Figure 1.5: Structures and core peptide sequences of bacillibactin and plantazolicin.

Bacillus species also produce various structurally and chemically diverse polyketides, including linear and cyclic compounds. Some polyketides contain additional amide-linked amino acid-based moieties. Due to the nature of the BGCs which produce and the manner in which they are constructed, these are referred to as hybrid polyketide/non-ribosomal peptides. The most commonly observed polyketides produced by *Bacillus* species are bacillaene, difficidin, and macrolactin (160). Bacillaene is a linear (hybrid) polyketide that contains Gly and Ala in its polyketide chain (Fig. 1.6) (168). It exhibits broad-spectrum antibacterial activity against species ranging from *E. coli*, *K pneumoniae*, *B. thuringiensis*, and *S. aureus*. Bacillaene antibacterial activity is linked with its ability to inhibit prokaryotic protein synthesis (169). However, its exact target in the protein synthesis is yet unknown. Difficidin is a phosphorylated cyclic polyketide that is widely known for antibacterial activity against both Gram-positive and Gram-negative bacteria (170). Similar to bacillaene, difficidin exerts antibacterial activity by inhibiting bacterial protein synthesis. Macrolactin is a cyclic polyketide produced by *Bacillus* species. It exhibits strong bioactivity as antibacterial, antifungal,

antiviral, anticancer, and anti-angiogenic (171). Zwittermicin is a hybrid polyketide-peptide produced by *Bacillus* species (172). It has broad spectrum antibiosis activity against phytopathogens and certain Gram-positive and Gram-negative organisms. Structurally, it is linear, containing 2,3-diaminopropionate that is produced by two amino acids, serine and ornithine.

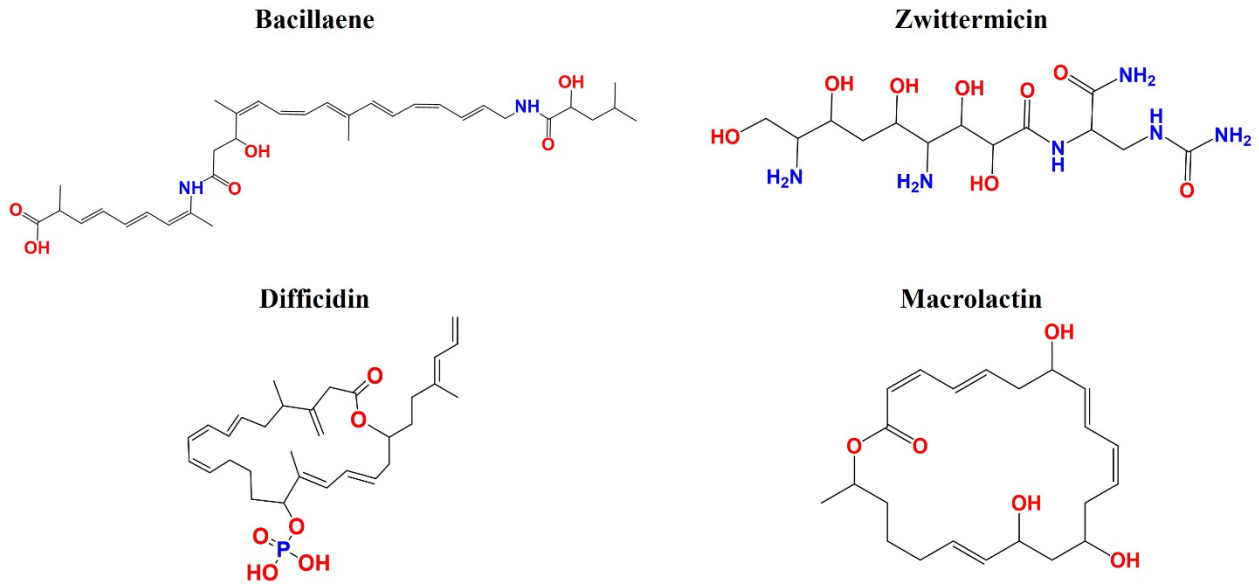


Figure 1.6: Structures of prominent antimicrobial polyketides produced by *Bacillus* species.

1.4.6. *Bacillus* secondary metabolite biosynthetic pathways

The biosynthesis of a natural product generally involves optimally-regulated multi-step enzymatic reactions in an assembly-line fashion (173). All core and accessory genes for the biosynthesis of a secondary metabolite are typically arranged within an operon. The cluster of genes in such an operon is collectively referred to as a biosynthetic gene cluster (BGC). Some of the most common BGCs identified in *Bacillus* species are non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), PK-NRPS hybrids, ribosomally-synthesized and post-

translationally modified peptides (RiPPs), and terpenes (174). An NRPS is a multi-modular mega-enzyme complex that generally carries the bulk of the chemical steps necessary to produce the core peptide product. Some of these peptides remain linear and some are cyclized. Note that the biosynthesis of secondary metabolites by NRPS does not require any of the cell's protein synthetic machinery (i.e., ribosomes, mRNA, tRNA, etc.). Each NRPS module consists of multiple domains capable of accepting, transferring, and/or modifying one amino acid residue. The major domains of an NRPS module include formylation (F), adenylation (A), thiolation, peptide carrier protein (PCP), condensation (C), termination, and/or thioesterase (TE) (150). Some NRPS also contain additional enzymes for derivatization of the core peptide product and/or cyclization to form the final product. Together, NRPS modules constitute an assembly line-like metabolic pathway as illustrated for production of one of the most commonly observed lipopeptides, surfactin (Fig. 1.7).

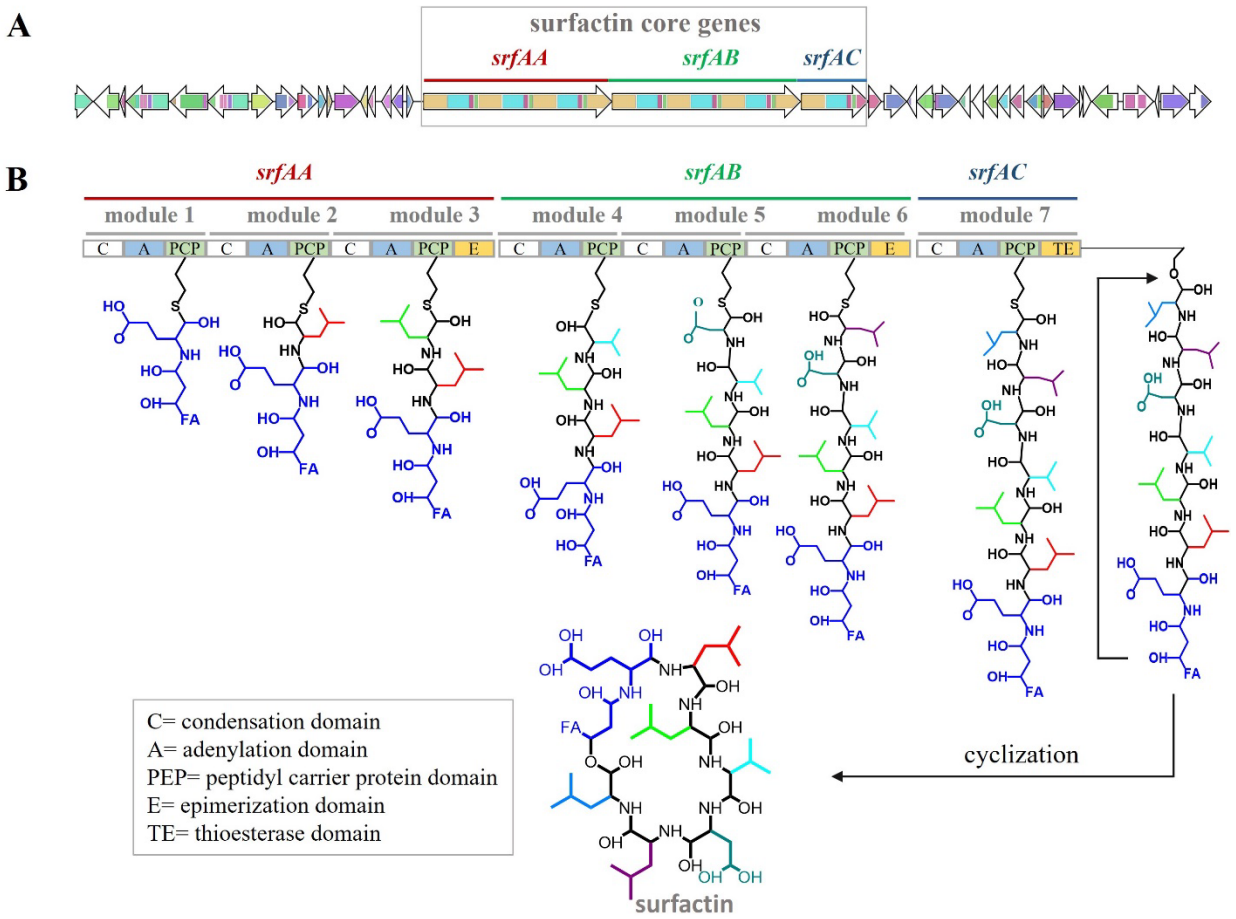


Figure 1.7: Biosynthetic steps catalyzed by the surfactin NRPS found in *Bacillus* species. The gene organization of the surfactin-producing BGC is shown (A), identifying its three core components, *srfAA*, *srfAB*, and *srfAC* as highlighted by the gray box. Biosynthetic steps for surfactin production (150) are shown (B). Surfactin-producing BGC contains a total of seven modules that are distributed within the three core domains, where each module contains necessary genes for incorporating one amino acid at a time during the biosynthetic process. Abbreviations corresponding to individual domains are shown inside the box on the lower left.

The production of multiple structurally diverse polyketides is also a feature commonly observed in *Bacillus* species. These products often exhibit some level of antibacterial activity (175). In a manner reminiscent of the NRPSs addressed above, such polyketides are produced by mega-enzyme PKS (polyketide synthase) complexes where each module contains multiple domains (176). The biosynthetic process for polyketide production is highly similar to that

necessary for standard, straight-chain, fully saturated fatty acids. Indeed, fatty acids themselves are polyketides. Biosynthesis catalyzed by PKSs involves a stepwise condensation of a starter unit such as acetyl-CoA or propionyl-CoA with an extender unit of malonyl-CoA or methylmalonyl-CoA (177). The major domains which comprise each module in biosynthesis are acyltransferase (AT), acyl carrier protein (ACP), β-ketoacyl-ACP synthase (KS), β-ketoacyl-ACP reductase (KR), dehydratase (DH), enoylreductase (ER), and thioesterase (TE) (178, 179). A proposed model for the biosynthesis of the polyketide difficidin requires seven core modules (Fig. 1.8).

In contrast to secondary metabolites generated by NRPS clusters, the RiPPs are peptides produced by modification of precursor polypeptides that are initially produced by ribosomes via standard cellular protein synthetic steps (180). In the first step of RiPP biosynthesis, the precursor peptide containing a leader peptide and a recognition sequence (RS) is produced by the translation of mRNA using the protein synthesis machine, ribosome (Fig. 1.9). The precursor peptide then goes through the removal of the leader peptide and recognition sequence by the biosynthetic enzymes. Finally, the initial core peptide is transformed into a matured peptide by a series of modifications, including cyclization, dehydration, prenylation, and cyclodehydration.

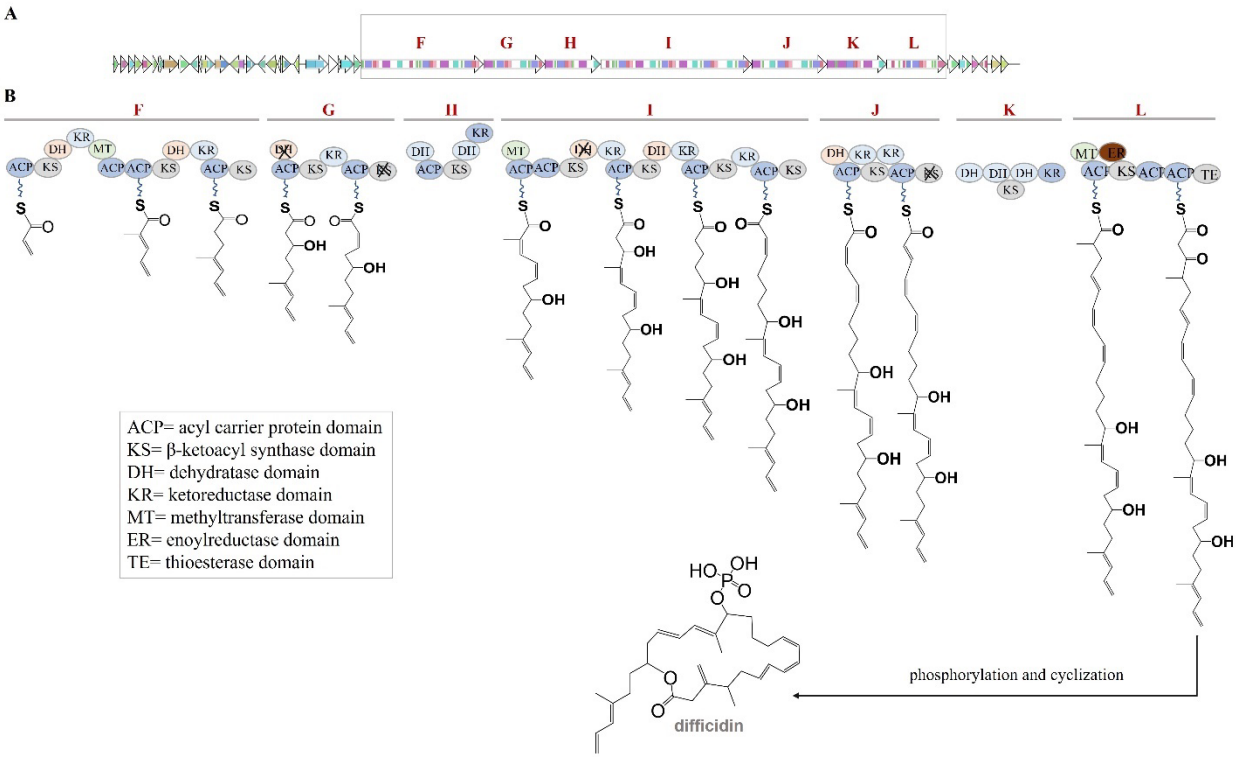


Figure 1.8: Proposed model of the biosynthetic pathway of a PKS BGC that produce difficidin in *Bacillus* species. The gene organization of the difficidin-producing BGC includes seven PKS core genes (F-L) marked off by the gray box (A). The proposed biosynthetic steps for difficidin production by these seven core modules are shown (B) (178). Difficidin-producing modules are subdivided into the necessary domains necessary for catalyzing the biosynthetic steps of the process. Domains that are unable to contribute their typical catalytic function are crossed out in the diagram. Abbreviations of associated domains are shown inside the box on the lower left.

Another commonly observed *Bacillus* secondary metabolite is terpene, consisting of a building block of isoprene (181). The isoprene unit is provided in the form of dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) (Fig. 1.9) (182). These units are produced by two distinct metabolic pathways: the mevalonate (MVA) pathway and the non-mevalonate (MEP) pathway. In terpene biosynthesis, these two units are condensed by geranyl pyrophosphate synthase to produce geranyl pyrophosphate (GPP), which is the precursor to all terpenes (182).

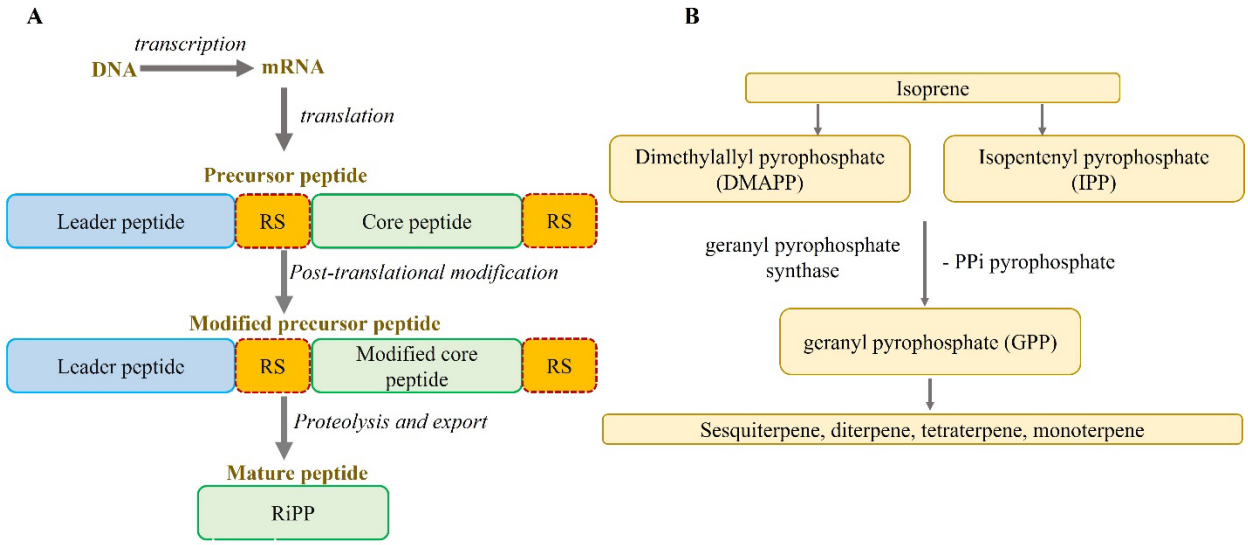


Figure 1.9: Schematics of the biosynthetic pathways of RiPPs (A) and terpenes (B) (180, 183). In the RiPP biosynthetic process, an initial precursor peptide is produced by ribosome-based translation. This initial intermediate contains a leader peptide, recognition sites (RS), and a core peptide. In subsequent steps, the core peptide is modified and released as a mature peptide through post-translational modifications and proteolysis. Terpene biosynthesis initiates with isoprene units and gets transformed to geranyl pyrophosphate through series of biosynthetic reactions.

1.5. Discovery pipelines

Since the early 20th century, bioassay-based fermentation has been the key strategy for discovering bioactive natural products. However, over time this process has become ineffective due to the frequent rediscovery of existing compounds. As a consequence, rate of novel bioactive compound slowed dramatically from the 1970s forward. Fortuitously, recent rapid advances in genome sequencing technology, robust bioinformatics algorithms, and computational capacity for "big data" has provided alternative pathways to address this challenge. In the past two decades, a staggering number of genomes have been sequenced covering organisms from all domains of life. The large proportion of these data are currently accessible in public databases. This has enabled mining of the whole genome sequences of target organisms to predict biosynthetic pathway of

antibiotic natural products. In addition, insight from the large-scale genomic analyses of target organisms expedites the identification of novel compounds or known compound scaffolds with novel derivatizations. These advances have already started to increase novel compound discovery from biological sources, a trend that is expected to accelerate into the future. This is due at least in part to diminishing redundant and unproductive compound rediscovery.

1.5.1. Bioassay-based chemical screening

Nearly a century ago, the discovery of penicillin from the fungus *Penicillium rubens* started the golden age of natural product-based antibiotic discovery. Since then, numerous antibiotics have been discovered in microorganisms and plants. Some major classes of microbiome-derived antibiotics are tetracyclines, macrolides, glycopeptides, cycloserines, bacitracin, polymyxins, mupirocin, and cephalosporins (19). The first successful antibiotic discovery platform, known as the "Waksman platform", launched the discovery of a substantial amount of antibiotics between 1940 and 1960 (184). This discovery platform consists of a dual-plate antibiosis assay of a bacterium against target pathogens to measure zones of growth inhibition (18). Since the 1960s, the rate of antibiotic discovery based on this platform has slowed significantly. This is due to the increased rate of rediscovery of existing compounds and the lack of novelty in the bioactive compound. This gap in novel antibiotic discovery has increased demands for antibiotic discoveries against pathogenic microorganisms (19). Further, antibiotic resistance and the emergence of new pathogens have continued to increase in the past three decades (185–188). These situations warrant exploring an alternative and sustainable approach to discovering antibiotic natural products with novel modes of action.

1.5.2. Genome mining

The natural products produced in bacteria, fungi, and plants are encoded in one or proximally located multiple operons that are widely known as biosynthetic gene clusters (BGCs). Due to the advances in computational speed/capacity and the development of robust bioinformatics algorithms/programs to capitalize on them, such BGCs can be reliably identified from whole-genome sequence and their putative products predicted with high accuracy (23). Some of the most comprehensive bioinformatics suites of programs for predicting BGCs are antiSMASH, NP.searcher, SMURF, ClustScan, eSNaPD, ClusterFinder, EvoMining, NRPS-PKS/SBPKS, NaPDoS, BAGEL, and RODEO (24). Table 1.3 shows prominent bioinformatics programs currently available for bacterial genome mining. Among these, the most popular bioinformatics program for predicting bacterial BGCs and secondary metabolites is antiSMASH (189). It is a comprehensive pipeline for identifying a variety of BGCs in bacterial genomes, including polyketides (PKs), non-ribosomal peptides (NRPs), terpenes, ribosomally-synthesized and post-translationally modified peptides (RiPPs), etc. Other prominent bioinformatics programs, including BiG-SCAPE, BiG-MAP, and BiG-SLiCE have been developed recently in parallel to simultaneously compare, analyze, and dereplicate a large number of BGCs. These comprehensive toolsets have enabled dereplication and selection of novel metabolites or the existing metabolites with novel derivatizations. The capacity to capitalize on these advances would be severely diminished without commensurate innovations in analytical instrumentation and procedures including but not limited to chromatography (e.g., HPLC), mass spectrometry, and NMR. Simultaneous technological advances on both fronts has already enabled the expedited discovery of novel natural products, and this is a trend that is expected to continue into the future. (22, 23, 190–193).

Table 1.3: Current state-of-the-art bioinformatics toolset for bacterial genome mining (23, 24).

Bioinformatics tool	Key features	Available at	Year Released
antiSMASH	Integrate multiple BGC prediction tools/algorithms: ClusterFinder, NaPDoS, RODEO.	https://antismash.secondarymetabolites.org/	2011
NP.searcher	Predict 2D and 3D structure of NRPS/PKS.	https://dna.sherman.lsi.umich.edu/	2009
ClustScan	Employ pHMMs of signature genes for BGC prediction.	obtain by request at novalis@novalis.hr	2008
eSNaPD	Uncover biosynthetic diversity from metagenomic data.	http://esnapd2.rockefeller.edu/	2014
ClusterFinder	Prediction is based on Pfam domain frequencies.	https://github.com/petercim/ClusterFinder	2014
EvoMining	Genome mining based on evolutionary principles.	https://github.com/nselem/evomining	2019
NRPS-PKS/ SBPKS	Model 3D structures of individual PKS catalytic domains.	http://202.54.249.142/~pksdb/sbspks/master.html	2010
NaPDoS	Phylogenic approach for domain analysis, various query types including genome contig.	https://npdomainseeker.sdsc.edu/	2012
BAGEL	Single-input whole-genome analysis for bacteriocin and RIPP BGC detection.	http://bagel.molgenrug.nl/	2013
RODEO	Combine hidden Markov model-based analysis, heuristic scoring, and machine learning.	http://ripp.rodeo/	2017
PRISM	A comprehensive platform for the prediction of the chemical structures of gnomically encoded antibiotics, including all classes of	https://prism.adapsyn.com/	2015
RiPPER	Identification of RiPP precursor peptides and biosynthetic gene clusters.	https://github.com/streptomyces/ripper	2021
PKMiner	Genome mining for type II polyketide synthases.	http://pks.kaist.ac.kr/pkminer	2012
ARTS	Specific and efficient genome mining for antibiotics with interesting and novel targets.	https://arts.ziemertlab.com	2017
CLUSEAN	Automated analysis of bacterial secondary metabolite biosynthetic gene clusters.	https://bitbucket.org/tilmweber/clusean	2009

1.5.3. Bioassays and chemical analyses guided by genome mining

In order to increase the rate of antibiotic discovery, combined bioassays and chemical analyses guided by genomic analysis have been implemented in the past two decades (194, 195). Implementation of this approach has been inspired by recent technological advances in genome sequencing, computational algorithm, and analytical instrumentation (31, 196, 197). These advancements have enabled the sequencing of a large number of whole genomes from various organisms on one hand and improved the resolutions and capabilities of analytical instruments such as MS, NMR, and HPLC on the other (22, 198–204). Currently, large-scale genomic analysis allows the prediction of the biosynthetic pathway of antibiotic natural products, *apriori.* This assists in focusing on natural products with novel properties/scaffolds from the target organism before performing bioassay-based chemical analysis. In subsequent steps, the predicted novel natural product is extracted and isolated from the culture of the target organism and analyzed using analytical instruments, followed by screening for antibiosis activity. In parallel, data from both genomic and chemical analysis are combined and analyzed simultaneously using comprehensive bioinformatics platforms such as SMART algorithm, DP4-AI machine learning algorithm, MixONat, ZODIAC, CANOPUS, Retip, MetFID, MASSST, Spec2Vec, FBMN, and BMDMS-NP (26). This combined approach is anticipated to reduce the rate of rediscovery of existing antibiotics and speed up the discovery of novel antibiotic natural products.

1.5.4. Functional, synthetic, and meta-genomic approaches

Fueled by the rapid advances in genome sequencing and bioinformatics algorithms, new technologies such as functional, synthetic, and meta-genomics have been developed to expedite novel antibiotic discovery. The functional genomics approach utilizes function-related aspects of

genomes such as gene (and protein) and whole biosynthetic pathways to determine the function, novelty, and related properties of a natural product (205–207). This technology has allowed focusing on a specific protein or metabolite pathway to understand the biological process in a specific context. In addition, it has enabled the transfer of BGCs of a whole biosynthetic pathway of a natural product into a heterologous system to understand the intricate enzymatic steps within the pathway by employing PCR and CRISPR technologies (208, 209). Currently, existing genomic technologies have already enabled the synthesis of large DNA, gene clusters of a biosynthetic pathway, or the whole genome of an organism using synthetic genomics on a large scale (210, 211). This approach has allowed genetic modification of existing genetic code or (custom) design of a novel genetic code to produce a novel phenotype. The metagenomics approach has allowed the recovery of the genetic material of target organisms directly from the environment that is otherwise inaccessible outside of their natural conditions (212, 213). In this approach, DNA sequences are extracted and cloned directly from the natural environment from as-yet-unculturable microorganisms (212, 214). The metagenomic approach is expected to significantly enhance the pace of antibiotic discovery from those inaccessible microbial communities commonly known as microbial "dark matter" (215, 216).

1.6. *Bacillus* cytochromes P450: Catalysis and contribution in secondary metabolite biosynthesis

Cytochromes P450 (CYPs or P450s) are heme-containing monooxygenase enzymes comprising of one of the largest enzyme superfamilies (217, 218). Their "P450" designation is derived from the appearance of a strong UV-visible absorption band at 450 nm when they are bound with CO in their reduced state (i.e., Fe^{II}-CO complex) (219). Currently, more than half a

million P450 enzymes with distinct amino acid sequences have been deposited in the UniProt database (220). P450 enzymes have been identified in all kingdoms of life, including Eukarya, Archaea, and Bacteria (221). All P450s are divided into families and subfamilies. These are first designated with the CYP root symbol to identify the superfamily followed by a number representing the gene family (e.g., CYP3). A capital letter follows which identifies the subfamily (e.g., CYP3A), and the individual gene is specified with the final number (e.g., CYP3A4) (222). The enzymes of the P450 superfamily vary substantially in their structure, preferred substrates, and function among the various organisms that utilize them (159). In humans, a total of 57 P450 enzymes catalyze critical reactions necessary for the biosynthesis of cholesterol and the specialization of that scaffold for production of steroid hormones such as estrogen and testosterone. CYPs are also central to the metabolism of vitamin D, many drug compounds, and xenobiotics (223). Here the action of P450s is often to derivatize otherwise highly non-polar compounds with oxygen-bearing substituents (e.g., -OH), making them more soluble and facilitating their excretion. Prominent human P450s include CYP1A2, CYP2C9, CYP2C19, CYP3A4, CYP5A1, CYP8A1, CYP7A1, CYP11A1, CYP17A1, and CYP21A2 (223). The distribution of P450s in bacteria is highly variable. For example, there are no P450 genes in *E. coli*, but there are 20 P450s in *Mycobacterium tuberculosis.* Several P450s have been recently discovered in giant viruses isolated from the deep ocean, terrestrial sources, and human patients (224).

The catalytic process of P450 enzymes involves the tightly coordinated delivery of electrons to the heme-dependent reaction center by either a fused or independent P450-reductase (225, 226). The catalytic transformation of a given substrate takes place in the distal site of the

heme-bearing CYP (P450) domain. This designation arises from the assignment of the heme's so-called proximal side as that face of the cofactor where the protein-derived Cys ligand coordinates the heme's iron. The so-called distal side of the heme is on the opposite side of the heme plane (226). The canonical catalytic mechanism of P450 enzymes involves the insertion of one oxygen atom into an otherwise highly non-polar and unreactive organic substrate (RH). Molecular oxygen (O_2) is the source of the inserted oxygen atom. The O_2 must be bound to and activated by the P450 heme center, and this requires an input of electrons derived from NAD(P)H as an electron donor. This process is facilitated by the cytochrome P450 reductase. It is important to note that a broad diversity of reactions outside of simple hydroxylations are catalyzed by P450s. These include epoxidation, demethylation, and ring coupling (227).

In bacteria, enzymes from the P450 superfamily are one of the most diverse enzymes in their structures and functions (159, 228). Due to their ability to catalyze the oxidative transformation of a broad range of substrates, they are known as "promiscuous enzymes" (227, 228). Bacterial P450s play crucial roles in many biosynthetic and/or biodegradative processes that include broad diversity of secondary metabolite biosynthesis and chemical transformations. Prominent reactions known to be catalyzed by bacterial P450s are hydroxylation, aromatic dehalogenation, ring formation, ring coupling, ring contraction, intramolecular rearrangement, heteroatom release, epoxide formation, dimer coupling, demethylation, and decarboxylation (229).

Many bacterial P450s are involved in oxidative transformation of broad range of physiologically important natural products that have potential applications as antibiotics, immunosuppressants or anticancer agents. These P450s are often associated with several classes of natural products generated through secondary biosynthetic pathways, including polyketides, ribosomal and non-ribosomal peptides, alkaloids, terpenes, steroids, and fatty acids. In two

examples from terpene biosynthesis, P450s catalyze an oxidative deamination reaction in clavulanic acid biosynthesis as well as hydroxylation reaction in di- and sesquiterpene biosynthesis (230). Some P450s catalyze oxygenation in monoterpenes, diterpenes, and steroids to facilitate their biodegradation (159). Many P450s are reported to be involved in polyketide biosynthetic pathways, with catalytic roles in simple oxidative transformations to complex, multi-site transformations. Some important P450-associated bacterial-generated polyketides are erythromycin A, bacillaene, pimaricin, candicidin, and griseoviridin (231–235). Several non-ribosomal peptides have been reported to be modified through P450-catalyzed reactions. These include vancomycin, balhimycin, teicoplanin, thaxtomin, nocardicin A, and rufomycin (236–240). Some P450s have been reported to be involved in biosynthesis of ribosomal peptides, including microbisporicin bottromycin, thiostrepton, and thiomuracin (241–244).

P450s have a highly conserved tertiary structure with a characteristic protein fold where only the protein-derived proximal ligand to the heme iron (a cysteine) is conserved across the whole superfamily (218). The core scaffold consists of twelve helices (~40% residues), labeled A-L and five antiparallel β pairs (10% residues) (Fig. 1.10) (245). Spatially, the secondary structural elements subdivide the enzyme scaffold in α- and β-rich domains, where the helical rich domain contains C-L helices and the helical poor domain contains A and B helices.

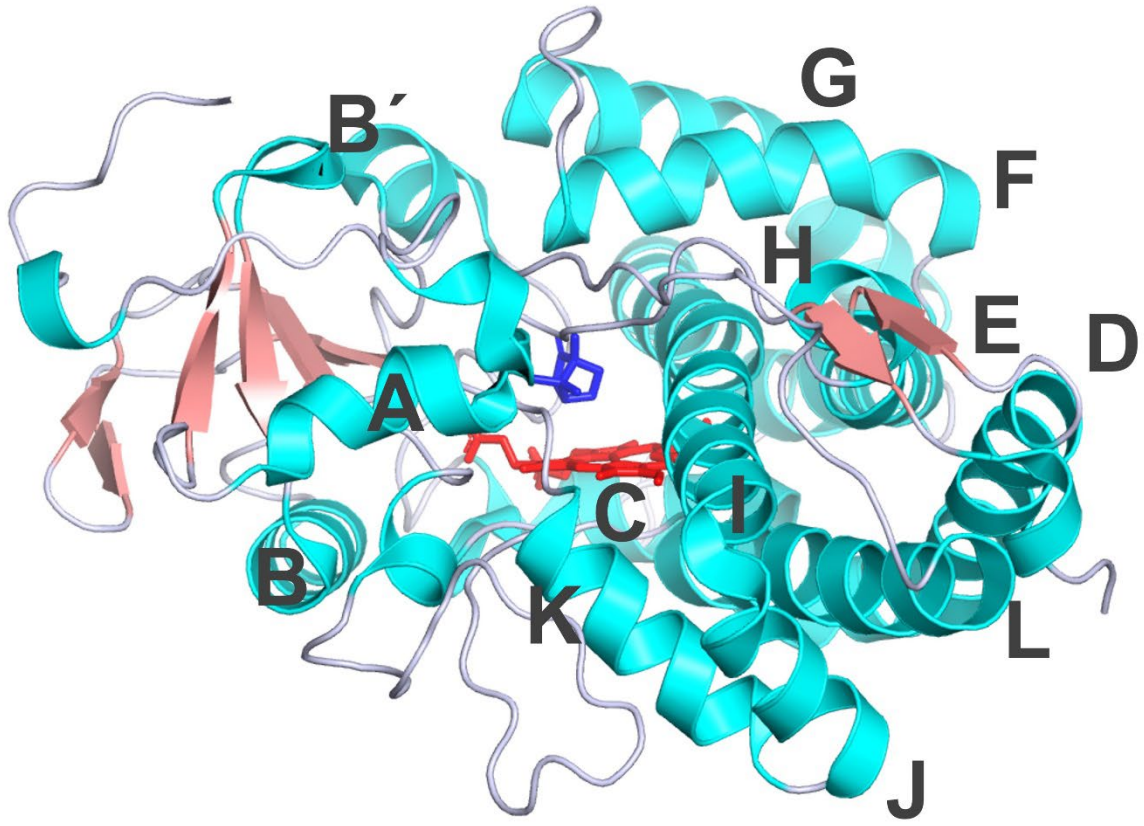


Figure 1.10: Structure of a cytochrome P450. The crystal structure of *Pseudomonas putida* CYP101A1 in complex with its preferred substrate, camphor (PDB: 1DZ4) (246).

Typically, P450 core structure consists of a four-helix bundle, D, E, I, and L that produces a triagonal prism-shaped structure (247). The iron-containing heme prosthetic group is bound to an invariant residue cysteine proximally situated at the N-terminus of L-helix. Within bacterial P450s, the cysteine-containing GXXXC motif in the heme-binding loop is highly conserved. An important feature of the P450 scaffold is that it contains a long I-helix running over the distal surface of the heme. In addition, the interhelical BC- and FG-loops are highly flexible and known

to accommodate substrate entry to the active site center. It is posited that the swinging of the F- and G-helices transiently exposes the active site for substrate entry and product exit (248).

1.6.1. Catalytic mechanism

The catalytic mechanism of the cytochrome P450 enzyme is highly complex and varies significantly depending on the type of reaction being catalyzed (226, 249). Even after an extensive investigation of nearly sixty years, some aspects of their catalytic mechanism are still unknown (226). Generally, the catalytic process begins with the substrate binding to the active site of the P450 enzyme (Fig. 1.11). Upon substrate binding, a water molecule from the active site is displaced, producing a transition from a hexacoordinate low-spin heme iron to a pentacoordinate high-spin heme iron (250). The shift in heme coordination triggers a positive change in the reduction potential of the Fe^{3+} center. This event enables electron transfer from NAD(P)H via the P450-reductase system, reducing the Fe^{3+} heme to its Fe^{2+} state (251). The facile binding of molecular oxygen to the Fe^{2+} heme produces an Fe^{II}-O_2 complex. Subsequently, the transfer of a second electron from NAD(P)H produces a short-lived peroxide-bound state. After the addition of a proton, heterolytic scission of the O-O bond produces H_2O and a highly reactive ferryl (i.e., Fe^{IV}=O[porphyin$^{\bullet+}$]) species referred to as "compound I". This highly reactive compound I then abstracts another proton, leading to the formation of ferryl-hydroxo compound (referred to as "compound II") (252). Finally, the reaction of the substrate radical with the hydroxyl group of compound II generates the hydroxylated product. The release of the product from the active site allowed a water molecule to reoccupy the active site to coordinate with Fe^{3+}. This event restores the resting state and completes the catalytic cycle.

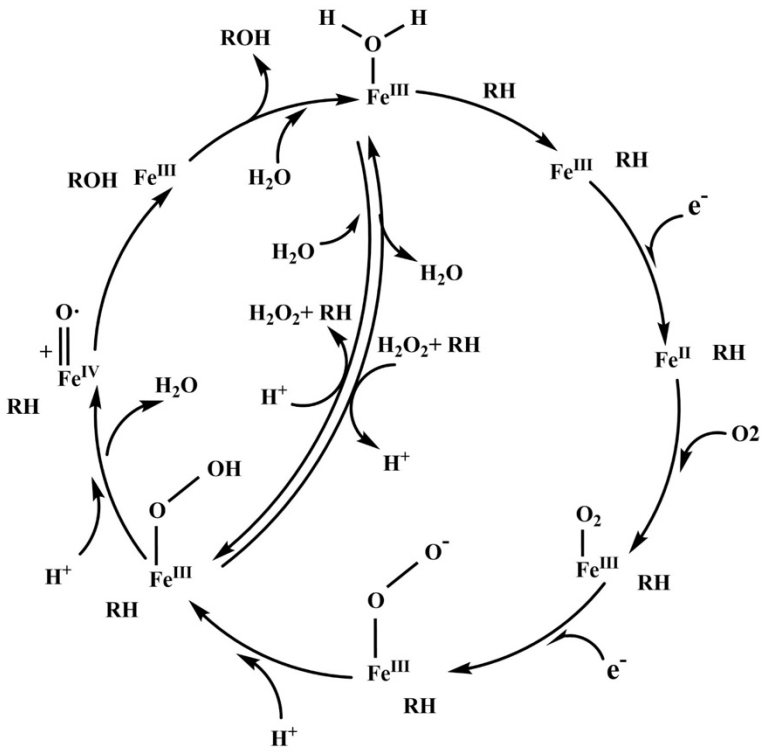


Figure 1.11: A generalized scheme of the P450 catalytic mechanism (159).

1.6.2. *Bacillus* P450s

Bacillus cytochromes P450 have been of great interest since the discovery of a full-length $P450_{BM3}$ in the 1970s (217, 253–255). Numerous studies have been conducted on $P450_{BM3}$ as a tool to understand enzyme catalysis and engineer novel variants with enhanced and/or useful properties for application in pharmaceutical and biotechnology, employing directed evolution or rational design (256–259). This enzyme is known to catalyze the hydroxylation of medium to long-chain fatty acids (257). However, the degree to which $P450_{BM3}$ catalyzes this reaction *in vivo,* and correspondingly, its exact physiological role is yet unknown. To date, only a handful of P450s have been identified in secondary metabolite-producing BGCs from *Bacillus* species. These may

contribute to the functional modification of respective secondary metabolites (159, 260). Examples of such P450s enzymes are CypX, PksS, and CYP109 (Fig. 1.12) (231, 261, 262).

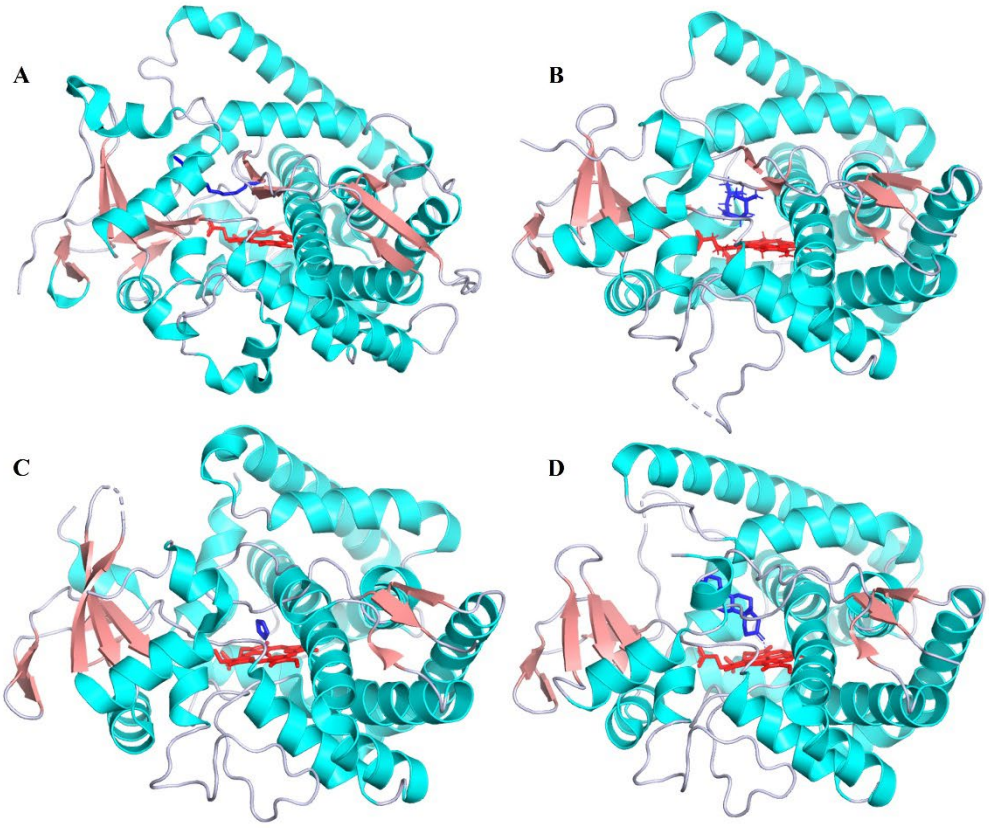


Figure 1.12: Crystal structures of *Bacillus* P450s. (A) CYP102 P450 domain ($P450_{BM3}$), PDB: 4ZFA (263), (B) CypX (CYP134), PDB: 7OW9 (264), (C) PksS, PDB: 4YZR (265), and (D) CYP09 (the versatile enzyme), PDB: 5L94 (266).

However, their distributions among biosynthetic pathways, sequence conservation, tertiary structure, substrate preference, catalytic mechanism, and exact physiological roles are yet unknown. New genomic, computational, and biotechnological tools and the data they have generated make it an opportune time to investigate BGC-affiliated P450s through *in silico* gene

comparison, structural modeling, molecular docking, molecular dynamics simulation, and *in vitro* P450-catalyzed oxidative transformations of candidate substrates.

A previous study reported that cytochrome P450 genes are frequently observed as accessory genes in the biosynthetic gene clusters (BGCs) of *Bacillus* species (260). These P450s are anticipated to participate in the functional modification of associated secondary metabolites or their precursors (267, 231). The *PKs*S gene codes one such P450. It was first identified in the *B. subtilis* (strain 168) as part of a polyketide-producing BGC (231). Its proposed catalytic function is the hydroxylation of a polyketide called dihydrobacillaene (231). CypX (CYP134A1) is another *B. subtilis* P450 found within a BGC. Its putative function involves hydroxylation of various cyclodipeptides (261). The CYP109 enzyme, also discovered in *B. subtilis,* is characterized as a "versatile enzyme" due to its ability to oxidize various substrates, including fatty acids, n-alkanes, primary alcohols, terpenoids, and the steroid testosterone (262).

1.6.3. P450$_{BM3}$

The P450$_{BM3}$ enzyme (CYP102A1) was discovered over forty years ago in *Bacillus megaterium* (254). It is referred to as P450$_{BM3}$ since it was the third P450 enzyme isolated from the *B. megaterium* species. It is a full-length enzyme comprised of a CYP domain as well as a fused P450-reductase domain. Due to its catalytic role in the oxidative transformation of fatty acid substrates, the P450$_{BM3}$ is commonly referred to as "fatty acid hydroxylase" (257). However, the catalytic ability of its laboratory variants has been extended to include the oxidation of many non-natural substrates, including aniline, chlorostyrene, fluorene, hexane, indole, β-lactone, phenanthrene, and many others (257). For fatty acid substrates, the hydroxylation of the substrate takes place exclusively at the fully reduced alkyl end (i.e., the ω-1, ω-2, or ω-3 positions) (268).

The $P450_{BM3}$'s substrates include 12 - 20 carbon fatty acids that can be saturated, unsaturated, or branched-chain. Its preference is for pentadecanoic acid (257). The full-length $P450_{BM3}$ is a continuous ~120-kDa polypeptide, consisting of a 55 kDa CYP domain fused with a 65 kDa reductase domain (257). The CYP domain contains the heme prosthetic group which marks monooxygenase active site, the reductase domain contains two prosthetic flavin groups, one FAD (flavin adenine dinucleotide) and FMN (flavin mononucleotide), in an equivalent ratio (257). The flavins facilitate electron transfer from the hydride-only NAD(P)H donor to the CYP domain heme center which accepts two sequential one-electron transfers. A crystal structure for the full-length $P450_{BM3}$ has not been obtained. This is likely due to difficulties in crystallizing the intact $P450_{BM3}$ with its constituent cofactors and highly flexible linker which connects the CYP and reductase domains (269). Further, the exact physiological role of $P450_{BM3}$ is yet unknown even after extensive investigations over four decades.

1.6.4. CypX (CYP134A1)

The CypX (CYP134A1) was first isolated from *Bacillus subtilis* (261). Its proposed catalytic activity involves three-step oxidative transformations of a diketopiperazine, cyclo-^{L}leucyl-^{L}leucyl into pulcherriminic acid, a precursor of extracellular siderophore pulcherrimun (261). The oxidation of ^{L}leucyl-^{L}leucyl by cypX focuses on the diketopiperazine amine nitrogen atoms, converting them to the corresponding N-oxides, concomitant with the aromatization of the diketopiperazine ring either via hydroxylation and elimination of water or directly via an electron transfer reaction (261). Another study has shown that CypX could catalyze the oxidative transformation of steroidal substrates such as androsta-1,4-diene-3,17-dione, methyltestosterone, and progesterone (270). The CypX enzyme is one where the CYP domain is a stand-alone protein

that requires a separate reductase system for electron delivery. Accordingly, CypX has a molecular weight of about 45 kDa. CypX also bears the usual Cys thiolate-ligated heme center. Its catalytic process requires a partner reductase system comprising ferredoxin reductase and ferredoxin. At present, its precise catalytic mechanism and the range of its substrate preference for dipeptide homologs are not well understood.

1.6.5. PksS

The *PksS* gene was first reported in 1993 as part of the bacillaene-producing BGC in *Bacillus subtilis* (267). However, the link between the PksS enzyme and bacillaene biosynthesis was not confirmed until 2006 (178). This is the first P450 from a *Bacillus* species reported to be involved in a polyketide biosynthetic pathway. The molecular weight of PksS is approximately 47 kDa. In light of this, it is not surprising to find that the PksS consists of a stand-alone CYP domain containing a Cys thiolate-ligated heme center, and it requires an external reductase system (e.g., ferredoxin/ferredoxin reductase) for catalysis. Reddick *et al.* reported that dihydrobacillaene, not bacillaene, is as the preferred substrate of PksS (231). However, the substrate preference range and the catalytic mechanism of PksS remain unknown. This is likely due to the difficulty with performing substrate-binding and/or substrate-dependent kinetic studies. Such investigations are complicated by the notorious instability of bacillaene and dihydrobacillaene under ambient conditions. At present, elucidating the enzyme's catalytic mechanism, structure, and substrate preference range have not appeared in the literature.

1.6.6. CYP109 "the versatile monooxygenase"

The CYP109 was first identified in the complete genome sequence of *Bacillus subtilis* strain 168 in 1997 (114, 262). Since then, several homologs of CYP109 have been identified,

including A1, B1, C1, C2, D1, and E1, with the sequence identity between homologs ranging from 33 to 47% (266, 271, 272). This enzyme is commonly referred to as the "versatile monooxygenase" due to its ability to bind and catalyze the oxidative transformation of a broad range of substrates, including fatty acids, n-alkanes, primary n-alcohols, terpenoids like (+)-valencene α- and β-ionone, and the steroid testosterone (262). A CYP109 subfamily, CYP109B1, catalyzes the oxidative transformation of fatty acid substrates at the subterminal position. For the terpenoid α-ionone, CYP109B1catalyzes a regioselective hydroxylation reaction to yield 3-hydroxy-α-ionone and 4-hydroxy-β-ionone. In another subfamily, CYP109E1, the enzyme catalyzes the hydroxylation of testosterone, and it has been observed to bind testosterone and corticosterone (266). The molecular weight of the heme-bearing CYP109 enzyme is approximately 47 kDa. Following the pattern of the CypX and PksS, it requires an external ferredoxin/ferredoxin reductase system in order to carry out its catalytic mechanism.

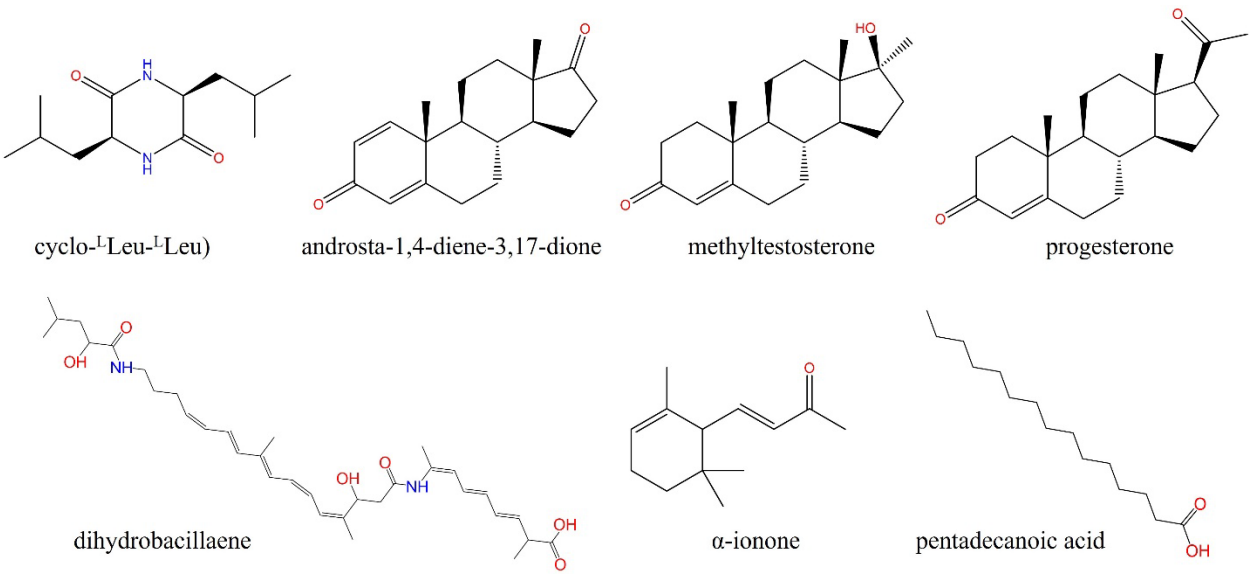


Figure 1.13: Examples of substrates of P450s from *Bacillus* species.

Chapter 2: Broad antibiosis activity of *Bacillus velezensis* and *Bacillus subtilis* is accounted for by a conserved capacity for lipopeptide biosynthesis

2.1. Abstract

Bacillus species produce diverse secondary metabolites that are of great interest in preventing disease due to plant pathogens. In this study, we evaluated 288 strains across 17 *Bacillus* species for antibiosis activity against the root-associated plant-pathogenic oomycete, *Phytophthora nicotianae*. Fifty-nine (20%) out of 288 strains exhibited strong inhibition, while 41 and 188 exhibited weak and no inhibition, respectively. Bioactivity was highly concentrated among only five species (*B. pumilus*, *B. safensis, B. altitudinis, B. velezensis*, and *B. subtilis*), accounting for 56 out of the 59 strongly inhibitory strains. These observations prompted additional investigation into the common biosynthetic gene clusters (BGCs) and secondary metabolites produced by these organisms. Of the 59 *P. nicotianae*-inhibitory strains, all the *B. velezensis* strains and all but one of the *B. subtilis* strains also showed strong antibiosis activity against three fungal pathogens, *Fusarium oxysporum*, *Fusarium graminearum*, and *Rhizoctonia solani* and were thus designated as *generalists*. Conversely, none of the *B. pumilus*, *B. safensis*, and *B. altitudinis* strains showed any inhibition of *F. oxysporum* or *F. graminearum* and only weak inhibition of *R. solani*, earning them the designation *Oomycete specialists*. Strains from specialist and generalist species were predicted to contain, on average, 9 to 13 BGCs, whereas strains from the 12 non-inhibitory species carried only 5 to 8 BGCs. Strikingly, lipopeptide BGCs (1 – 3 on average) were prominently represented within the five bioactive species and virtually absent from the 12 non-inhibitory species. Surfactin was evenly distributed across all specialists and generalists. However, among the generalists *B. subtilis* strains also carried a fengycin BGC, and *B. velezensis* encoded a

novel iturin and fengycin BGC. Iturin (including bacillomycin L), fengycin, and surfactin were the most commonly observed lipopeptide BGCs among the most bioactive species, with many strains containing all three. These lipopeptides from strongly inhibitory *B. velezensis* JJ334 were isolated, identified, and characterized by LC-MS. Each lipopeptide was also evaluated for its antibiosis properties by disk diffusion and microtiter plate-based antibiosis assays. Fengycin and bacillomycin L produced potent antibiosis activity against oomycetes and fungi as compared to surfactin while fengycin was the strongest inhibitor among lipopeptides evaluated. Six to ten derivatives of each lipopeptide were observed, accounted for primarily by the variable length of the fatty acid side chain. This study is the first of its kind on the evaluation of diverse *Bacillus* species against *P. nicotianae* and fungi in correlation with conserved capacity for the biosynthesis of potent inhibitory lipopeptides.

2.2. Introduction

Plant-pathogenic fungi and oomycetes are the cause of serious and intractable diseases that result in multiple billions of dollars in annual crop losses (273). Wheat stem rust, rice blast, corn smut, soybean rust, and potato late blight destroy five of the most important global cash crops, threatening global food security especially in developing nations (4). The agricultural impacts of climate change combined with the overuse of synthetic fungicides have simultaneously produced an increase in fungal diseases and fungicide resistance (38, 274, 275). Due to these concerns, biological agents have been considered as an alternative and promising strategy for disease control and crop management (276). Studies have shown bacteria from the genera *Bacillus*, *Streptomyces*, *Pseudomonas* can be effective biocontrol agents against various plant pathogens (277, 278). In particular, *Bacillus* species have shown strong bioactivity against plant pathogens (15, 279, 280).

Currently, several *Bacillus*-based biological agents are commercially distributed worldwide, including Avogreen (*B. subtilis*), Ballad Plus i Sonata (*B. pumilus*), RhizoVital 42 (*B. amyloliquefeciens*), EcoGuard TM (*B. licheniformis*), and Botrybel (*B. velezensis*) (9–12).

Bacillus species inhabiting the soil and plant rhizosphere, are particularly well-suited as biological agents due to their adapted physiology in that environment (15). They are known as plant growth-promoting rhizobacteria (PGPR) for their contributions to plant growth and disease biocontrol (8). Factors contributing to these aspects include abilities in plant-root colonization and production of allelochemicals such as siderophores, antibiotics, biocidal volatiles, lytic and detoxification enzymes (278). The most common antimicrobials produced by *Bacillus* species are peptides, polyketides, betalactones, fatty-acid derivatives, and lytic enzymes (17). These compounds are known as secondary (or specialized) metabolites, often encoded by biosynthetic gene clusters (BGCs).

The most common bioactive secondary metabolites produced by *Bacillus* species are non-ribosomal and ribosomal-peptides and polyketide-derived macrolides (17). The non-ribosomal peptides (NRP) are produced by multimodular BGCs called non-ribosomal peptide synthetases (NRPS) that accept proteinogenic or modified amino acids as substrates (282). The most extensively studied *Bacillus* NRPs are cyclic lipopeptides (e.g., surfactin, fengycin, iturin, etc.), and siderophores (e.g., bacillibactin) (17). In particular, *Bacillus* lipopeptides are known to be strongly antagonistic against plant pathogenic fungi and oomycetes (283). Structurally, these lipopeptides are small peptides (5 - 12 amino acids) consisting of a cyclic lactone ring with a linked β-amino or β-hydroxy fatty acid of variable carbon chain length. The bioactivity of each lipopeptide varies significantly, ranging from broad to narrow-spectrum antifungal or antibacterial

activity, which may depend on the chemical properties of constituent amino acids as well as fatty acid chain length and branching (283). The common modes of action of these lipopeptides are cell lysis, cell membrane leakage, inhibition of enzymes, and inhibition of protein synthesis of target pathogens (284).

Past studies have demonstrated an excellent biological activity of several *Bacillus* species against plant pathogens in correlation with their abilities to produce single or multiple secondary metabolites (147, 156, 285–288). However, the bioactivity of the *Bacillus* strain and derived natural products are mostly studied either in a single strain or a few strains within the same species. Consequently, a comparative overview of the conserved roles of such natural products in correlation with their diversity and extent of expression in diverse *Bacillus* species is missing from the literature. In particular, the breadth of *Bacillus* antibiosis activity across various plant pathogens such as oomycete and fungi in connection with the ability for secondary metabolite production is far less understood.

To address this gap, we carried out a comparative evaluation of the antibiosis activity of 288 strains of diverse *Bacillus* species in connection with their ability to produce antimicrobial secondary metabolites. Antibiosis screening of these *Bacilli* against plant pathogenic oomycete, *Phytophthora nicotianae* identified 59 (20%) strongly inhibitory strains. These were further evaluated for broad-spectrum inhibitory properties against three plant pathogenic fungi: *Fusarium oxysporum*, *Fusarium graminearum*, and *Rhizoctonia solani*. Genomic analysis of these *Bacilli* showed a striking strong conservation of three lipopeptide BGCs (iturin/bacillomycin L, fengycin, and surfactin) among strongly inhibitory but not non-inhibitory *Bacillus* species. All three (bacillomycin L, fengycin, and surfactin) were produced, extracted, and isolated from a

representative strong inhibitory *B. velezensis* strain. Characteristic chemical properties and antibiosis activity of each purified lipopeptide were further evaluated using UV-vis absorption spectroscopy, liquid chromatography, high-resolution mass spectrometry, and plate-based antibiosis assays.

2.3. Materials and Methods

2.3.1. Strains, chemicals, and culture conditions

In the present study, 288 PGPR strains of diverse *Bacillus* species obtained from various plant rhizospheres were evaluated as potential biological agents. For genomic analysis, Illumina-generated draft genome sequences were analyzed, trimmed, and assembled using CLC Genomic Workbench (289). Genome quality was further evaluated by CheckM v1.1.3. (290), and 288 genome sequences with returned sequence completeness of greater than 70% were included in this study. Twenty-nine (out of 288) *Bacillus* strains with the greatest biocontrol potential were further sequenced at Nanopore for single-contig complete genome sequence. The taxonomy of all *Bacillus* strains was confirmed by the top hit of average nucleotide identity (ANI) of whole-genome sequence at Microbial Genomes Atlas (MiGA) webserver (291). All 288 *Bacillus* strains were taxonomically distributed into 17 species that include "*Bacillus* (others)" representing a collection of 30 strains from less commonly observed species contributing < 5 strains. A phylogenetic tree was constructed using 16S rRNA sequence of each type strain matching all 17 species obtained from EZbioCloud database (292). For routine bacterial growth, the cells were cultured at 37 °C in tryptic soy broth (TSB) medium, supplemented with 1.5% Bacto agar (if required).

2.3.2. Antibiosis assay of *Bacillus* strains against *P. nicotianae* and fungal pathogens

Bacillus strains were screened for their abilities to inhibit the growth of the root-associated plant-pathogenic oomycete, *P. nicotianae* in a plate-based assay. *P. nicotianae* was grown in a V8 agar medium (180 mL/L V8 juice, 2 g/L $CaCO_3$, and 15 g/L Bacto agar) while the *Bacillus* strains were grown in TSB (tryptic soy broth) medium supplemented with 1.5% Bacto agar. Assay plates were prepared using the V8 agar medium in which bacterial colonies were transferred into a well (diameter = 10 mm) containing TSB-agar at the edge of the plate and the *P. nicotianae* was transferred as a plug to the center of the plate. The growth inhibition of *P. nicotianae*'s hyphae due to the presence of *Bacillus* colonies was recorded after a 7-10 day incubation. The inhibitory responses of *Bacillus* strains were classified as strong, weak, and no inhibition based on the measurement of the zone of inhibition (ZOI) and morphological changes of both *Bacillus* strains and *P. nicotianae* being evaluated. A strong inhibition was assigned for a clear zone of inhibition (ZOI) of 5-15 mm with complete elimination of *P. nicotianae*'s hyphae in the interface of *Bacillus* colonies and *P. nicotianae*, while no inhibition was assigned when the *P. nicotianae*'s hyphae spread over the *Bacillus* colonies with no observable ZOI. Exhibition of a less clear ZOI of 2-7mm with substantial reduction of *P. nicotianae*'s hyphae was assigned as weak inhibition.

Fifty-nine *P. nicotianae*-inhibitory *Bacillus* strains were further evaluated for antibiosis against *F. oxysporum*, *F. graminearum*, and *R. solani*. Each organism was assayed and evaluated in the same condition as described above for *P. nicotianae* and resulting antibiosis response was similarly classified as strong, weak, and no inhibition based on ZOI and morphological changes of organisms evaluated.

2.3.3. Calculation of bioactivity index

Conservation of antibiosis activity expressed by various *Bacillus* species was calculated on the basis of the distribution of strong, weak, and no inhibition among strains within each species. A term "bioactivity index" accounting for such conservation of antibiosis activity was calculated using a weighted-average score of 1 for strong inhibition, 0.5 for weak inhibition, and 0 for no inhibition using the following equation:

$$Bioactivity\ Index = \frac{(\#\text{strong inhibitor strains} \times 1) + (\#\text{weak inhibitor strains} \times 0.5) + (\#\text{ noninhibitory strains} \times 0)}{\#\text{ of total strains within a species}}$$

As defined, bioactivity index of 1 indicates the highest expression of antibiosis activity, where all tested strains of a given species show strong antibiosis activity. Conversely, a 0 indicates no expression of antibiosis activity (i.e., no tested strains of a given species showed antibiosis activity).

2.3.4. Genome mining and bioinformatics analyses of *Bacillus* strains

Genome sequences of all 288 *Bacillus* strains were analyzed by antiSMASH v.5 (antibiotics and secondary metabolite analysis shell) (189) to predict biosynthetic gene clusters (BGCs) and secondary metabolites. Predicted BGCs were further dereplicated based on respective BGCs from single contig complete genome sequences to eliminate duplicated and/or fragmented BGCs. To infer conservation in sequence and putative function, predicted BGCs were grouped into networks of clusters based on sequence similarities using BiG-SCAPE v.0.0.0r (Biosynthetic Gene Similarity Clustering and Prospecting Engine) (293). Finally, network distances generated by BiG-SCAPE analysis were visualized and annotated using Cytoscape 2.8 (294).

2.3.5. Extraction of secondary metabolites from bioactive *Bacillus* strains

Representative strains from five bioactive *Bacillus* species were selected for producing secondary metabolites that may be responsible for antibiosis activity. In order to produce secondary metabolites, the *Bacillus* strains were grown in Landy medium (glucose, 20 g/L, yeast 1g/L, L-glutamic acid 5 g/L, KCl 0.5 g/L, MgSO4 0.5 g/L, KH_2PO_4 1 g/L, L-phenylalanine 3 mg/L, $MnSO_4$ 5 mg/L, $FeSO_4$ 0.15 mg/L, $CuSO_4$ 0.16 mg/L, pH 7.0) for 72h at 30 °C with constant agitation (175 rpm). To pellet cells, liquid cultures were centrifuged at 6,000 × g for 40 mins. The pH of harvested cell-free supernatant was adjusted to 2.0 by dropwise addition of concentrated HCl with constant stirring. Following overnight incubation at 4 °C, the precipitate was collected by centrifugation at 6,000 ×g for 50 mins. The precipitate was extracted twice using 100% MeOH. The pooled MeOH extract was dried under a constant flow of N_2 (*g*), and the dried residue was redissolved in MeOH, filtered with a 0.2 µm Acrodisc syringe filter (Pall Corporation, Ann Arbor, MI), and stored at -20 °C until evaluated.

2.3.6. Evaluation of antibiosis of total extracts against *P. nicotianae*

The antibiosis assay of the total extract from each *Bacillus* strain against *P. nicotianae* was carried out using a disk diffusion method (288). The assay was conducted using a V8 agar plate. Freshly grown *P. nicotianae* was transferred as a plug (diameter = 5 mm) to the center of the assay plate and allowed to grow for 72 h. Ten µL of the total extract was added onto a sterilized filter disk (diameter = 6 mm) and then placed at the edge of the assay plate. The growth inhibition of *P. nicotianae* hyphae surrounding the filter disk was measured and recorded after 5-7 days of incubation at 25°C.

2.3.7. Evaluation of chemical properties of total extract by UV-vis absorption and LC-MS

UV-vis absorption spectra of *Bacillus* strain total extract were evaluated for characteristic absorption features. To detect compounds at 220 nm, 20 µL of total extract were separated through a ZORBAX SB-C18 column (5 µm, 4.6 × 150 mm) for 85 mins at a flow rate of 0.20 ml/min using an Agilent Infinity 1100 HPLC system (Santa Clara, CA). Solvent A (100% water) and B (100% acetonitrile), each containing 0.1% trifluoroacetate (v/v) were used with the following elution gradient for solvent B: 40% at 0 min, 55% at 15 min, 75% at 40 min, 100% at 60 min, and 75% at 77 min. To identify compounds in the total extract, 0.2 µL of the extract was separated by LC through an Acquity UPLC BEH C18 (2.1 × 50 mm, 1.7 µm) column and eluted onto an Thermo Fisher Exploris 120 orbitrap LC-MS (Milford, MA). The LC separation was run for 20 mins with a column temperature of 40 °C and a flow rate of 0.20 mL/min. The mobile phase gradient was created using water (A) and acetonitrile (B), each containing 0.1% formic acid (v/v) such that solvent B was at 40% at 0 min and ramped to 100% at 14 min, and then back to 40% B at 16 min.

Ions were generated using both positive and negative ionization modes employing an electrospray ionization (ESI) source. In addition, fragment ions from the precursor ions were simultaneously produced in a high-stage MS^n analyzer. The identity of compounds from the total extract were initially confirmed by the parent ions generated by both positive and negative ionization modes. The mass spectra of each fragmented ion generated by the MS^n analyzer from the corresponding precursor ion produced in the positive ionization mode were used for the unambiguous identification of each compound. Further, characteristic fragment ions generated from lipopeptide core peptides were used as diagnostic ions for the identification of lipopeptide derivatives. For fengycin derivatives, two reporter fragment ions (A and B), generated by the cleavage of Orn2-Tyr3 (A) and Glu1-Orn2 (B) bonds from fengycin core peptide (Glu1-Orn2-

Tyr3-Thr4-Glu5-Ala/Val6-Pro7-Gln8-Tyr9-Ile/Val10) were used as diagnostic ions for unambiguous identification (145, 146, 295, 296). For surfactin, the fragment ions generated by the cleavage of Glu1-Leu/Ile2 bond from the core peptide (Glu1-Leu/Ile2-Leu3-Val4-Asp5-Leu6-Leu/Ile7) and the remaining Glu1-fatty acid tail were used to determine the derivatives and length of fatty acid tail (297–299). Similarly, the fragment diagnostic ions generated by the cleavage of Asn-Tyr (278.11) and Asn-Tyr-Asn (392.15) fragments from the core peptide and fragmented fatty-acid tail were used for unequivocal identification of bacillomycin L derivatives (143).

2.3.8. Isolation of bioactive secondary metabolites using HPLC

Agilent Infinity 1100 LC system was used for isolating bioactive compounds from the total extract of *Bacillus* strains. 100µl of the total extract was injected and eluted through ZORBAX (Santa Clara, CA) SB-C18 column (5 mm, 4.6 × 150 mm) for 85 min. at a flow rate of 0.20 mL/min. The mobile phases were water (A) and acetonitrile (B), each containing 0.1% trichloroacetic acid (v/v) with the following gradient for solvent B: 40% between 0-8 min., 55% between 15-30 min., 75% between 40-50 min., 100% between 60-75 min., 75% at 55 min., and 40% at 79 min. The compounds were detected at 220, 275, 375, and 450 nm by a diode array detector coupled with a full-spectrum (220 - 500 nm) analysis. Fifteen to 20 fractions were collected and pooled and concentrated from five consecutive runs. The purity of compounds in each fraction was evaluated by UV-vis absorption as well as by MS analyses.

2.3.9. Antibiosis evaluation of isolated lipopeptides and target pathogens

Purified bacillomycin L, fengycin, and surfactin were evaluated for bioactivity against *P. nicotianae, F. oxysporum*, *F. graminearum*, and *R. solani* using a disk diffusion assay (288). The

assay was conducted onto a V8 agar plate wherein a freshly grown target pathogen was transferred as a plug (d = 5 mm) to the center of the assay plate and allowed to grow for 72 h. Ten µL of the purified lipopeptide was added into a sterilized filter disk (d = 6 mm) and then transferred onto the edge of the assay plate. The growth inhibition of *P. nicotianae*'s hyphae surrounding the disk was measured and recorded after 5 - 7 days. In order to determine the inhibitory strength of purified bacillomycin L, fengycin, and surfactin, a quantitative bioassay was carried out using a 96-well microtiter-based plate assay against *P. nicotianae* and *F. oxysporum* as described by Romano *et al*., with minor modifications (300). Both *P. nicotianae* and *F. oxysporum* were grown in a diluted V8 medium (80 mL/L of V8 juice, 1 g/L of $CaCO_3$, and 6 g /L Bacto agar). Freshly grown plugs of target pathogens were macerated by passing through a 22-gauge needle attached to a 10 mL syringe and further homogenized by vortexing for 2 min. Twenty µL of homogenized culture macerate, 160 µL of diluted V8 broth (80 mL/L of V8 juice and 0.5 g/L $CaCO_3$), and 20 µL of lipopeptide with desired concentration were then loaded into a 96-well microtiter plate using wide-orifice tips. Plates were incubated for 2 days at 25°C and the growth of the target organism was determined spectrophotometrically at 600 nm by a microtiter plate reader (Biotek Instruments, Highland Park, VT).

2.4. Results

2.4.1. Concentration of *Phytophthora nicotianae* antibiosis activity among five *Bacillus* species

A library of 288 PGPR strains representing 17 *Bacillus* species was evaluated for antibiosis activity against a root-associated plant-pathogenic oomycete, *P. nicotianae*. Table 2.1 shows the number of strains belonging to each *Bacillus* species analyzed for this study.

Table 2.1: Diversity of PGPR strains of *Bacillus* species evaluated for biocontrol ability.

***Bacillus* species**	**Number of strains**
B. megaterium	54
B. velezensis	50
Bacillus (other)[1]	30
B. safensis	26
B. drentensis	19
B. pumilus	17
B. toyonensis	15
B. thuringiensis	14
B. altitudinis	14
B. weihaiensis	11
B. subtilis	7
B. acidiceler	6
B. niacin	5
B. vireti	5
B. pseudomycoides	5
B. wiedmannii	5
B. firmus	5
Total	**288**

[1]Species containing less than five strains were collectively grouped as *Bacillus* (other); these were: *B. humi (4), B. dafuensis* (4), *B. sp. (*2), *B. simplex* (2), *B. luti* (2), *B. bingmayongensis* (2*), B. selenatarsenatis* (2), *B. taxi (2)*, *B. glycinifermentans* (2), *B. paramycoides* (2), *B. dielmoensis* (1), *B. zeae* (1), *B. cereus* (1), *B. circulans* (1), *B. solisilvae* (1*),* and *B. asahii* (1).

Antagonism against *P. nicotianae* was classified as strong, weak, or non-inhibitory based on measurement of zones of inhibition (ZOI) as well as morphological changes to *P. nicotianae* and the *Bacillus* strain being evaluated. Defining characteristics of these levels of inhibition are given in *Materials and Methods*, and a representative antibiosis assay plate illustrating all three levels of inhibition is shown in Fig. 2.1.

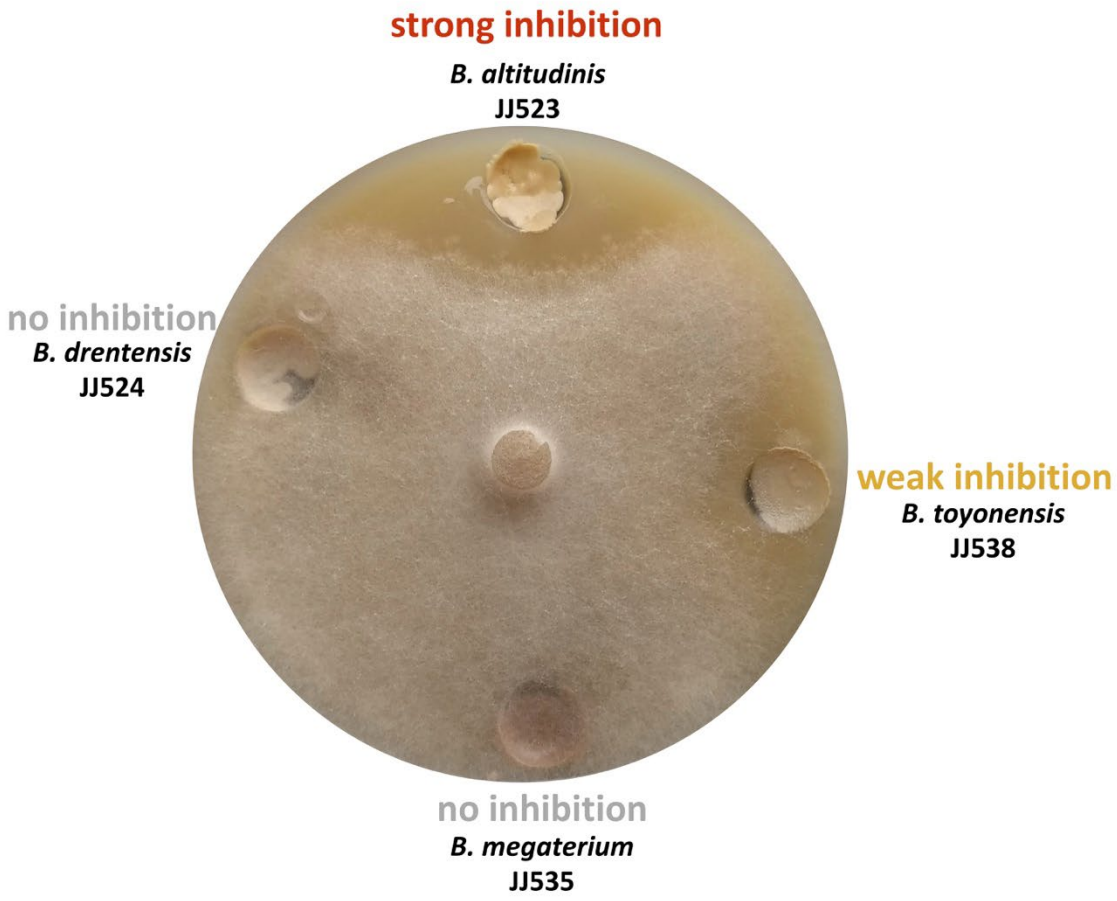


Figure 2.1: *Bacillus* antibiosis against *Phytophthora nicotianae*. A representative antibiosis assay plate showing strong, weak, and no inhibition of *Bacillus* species against the plant-pathogenic oomycete, *P. nicotianae*.

Fifty-nine (20%) of the 288 strains exhibited strong inhibition, while 41 strains were weak inhibitors, and 188 showed no pathogen inhibition at all. Fifty-six out of the 59 strongly inhibitory strains belonged to only five *Bacillus* species: *B. pumilus*, *B. safensis, B. altitudinis, B. velezensis*, and *B. subtilis* (Fig. 2.2A). Interestingly, these five species are more closely phylogenetically related as compared to the other species (Fig. 2.2A), suggesting that the common factors contributing to antibiosis activity may be phylogenetically conserved. For each of these species, at least 40% of strains tested exhibited some level of inhibition (strong or weak): *B. velezensis*

(74%), *B. pumilus* (100%), *B. safensis* (65%), *B. subtilis* (57%), and *B. altitudinis* (43%). Accounting for the overall percentage of inhibitory strains as well as the relative contribution of strong *vs* weak *vs* non-inhibitory strains, a bioactivity index (ranging from 0 to 1) was calculated (see *Materials and Methods*); these five species returned values of 0.39 (*B. altitudinis*), 0.50 (*B. subtilis*), 0.56 (*B. safensis*), 0.61 (*B. velezensis*), and 0.85 (*B. pumilus*) (Fig. 2.2B). Inhibition was sparsely distributed among *B. toyonensis*, and *B. thuringiensis* strains, generating bioactivity indices of 0.27, and 0.21, respectively. Finally, all other species tested showed bioactivity indices ≤ 0.10: *B. firmus* (0.1)*, Bacillus (other)* (0.07), *B. drentensis* (0.03)*, B. megaterium* (0.02), *B. niacin*i (0), *B. vireti* (0), *B. pseudomycoides* (0), *B. weihaiensis* (0), *B. wiedmanni* (0)*, and B. acidiceler* (0) (Fig. 2.2B).

2.4.2. *P. nicotianae*-inhibitory *Bacillus* species divide into specialists and generalists

The fifty-nine strong inhibitors of *P. nicotianae* were further evaluated for antibiosis activity against three fungal pathogens: *F. graminearum, F. oxysporum*, and *R. solani.* Interestingly, strains from *B. velezensis* and *B. subtilis* exhibited strong antibiosis activity against all three fungal pathogens, whereas strains from *B. pumilus*, *B. safensis*, *B. altitudinis*, *B. toyonensis* and *Bacillus (other)* exhibited either weak inhibition or were non-inhibitory (Fig 2.3A). Specifically, all strains from *B. pumilus*, *B. safensis*, and *B. altitudinis* exhibited no inhibition against both *F. oxysporum* and *F. graminearum*. Further, only weak inhibition was observed against *R. solani*. As shown in Fig. 2.2A, *B. velezensis* and *B. subtilis* are phylogenetically closely related; similarly, *B. pumilus*, *B. safensis*, and *B. altitudinis* are closely related to one another. Due to their ability to exert strong antagonism against *P. nicotianae* and all three pathogenic fungi, we classified

B. velezensis and *B. subtilis* as generalists for their broad-spectrum bioactivity. Conversely, we classified *B. pumilus*, *B. safensis*, and *B. altitudinis* as specialists.

2.4.3. Number, type, and distribution of BGCs are distinct between strongly inhibitory and non-inhibitory *Bacillus* species

We surmised that strains strongly inhibitory against *P. nicotianae* were likely to possess conserved factors which would account for their antibiosis activity, and these factors would be absent from the non-inhibitory species. In order to evaluate this hypothesis, draft genome sequences of 288 strains were analyzed for the presence of predicted BGCs using antiSMASH (v.5) (189). A total of 2,442 BGCs (average 8 per strain) were predicted across the 288 genomes, including 1,259 known and 1,183 as yet unknown BGCs. Based on the structural and chemical properties of predicted secondary metabolites, BGCs were grouped into six functional classes: non-ribosomal peptide synthetases (NRPS), polyketides (PKS), a hybrid of PKS and NRPS (PKS-NRPS), ribosomally-synthesized and post-translationally modified peptides (RiPP), terpenes, and BGCs outside these five classes (other).

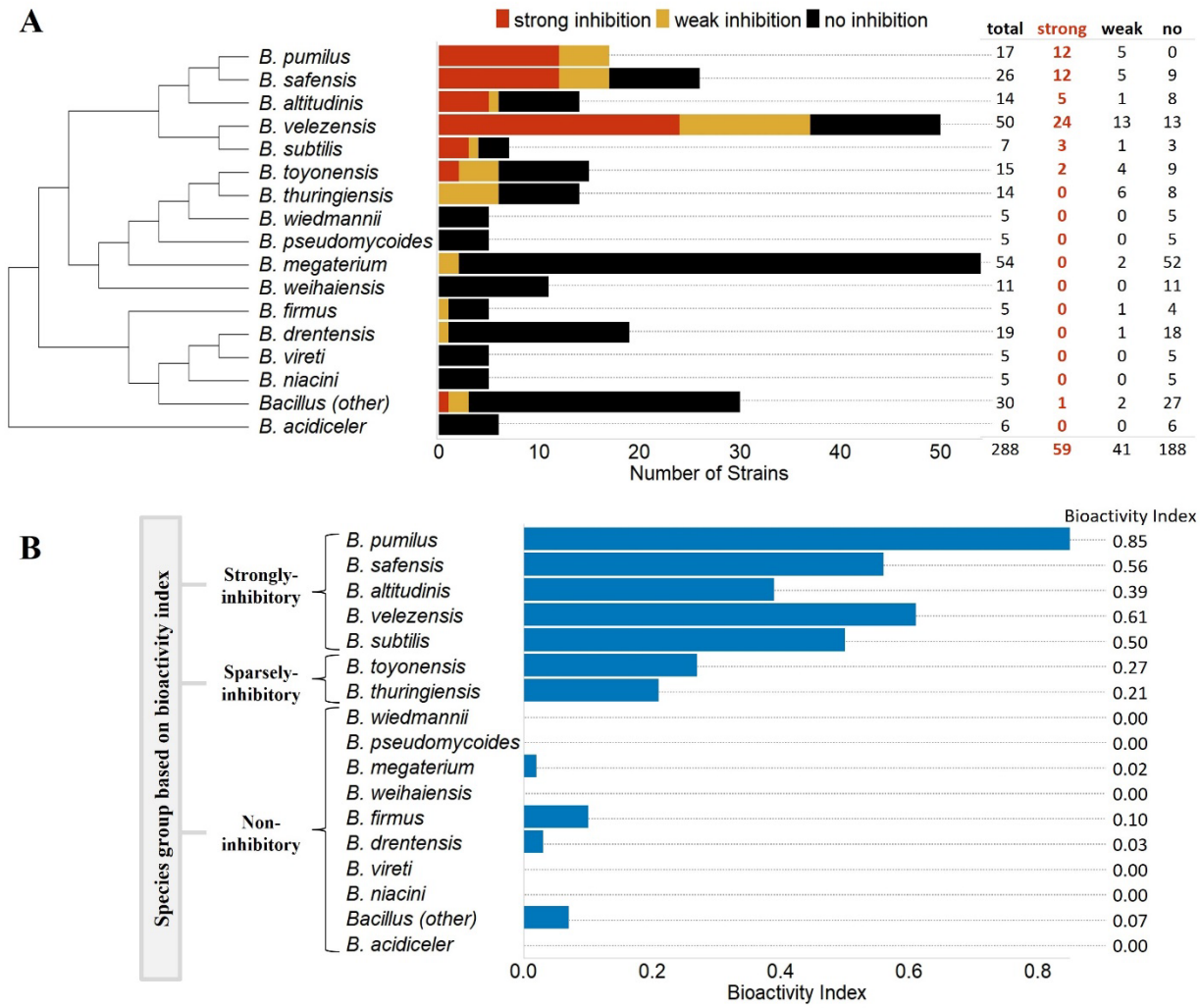
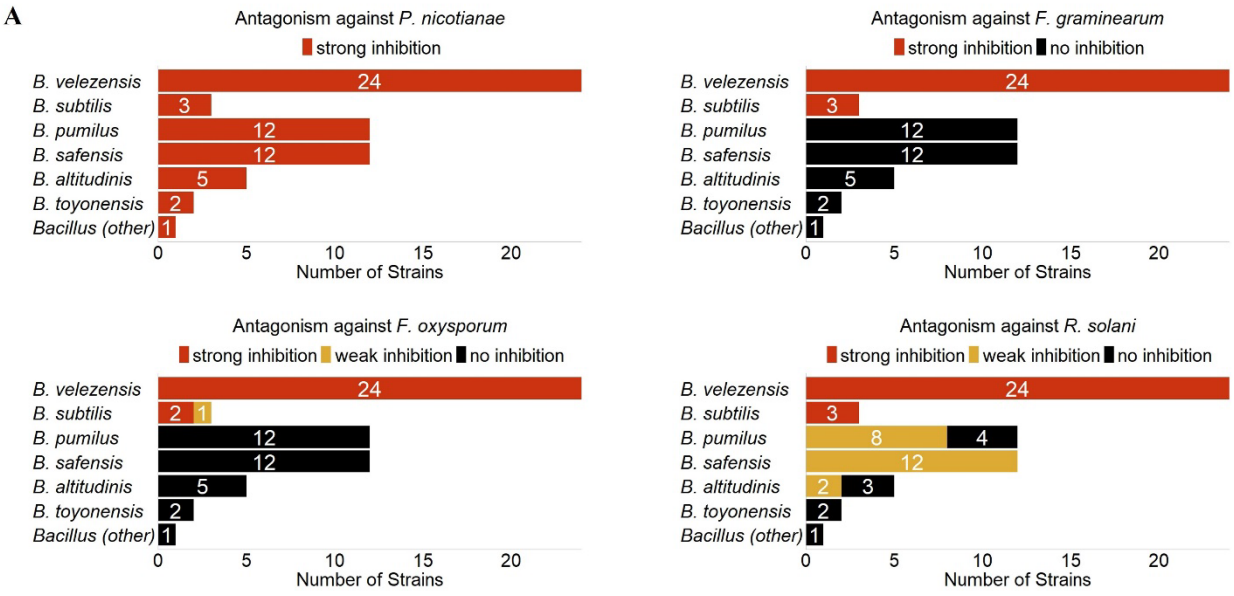


Figure 2.2: Antibiosis activity of *Bacillus* species against *Phytophthora nicotianae*. The antibiosis activity of 288 strains across 17 *Bacillus* species against the plant-pathogenic oomycete, *P. nicotianae* (A) *Bacillus* antagonism was classified as strong, weak, or no inhibition. Species denoted as "*Bacillus* (other)" contained strains from less-commonly observed species, each contributing fewer than five strains. The degree to which antibiosis activity against *P. nicotianae* was expressed across strains from given *Bacillus* species was expressed as a bioactivity index (BI) (B). Scores for BI range from 0 to 1, where 0 would indicate that no strains within a species demonstrated any antibiosis activity, and 1 would indicate that all strains within a species demonstrated strong antibiosis activity against *P. nicotianae* (see *Materials and Meth*ods). Five species were classified as strongly inhibitory based on BI values from 0.39 to 0.85, two were classed as sparsely inhibitory with values from 0.20 to 0.30, and the rest were regarded as non-inhibitory with BI values less than or equal to 0.10.

A



B

Bacillus species	Species group based on breadth of inhibition
B. velezensis *B. subtilis*	generalists
B. pumilus ***B. safensis*** ***B. altitudinis***	**specialists**

Figure 2.3: Antibiosis activity of fifty-nine *P. nicotianae*-active *Bacillus* species. Antibiosis response of *P. nicotianae*-inhibitory *Bacillus* strains are shown against three root-associated plant-pathogenic fungi, *F. graminearum, F. oxysporum*, and *R. solani* (A). Based on the breadth of their inhibition against plant pathogens, the five strongly *P. nicotianae*-inhibitory *Bacillus* species were further subdivided into *generalists* and *specialists* (B).

Bacillus species belonging to the strong inhibitory species group had a higher number of BGCs per strain (8.9 to 13.4) with *B. velezensis* carrying the largest number of BGCs per strain. By comparison, the sparsely inhibitory species group contained 9.0 to 9.5 BGCs per strain, and the non-inhibitory species group contained 4.2 to 8.4. Consistent with these observations, the correlation between the overall number of BGCs and bioactivity was strong (Pearson r = 0.80; p = 0.0001). In addition, a striking distinction was observed in the representation of NRPS clusters among the strongly and sparsely inhibitory groups on one hand (1.9 – 4.3 NRPS/strain) *versus* the

non-inhibitory group on the other where the vast majority of these species showed no NRPS BGCs (Fig. 2.4A). Notably, a distinction was also noted between the strong-inhibition generalists (*B. velezensis* and *B. subtilis*) with an average 3.0 to 4.3 per strain, while the strong-inhibition specialists (*B. pumilus*, *B. safensis*, and *B. altitudinis*) only carried 1.9 to 2.3 NRPS per strain. As a contrast, PKS, RiPP, and terpene BGCs were relatively evenly distributed across all of the *Bacillus* species evaluated, regardless of the level or breadth of inhibition exhibited. Accordingly, correlations between NRPS content and bioactivity were relatively strong (*Pearson r* = 0.63, p = 0.0069) while those between RiPPs and Terpenes were quite poor (*Pearson* r = -0.24, p = 0.3640; *Pearson* r = 0.01, p = 0.9620, respectively) (Fig. 2.5). The PKS and other BGCs occupied the middle, showing moderate correlations with bioactivity index (*Pearson* r = 0.42, p = 0.0953, and *Pearson* r = 0.58, p = 0.0153, respectively). Interestingly, the tightest correlation parameters between bioactivity index and BGC type were observed for the PKS-NRPS clusters (*Pearson* r = 0.79, p = 0.0002); however, one species from the strong inhibitory group (*B. altitudinis*) and one from the sparsely inhibitory group (*B. toyonensis*) carried no PKS-NRPS BGCs at all.

In order to compare their distribution among the bioactive *Bacillus* species based on their structural and chemical properties, NRPS clusters were further divided into four subtypes: lipopeptides, siderophores, other, and unknown. All five species belonging to strongly inhibitory group had at least one and up to 2.8 lipopeptide BGCs per strain on average (Fig. 2.4B). In contrast, lipopeptides were nearly completely absent from species belonging to non-inhibitory and sparingly inhibitory species groups. Siderophores were relatively common among species from strongly and sparsely inhibitory groups, while NRPS classified as “other” were almost completely absent from inhibitory species and the distribution of unknown NRPS was sporadic.

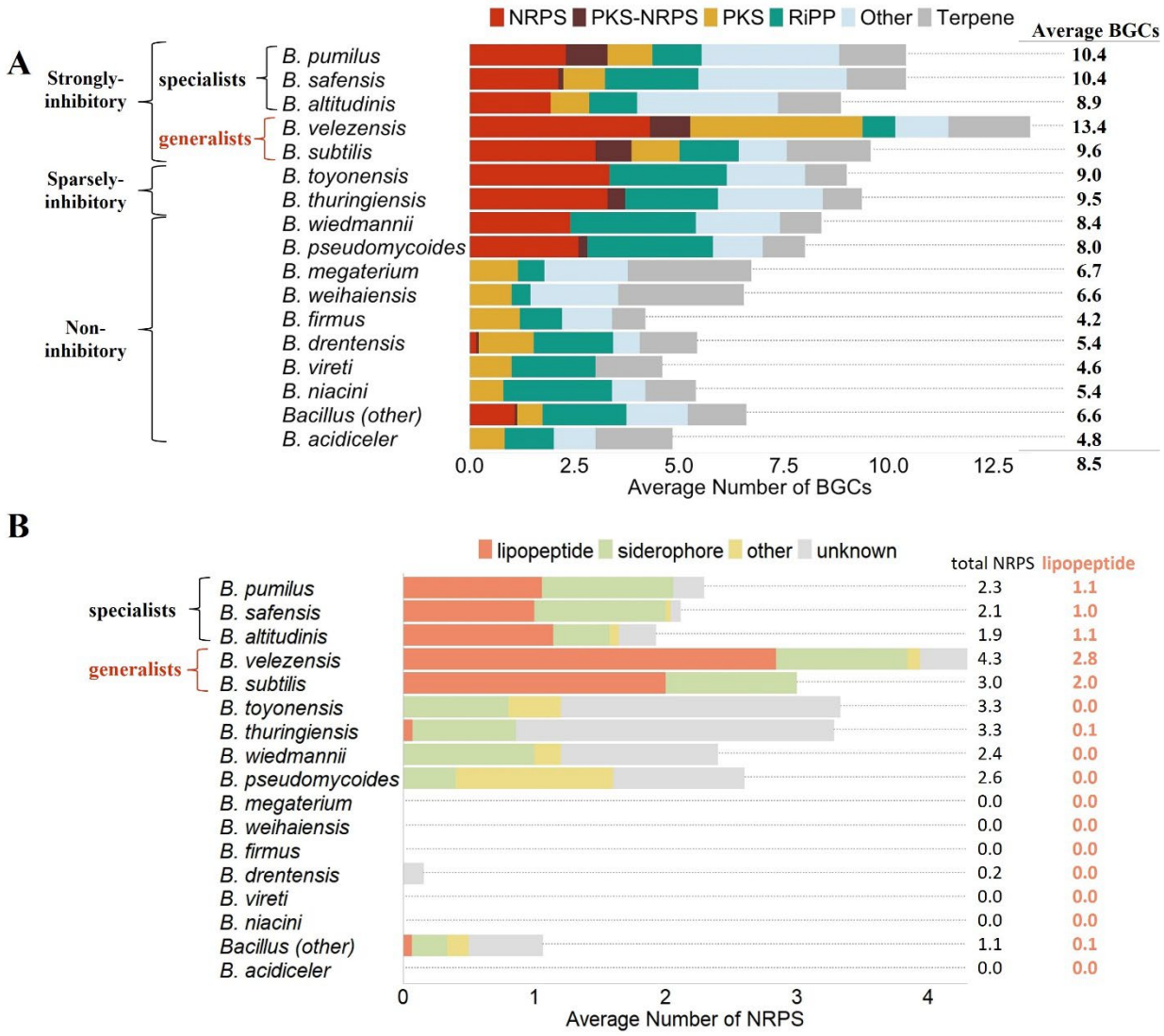
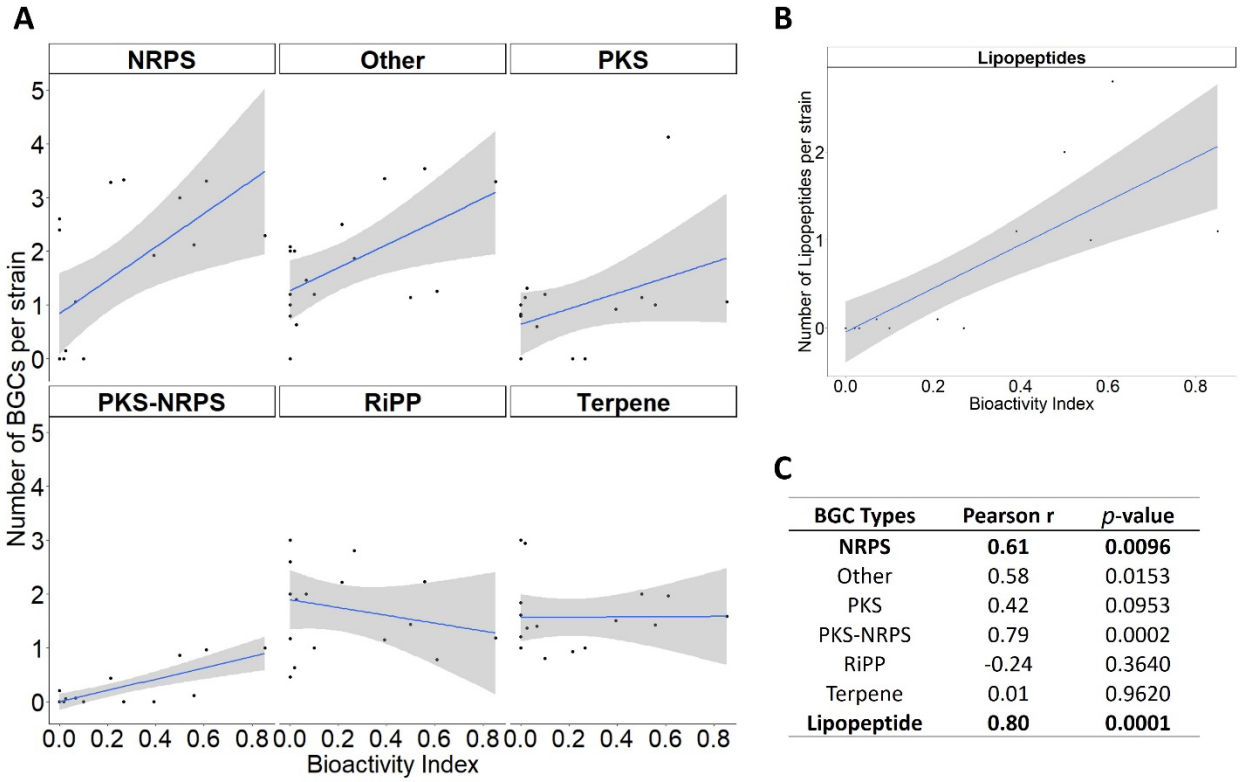


Figure 2.4: *Bacillus* BGC diversity as related to strength and breadth of antibiosis activity. The average numbers of BGCs by type on a per strain basis among *Bacillus* species are shown (A). All BGCs were classified into six functional groups based on the structural and chemical properties of the secondary metabolites predicted by antiSMASH. *Bacillus* species grouped into three classes based on the bioactivity index are designated next to the species, and species groupings into "generalist" *vs* "specialist" based on the breadth of inhibition against various organisms also are indicated (see *Materials and Methods*). The distribution of specific types of NRPS BGCs across *Bacillus* species are also shown (B). The NRPS BGCs were subdivided based on structural and chemical properties of the predicted secondary metabolite: lipopeptide, siderophore, other, and unknown.

These data suggest that lipopeptides, nearly exclusively produced by strains strongly antagonistic to *P. nicotianae*, may be substantial contributors to strong antibiosis activity. Interestingly, species from the generalist group contained two to three lipopeptides per strain while species from the specialist group contained only one lipopeptide per strain. This indicates that a diverse set of lipopeptides may produce a synergistic effect that contributes to the broad-spectrum antibiosis activity by generalists.



BGC Types	Pearson r	*p*-value
NRPS	**0.61**	**0.0096**
Other	0.58	0.0153
PKS	0.42	0.0953
PKS-NRPS	0.79	0.0002
RiPP	-0.24	0.3640
Terpene	0.01	0.9620
Lipopeptide	**0.80**	**0.0001**

Figure 2.5: Correlations of bioactivity index and the number of BGCs from each of six major classes (A) and lipopeptide BGCs in particular (B). The average number of a given BGCs per strain within a given species were plotted against the bioactivity index expressed by that same species. The correlational parameters (*Pearson* r and *p*-value) for the six general BGC classes and specific lipopeptides are shown (C).

2.4.4. Specific lipopeptide BGCs are highly conserved among generalists versus specialists

Biosynthetic gene clusters from all 288 *Bacillus* strains were further analyzed (BiG-SCAPE v.0.0.0r) (293) for gene cluster similarities to determine the extent to which secondary metabolite biosynthesis is conserved. Similarity analysis showed that BGCs identified to produce a common putative metabolite across generalists, specialists, and non-inhibitory strains tended to segregate into separate clusters corresponding to these groups (Fig. 2.6). This was particularly striking among lipopeptide BGCs where the differences in the structure of the gene clusters were highly distinct between antibiosis-based species groupings. Only generalists carried BGCs with modules for synthesis of the fengycin core decapeptide (*fenA – fenE*). Interestingly, the fengycin BGC from *B. velezensis* also contained *ituA*, *ituB*, and *ituC*, the modules necessary for the synthesis of an iturin core heptapeptide (Fig. 2.7A). Although *Bacillus* strains from specialist and non-inhibitory species also carried a gene cluster identified to bear similarity to a fengycin BGC, none of these contained modules for core peptide synthesis, but only genes supporting the synthesis of a putative betalactone. Interestingly, the specific structures of this BGC from specialists on one hand and non-inhibitory *Bacillus* species on the other were distinct (Fig. 2.7A). Similarly, generalist and specialist *Bacillus* species groups all carried a BGC for production of a surfactin-like lipopeptide (Fig. 2.6); no such BGC was identified in any strains from non-inhibitory species. Invariably, generalists (*B. velezensis* and *B. subtilis*) carried a BGC identified as surfactin (Fig. 2.6) characterized by three core genes with modules for the synthesis of a heptapeptide (*srfAA – srfAC*) (Fig. 2.7B). A separate bi-lobed cluster was observed for specialists with high similarity scores (85%) for lichenysin, a surfactin-like lipopeptide (Fig. 2.6). The typical structure for this BGC contained two core genes in addition to the three required for heptapeptide synthesis (Fig. 2.7B). These data suggest that the two to three highly conserved lipopeptides produced by

generalists may contribute to a broad antifungal/antioomycete activity while the single surfactin or surfactin-like BGC carried by specialists may only enable strong anti-oomycete activity. We have observed that some strains belonging to generalists or specialists species exhibited no (or weak) inhibition (see Fig. 2.2A) even though they appear to carry lipopeptide BGCs. It is possible that these strains may ultimately be unable to produce one or more of these lipopeptides due to missing core genes, nonsense mutations, frameshift mutations, or altered gene regulation, as demonstrated by Kiesewalter *et al* (284).

It should also be noted that other clusters of BGCs which putatively generate metabolites with antimicrobial properties segregate along the lines of generalists *versus* specialists *versus* non-inhibitors as well. Many of these BGCs belong to NRPS, PKS-NRPS, or PKS classes. For example, unique bacillibactin BGCs were identified for each of the three groups, and unique bacilysin clusters were each observed for generalists and specialists while non-inhibitors appeared to lack such a cluster altogether. In a similar manner, a bacillaene BGC is found only in generalists, and zwittermicin BGCs were only observed in specialists. Notably, the PKS BGCs for macrolactin and difficidin were only observed in *B. velezensis*.

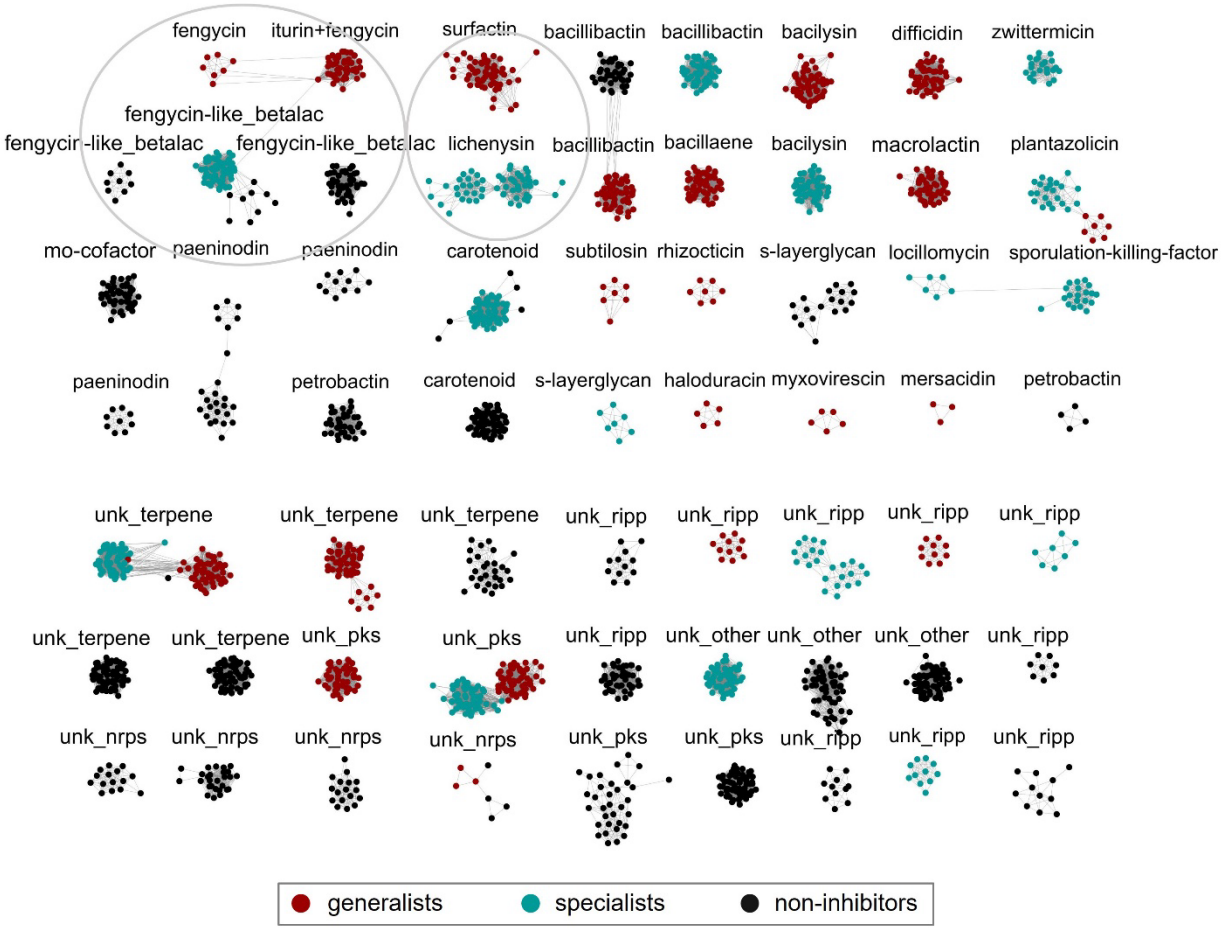


Figure 2.6: Similarity analysis of *Bacillus* BGCs. All 2,442 BGCs identified were grouped based on the distance matrices of gene cluster estimated by BiG-SCAPE (v.0.0.0r) (293) analysis. Clusters containing three or more nodes (2,055 BGCs) are shown here and each cluster is labeled according to the secondary metabolite predicted by antiSMASH v.5 (189). The color of nodes is according to the breadth of observed antibiosis activity (generalist – red; specialist – teal; non-inhibitor – black) for the strain containing the BGC identified. The diversity of BGCs connected with fengycin and surfactin production are circled. Abbreviations for cluster labels are as follows: fengycin-like betalactone (fengycin-like_betalac), molybdenum cofactor (mo-cofactor), unknown RiPP (unk_ripp), unknown NRPS (unk_nrps), unknown PKS (unk_pks), and unknown other BGC (unk_other).

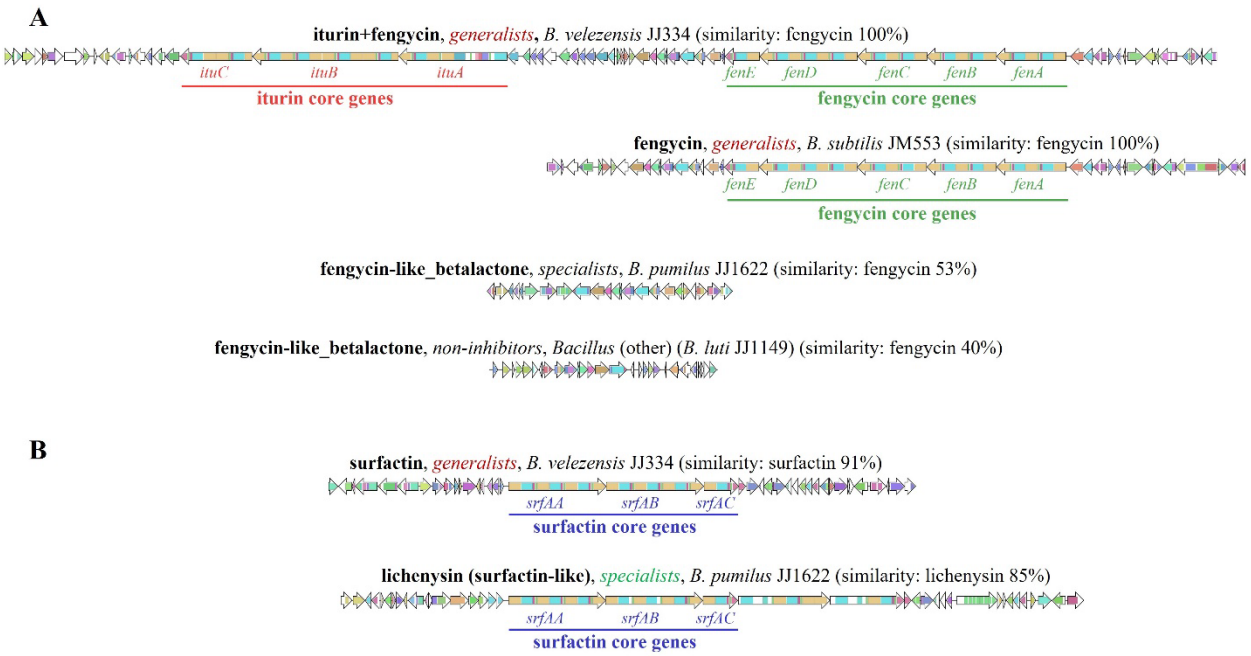


Figure 2.7: Comparison of lipopeptide BGC gene organization. Representative BGCs from four distinct types of clusters identified as having fengycin similarity are shown (A). These four include an iturin and fengycin tandem cluster identified exclusively in *B. velezensis* strains, a fengycin-only cluster identified exclusively in *B. subtilis* strains, and two fengycin-like betalactone BGCs, neither of which contain core genes for the production of a fengycin metabolite. One type is observed in specialists, and the other is observed in non-inhibitors. Representative BGCs from two distinct types of clusters identified as having surfactin (or lichenysin) similarity are shown (B). Both types contain core genes for the production of surfactin or a surfactin-like lipopeptide. The BGC identified exclusively in generalists contains only the core-gene modules for the production of a surfactin-like heptapeptide. The BGC identified exclusively in specialists contains two additional core-gene modules.

2.4.5. Antibiosis generalists produce at least two out of three lipopeptides: iturin, fengycin, and surfactin

Lipopeptides were produced by and extracted from five strains, each representing a *Bacillus* species with strong *P. nicotianae* inhibitory activity: JJ334 (*B. velezensis*), JM553 (*B. subtilis*), JJ1622 (*B. pumilus*), JJ1244 (*B. safensis*), and JJ1138 (*B. altitudinis*). Each metabolite extract was evaluated by UV-vis and LC elution profile (Fig. 2.8). Consistent with the structure

and predicted products of its BGCs, mass spectrometric screening of *B. velezensis* strain JJ334 extracts showed the production of bacillomycin L (an iturin), fengycin, and surfactin. Fengycin and surfactin were identified in extracts of the *B. subtilis* strain, JM553. This also was consistent with the fengycin BGC observed across *B. subtilis* strains which contained core genes for only fengycin production but not an iturin. Interestingly, only surfactin was identified in JJ1622 (*B. pumilus*), JJ1244 (*B. safensis*), and JJ1138 (*B. altitudinis*) extracts. This was consistent with the production of surfactin-like (lichenysin) compound predicted by antiSMASH. Fractionation of JJ334 extracts by HPLC followed by LC-MS analyses of lipopeptide fractions showed that bacillomycin L derivatives eluted between 1.5 and 2.8 min., while fengycin and surfactin derivatives eluted from 8.0 to 10.0 min., and 13.4 to 17.4 min., respectively (Fig. 2.9A). Mass spectrometric analyses identified the presence of six derivatives of bacillomycin L, ten of fengycin, and eight of surfactin in the purified lipopeptide fractions (Fig. 2.9B).

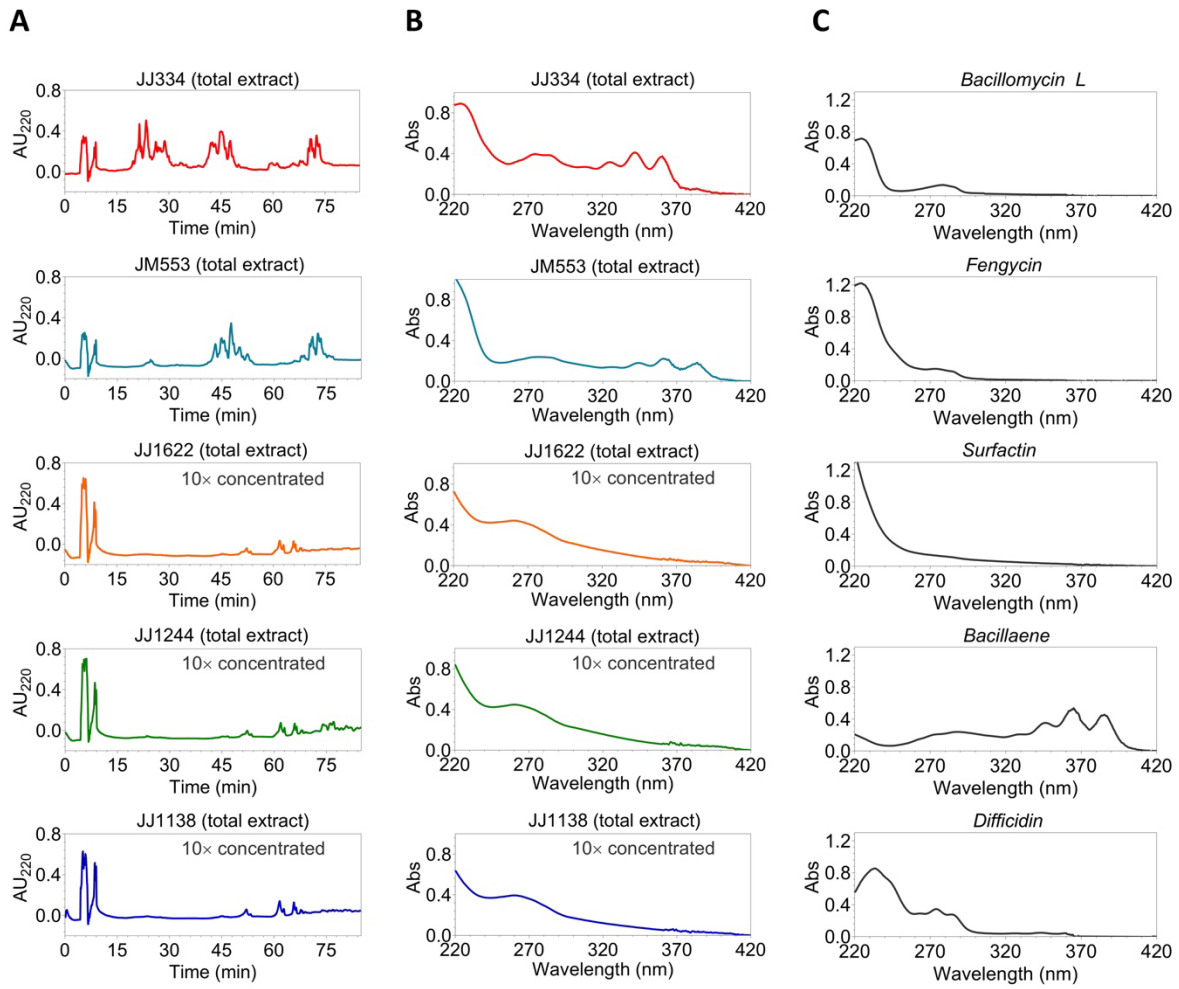


Figure 2.8: Separation and spectral properties of total extracts from representative bioactive *Bacillus* species. Typical LC chromatogram of total extraction of secondary metabolites produced by *B. velezensis* JJ334, *B. subtilis* JM553, *B. pumilus* JJ1622, *B. safensis* JJ1244, and *B. altitudinis* JJ1138 are shown (A). Elution of metabolites was monitored at 220 nm. Secondary metabolites in each total extract were separated by passing through a ZORBAX SB-C18 column (5 µm, 4.6 × 150 mm) using an Agilent 1100 HPLC system. Twenty µL of each extract was injected and eluted for 85 minutes at a flow rate of 0.20 mL/min using the solvent H_2O (A) and CH_3CN (B), each containing 0.1% trifluoracetic acid (TFA) (v/v). Compounds were eluted with 40% of solvent B between 0-8 min., 55% between 15-30 min., 75% between 40-50 min., 100% between 60-75 min., 75% at 77 min., and finally 40% between 79- 85 min. Secondary metabolites were detected at 220. The UV-vis absorption spectra of total extracts from cultures of each representative strain are shown (B). Diode array-captured spectra LC separations of five purified compounds: Bacillomycin L, fengycin, surfactin, bacillaene and difficidin (C).

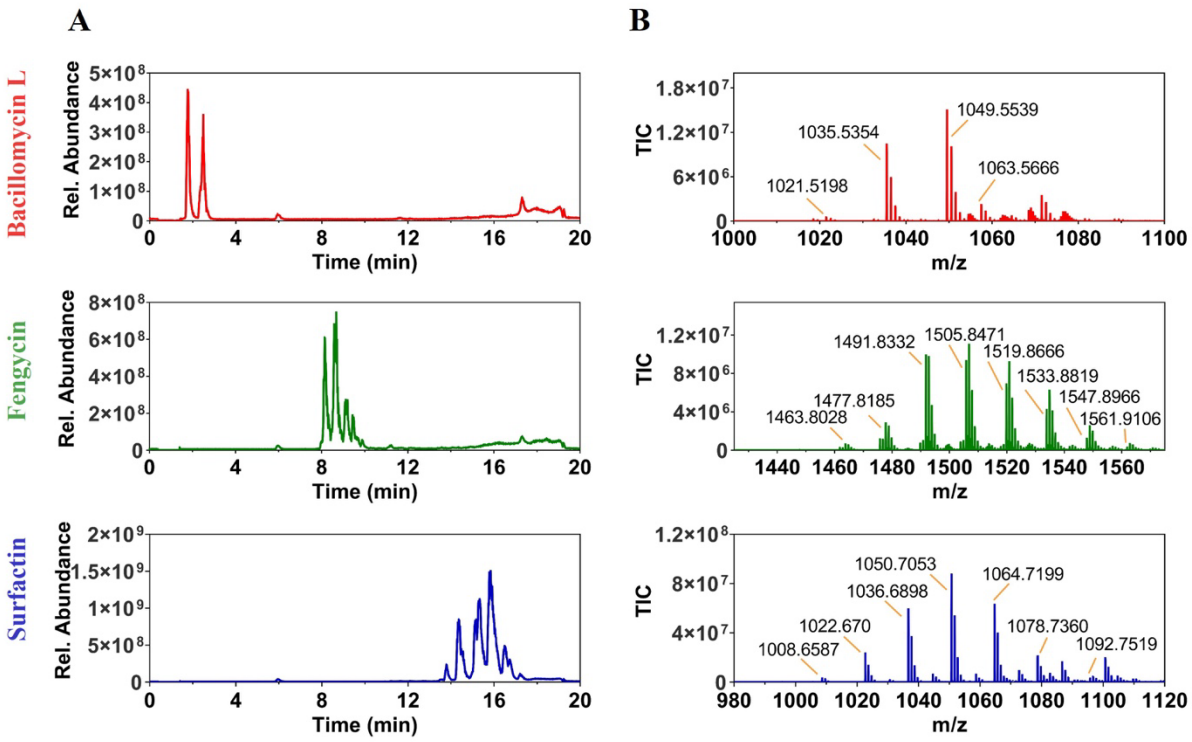


Figure 2.9: Chromatograms (A) and mass spectra (B) for bacillomycin L (an iturin), fengycin, and surfactin. Fractions containing each lipopeptide were obtained from acid-methanolic extracts of *B. velezensis* JJ334. Mass spectra were obtained in positive-ion mode, capturing $[M+H]^+$ and/or $[M+Na]^+$ ions of multiple derivatives of bacillomycin L, fengycin, and surfactin. Each lipopeptide was analyzed using a Thermo Fisher Orbitrap Exploris 120 LC-MS instrument and was separated by passing through an Acquity UPLC BEH C18 (1.7 µm, 2.1 × 50 mm) column for 20 minutes. 0.2 µL of sample was injected and the separation was performed at a flow rate of 0.20 mL/min. The mobile phases were H_2O (A) and CH_3CN (B), each containing 0.1% formic acid (v/v). The LC elution was started with 40% of solvent B, ramped up to 100% at 14 min., and then back to 40% at 16 min.

Extracts corresponding the purified fraction of each lipopeptide were analyzed by LC-MS^n employing positive and negative ionization modes (see *Materials and Methods*). Each lipopeptide compound was confirmed by the fragmented ions generated by MS^n analyses. Bacillomycin L (six) and surfactin (eight) derivatives were identified each varying by the carbon chain length of the

fatty-acid side chains. For both lipopeptides, derivatives were detected in both positive and negative ionization modes as protonated, deprotonated, double-protonated, and/or sodium-adduct forms (Table 2.2).

Table 2.2: Ions of bacillomycin L, fengycin, and surfactin detected by LC-MS^n.

Lipopeptides[1]	**Derivatives**[2,3]	**Molecular Form.**	**$[M+H]^+$**	**$[M-H]^-$**	**$[M+2H]^{2+}$**	**$[M+Na]^+$**	**Error (ppm)**
Bacillomycin L	C13	$C_{45}H_{70}N_{10}O_{16}$	1007.5044	1005.487	504.2562	1029.4858	0.03
	C14	$C_{46}H_{72}N_{10}O_{16}$	1021.5198	1019.504	511.2637	1043.5019	0.69
	C15	$C_{47}H_{74}N_{10}O_{16}$	1035.5354	1033.520	518.2717	1057.5178	0.50
	C16	$C_{48}H_{76}N_{10}O_{16}$	1049.5539	1047.536	525.2793	1071.5327	0.95
	C17	$C_{49}H_{78}N_{10}O_{16}$	1063.5666	1061.550	532.2871	1085.5472	0.53
	C18	$C_{50}H_{80}N_{10}O_{16}$	1077.5820	1075.566	539.295	1099.5629	1.50
Fengycin	C14 A[4]	$C_{70}H_{106}N_{12}O_{20}$	1435.7730	1433.749	718.3893	–	2.17
	C15 A/C16	$C_{71}H_{108}N_{12}O_{20}$	1449.7878	1447.772	725.3971	–	1.69
	C16 A/C17	$C_{72}H_{110}N_{12}O_{20}$	1463.8028	1461.789	732.4049	–	0.93
	C17 A	$C_{73}H_{112}N_{12}O_{20}$	1477.8185	1475.804	739.4134	–	1.30
	C16 [illegible]	$C_{74}H_{114}N_{12}O_{20}$	1491.8332	1489.816	746.4210	–	0.61
	C17 B	$C_{75}H_{116}N_{12}O_{20}$	1505.8471	1503.832	753.4286	–	0.50
	C18 B	$C_{76}H_{118}N_{12}O_{20}$	1519.8666	1517.847	760.4343	–	0.34
	C20 B2	$C_{77}H_{120}N_{12}O_{20}$	1533.8819	1531.863	767.4439	–	0.17
	C21 B2	$C_{78}H_{122}N_{12}O_{20}$	1547.8966	1545.879	774.4518	–	0.30
	Feng. ND[8]	$C_{79}H_{124}N_{12}O_{20}$	1561.9106	1559.894	781.4584	–	1.42
Surfactin	C12	$C_{50}H_{87}N_7O_{13}$	994.6428	992.6281	497.8256	1016.6242	0.12
	C13	$C_{51}H_{89}N_7O_{13}$	1008.6587	1006.642	504.8333	1030.6397	0.33
	C14	$C_{52}H_{91}N_7O_{13}$	1022.6740	1020.657	511.8408	1044.6560	0.61
	C15	$C_{53}H_{93}N_7O_{13}$	1036.6898	1034.673	518.8463	1058.6716	1.53
	C16	$C_{54}H_{95}N_7O_{13}$	1050.7053	1048.688	525.8563	1072.6881	1.12
	C17	$C_{55}H_{97}N_7O_{13}$	1064.7199	1062.705	532.8642	1086.7013	0.79
	C18	$C_{56}H_{99}N_7O_{13}$	1078.7360	1076.721	539.8720	1100.7178	0.31
	C19	$C_{57}H_{101}N_7O_{13}$	1092.7519	1090.737	546.8800	1114.7327	0.40

[1]Each lipopeptide was produced by and extracted from *Bacillus velezensis* strain JJ334.
[2]MS data supporting lipopeptide structure assignments are available in Supplementary Material.
[3]Lipopeptide derivatives; C12 – C21 refers to fatty acid carbon chain length
[4]Fengycin A core decapeptide sequence: Glu-^{D}Orn-^{D}Tyr-^{D}*allo*Thr-Glu-**^{D}Ala**-Pro-Gln-Tyr-**Ile**.
[5]Fengycin A2 core decapeptide sequence: Glu-^{D}Orn-^{D}Tyr-^{D}*allo*Thr-Glu-**^{D}Ala**-Pro-Gln-Tyr-**Val**.
[6]Fengycin B core decapeptide sequence: Glu-^{D}Orn-^{D}Tyr-^{D}*allo*Thr-Glu-**^{D}Val**-Pro-Gln-Tyr-**Ile**.
[7]Fengycin B2 core decapeptide sequence: Glu-^{D}Orn-^{D}Tyr-^{D}*allo*Thr-Glu-**^{D}Val**-Pro-Gln-Tyr-**Val**.
[8]Feng. ND: Fengycin not fully determined; *m/z* values are consistent with C21 B or C22 D, but MS^n data are inconclusive.

Ten fengycin derivatives were identified as $[M+H]^+$, $[M-H]^-$, and $[M+2H]^{2+}$ ions. Two were fengycin A and two were fengycin A2. The amino acid at position six in the core peptide of both types was Ala, whereas the former had Ile at position ten and the latter had Val at the same position. Three were fengycin B where the amino acid at position six in the core peptide was Val. The remaining three derivatives were fengycin B2 where the amino acid at position six and position ten in the core peptide was Val. Here as well, derivatives within fengycin A, A2, B and B2 varied either by fatty acid carbon chain length or amino acid in the core peptide (Table 2.2). To our knowledge, this diversity of production of lipopeptide derivatives from a single *Bacillus* species has not yet been observed.

2.4.6. Inhibitory capacity of bacillomycin L, fengycin, and surfactin against *P. nicotianae* and *F. oxysporum*

Fractions containing bacillomycin L, fengycin, and surfactin were evaluated for their ability to inhibit the growth of *P. nicotianae, F. oxysporum, F. graminearum, and R. solani* by a disk diffusion method. Interestingly, each lipopeptide fraction exhibited strong inhibition against all four target pathogens (Fig. 2.10), confirming strong anti-oomycete/fungal activity. Each lipopeptide was further evaluated for inhibitory effect against *P. nicotianae* and *F. oxysporum* by a microtiter plate assay. The maximum inhibition of *P. nicotianae* growth produced by bacillomycin L and fengycin was observed at a concentration of 50 µg/mL each (Fig. 2.11A), where 56% and 59% growth inhibition were detected, respectively. For surfactin, similar levels of *P. nicotianae* growth inhibition (~40%) were observed at concentrations ranging from 12.5 to 50 mg/mL. Growth inhibition of *F. oxysporum* was also observed for all three lipopeptides (Fig. 2.11B). Bacillomycin L and fengycin showed similar inhibitory potency against this organism,

with significant growth inhibition observed at lipopeptide concentrations 25 mg/mL and producing 85% and 87% growth inhibition at 100 mg/mL, respectively. The maximum inhibitory effect of surfactin (~40%) against *F. oxysporum* was observed at a concentration of 50 µg/mL.

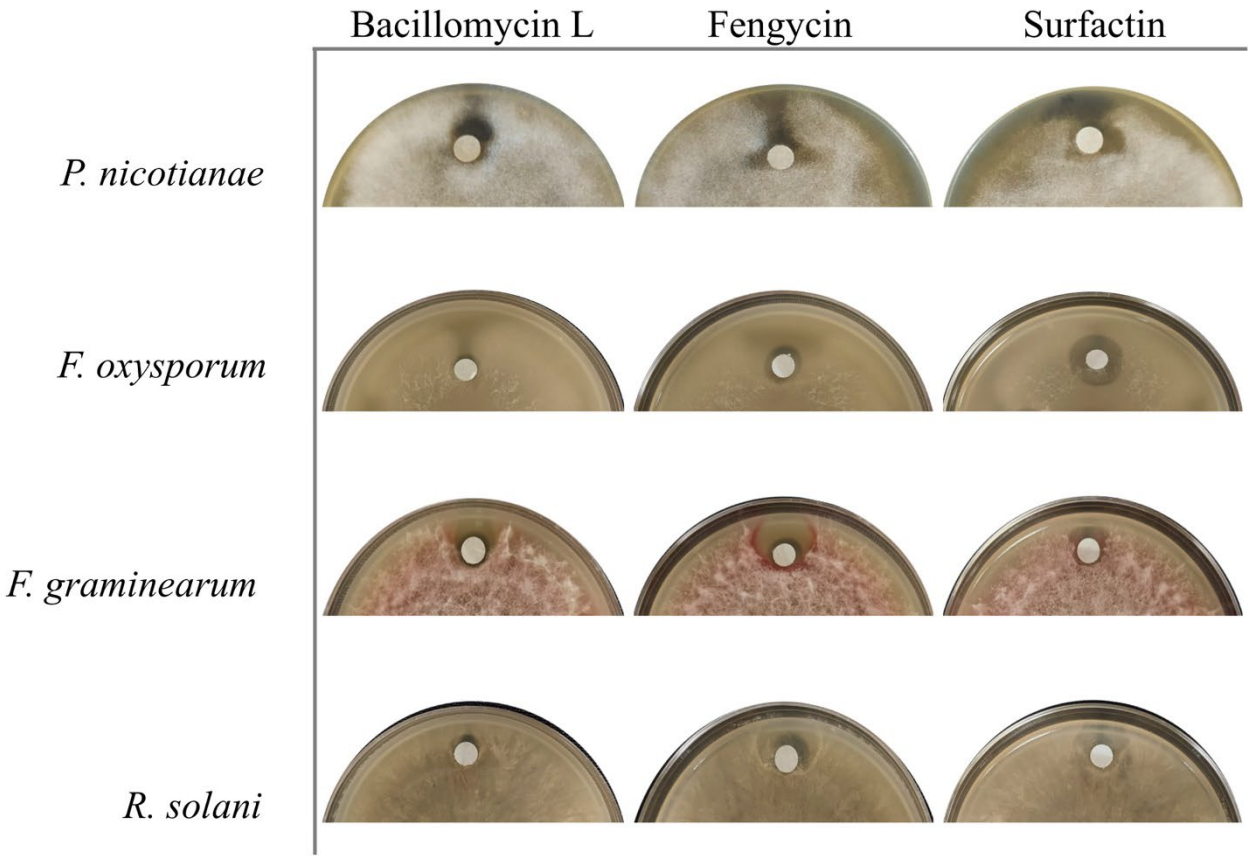


Figure 2.10: Plate-based evaluation of antibiosis exerted by isolated lipopeptides against *P. nicotianae* and fungal pathogens. A disk-diffusion assay was performed to observe antibiosis activity of bacillomycin L, fengycin, and surfactin against *P. nicotianae* and three fungal pathogens: *F. oxysporum*, *F. graminearum*, and *R. solani*.

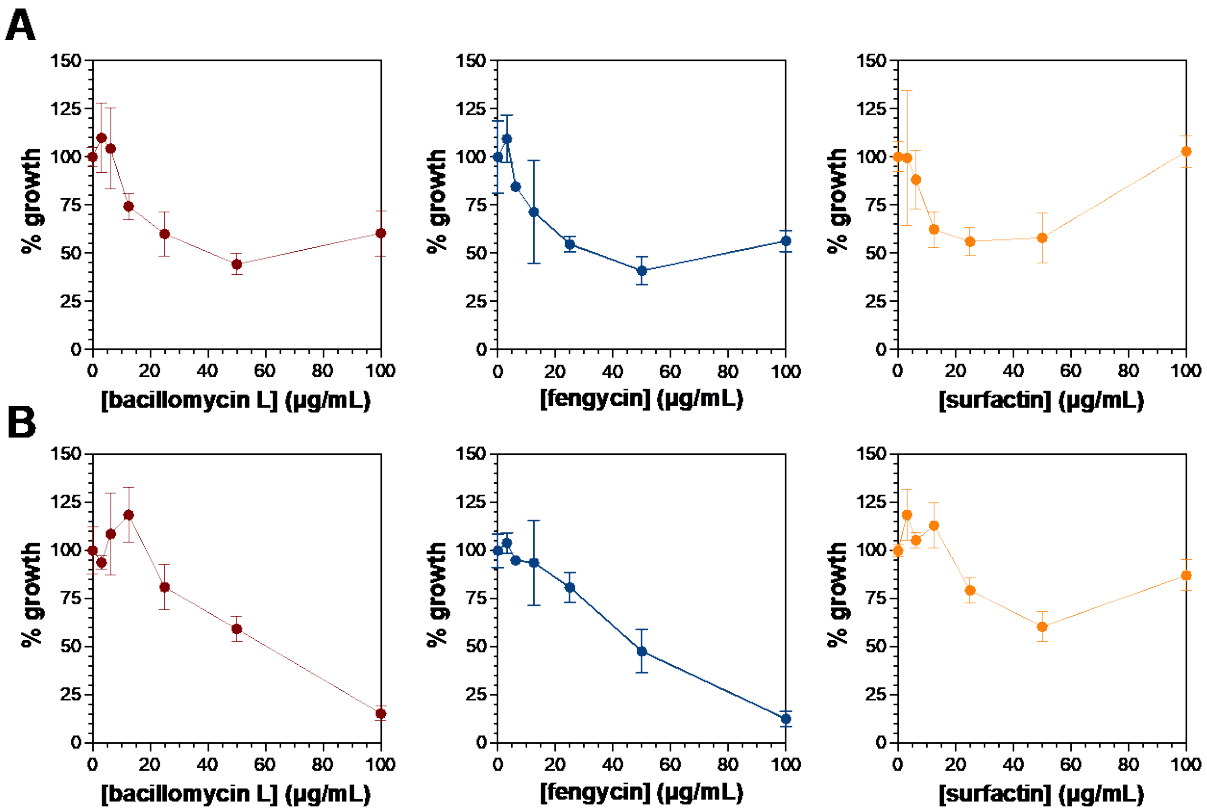


Figure 2.11: Inhibition of *P. nicotianae* (A) and *F. oxysporum* (B) by bacillomycin L, fengycin, and surfactin. Growth inhibition of *P. nicotianae* (A) and *F. oxysporum* (B) was evaluated using a microtiter plate-based assay. The bacillomycin L, fengycin, and surfactin utilized to evaluate inhibition were extracted and isolated from cultures of *B. velezensis* strain JJ334.

2.5. Discussion

Several studies have demonstrated that *Bacillus* species inhibit plant pathogens and enhance plant growth by producing various secondary metabolites (15, 17, 283, 284). However, most studies have been conducted on a single strain or a few strains within a species against one to three pathogenic organisms. To date, the most extensive study in this area analyzed genomic and chemical diversity of 23 strains of *B. subtilis* against three fungal pathogens (284). Studies of this kind naturally limit the scope of investigation on the conservation of antibiosis activity among

species and the breadth of antibiosis against different pathogens. This also limits investigating whether the diversity and variation of secondary metabolites produced by different *Bacillus* species correlate with the observed bioactivity. To address this gap, we evaluated the antibiosis activity of 288 strains across 17 different *Bacillus* species against *P. nicotianae* and compared *Bacillus* antibiosis response with their secondary metabolites to identify conserved factors contributing to antibiosis activity. Strains strongly inhibitory against *P. nicotianae* were evaluated further for their ability to inhibit three fungal pathogens, *F. oxysporum*, *F. graminearum*, and *R. solani*.

Our results demonstrate that *Bacillus* antibiosis against *P. nicotianae* is highly concentrated among five phylogenetically related species: *B. pumilus*, *B. safensis*, *B. altitudinis*, B. *velezensis*, and *B. subtilis*. *Bacillus* antibiosis conservation among strains within each species was expressed as a species-specific bioactivity index (ranging from 0 - 1) and compared against the BGCs carried by these species as determined from genome sequence analyses. This comparison showed a strong correlation between NRPS (especially lipopeptide) BGCs and the bioactivity index, suggesting that lipopeptide BGCs identified in strongly inhibitory *Bacillus* species are conserved factors that may account for a substantial portion of antibiosis activity against eukaryotic pathogens (fungi and oomycetes). In terms of lipopeptide BGC content, there was a clear distinction between *B. velezensis* and *B. subtilis* strains on one hand and *B. pumilus*, *B. altitudinis*, and *B. safensis* strains on the other; on average, strains of the former carried two to three times more lipopeptide BGCs than species from the latter group.

The breadth of antibiosis activity followed the same line of demarcation observed with lipopeptide BGC content. Indeed, further evaluation of strong *P. nicotianae* antagonists from *B.*

velezensis and *B. subtilis* revealed that all but one *B. subtilis* strain exerted strong antibiosis activity against all three fungal pathogens evaluated (*F. graminearum, F. oxysporum, and R. solani*). Even the one *B. subtilis* strain (JM553) that was observed to be a strong inhibitor of *R. solani* and *F. graminearum* and showed weak inhibition of *F. oxysporum*. Conversely, none of the *P. nicotianae*-antagonistic strains of *B. pumilus*, *B. safensis*, or *B. altitudinis* were able to inhibit either *Fusarium* species examined, and only weak inhibition was detected against *R. solani*.

The distribution of lipopeptide BGCs across these species showed distinct patterns that matched up well with the antibiosis phenotypes. The most prolific lipopeptide producing species was *B. velezensis*, the strains of which consistently carry three such BGCs: an iturin, a fengycin, and a surfactin. Interestingly, the iturin and the fengycin core genes occupy the same BGC. Consistent with previous observations (143), our data indicate that bacillomycin L (e.g., *B. velezensis* JJ334) and iturin A (e.g., *B. velezensis* JJ951) are the two most common iturins generated by strains from this species. A broad panel of fengycins (i.e., multiple fatty acid derivatives of fengycin A, A2, B, and B2) are produced from the corresponding *B. velezensis* BGC. By comparison, *B. subtilis* strains reliably carry two lipopeptide BGCs, a fengycin-only cluster and a surfactin cluster. Finally, strains from all three of the specialist *Bacillus* species, *B. pumilus*, *B. altitudinis*, and *B. safensis*, reliably carry, on average, one lipopeptide-generating BGC, and this BGC was predicted to be lichenysin. Though its overall structure is distinct from that of the closely related surfactin BGCs observed in the *B. velezensis* and *B. subtilis* strains, the amino-acid sequence of the core peptide predicted from the modules of the cluster's first three core genes is Glu-Leu-^{D}Leu-Val-Asp-^{D}Leu-Ile. This is consistent with the surfactin-like group of lipopeptide derivatives observed from these species by LC-MS^{n} analyses. Taken together, these results suggest

that production of surfactin alone may be sufficient to exert anti-*P. nicotianae* activity while antifungal activity requires production of at least fengycin in addition to surfactin. Indeed, Khan *et al.* have suggested that strong antifungal activity of *Bacillus* species is due to the synergistic actions of bacillomycin, fengycin, and surfactin (301).

The effectiveness of surfactin as a unilateral agent of antibiosis has varied in the literature, depending largely on the identity of the target organism(s). One study reported that surfactin directly contributed to the biological activity of *B. subtilis* GBL191 against an oomycete, *Plasmopara viticola* (154). As a contrast, Wang *et al.* reported that surfactin alone did not inhibit the mycelial growth of the oomycete *Phytophthora infestans* even at the concentration as high as 25 mg/mL (288). Another study showed that surfactin produced antibacterial activity against a Gram-negative bacteria, *Pseudomonas syringae* with a minimum inhibitory concentration of 25 mg/mL (302). Although the strong anti-*P. nicotianae* activity of *B. pumilus*, *B. safensis*, and *B. altitudinis* may not be fully explained solely by their ability to produce surfactin, we have observed that isolated surfactin is able to inhibit *P. nicotianae.* Although its maximum effect against the organism was observed at 50 mg/mL, it should be noted that very little difference in growth inhibition was observed between 12.5 and 50 mg/mL (see Fig. 7A). They may produce other compounds with antagonistic activity such as betalactone and/or bacillibactin along with surfactin to synergistically contribute to anti-*P. nicotianae* activity. Future studies are anticipated to isolate these compounds and evaluate their antibiosis activity alone or collectively against *P. nicotianae*.

Bacillomycin L (an iturin) and fengycin are known as strongly antagonistic against plant-pathogenic fungi (147, 303, 304). On the other hand, surfactin is known as a broad-spectrum antimicrobial that exhibits bioactivity against fungi, oomycetes, and bacteria (283, 288, 305). We

identified a total of six derivatives of bacillomycin L, ten derivatives of fengycin, and eight derivatives of surfactin from *B. velezensis* JJ334. We observed variation in core peptide structure in the case of fengycin, and we detected lengths of fatty acid side chains varying from 12 to 21 carbons, depending on the lipopeptide in question. Very few *Bacillus* species are known to produce all three lipopeptides (143, 306) and when they are observed identification of only a few derivatives (2 to 5) of each lipopeptide is typical (147, 307, 308). Previous reports have suggested that the bioactivity of lipopeptides may increase along with the length of the fatty acid chain and the number of derivatives, since more derivatives may produce a synergistic effect that contributes to strong bioactivity (17, 305). In general, strong bioactivity was observed for bacillomycin L and fengycin whereas weak activity was observed for surfactin against oomycetes and fungi (143, 156, 283, 288, 309, 310). What stands out in this study is that the *B. velezensis* JJ334 strain produced a total of 24 lipopeptide derivatives with varying lengths of fatty acids, an impressive ability to produce that many lipopeptides from a single *B. velezensis* strain.

In summary, antibiosis activity of 288 PGPR *Bacillus* strains against *P. nicotianae* showed that antibiosis was highly concentrated among five species. Specifically, 56 out of 59 inhibitory strains identified came from *B. velezensis*, *B. subtilis*, *B. pumilus*, *B. safensis*, or *B. altitudinis*. Strains from closely related *B. velezensis* and *B. subtilis* also showed strong antifungal activity against multiple species. Conversely, the *B. pumilus*, *B. safensis*, and *B. altitudinis* strains able to inhibit *P. nicotianae* showed little if any ability to inhibit fungi. Consistent with genomic analysis, 2 to 3 lipopeptides (either fengycin and surfactin or bacillomycin/iturin, fengycin, and surfactin) were identified in the total extracts of *B. subtilis* and *B. velezensis*, respectively. Only surfactin was identified from *B. pumilus*, *B. safensis*, and *B. altitudinis*. All three lipopeptides showed

antibiosis activity against both *P. nicotianae* and *F. oxysporum*, however, the most potent activities were observed for fengycin and bacillomycin L as compared to surfactin. The broad antibiosis activity of *B. velezensis* and *B. subtilis* is likely accounted for by their ability to produce fengycin (*B. subtilis*) and iturin + fengycin (*B. velezensis*) along with surfactin.

Chapter 3: Toward identifying promising *Bacillus* strains as commercially viable biological agents against plant pathogens

3.1. Abstract

Plant-pathogenic fungi and oomycetes pose a serious threat to plant protection and crop production worldwide. Biological agents are rapidly gaining favor over synthetic pesticides as effective and sustainable strategies to control such pathogens. The plant growth-promoting rhizobacteria (PGPR) from the genus *Bacillus* are physiologically well-adapted and well-suited as biopesticides and biofertilizers. In the present study, we evaluated a representative strain from *B. pumilus*, *B. safensis, B. altitudinis*, *B. subtilis,* and *B. velezensis* against root-associated plant-pathogenic fungi: *Fusarium oxysporum*, *F. graminearum*, and *Rhizoctonia solani*. *B. velezensis* and *B. subtilis* exhibited strong antifungal activity against each fungus being evaluated. As a contrast, *B. altitudinis*, *B. pumilus*, and *B. safensis* showed no antifungal activity at all. Comparisons of growth phenotype across the selected strains showed that *B. velezensis* and *B. subtilis* produce a more robust biofilm than the other three species. Interspecies interaction studies showed that *B. velezensis* had a greater capacity to inhibit other *Bacillus* species and also showed greater resistance to inhibition by the four other species. Accordingly, we focused on 18 *B. velezensis* strains to evaluate intraspecies inhibition and resistance as potential markers for the most promising strains for commercial deployment against plant pathogens. The selected *B. velezensis* strains showed a range of antagonistic/inhibitory effects against the other strains as well as a range of capacities to resistant inhibition by other strains. The AB01 and JJ951 strains were the most antagonistic, and AB01, JJ951, AP46, and JJ747 were the most resistant as compared to other strains. The antifungal activity of each strain against *F. oxysporum*, *F. graminearum*, and *R. solani* produced variable antibiosis responses depending on the fungus being evaluated. Strikingly, AP215 produced the greatest level of inhibition across all fungi evaluated, but significant antifungal activity was also observed from AB01, JJ1284, AP52, JM204, AP81, AP202, JM199, and JJ747. 16S rRNA-based phylogenetic analyses revealed that AB01, AP215, JJ951, and JJ1284 were

closely related to one another. Similarly, AP202, JM199, and JJ747 are closely related to each other. A comparison of biosynthetic gene clusters among all 18 strains showed highly similar secondary metabolite profiles. Secondary metabolites were extracted from the culture media of all 18 *B. velezensis* strains using identical conditions and evaluated using UV-vis absorption, HPLC elution profile, and LC-MS identification. A comparison of secondary metabolite production based on the extracted ion chromatogram peak area of six antimicrobial lipopeptides and polyketides showed that AB01 and JJ951 produced the largest quantities of these secondary metabolites relative to other strains. In summary, *B. velezensis* possesses the most potent biocontrol ability among five *Bacillus* species, all of which showed strong antibiosis activity against one or more fungal/oomycete pathogens. Among *B. velezensis* strains evaluated, AB01, JJ951, and AP215 showed the most robust production of antimicrobial metabolites, and accordingly, the broadest and most potent antagonistic capacity.

3.2. Introduction

Crop-destroying fungi are a potential threat to global food security (311). Fueled by climate change and emerging fungi are becoming more detrimental to crop management and food production (4). Plant-pathogenic fungi, including *F. oxysporum*, *F. graminearum*, and *R. solani* cause devastating diseases in economically important plants, including but not limited to, wheat, rice, tomato, and potato (312). These pathogens are hard to control with traditional synthetic fungicides, and the development of fungicide resistance has become a substantial problem (313–315). Due to the challenges in controlling these fungi, the discovery of novel, effective, and sustainable treatments have become increasingly active area of research and development.

Inorganic and organic fungicides have been used to control plant diseases for more than a century (67). Decrease in fungicide sensitivity was being reported as early as the 1960s (316). The development of fungicide resistance is primarily associated with the mode of action of the

fungicidal compound, the reproductive ability of the target fungus, and the frequency of use of the fungicide (317). Several studies have reported an accelerated number of cases of pesticide resistance and resulting difficulties in crop management (4, 32, 318). A recent case study on the mechanism and evolution of fungicide resistance in *Botrytis* has demonstrated rapid and persistent development of resistance against benzimidazoles, quinone outside inhibitors (QoIs), dicarboximides, anilinopyrimidines, and succinate dehydrogenase inhibitors (SDHIs) (319). To maintain adequate protection against *Botrytis,* the implementation of strict resistance management measures and alternative strategies with non-chemical products were highly recommended. Further, *F. graminearum* has shown resistance against azole fungicides, whereas *F. oxysporum* has shown resistance against methyl benzimidazole carbamate (320, 315). Also, *R. solani* has shown resistance against the succinate dehydrogenase inhibitor thifluzamide (321). Due to the prevalence of fungicide resistance, biological agents have become increasingly attractive as potential alternatives to chemical pesticides for controlling fungal pathogens (322–324).

Plant growth-promoting rhizobacteria (PGPR) from the genus *Bacillus* have the ability to increase plant growth and suppress plant disease via several direct and indirect mechanisms (79). Direct plant growth is primarily facilitated by the biosynthesis of various compounds and growth factors such as nutrients and hormones that directly participate in nitrogen fixation, phosphate solubilization, iron mobilization (by siderophores), and provision of various hormones (78). On the other hand, indirect plant growth promotion generally occurs by suppressing plant pathogens through the production of antibiotics and lytic enzymes, competition for nutrients, or by triggering induced systematic resistance (ISR) against plant pathogens (85, 86, 325). The most commonly observed *Bacillus* secondary metabolites that are antagonistic to plant pathogens are non-

ribosomal and ribosomal peptides, siderophores, polyketides, and volatile organic compounds (17). Among these, the most extensively studied non-ribosomal peptides with antifungal activity are lipopeptides such as iturin, fengycin, and surfactin (326). Previous studies showed broad to narrow-spectrum antibiotic activity of each lipopeptide across various phytopathogens, including bacteria, fungi, and oomycetes (288, 327). For the past three decades, *Bacillus* species and derived antimicrobial lipopeptides have been of great interest to develop commercially viable biological agents against plant pathogens.

Studies have shown that PGPR *Bacillus* strains can increase crop yields by as high as 57%, depending on the crop types (328). In addition, some PGPR *Bacillus* can produce heat and desiccation-tolerant endospores and have high cell viability and prolonged shelf life (329). Despite the many advantages of some *Bacillus* species for development as commercially deployable biological agents, only a few *Bacillus* strains have been reported as biological agents for commercial use. This is due to their inefficacy in plant growth and disease suppression under ambient field conditions (330). To better understand their efficacy in laboratory conditions, several comprehensive frameworks have been developed to evaluate them for commercially viable biological agents (78, 331). Although their efficacy in laboratory vs field conditions is less understood, evaluation of their antibiosis activity, antimicrobial production, and growth phenotype is the first step in that direction (332).

In the present study, we evaluated antibiosis activity of five *Bacillus* species against plant-pathogenic fungi, *F. oxysporum*, *F. graminearum*, and *R. solani*, followed by their interspecies interaction and biofilm formation. These experiments showed *B. velezensis* produces the strongest antifungal activity as well as interspecies inhibition and resistance activity. The *B. velezensis*

strains also produced more robust biofilms and they reliably generated a broader range and larger quantities of antimicrobial secondary metabolites than the remaining four species. This result prompted additional investigations to evaluate the genomic, antibiotic, and chemical properties of 18 *B. velezensis* strains for identifying promising strains for commercially viable biological agents. All 18 strains were also evaluated for antifungal activity, intraspecies inhibition and resistance activity, and ability to produce biofilm and antimicrobial secondary metabolites.

3.3. Materials and Methods

3.3.1. *Bacillus* species, strains, and culture conditions

Representative strains from five *Bacillus* species that previously exhibited strong anti-*Phytophthora nicotianae* activity were evaluated for antifungal activity against three root-associated fungal pathogens. The *Bacillus* species were *B. velezensis* (JJ334), *B. subtilis* (JM553), *B. pumilus* (JJ1622), *B. safensis* (JJ1244), and *B. altitudinis* (JJ1138), and the fungal pathogens were *F. oxysporum*, *F. graminearum*, and *R. solani*. Each strain was further evaluated for interspecies inhibition and resistant activity, ability for producing robust biofilm and antimicrobial secondary metabolites. In addition, 18 strains from the *B. velezensis* species were selected to identify promising strains for commercially viable biological agents. All 18 strains were evaluated for antifungal activity, intraspecies inhibition and resistant activity, and ability to produce biofilm and antimicrobial secondary metabolites. In order to compare their ability to produce secondary metabolites and overall genomic properties, the genomic DNA of all 18 *B. velezensis* strains was sequenced at Nanopore for a single-contig complete genome sequence. The taxonomy of each strain was confirmed by the best hit of known species based on the average nucleotide identity (ANI) of the whole-genome sequence at the MiGA (Microbial Genomes Atlas) webserver (291).

For routine bacterial growth, *Bacillus* strains were cultured in tryptic soy broth (TSB) at 37 °C at constant agitation (170 rpm) (supplemented with 1.5% agar, if required).

3.3.2. Antibiosis assay of *Bacillus* species against fungal pathogens

Each *Bacillus* strain was evaluated for its ability to inhibit the growth of three fungal pathogens: *F. oxysporum*, *F. graminearum*, and *R. solani*. For antifungal assays, each target pathogen was grown in a V8-agar medium (V8 juice 82 ml/L, $CaCO_3$ 0.5 g/L, and agar 15 g/L) at room temperature. Each *Bacillus* strain was grown in TSB (tryptic soy broth) at 37 °C for 24 hours, and the cell density was normalized to OD600=1 based on the measurement of optical density at 600 nm. Five µL of each *Bacillus* culture was transferred onto the edge of the assay plate, and the 72-hours-grown target pathogen was transferred as a plug at the center of the assay plate. The growth inhibition of the target pathogen due to the presence of *Bacillus* colonies was recorded after 7-10 days. The antagonistic responses of *Bacillus* strains against the target pathogen were classified as strong, weak, and no inhibition based on the measurement of the zone of inhibition (ZOI). The strong inhibition was assigned for displaying a clear ZOI with complete elimination of hyphal growth of target pathogens around *Bacillus* colonies. No inhibition was assigned when the target pathogens were over-grown on *Bacillus* colonies with a display of no ZOI. Exhibition of less clear ZOI with partial elimination of hyphal growth of target pathogens around *Bacillus* colonies was assigned as weak inhibition.

All 18 *B. velezensis* strains were evaluated against three fungal pathogens as described above. In order to generate a quantitative antibiosis score, the antibiosis assay of each strain was replicated at least three times. The extent of inhibition, the interactions between *Bacillus* colonies and target pathogens, and their growth morphologies were recorded and taken into consideration

for assigning the inhibition score. The antagonistic response of each *Bacillus* strain against the target pathogen was scored from 0 to 4 based on the measurement of ZOI. Further, the percent inhibition of each strain was calculated based on estimated antibiosis score. Finally, a cumulative percent antibiosis score of each strain was calculated by accounting for the average percent inhibition against all fungi.

3.3.3. *Bacillus* biofilm formation

In order to evaluate *Bacillus* growth morphology and biofilm formation, each *Bacillus* strain was initially grown for 24 hours in TSB medium with constant agitation (170 rpm) at 37 °C. Subsequently, *Bacillus* cell density was normalized to OD600=1. In order to produce biofilm, 5 µL of cell culture was transferred onto a plate containing glucose 20 g/L, yeast 1 g/L, L-glutamic acid 5 g/L, KCl 0.5 g/L, $MgSO_4$ 0.5 g/L, KH_2PO_4 1g/L, L-phenylalanine 3 mg/L, $MnSO_4$ 5 mg/L, $FeSO_4$ 0.15 mg/L, $CuSO_4$ 0.16 mg/L, pH 7.0, and agar 15 g/L. Finally, plates were incubated for 62-70 hours at 30 ° to produce biofilm.

3.3.4. *Bacillus* interspecies and intraspecies interaction assay

In order to identify *Bacillus* species with strong antibacterial and resistant activity against other *Bacillus* species, interspecies interaction assay was performed for each species against all other species, including itself. For the assay, each *Bacillus* strain was grown for 24 hours in TSB medium, and the cell density was adjusted to OD_{600}=1.0. The assay plate was prepared using the agar-overlay technique as follows: 10 ml of TSB medium containing 1.5% agar were plated and used as the bottom layer, whereas 5 ml of TSB medium containing 0.5% agar pre-inoculated with the target strain in 1:200 dilutions were plated on top of the bottom later. The assay plate was dried at room temperature for 30 minutes. 7.5 µL of focal strain was spotted onto the double-layered

assay plate and incubated at 37 °C for 36-40 hours. The interspecies interaction/inhibition was evaluated by measuring the ZOI at the intersection of focal colonies and bacterial lawn of the target strain.

To evaluate the intraspecies interaction of 18 *B. velezensis* strains, each strain was assayed and evaluated against all 18 strains using the method described above. The intraspecies interaction experiment was replicated at least three times in order to generate a quantitative inhibition score. The intraspecies inhibition for each focal strain against a target strain was further classified into five types (score 0 to 4) based on the measured ZOI and morphological changes of focal/target strains: no inhibition (0), very weak inhibition (1), weak inhibition (2), strong inhibition (3), and very strong inhibition (4). Subsequently, cumulative % inhibition for each focal strain was calculated using the inhibition score of each focal strain against all target strains. Conversely, the cumulative % resistance activity for each target strain was calculated from the inverted percent inhibition derived from the percent inhibition of each focal strain as follows:

$$\text{Cumulative \% inhibition of a focal strain} = \frac{\Sigma \text{ percent inhibition against each target strain}}{18}$$

$$\text{Cumulative \% resistance of a target strain} = \frac{\Sigma\,(100 - \text{percent inhibition of each focal strain})}{18}$$

3.3.5. Genomic and bioinformatics analyses of *Bacillus* strains

Genome sequence of each *Bacillus* strain was analyzed by antiSMASH v.5 (189) to predict biosynthetic gene clusters (BGCs) and associated secondary metabolites. Predicted BGCs were further grouped into clusters of BGCs based on their gene sequence similarities using BiG-SCAPE v.0.0.0r. (293). The network distances generated by BiG-SCAPE analysis were further visualized

and annotated using Cytoscape 2.8 (294). In order to compare phylogenetic relationships, the 16S rRNA sequence from each strain was predicted using barrnap (333) from the whole-genome sequence. The predicted 16S rRNA sequences were used to construct a phylogenetic tree using MEGA X (334).

3.3.6. Extraction of secondary metabolites from *Bacillus* strains

In order to produce the secondary metabolites, the *Bacillus* strains were cultured in the Landy medium for 72 hours with constant agitation (170 rpm) at 30 °C. Following the centrifugation of cell culture at 6,000 × g for 50 minutes, the supernatant was collected, and its pH was adjusted to 2.0 by dropwise addition of concentrated HCl with constant stirring. Subsequently, acidic supernatant was incubated overnight at 4 °C, and secondary metabolite precipitates were isolated by centrifugation at 6,000 × g for 50 minutes. The precipitate was collected and redissolved in 100% MeOH. The MeOH was evaporated under the flow of dry nitrogen and the dried residue was redissolved in MeOH by vortexing for 2 minutes. Finally, secondary metabolites were extracted twice using 100% MeOH, filtered through a 0.2 µm Acrodisc syringe filter, and stored at -20 °C until use.

3.3.7. Analysis of secondary metabolites using UV-vis absorption, HPLC, and LC-MS

Characteristic absorption feature of secondary metabolites produced by each *Bacillus* strain was recorded using a UV-vis absorption (220-600 nm) spectrophotometer. Secondary metabolites in each total extract were separated by passing through a ZORBAX SB-C18 column (5 µm, 4.6 × 150 mm) using an Agilent 1100 HPLC system. Twenty µL of each extract was injected and eluted for 85 minutes at a flow rate of 0.20 mL/min using the solvent H_2O (A) and CH_3CN (B), each containing 0.1% trifluoracetic acid (TFA) (v/v). Compounds were eluted with 40% of solvent B

between 0-8 min., 55% between 15-30 min., 75% between 40-50 min., 100% between 60-75 min., 75% at 77 min., and finally 40% between 79- 85 min. Secondary metabolites were detected at 220, 275, 375, and 450 nm by a diode array detector coupled with a full-spectrum (220 - 500 nm) analysis.

In order to compare secondary metabolite production, extracts of all 18 *B. velezensis* strains were further analyzed using a Thermo Fisher Orbitrap Exploris 120 LC-MS instrument. Secondary metabolites were separated by passing through an Acquity UPLC BEH C18 (1.7 µm, 2.1 × 50 mm) column for 20 minutes. 0.2 µL of sample was injected and the separation was performed at a flow rate of 0.20 mL/min. The mobile phases were H_2O (A) and CH_3CN (B), each containing 0.1% formic acid (v/v). The LC elution was started with 40% of solvent B, ramped up to 100% at 14 min., and then back to 40% at 16 min. In order to identify secondary metabolites, ions were generated using an electrospray ionization source employing both positive and negative ionization modes. For unambiguous identification of secondary metabolites, each precursor ion was simultaneously fragmented using a multi-stage MS^n analyzer. The extent of antimicrobial secondary metabolite production from all 18 *B. velezensis* strains was calculated from the peak area in extracted ion chromatogram for three lipopeptides and three polyketide compounds. The lipopeptides were iturin, fengycin, and surfactin and the polyketides were bacillaene, difficidin, and macrolactin W.

3.4. Results

3.4.1. *B. velezensis* and *B. subtilis* are strong inhibitors of fungi

Our previous evaluation of the P. nicotianae antibiotics activity of 288 strains across 17 *Bacillus* species revealed that antibiosis is highly conserved among five species: *B. pumilus*, *B. safensis, B. altitudinis, B. subtilis,* and *B. velezensis.* This finding has prompted further investigation into their antifungal activity, interspecies inhibition and resistance activity, and production of robust biofilm and antimicrobial secondary metabolites. Representative strain from *B. pumilus* (JJ1622), *B. safensis* (JJ1244), *B. altitudinis* (JJJ1138), *B. subtilis* (JM553), and *B. velezensis* (JJ334) were evaluated for antibiosis activity against three fungal pathogens, *F. oxysporum*, *F. graminearum*, and *R. solani*. Antagonism between each strain and target pathogen was grouped into three classes based on the measured zone of inhibition: strong, weak, and no inhibition (see *Materials and Methods*). Although each strain exhibited strong antibiosis activity against *P. nicotianae*, only *B. velezensis* and *B subtilis* exhibited strong antifungal activity against all fungi, whereas strains from the remaining three species were non-inhibitors (Fig. 3.1). Interestingly, *B. velezensis* (JJ334) produced larger zones of inhibition (ZOI) than *B. subtilis* (JM553) against each fungus being evaluated, indicating *B. velezensis* is more effective inhibitor of fungi than *B. subtilis*.

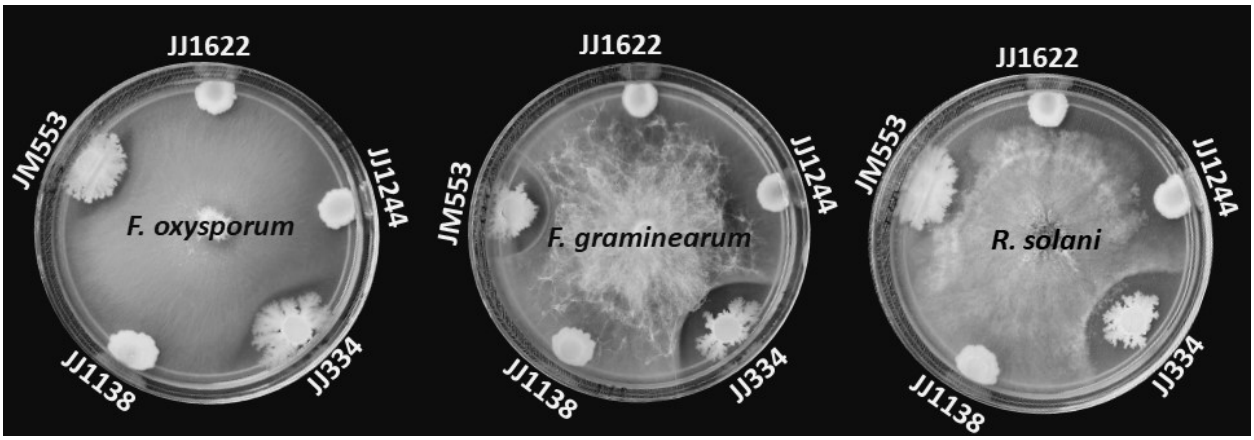


Figure 3.1: Antifungal activity of five *Bacillus* species. Antibiosis responses of *B. pumilus* (JJ1622), *B. safensis* (JJ1244), *B. altitudinis* (JJ1138), *B. subtilis* (JM553), and *B. velezensis* (JJ334) against *F. oxysporum*, *F. graminearum*, and *R. solani*. *Bacillus* colonies are grown at the edge of assay plates, whereas the target fungus is allowed to grow from the center to the edges of the plates to encounter the *Bacillus* colonies.

3.4.2. Robust biofilm production by *Bacillus velezensis* and *Bacillus subtilis*

Biofilm formation in *Bacillus* species enhances the production of secondary metabolites, enables rapid adaptation, increase long-term viability, and improves plant root colonization and plant protection (127, 130). Therefore, robust biofilm formation is a desirable phenotype for considering *Bacillus* species as a commercial biological agent against plant pathogens. Both *B. velezensis* JJ334 and *B. subtilis* JM553 produced a rigid-textured multilayer structure, indicating an excellent ability for biofilm formation (Fig. 3.2). However, *B. pumilus* JJ1622, *B. safensis* JJ1244, and *B. altitudinis* JJ1138 strains failed to form a highly complex biofilm. Instead, they produced smaller, stagnated colonies with limited swarming motility.

Figure 3.2: *Bacillus* biofilm formation. Growth morphology and biofilm formation observed for *B. velezensis* JJ334, *B. subtilis* JM553, *B. pumilus* JJ1622, *B. safensis* JJ1244, and *B. altitudinis* JJ1138.

It has been shown that the surfactin and fengycin contribute to biofilm formation in *B. subtilis* (335, 336). Li *et al*. reported that the polyketide bacillaene enhance biofilm production in *B. methylotrophicus* (337). Our observation suggests that the high-quality biofilm produced by *B. velezensis* and *B. subtilis* may be due to their ability to produce biofilm-forming secondary metabolites such as lipopeptides and polyketides. The lack of complex biofilm in the remaining three species may be due to the absence of those secondary metabolites.

3.4.3. *B. velezensis* is a more capable interspecies competitor than other *Bacillus* species

In order to compare interspecies inhibition and resistance across the five *Bacillus* species, each representative strain was challenged by all of the representative strains, including itself. None of the strains exhibited any inhibition against themselves, indicating that each has sufficient

intrinsic self-resistance mechanisms (Fig. 3. 3A). However, *B. velezensis* JJ334 exhibited strong inhibition against *B. pumilus* JJ1622, *B. safensis* JJ1244, and *B. altitudinis* JJ1138 but no inhibition to *B. subtilis* JM553. In contrast, JM553, JJ1622, and JJ1138 were non-inhibitory against all other strains, while JJ1244 exhibited weak inhibition against JJ334, JJ1622, and 1138.

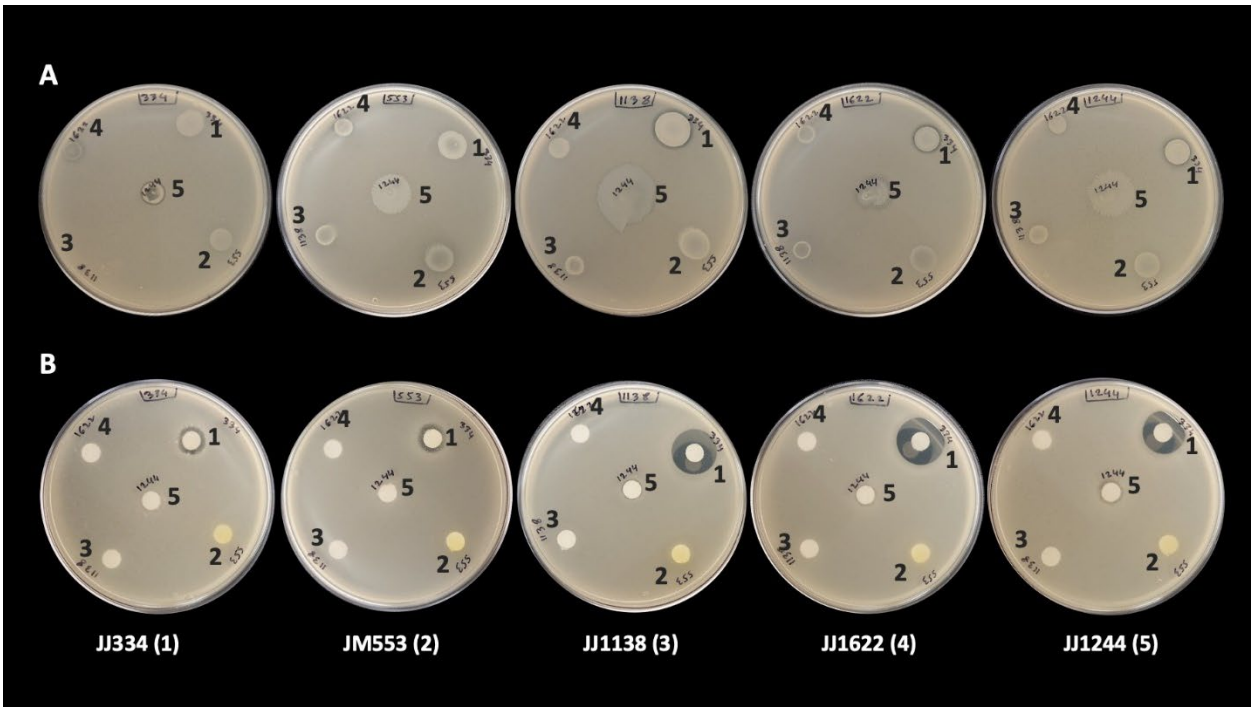


Figure 3.3: *Bacillus* interspecies interactions observed for intact cells (A) and total extracts (B). Each *Bacillus* species representative strain was challenged by all the other representative strains, including itself. Each plate was layered with the target species indicated in white under each image. Each plate was then inoculated with intact cells from each representative as the focal strain (A) or treated with filter disks containing total extract from each representative strain (B) according to the following pattern: *B. velezensis* JJ334 (1), *B. subtilis* JM553 (2), *B. altitudinis* JJ1138 (3), *B. pumilus* JJ1622 (4), and *B. safensis* JJ1244 (5).

Extracted secondary metabolites from the culture of each *Bacillus* species were further evaluated against all *Bacillus* species (Fig. 3.3B). The extract from JJ334 showed strong inhibition against JJ1622, JJ1244, and JJ1138 and weak inhibition against JM553 and itself. Importantly, the extracts from the remaining four species exhibited no inhibition against any other *Bacillus* species.

The observed strong interspecies inhibition ability of the *B. velezensis* JJ334 strain may be due to its ability to produce antibacterial secondary metabolites. Based on this result, it is clear that *B. velezensis* has the highest potential for antibacterial activity as compared to the other four species, and it appears well-equipped to compete with other organisms, even those closely related to itself.

B. velezensis **carried a larger and more diverse set of BGCs**

Bacillus polyketides and ribosomal peptides are commonly known as antibacterial, whereas the non-ribosomal peptides are known as antifungal/anti-oomycete secondary metabolites (147, 156, 167, 169, 171, 304). BGC analysis (antiSMASH v.5) of five *Bacillus* species identified on average 11 BGCs per species. Among these five species, *B. velezensis* carried the highest number of BGCs (13). *B. pumilus* and *B. safensis* carried 11 BGCs, and both *B. altitudinis* and *B. subtilis* carried 10 BGCs. Importantly, *B. velezensis* contained the highest number of antimicrobial secondary metabolite-producing BGCs, including three polyketides (bacillaene, difficidin, and macrolactin W), three non-ribosomal lipopeptides (bacillomycin L, fengycin, and surfactin) and a ribosomal peptide (plantazolicin) (Fig. 3.4). In contrast, *B. subtilis* contained only one polyketide (bacillaene) and two lipopeptides (fengycin and surfactin) BGCs. The remaining three species lacked any polyketide BGC but carried surfactin-like and fengycin-like betalactone BGCs. It is important to note that a bacillibactin (siderophore) BGC was identified in all five species. All representative strains, except *B. altitudinis* JJ1138, contained a bacilysin BGC. *B. pumilus*, *B. safensis*, and *B. altitudinis* all carried highly similar carotenoid BGCs, whereas *B. altitudinis*, *B. pumilus*, and *B. subtilis* carried a locillomycin, zwittermicin A, and subtilosin A, respectively.

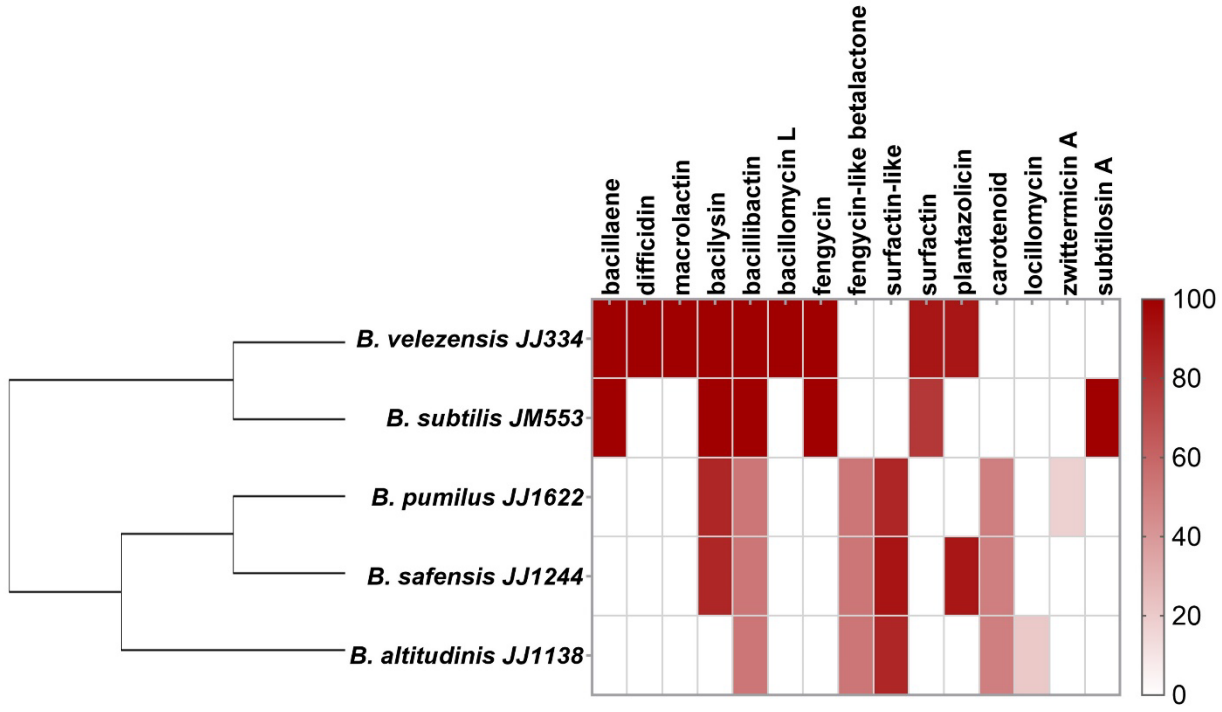


Figure 3.4: Diversity of predicted secondary metabolites among five *Bacillus* species. The 16S rRNA-based phylogenetic relationship of five species is shown on the left. The color of each box in the heatmap represents the percent similarity (0 to 100) of the BGC identified to its archetypal representative as uncovered through antiSMASH analyses. The heatmap scale is presented on the right.

Interestingly, 16S rRNA-based phylogenetic analyses showed that *B. velezensis* and *B. subtilis* formed one clade while *B. pumilus*, *B. safensis*, and *B. altitudinis* grouped into another (Fig. 3.4). This is consistent with the BGC types identified among these five species, where the former two species had five nearly identical BGCs. Similarly, *B. pumilus, B. safensis,* and *B. altitudinis* had four identical BGCs and three of these were completely distinct from those identified in *B. subtilis* and *B. velezensis*. The distinct BGC distribution among the two species groups (*B. velezensis* and *B. subtilis* vs *B. pumilus, B. safensis,* and *B. altitudinis*) is consistent with the observed antibiosis response. These data suggest that the expression of phylogenetically-

related common BGCs identified in *B. velezensis* and *B. subtilis* may be responsible for the observed broad-spectrum antibiosis response.

Accordingly, the observed antibacterial and antifungal activity of all five *Bacillus* species can be explained by their ability to produce various secondary metabolites. The strong antifungal activity of *B. velezensis* and *B. subtilis* may be due to their ability to produce two or three cyclic lipopeptides (fengycin, surfactin, and/or bacillomycin L). Similarly, the strong interspecies antibiosis activity of *B. velezensis* may be due to its ability to produce the three antibacterial polyketides (bacillaene, difficidin, and macrolacin W) and one ribosomal peptide (plantazolicin). Note that all three lipopeptides are known as antifungal compounds, and three polyketides and the ribosomal peptide have been reported as antibacterial compounds (156, 169, 171, 303, 338). The relatively strong antifungal activity of *B. velezensis* as compared to the *B. subtilis* may be due to its ability to produce three lipopeptides (fengycin, surfactin, and bacillomycin L). The lack of antibacterial activity expressed by *B. velezensis* even though it carries a bacillaene BGC suggests that either bacillaene alone cannot exert strong interspecies inhibition, or *B. subtilis* does not produce bacillaene in quantities sufficient to produce such an effect. Finally, the weak antibacterial activity of *B. safensis* may be due to its ability to produce the narrow-spectrum antibacterial plantazolicin.

3.4.4. *Bacillus velezensis* produced the highest number of antimicrobial secondary metabolites

Secondary metabolites were produced and extracted from the culture media of all five representative *Bacillus* species in order to compare the extent and breadth of secondary metabolite production. Metabolites were identified by high-resolution LC-MS analysis followed by

evaluation of fragment ions simultaneously generated by an MS^n analyzer. The LC profile as observed by monitoring the total ion current (TIC) detected in positive and negative ionization mode showed a highly similar elution profile for the extracts of *B. velezensis* and *B. subtilis*, except that corresponding to *B. velezensis* had an additional cluster of peaks which eluted between 2 to 6 mins (Fig. 3.5). Both MS and MS^2 analyses confirmed that *B. velezensis* JJ334 had six bacillomycin L (an iturin) derivatives which eluted between 2 to 6 mins. Compounds from the extracts of *B. velezensis* JJ334 and *B. subtilis* JM553 that eluted between 7 and 11 mins were fengycin derivatives that varied either in the fatty acid side chain or in amino acids in the core peptide. Surfactin derivatives with varied fatty acid chains were identified in the extract of all five *Bacillus* species; these were observed to elute between 14 to 20 mins. Among the five species, the highest number of antimicrobial compounds were identified in *B. velezensis* extracts. These included three polyketides (bacillaene, difficidin, macrolactin W), a ribosomal peptide (plantazolicin), and multiple derivatives of three lipopeptides (iturin, fengycin, and surfactin). Taken together, the largest number of known antimicrobial compounds were identified in *B. velezensis* extracts. As shown in the total-ion-current elution profiles for JJ334 culture extracts (Fig. 3.6), three polyketides or polyketide/NRPS hybrids (bacillaene, difficidin, macrolactin W), a ribosomal peptide (plantazolicin), and multiple derivatives of three lipopeptides (iturin, fengycin, and surfactin) were identified.

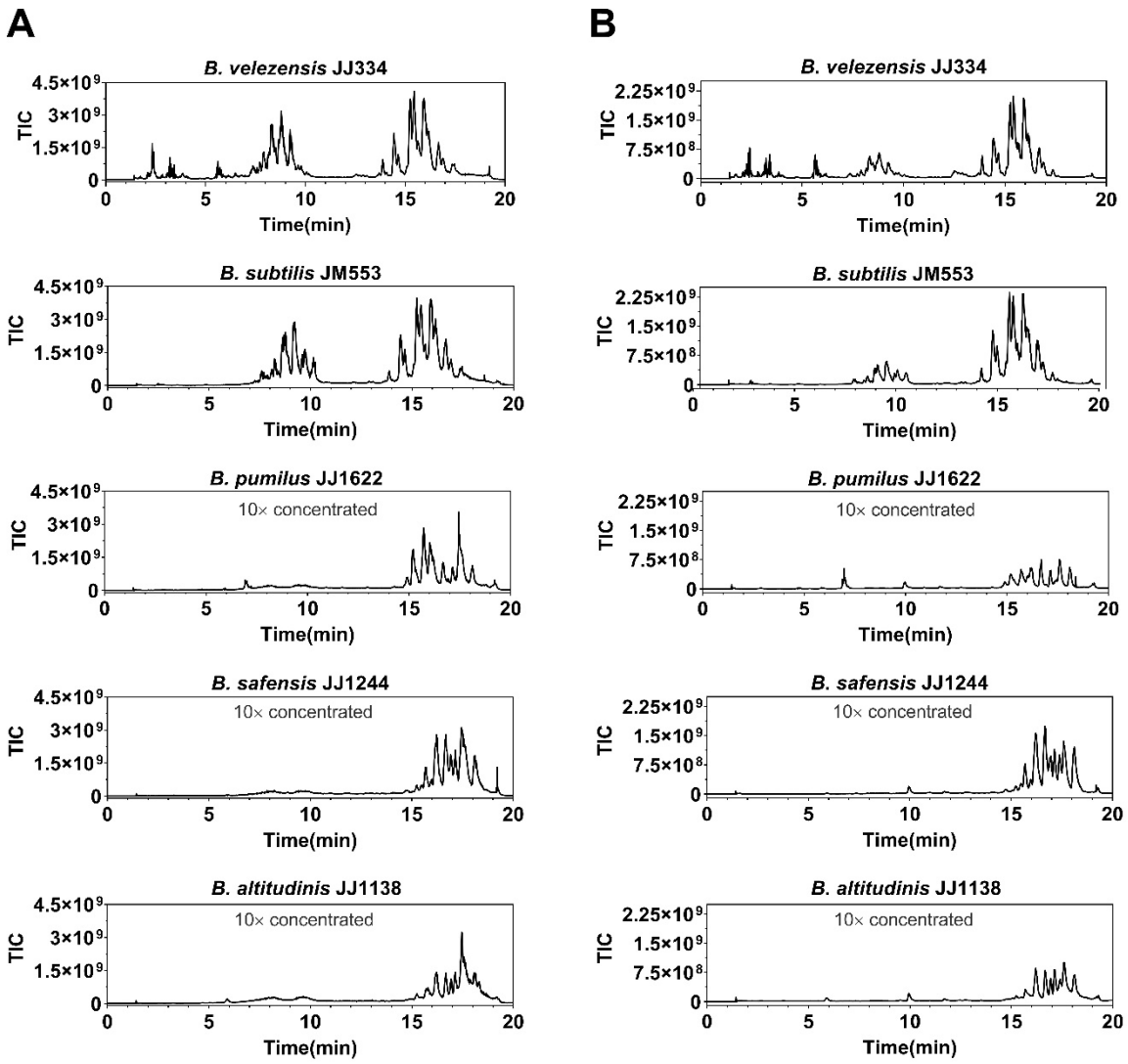


Figure 3.5: LC elution profile of *Bacillus* secondary metabolites generated by (A) positive ion (ESI+) mode and (B) negative ion (ESI-) mode. LC elution profile of secondary metabolite extracts from the culture of five *Bacillus* species showing total ion current (TIC). Note that chromatograms of JJ1622, JJ1244, and JJ1138 are obtained from the extracts that are 10× concentrated than JJ334 and JM553 extracts. Secondary metabolite extracts were analyzed using a Thermo Fisher Orbitrap Exploris 120 LC-MS instrument and were separated by passing through an Acquity UPLC BEH C18 (1.7 µm, 2.1 × 50 mm) column for 20 minutes. 0.2 µL of sample was injected and the separation was performed at a flow rate of 0.20 mL/min. The mobile phases were H_2O (A) and CH_3CN (B), each containing 0.1% formic acid (v/v). The LC elution was started with 40% of solvent B, ramped up to 100% at 14 min., and then back to 40% at 16 min.

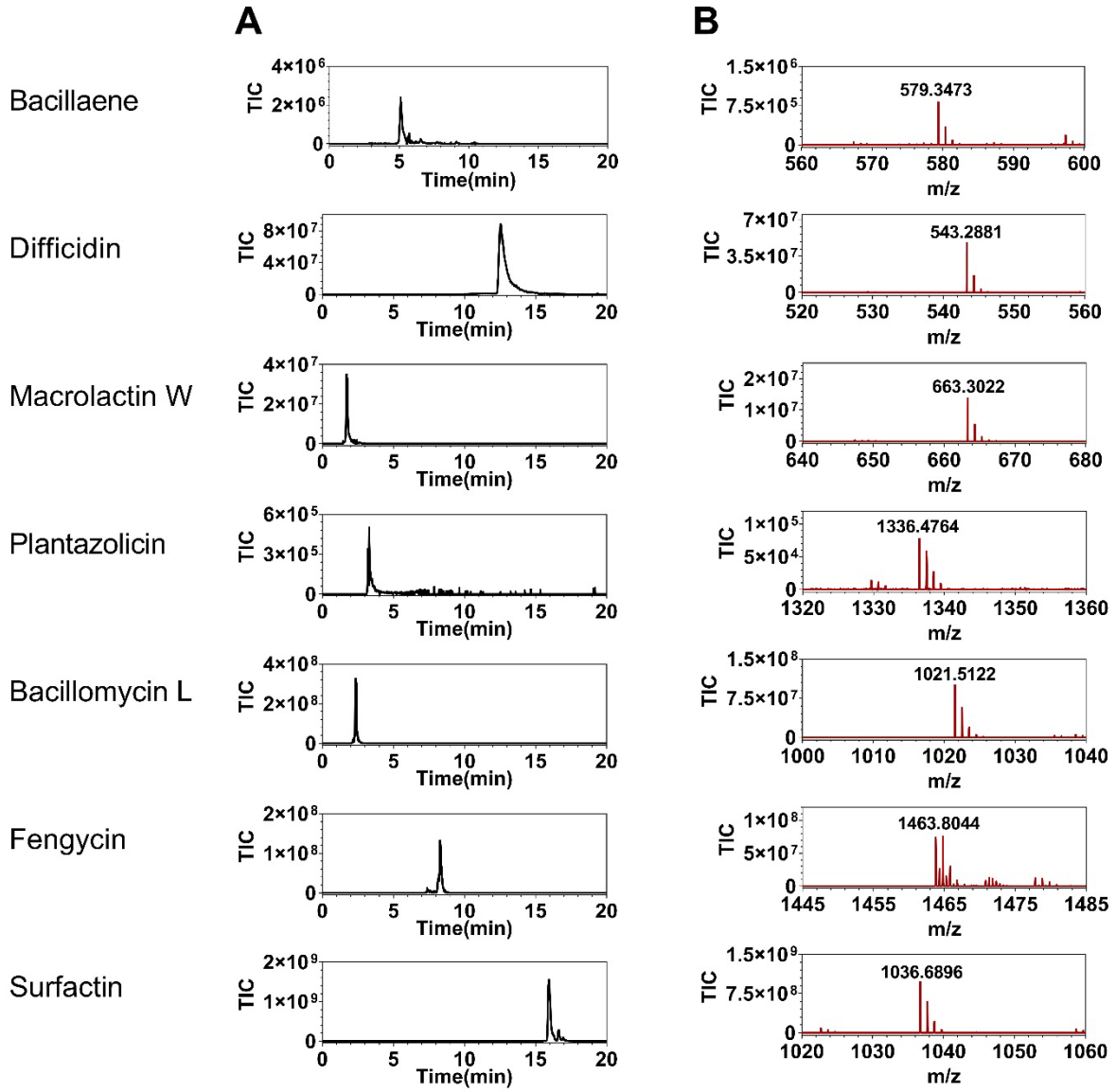


Figure 3.6 (A) Extracted ion chromatograms (EIC) and (B) molecular ion peak mass spectra of seven antimicrobial compounds produced by *Bacillus velezensis* JJ334. Ions of bacillaene, difficidin and macrolactin W are generated using negative ion mode (ESI-) while the rest are generated using positive ion mode (ESI+). EIC was generated from the precursor ion of respective compound using FreeStyle software from the LC total ion chromatogram produced by Thermo Fisher Orbitrap Exploris 120 LC-MS instrument (Acquity UPLC BEH C18 (1.7 µm, 2.1 × 50 mm) (see *Materials and Methods*). For unambiguous identification of secondary metabolites, each precursor ion was simultaneously fragmented using a multi-stage MS^n analyzer and each compound was confirmed by representative fragment ions generated form the precursor ion.

Table 3.1: Secondary metabolites identified from five *Bacillus* species.

Compounds	Producer Species[1]	M.F.[2]	$[M+H]^+$	$[M-H]^-$	Error (ppm)
Bacillaene	*B. vel., B. sub.*	$C_{34}H_{48}N_2O_6$		579.3473	1.91
Difficidin	*B. vel.*	$C_{31}H_{45}O_6P$		543.2881	1.28
Oxydifficidin	*B. vel.*	$C_{31}H_{45}O_7P$		559.2831	1.41
Macrolactin W	*B. vel.*	$C_{34}H_{48}O_{13}$		663.3022	0.82
Plantazolicin	*B. vel, B. saf.*	$C_{63}H_{69}N_{17}O_{13}S_2$	1336.4764	1334.4613	1.43
Bacillomycin L	*B. vel.*	$C_{45}H_{70}N_{10}O_{16}$	1007.5045	1005.4879	0.03
	B. vel.	$C_{46}H_{72}N_{10}O_{16}$	1021.5192	1019.5049	0.69
	B. vel.	$C_{47}H_{74}N_{10}O_{16}$	1035.5351	1033.5203	0.50
	B. vel.	$C_{48}H_{76}N_{10}O_{16}$	1049.5504	1047.5366	0.95
	B. vel.	$C_{49}H_{78}N_{10}O_{16}$	1063.5662	1061.5507	0.53
	B. vel.	$C_{50}H_{80}N_{10}O_{16}$	1077.5809	1075.5669	1.50
Fengycin	*B. vel.*	$C_{70}H_{106}N_{12}O_{20}$	1435.7730	1433.7492	2.17
	B. vel., B. sub.	$C_{71}H_{108}N_{12}O_{20}$	1449.7894	1447.7728	1.69
	B. vel., B. sub.	$C_{72}H_{110}N_{12}O_{20}$	1463.8044	1461.7891	0.93
	B. vel., B. sub.	$C_{73}H_{112}N_{12}O_{20}$	1477.8195	1475.8048	1.30
	B. vel., B. sub.	$C_{74}H_{114}N_{12}O_{20}$	1491.8349	1489.8163	0.61
	B. vel., B. sub.	$C_{75}H_{116}N_{12}O_{20}$	1505.8505	1503.8320	0.50
	B. vel., B. sub.	$C_{76}H_{118}N_{12}O_{20}$	1519.8625	1517.8474	0.34
	B. vel.	$C_{77}H_{120}N_{12}O_{20}$	1533.8819	1531.8637	0.17
	B. vel.	$C_{78}H_{122}N_{12}O_{20}$	1547.8966	1545.8792	0.30
	B. vel.	$C_{79}H_{124}N_{12}O_{20}$	1561.9106	1559.8946	1.42
Surfactin	All[3]	$C_{50}H_{87}N_7O_{13}$	994.6383	992.6281	0.12
	All	$C_{51}H_{89}N_7O_{13}$	1008.6592	1006.6429	0.33
	All	$C_{52}H_{91}N_7O_{13}$	1022.6739	1020.6577	0.61
	All	$C_{53}H_{93}N_7O_{13}$	1036.6896	1034.6730	1.53
	All	$C_{54}H_{95}N_7O_{13}$	1050.7054	1048.6884	1.12
	All	$C_{55}H_{97}N_7O_{13}$	1064.7209	1062.7056	0.79
	All	$C_{56}H_{99}N_7O_{13}$	1078.7299	1076.7212	0.31
	B. vel., *B. saf.*	$C_{57}H_{101}N_7O_{13}$	1092.7523	1090.7370	0.40

[1]Producer species: *B. vel. = B. velezensis*, *B. sub.= B. subtilis*, *B. pum. = B. pumilus*, *B. saf. = B. safensis*, *B. alt. =B. altitudinis.*

[2]M.F: Molecular Formula.

[3]All: Representative strains from all five *Bacillus* species.

Consistent with the secondary metabolite prediction (see Fig. 3.4), three different antibacterial polyketides (bacillaene, difficidin, and macrolactin W) were produced by *B. velezensis* JJ334*,* whereas only bacillaene was produced by *B. subtilis* JM553. Note that all these polyketides were detected in the negative ionization mode. Plantazolicin was produced by both *B. velezensis* JJ334 and *B. safensis* JJ1244 as predicted by BGC analysis. Six bacillomycin L derivatives were detected in the *B. velezensis* JJ334 extract as both protonated and deprotonated ions, whereas seven and six fengycin derivatives were detected as protonated and deprotonated ions in *B. velezensis* JJ334 and *B. subtilis* JM553 extracts, respectively. Eight to nine surfactin derivatives were delectated in the extract of all five *Bacillus* species

3.4.5. Intraspecies inhibition and resistance among 18 *B. velezensis* strains

A total of 18 *B. velezensis* strains were evaluated for intraspecies interactions in order to compare how they inhibit and resist one another. Each target strain was challenged by all strains (focal strains), including itself. The percent antibiosis score for each strain was determined based on the measured zone of inhibition and morphological changes of the focal/target strain (see *Materials and Methods*). Each strain was non-inhibitory against itself, with the minor exceptions of AP202 and AP52, where only 8% self-inhibition was observed (Fig. 3.7). Strikingly, the observed inhibitory response was highly similar for strains that were phylogenetically more closely related. For instance, closely related AB01 and JJ951 exhibited strong inhibition against virtually all strains but themselves. Strain AP215 also exhibited strong to moderate inhibition against all strains but was unable to inhibit AP46 and itself. Similarly, closely related strains AP46, AP202, JM199, JJ747, and JJ947 exhibited strong inhibition against most strains. As a contrast, closely

related strains AP52, JM204, and JJ213, along with JJ1284 showed either weak or no inhibition against almost all strains except JJ1043.

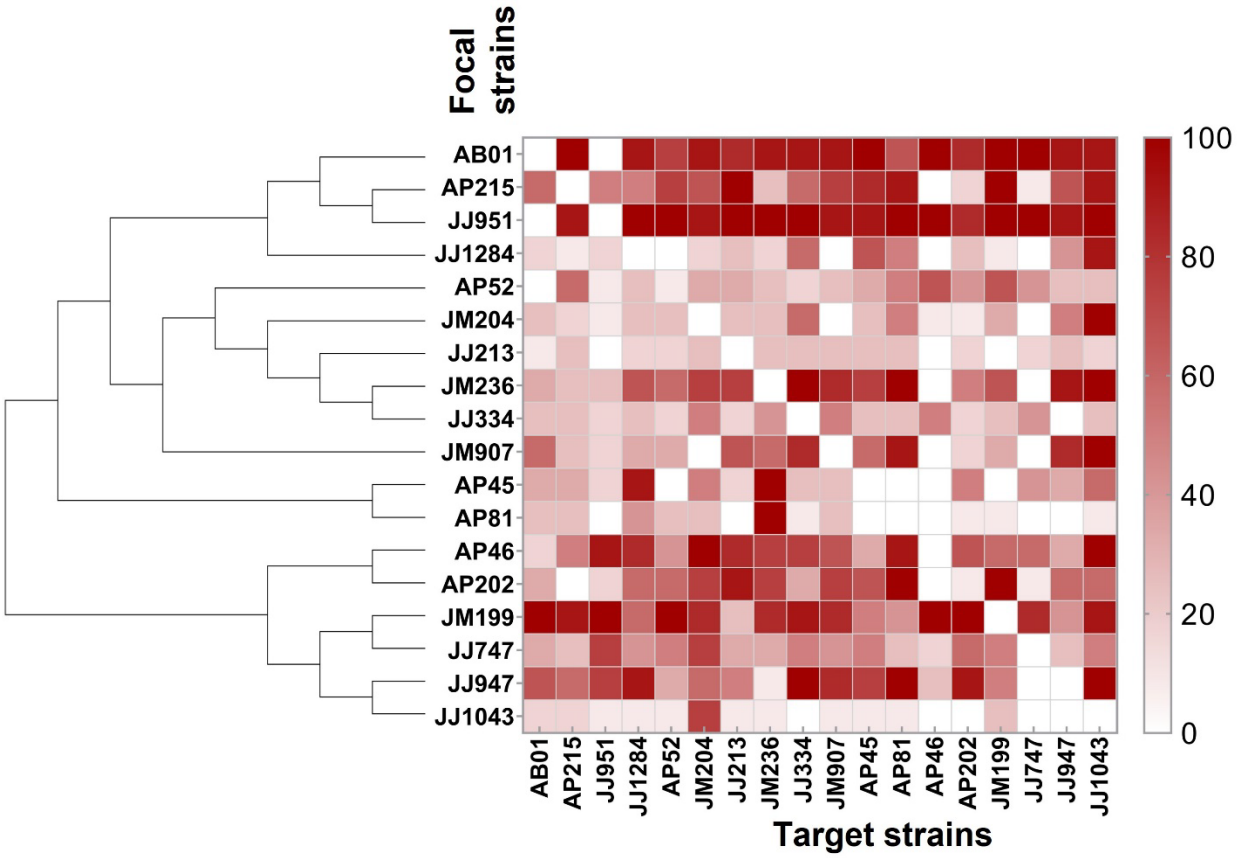


Figure 3.7: Overview of *Bacillus velezensis* intraspecies inhibition. Percent inhibition score of each focal strain against each target strain is displayed as a heatmap (the scale is shown on the right). The 16S rRNA-based phylogenetic relationships of all 18 *B. velezensis* focal strains is shown to the left of the vertical axis. Each box in the heatmap is colored based on the average inhibitory response (0 to 100%) from the evaluation of three replicates. The standard deviation for intraspecies inhibition is in the range of 0 to 29%. Heatmap boxes next to each focal strain (horizontal direction) indicates percent inhibition against each target strain, where darker red corresponds to stronger inhibition. Conversely, boxes ascending vertically from each target strain indicates resistance of the target strain to inhibition from each focal strain, where lighter red corresponds to greater resistance.

Note that the JJ1043 strain did not show strong inhibition against any strain except JM204, whereas AP81 showed weak to no inhibition to virtually all strains except JM236. JM236 showed moderate to strong inhibition against all but AP46 and JJ747, whereas JJ334 showed at least some level of inhibition against all but JJ947.

3.4.6. *Bacillus velezensis* intraspecies interaction revealed several strains with high antibacterial and resistance activity

Inhibition and resistance abilities of each *B. velezensis* strain against all 18 strains were compared to identify stains with the highest inhibitory and resistance capacity within the species. The cumulative % inhibition of each focal strain was calculated based on the percent inhibition of each strain against all 18 strains (see *Materials and Methods*). Conversely, the cumulative % resistance of each target strain was calculated based on the inverted percent resistance of each focal strain against all 18 strains. As shown in Fig. 3.8, a wide range of inhibitory and resistance activity was observed across the *B. velezensis* strains. A few strains exhibited markedly higher cumulative % inhibition and % resistance compared to the others. Interestingly, only two strains produced inhibition in excess of 80%, AB01 (81%) and JJ951 (86%). Strains showing moderate inhibition of other strains (50-77%) included JM199 (74%), AP46 (63%), JJ947 (59%), JM236 (57%), AP215 (56%), and AP202 (51%). Note that the phylogeny of these 18 *B. velezensis* strains determined from16S rRNA showed that AB01, JJ951, and AP215 are closely related, while JM199, AP46, JJ947, and AP202 group together in a separate clade. This suggests that the closely related strains may carry conserved factors that are responsible for the observed competitive advantages imparted by antibiosis activity that some *B. velezensis* strains express over others.

Strains with less than 20% antibacterial activity were AP81 (17%), JJ213 (16%), and JJ1043 (11%).

Accounting for percent resistance of each target strain against all strains, only five strains exhibited values > 60%: AP46 (74%), JJ747 (72%), JJ951 (71%), AB01 (69%), and AP215 (63%). Consistent with the previous observation, the trait of resistance tended to appear among phylogenetically closely related strains. For example, AB01, AP215, and JJ951 on one hand and AP46 and JJ747 on the other showed relatively robust resistance. Alternatively, strains with the lowest percent resistance (<50%) were JJ334 (46%), JM204 (45%), AP81 (44%), and JJ1043 (33%). In summary, the intraspecies interaction showed phylogenetically closely related strains consistently exhibited stronger and broader inhibition *and* more effective resistance within *B. velezensis* species. This suggests that intraspecies inhibition and resistance may be linked with phylogenetically conserved capacity for producing antibacterial compounds and expressing resistance genes, respectively.

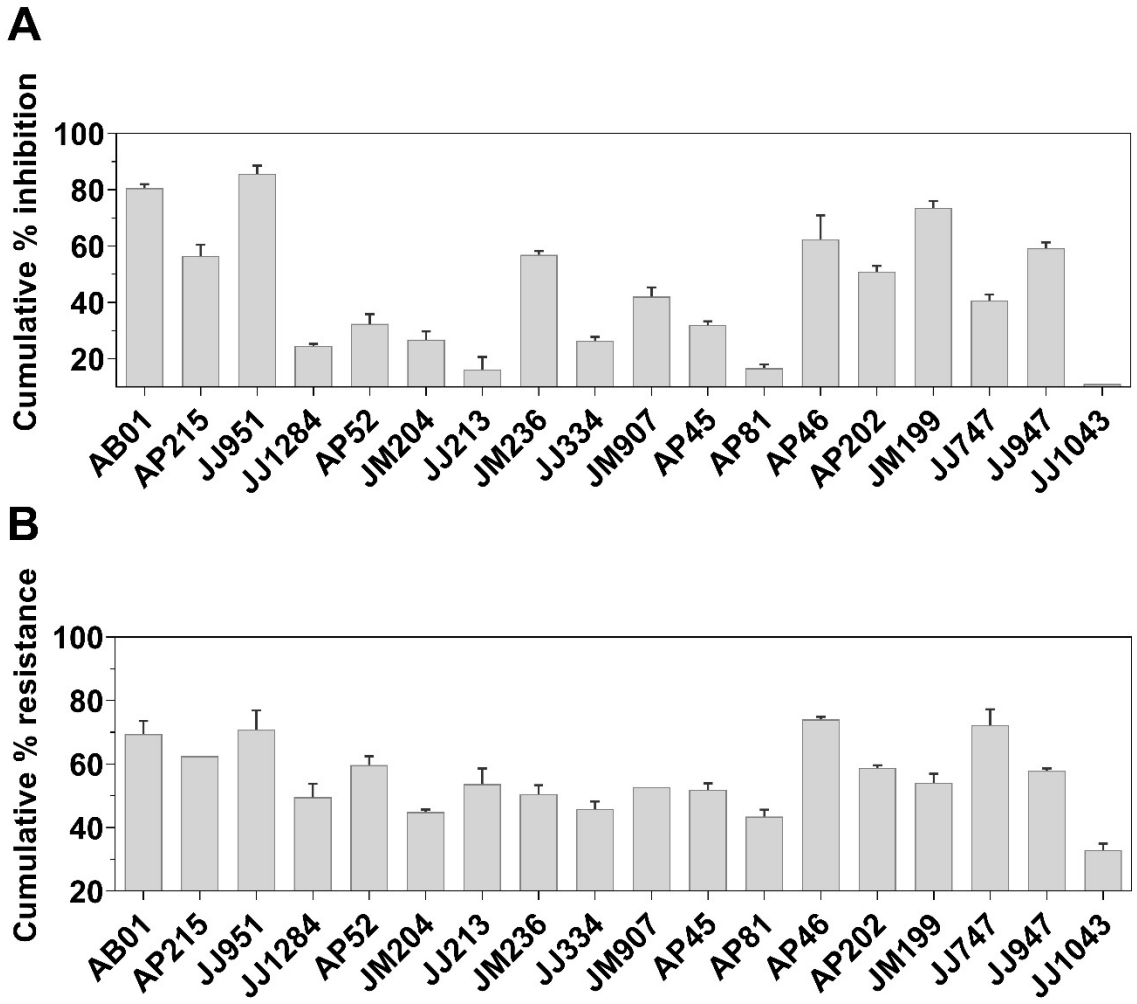


Figure 3.8: Quantitative intraspecies inhibition and resistance activity of *Bacillus velezensis* strains. (A) Cumulative % inhibition of each *B. velezensis* focal strain is determined from the average percent inhibition across all 18 strains (see *Materials and Methods*). Similarly, the (B) cumulative % resistance of each target strain is determined from the average inverted percent inhibition derived from the percent inhibition of all focal strains. Note that percent inhibition and resistance were calculated from at least three replicates of intraspecies interaction experiments, and the error bars represent standard deviation.

3.4.7. Antifungal activity of *B. velezensis* strains

Each *B. velezensis* strain was further evaluated for antifungal activity against three fungal pathogens: *F. oxysporum*, *F. graminearum*, and *R. solani*. The percent antifungal score was

calculated based on the measurement of zones of inhibition (see *Materials and methods*). The relative antifungal inhibition of each strain was consistent across three fungi, but the percent inhibition varied from one fungus to another (Fig. 3.9). This variation is due to the difference in sensitivity of each fungus towards *Bacillus* strains. Among the three fungi evaluated, *R. solani* was the most susceptible to *B. velezensis* inhibition and *F. oxysporum* was the least. Three strains that exhibited ≥ 75% inhibition against *F. oxysporum* were AP215 (92%), AP202 (88%), and AP52 (75%). Strains exhibiting less than 40% inhibition against *F. oxysporum* were AP46 (38%), JM907 (33%), JJ947 (25%), and JM236 (25%). For *F. graminearum*, five strains exhibiting ≥ 75% inhibition were AP215 (79%), JJ1284 (88%), AP52 (88%), AP204 (88%), and JJ747 (79%). Alternatively, strains exhibiting less than 60% inhibition against *F. graminearum* were JM236 (33%), JM907 (46%), and JJ947 (50%). Three strongly inhibitory strains against *R. solani* were AP215 (100%), JM907 (96%), and AP81 (96%), whereas the least three inhibitory strains with less than 65% inhibition against *R. solani* were JM236 (46%), AP202 (58%), and JJ947 (63%).

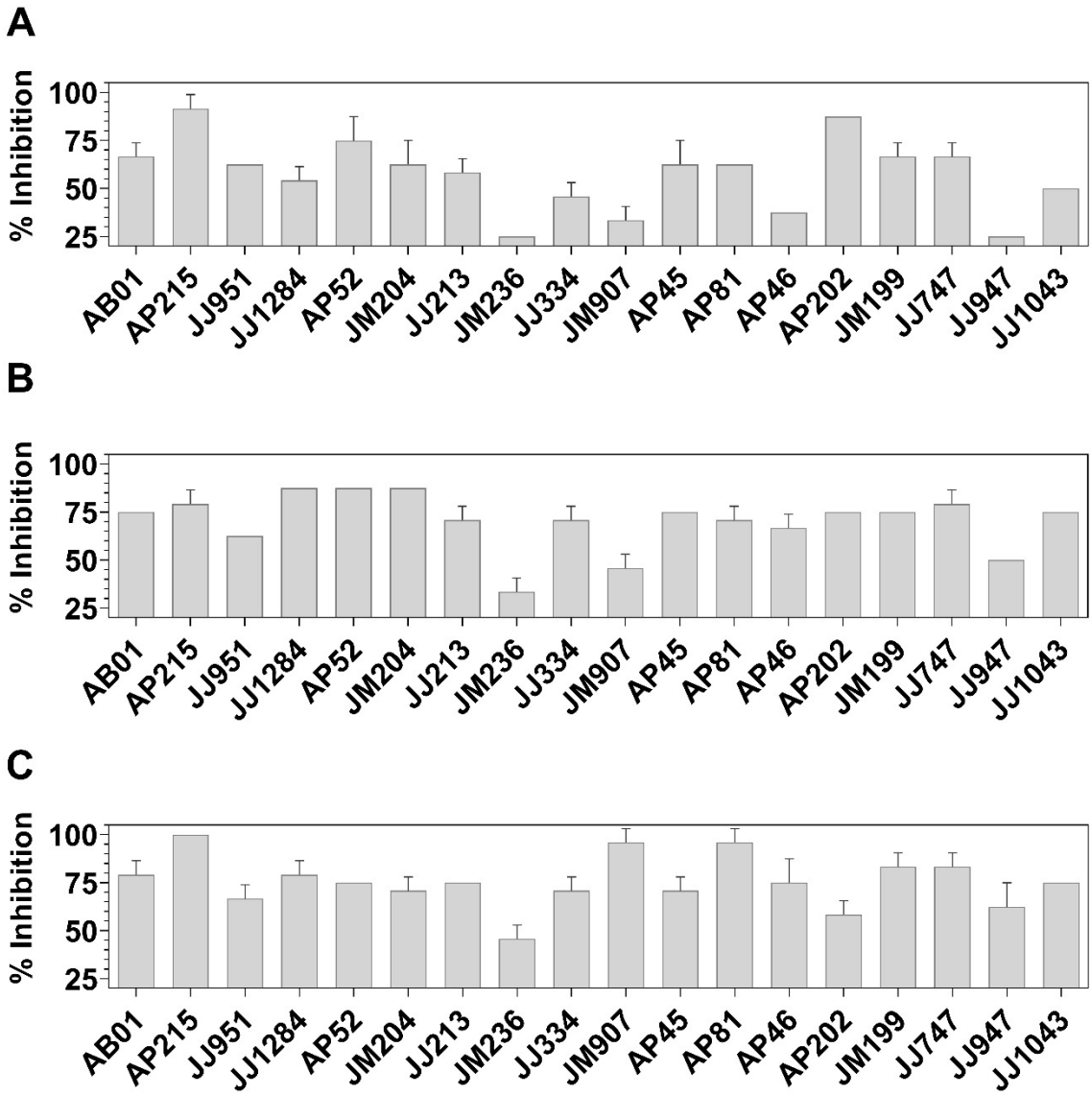


Figure 3.9: Overview of quantitative antifungal activity of 18 *B. velezensis* strains against (A) *F. oxysporum*, (B) *F. graminearum*, and (C) *R. solani*. The percent inhibition is calculated based on the measurement of the zones of inhibition (ZOI) of each strain against the target fungus. Antibiosis assays were replicated at least three times, and the error bars represent standard deviation.

Accounting for antifungal activity across all three fungi, four *B. velezensis* strains that produced higher than 75% inhibition were AP215 (90%), AP52 (79%), AP81 (76%), and JJ747

(76%) (Fig. 3.10). Strains with percent inhibition score ranging from 70-75% were JM199 (75%), JM204 (74%), AB01 (74%), JJ1284 (74%), and AP202 (74%). The two least inhibitory strains were JJ947 (46%) and JM236 (35%).

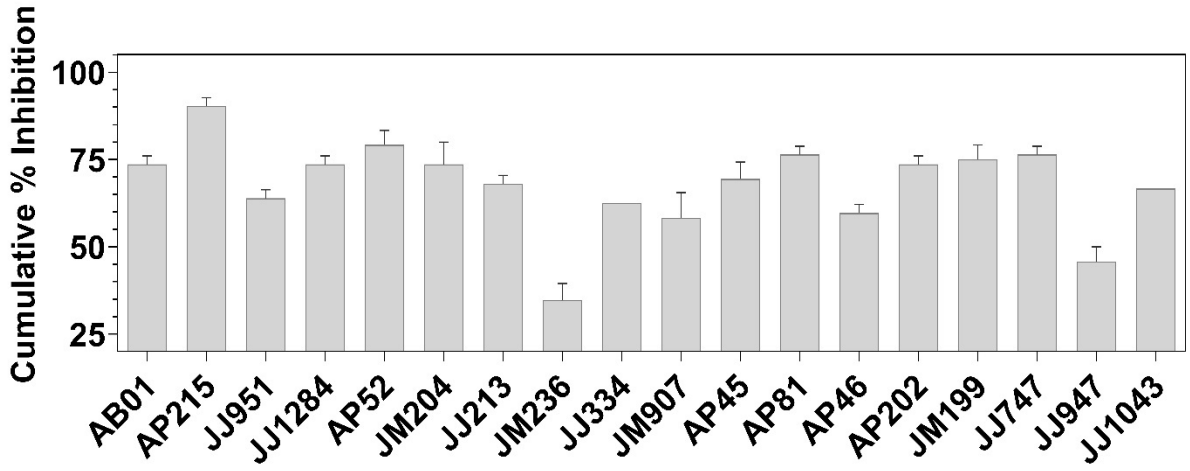


Figure 3.10: Antifungal activity of *B. velezensis* strains across three fungi. Cumulative percent inhibition of each *B. velezensis* strain across three fungal pathogens, *F. oxysporum*, *F. graminearum*, and *R. solani*. Note that cumulative percent inhibition was calculated from at least three replicates of antibiosis assays against each fungus, and the error bars represent standard deviation.

3.4.8. Representation of BGCs across 18 *B. velezensis* strains

All 18 *B. velezensis* strains carried highly similar BGCs with only a few exceptions. As determined by antiSMASH analyses, all strains carried BGCs for bacillaene, difficidin, macrolactin H, bacilysin, bacillibactin, fengycin, and surfactin. All 18 strains also carried a BGC for some form of iturin; fifteen of them were predicted to produce canonical iturin (also referred to as iturin A), while only three were predicted to produce an alternative iturin derivative, bacillomycin L. Only four strains, JJ1244, JM236, JJ334, and AP202 contained a plantazolicin BGC, while JJ1284 and JM236 contained a mersacidin BGC. Myxovirescin A was identified in

AP45 and AP81; rhizomide A was identified in AP46, JM199, and JJ747; the only strain identified to carry a subtilin BGC was JM199.

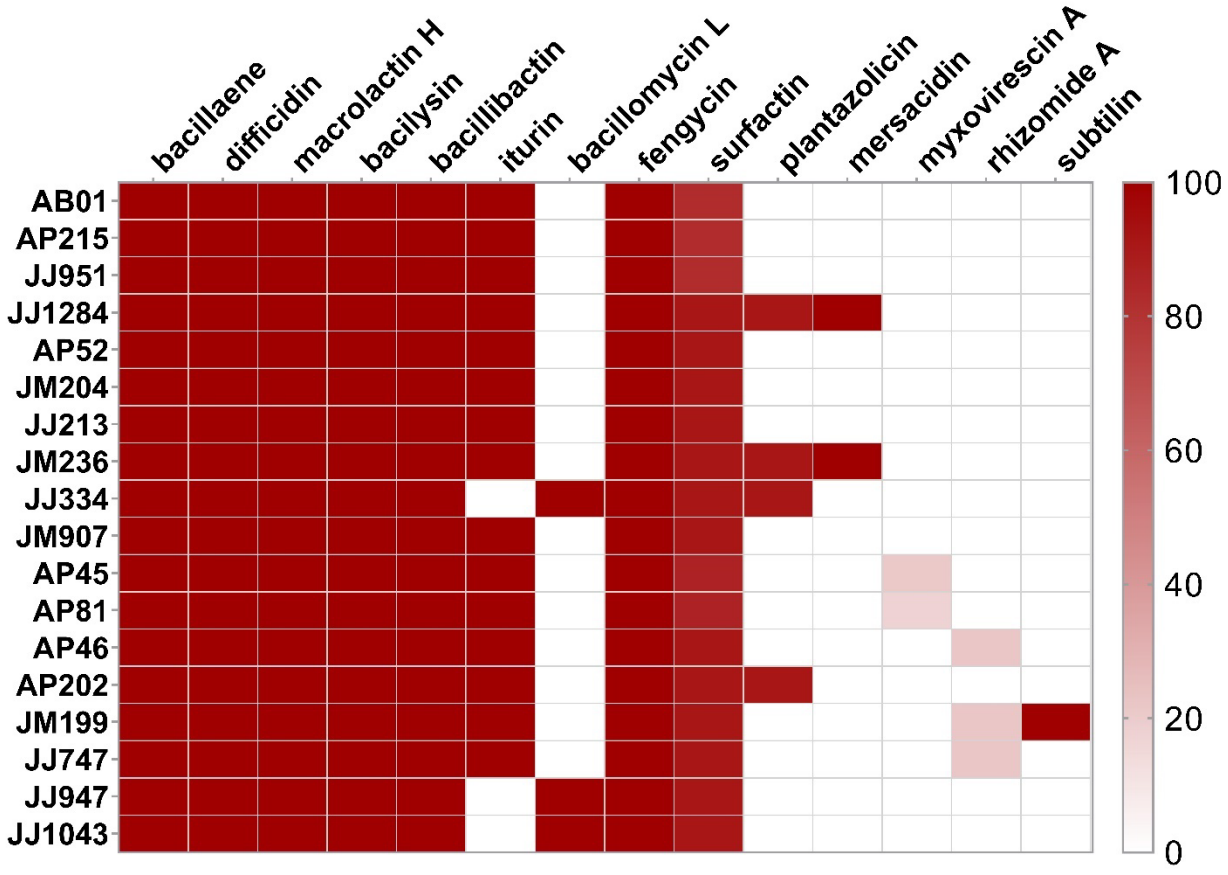


Figure 3.11: Distribution of secondary metabolites-producing BGCs in 18 *Bacillus velezensis* strains. The heatmap is generated using the percent similarity score predicted by antiSMASH v.5 for the secondary metabolites carried by all 18 *B. velezensis* strains (189).

It is important to note that this highly similar profile of secondary metabolite BGCs cannot fully explain the variation of antibacterial and/or antifungal activity across these 18 strains. At a minimum, identification of myxovirescin and rhizomide A BGCs were based on less than 25% similarity to the expected structures/sequences of these BGCs according to antiSMASH. In

addition, it is likely that there are differences in the levels of expression/production of these antimicrobial secondary metabolites that contributes to the differences in antibiosis properties (intraspecies, intrabacillus, antibacterial, antifungal, etc). Also, the expression of each secondary metabolite would be dictated by the regulatory genes associated with each BGC. In a similar manner, the resistance against each secondary metabolite may be controlled by the presence or expression of resistance genes present in the BGC.

3.4.9. Production of antimicrobial secondary metabolite by 18 *B. velezensis* strains

One factor that may account for the distinctions between *B. velezensis* strains in the breadth and/or intensity of antibiosis properties is the extent to which each strain is able to generate the secondary metabolites predicted by BGC content. To evaluate this hypothesis, culture media from all 18 *B. velezensis* strains were collected and extractions were performed to evaluate production of antimicrobial secondary metabolites (see *Materials and Methods*). Secondary metabolites were identified by LC-MS as for Fig. 6 and Table 1, and the relative quantity produced by each strain was determined from peak areas under the curves of extracted ion chromatograms (EIC). Relative quantities of three major polyketides (bacillaene, difficidin, and macrolactin W) and three lipopeptides (iturin, fengycin, and surfactin) were obtained (Fig. 3.12). In general, the three polyketides are regarded for their antibacterial activity while the three cyclic lipopeptides are best known for their activity against oomycetes and fungi (24, 44, 53–55). Although almost all strains produced some level of each polyketide, the production profile for each one was highly variable across the 18 strains (Fig. 12A). Bacillaene production was detected in most strains; however, it was most abundant for JJ334 followed by JM907, and JJ1043. Conversely, little if any was produced by JJ951, JJ1284, AP45, JM199, and JJ747. When difficidin was considered, a

completely different strain profile was observed. The most prolific difficidin-producing strains were AB01 and JJ951 followed closely by AP215. Difficidin was not detected from JJ1043. Macrolactin W production produced a third distinct strain profile. The most prolific macrolactin producer was JM907 followed by JJ213, JM199, and JJ947. Notable quantities of macrolactin W were also observed for AP46 and JJ334.

As with the polyketides, there were also unique strain profiles for each of the cyclic lipopeptides, albeit with less variation across strains (Fig. 12B). Consistent with BGC predictions from antiSMASH, bacillomycin L was identified as the iturin produced by only three strains (JJ334, JJ947, and JJ1043); when detected, iturin A was observed for the other *B. velezensis* strains. Iturin production appeared to be most abundant for JJ334 followed by JJ213 then AP46 and JJ947. Neither bacillomycin L nor iturin A were detected in appreciable amounts from AP45 or JJ747. Most strains were strong producers of fengycin, but JJ951, JJ213, and AP46 generated the strongest signals followed closely by AB01, AP215, JM204, and AP81. Interestingly, the lowest quantity of fengycin was observed in JJ1043 culture media extracts. Most strains were also strong producers of surfactin, though the specific strain profile was distinct. The most abundant surfactin production came from AB01 and JJ951, but AP215, JJ1284, JM204, JJ213, AP45, and AP81 all followed close behind. Notably, JM907 was not able to generate appreciable quantities of this cyclic lipopeptide.

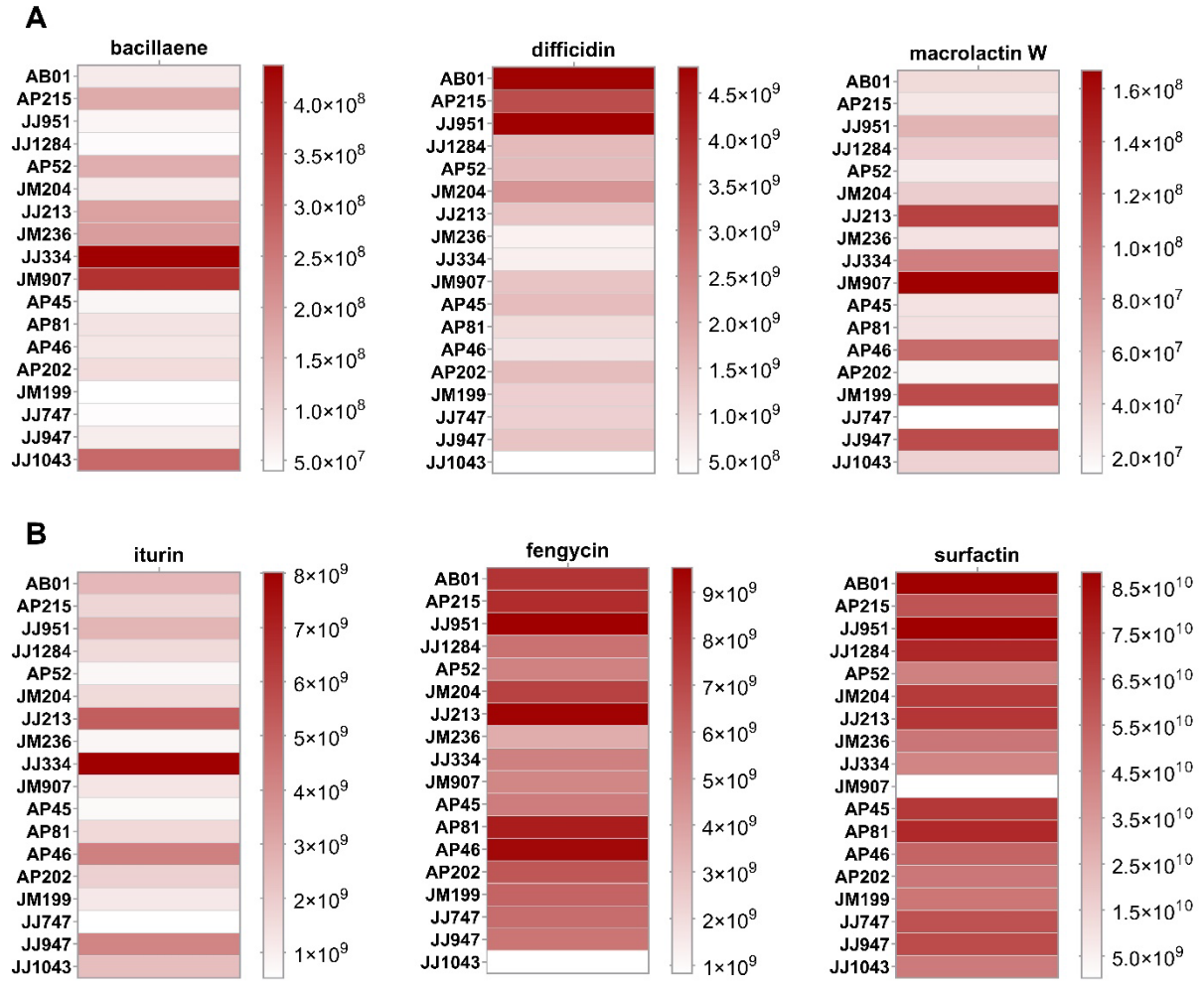


Figure 3.12: Comparison of antimicrobial secondary metabolites produced by 18 *Bacillus velezensis* strains. Peak areas are shown from the extracted ion chromatograms (EIC) of representative (A) polyketides (bacillaene, difficidin, and macrolactin W) and (B) lipopeptides (iturin, fengycin, and surfactin) produced by *B. velezensis* strains. Each box in the heatmap is colored based on the peak area calculated from EIC of the respective compound (the scale is shown on the right).

3.4.10. Comparison of growth phenotype of 18 *Bacillus velezensis* strains

All 18 *B. velezensis* strains were grown for 48 hours in order to compare the growth phenotype and evaluate biofilm formation. Fifteen out of the 18 strains produced robust and highly

complex biofilms (Fig. 3.13). The remaining three strains that failed to do so were JJ1043, JM204, and JM907.

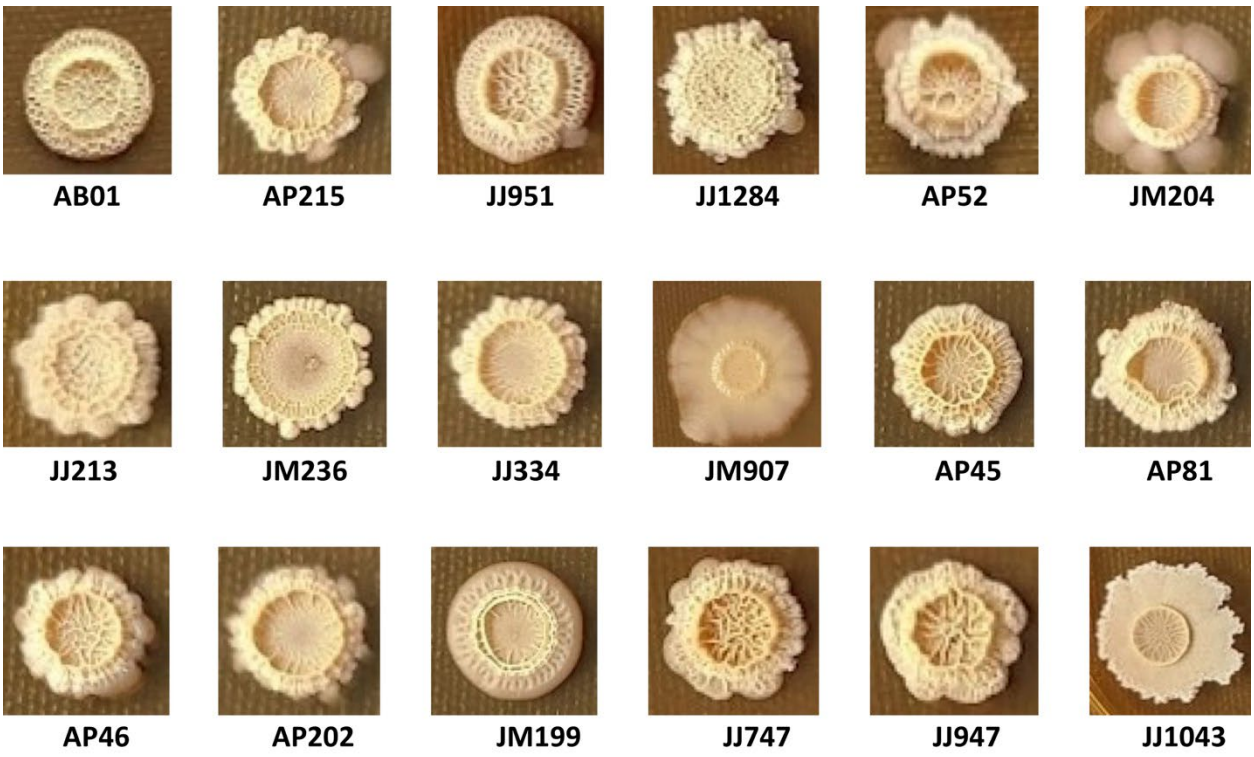


Figure 3.13: *Bacillus velezensis* growth morphology and biofilm formation.

3.5. Discussion

Bacillus species have been considered one of the most promising biocontrol agents since the 1980s (339). This is due to their ability to promote plant growth by assisting with nutrient uptake and modulating phytohormones, spore formation, biofilm formation, and survival in various field conditions, including extreme environments. Plant disease protection comes from their ability to produce various biomolecules, including lytic enzymes, antimicrobial compounds, and siderophores (15). Since the 1990s, several *Bacillus*-based bioproducts have been commercially used as active ingredients. The species which have been utilized for this purpose

include *B. subtilis*, *B. thuringiensis*, *B. amyloliquefaciens*, *B. licheniformis*, *B. megaterium*, *B. velezensis*, *B. cereus*, and *B. pumilus* (340, 341). Major factors contributing to *Bacillus* biocontrol and plant-growth abilities include production of antimicrobial secondary metabolites, siderophores, hormones, and formation of robust biofilms and endospores. Several commercial bioproducts of *B. amyloliquefaciens* and *B. spp.* have achieved 44-85% of biocontrol ability through the action of antifungal lipopeptides, *Fusarium* disease reduction, and plant root elongation (342). *Bacillus subtilis* commercial bioproducts have been reported to elongate plant roots and reduce *Fusarium* colony with a biocontrol ability of 75-82% (343, 344).

As of today, numerous studies have been performed to evaluate the biocontrol potential of *B. subtilis* and *B. velezensis* for their ability to control fungal and bacterial diseases. The secondary metabolites produced by these two species contribute to biofilm formation, antifungal activity, anti-oomycete activity, and disease suppression (345, 346). Fan *et al.* demonstrated that *B. velezensis* produces a higher number of bioactive metabolites that may contribute to its broad-spectrum biocontrol ability (347). Similarly, *B. subtilis* produces antifungal and antibacterial compounds and have an excellent ability for biofilm formation and root colonization (123, 348). These attributes enabled *B. subtilis* to be an effective biological agent against plant pathogens. To date, *B. pumilus*, *B. safensis*, and *B. altitudinis* are less explored, and their biocontrol ability against plant pathogens is far less understood. Previous studies reported that the antibacterial and antifungal activity of *B. pumilus* may associate with the production of extracellular enzymes, biosurfactants, and volatile organic compounds (VOCs) (349, 350).

In the present study, *B. velezensis* and *B. subtilis* exhibited strong antifungal activity, whereas *B. pumilus*, *B. safensis*, and *B. altitudinis* were non-inhibitors. *B. velezensis* produced

relatively larger ZOI than *B. subtilis* across all fungal pathogens. This can be explained by the fact that *B. subtilis* produces only fengycin and surfactin, whereas *B. velezensis* produces bacillomycin L (an iturin), fengycin, and surfactin. Previous studies reported strong antifungal activity of bacillomycin L and fengycin, whereas no or weak at best antifungal activity of surfactin (143, 156, 288, 351). The lack of antifungal activity of *B. pumilus*, *B. safensis*, and *B. altitudinis* can also be explained by their inability to produce no more lipopeptide than surfactin. The broad interspecies antibacterial activity of *B. velezensis* is likely due to its ability to produce the highest number of antibacterial secondary metabolites than the other four species being evaluated. Likewise, relatively higher interspecies resistance of *B. velezensis* may be linked with its ability to neutralize or excrete antibacterial compounds by expressing resistance genes when it is challenged by other bacteria. Previous studies showed that biofilm formation and swarming motility of *Bacillus* species are known to be mediated by the production of surfactin and fengycin (336, 352). This is consistent with the robust biofilm formation of *B. velezensis* and *B. subtilis* since both species can produce both fengycin and surfactin.

Our BGC comparison of 18 *B. velezensis* species showed highly similar BGC distribution across all 18 strains, with a few exceptions. All strains carried major antibacterial and antifungal compounds, including bacillaene, difficidin, macrolactin, iturin (or bacillomycin L), fengycin, and surfactin. However, we observed a diverse intraspecies inhibition activity within these strains. This may be partly explained by the variation of expression in antimicrobial secondary metabolites even though each strain has the ability to produce the same antibacterial compounds. Strikingly, strains with the highest antibacterial activity are phylogenetically related. This suggests that phylogenetically conserved regulatory factors for the expression of antibacterial secondary

metabolites may associate with the observed antibacterial activity. Conversely, the observed antibacterial resistance may be linked with the level of antibacterial production of each strain since higher antimicrobial-producing strain also needs an intrinsic mechanism for self-protection through the expression of resistance genes or associated regulatory genes. The variation of antifungal activity within *B. velezensis* strains may also be linked with the level of expression of antifungal lipopeptides. Further experiments must be performed to unequivocally determine what factors are responsible for the observed variation of antibacterial/fungal activity.

In summary, a comparison of five *Bacillus* species for antifungal activity, intraspecies inhibition and resistance activity, biofilm formation, and antimicrobial secondary metabolites production identified *B. velezensis* as the most promising species for developing commercially viable biological agents. Further evaluation of 18 *B. velezensis* on similar traits showed phylogenetically closely related AB01, JJ951, and AP215 performed relatively better than other strains. Further experiments on evaluating antifungal activity and plant growth promotion in green-house and field conditions must be performed to identify the most promising *B. velezensis* strains for commercially viable biological agents.

Chapter 4: Structural and functional insights into *Bacillus* cytochromes P450 associated with secondary metabolite biosynthesis

4.1. Abstract

Cytochrome P450 enzymes can catalyze broad diversity of reactions with a high degree of chemo-, regio-, and stereoselectivity. Their role in bacterial secondary metabolite biosynthesis is particularly interesting since they are often involved with various oxidative transformations potentially useful in pharmaceutical and biotechnological applications. To examine their distribution, sequence conservation, tertiary structure, and substrate preference, we carried out a comprehensive genomic analysis of 1,562 *Bacillus* strains. From a total of 5,051 P450 genes, we identified 614 CYPs encoded within biosynthetic gene clusters (BGCs) as "accessory genes". These BGC-affiliated P450s were distributed among six families: CYP107 (326), CYP134 (134), CYP113 (84), CYP109 (33), CYP102 (27), and CYP106 (2). Interestingly, enzymes from each P450 family were associated with a specific secondary metabolite biosynthetic pathway: bacillaene/fengycin (CYP107), cyclodipeptide (CYP134), difficidin (CYP113), bacillibactin (CYP109), and plantazolicin (CYP102). Their amino acid sequence conservation across the P450 family showed that CYP113 and CYP134 are the most highly conserved, whereas CYP107 and CYP109 are the least conserved. Based on this observation, we surmised that the phylogenetically-related amino acid sequence conservation is linked with substrate specificity with respective biosynthetic pathways. High-resolution homology models of a representative P450 from these families showed distinct structural features at the active site that may relate to substrate specificity. Molecular docking simulations of each P450 with putative substrates showed favorable binding of CYP113, CYP134, CYP109, CYP107, and CYP102 with the substrates difficidin, cyclodipeptide,

dihydroxybenzoate, bacillaene, and plantazolicin, respectively. Furthermore, a representative CYP102 from *Bacillus amyloliquefaciens* (*Ba*CYP102A2) was synthesized, cloned, and transferred into *E. coli* for expression, purification, and characterization. The identity of the *Ba*CYP102A2 was confirmed by the SDS-PAGE band at ~120 kDa, UV-vis Soret band at 419.5 nm, characteristic charge transfer bands, and type-I spectral shift upon substrate binding. The latter is a canonical substrate-dependent shift in the heme iron's coordination state from hexacoordinate low-spin to pentacoordinate high-spin. This was exploited to examine the substrate binding response of *Ba*CYP102A2 with fatty acid substrates. Sodium dodecyl sulfate (SDS) produced a sigmoidal response, whereas oleic acid (OA) produced a hyperbolic response. Steady-state kinetic responses of *Ba*CYP102A2 with respect to SDS, myristic acid (MA), and palmitic acid (PA) produced a sigmoidal response, whereas OA produced a hyperbolic response.

4.2. Introduction

Across the cytochrome P450 monooxygenase (CYP/P450) superfamily, a heme-bearing enzyme, the typical reaction catalyzed by these enzymes is a single-oxygen insertion into a broad variety of substrates (353–356). The P450s rely on a Cys thiolate-coordinated heme center, and a hallmark of their reactivity is to oxygenate otherwise unreactive and recalcitrant substrates. Accordingly, the cytochromes P450 have been one of the most extensively studied enzyme superfamilies in all of biology (357). P450s have been identified across all kingdoms of life, including the viruses, and are involved in various cellular processes ranging from the biosynthesis of metabolites to biodegradation of toxic compounds (355, 358). The P450s play an important role in the excretion of drugs and xenobiotics in humans and biosynthesis of primary and secondary metabolites, (e.g., steroids, fatty acids, hormones, terpenes, polyketides, and biotins) across all

kingdoms of life (247, 359–363). At present, more than half a million distinct P450 protein sequences are available in the UniProt database.

Currently, with the increasing availability of genome sequences in public databases and the rapid improvement of bioinformatics algorithms, a large number of genomic datasets are available and remain to be mined to infer key insights on the genomic contents, gene organizations, protein structures and their corresponding functions. Enzymes from the cytochrome P450 superfamily are particularly attractive within this context given the diversity of their structures, the range of substrates they act upon, and their distribution across all kingdoms of life. P450s in the *Bacillus* species have drawn substantial interest since the first identification of a full-length CYP102 ($P450_{BM3}$) in *Bacillus megaterium* ATCC 14581 in 1986 (364). Numerous studies have been performed *in vitro* and *in silico* on this P450 to understand not only the P450 catalytic mechanism more generally, but also this homolog's substrate specificity and kinetic responses, as well as its biotechnological applications (363, 365–371). A recent comparative study on 128 *Bacillus* P450s provided critical insights into the diversity of P450s and their association with secondary metabolite biosynthesis (260). However, their structures, substrate preferences, in vivo targets, and as such, their roles in biosynthetic processes remain unclear. Thus, analysis of the currently existing large number of *Bacillus* genome sequences is anticipated to identify P450s with novel structure, functional role, and/or catalytic mechanism (370).

Bacillus species inhabiting the soil and plant rhizosphere carry various classes of biosynthetic gene clusters (BGCs) for producing secondary metabolites. The most commonly observed *Bacillus* BGCs are non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), ribosomally-synthesized, and post-translationally modified peptides (RiPP), and terpenes

(371). These BGCs often carry a cytochrome P450 gene as "accessory gene." Enzymes encoded by these P450s are anticipated to participate in the functional modification of respective BGC-expressed secondary metabolites. Although P450s, including a few *Bacillus* enzymes have been subject of intense investigation for many years, the BGC-affiliated P450s have been neglected due to the challenges in their heterologous expression and the availability of natural substrates. Therefore, investigating the BGC-affiliated P450s in the ever-expanding publicly available *Bacillus* genome sequence data is particularly interesting to uncover P450s with novel structure, function, and catalytic mechanism.

Although P450s were discovered in *Bacillus* species more than four decades ago, their exact catalytic mechanism, physiological roles, and substrate specificity are far less understood. For example, at present, the actual physiological function and the full-length structure of *Bacillus* CYP102 ($P450_{BM3}$) are unknown. Although *Bacillus* PksS has been identified in the biosynthesis of a polyketide bacillaene, its physiological substrate and the catalytic mechanism remain unknown (362). *Bacillus* CYP109 (also known as "versatile enzyme") has a broad-range substrate tolerance and catalytic role in fatty acids and steroids (272). However, its exact catalytic mechanism and physiological role are far less understood. Further, *Bacillus* CypX enzyme (CYP134) has been identified to be associated with the oxidative transformation of cyclodipeptides through a multistep oxidation process (365). However, its catalytic mechanism and substrate specificity are unclear. Inspired by these research gaps in *Bacillus* P450s, we carried out extensive analyses on currently available *Bacillus* genomes and P450 sequences to understand structure, catalytic role, and diversity of BGC-affiliated P450.

In the present study, 1,562 *Bacillus* genomes were analyzed to identify and annotate cytochromes P450, especially those involved in secondary metabolite biosynthesis. Amino acid sequence conservation and phylogenetic analysis of 614 BGC-affiliated P450s were performed to understand the distribution and conservation of P450s among various secondary metabolite biosynthetic pathways. High-resolution homology models of representative enzymes from five predominant BGC-affiliated P450 families were constructed to evaluate structural features in relation to respective secondary metabolite biosynthesis. In addition, to understand enzyme-substrate interactions, molecular docking simulations of representative enzymes were carried out with the candidate substrates linked with respective secondary metabolite biosynthesis. Furthermore, a CYP102 enzyme from *Bacillus amyloliquefaciens* (*Ba*CYP102A2) predicted to involve in plantazolicin biosynthesis was selected for *in vitro* investigation. The *Ba*CYP102A2 gene was synthesized, cloned, and transferred into a heterologous expression system. Subsequently, the expressed *Ba*CYP102A2 enzyme was purified and further characterized using SDS-PAGE, UV-vis absorption, substrate-binding titration, and steady-state kinetic assays.

4.3. Materials and Methods

4.3.1. *Bacillus* genomes and P450 sequences

A total of 1,562 high-quality *Bacillus* whole-genome sequences were downloaded from the Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters (IMG-ABC) database (https://img.jgi.doe.gov/cgi-bin/abc/main.cgi) (372). In total, 6,921 P450 protein sequences were identified in 1,562 *Bacillus* genomes at the IMG-ABC database based on a BLAST search using the six most commonly observed *Bacillus* P450 protein sequences. All P450 sequences were downloaded and then subjected to the NCBI Web CD-Search Tool to filter out the P450 proteins

with ≥ 40% sequence similarity using Pfam database (PF00067) (373). Finally, 6,096 P450 protein sequences with ≥ 40% similarity with known homologs were selected for further analysis.

4.3.2. *Bacillus* P450 sequence analysis

To annotate P450 proteins into families and subfamilies, previously annotated homolog P450 sequences were used as a reference from the Cytochrome P450 Homepage database (374). The P450 protein annotation was carried out based on the rules of the International P450 Nomenclature Committee (222). Following these rules, the P450 sequence identity of ≥ 40% with a known homolog was assigned to the same family and the sequence identity of ≥ 55% with a known homolog was assigned to the same subfamily. Based on this evaluation, returned 5,051 P450 protein sequences with ≥ 40% sequence identity with known homologs were selected for the analysis.

4.3.3. *Bacillus* P450-containing BGCs analysis

All 1,562 *Bacillus* genome sequences were analyzed by antiSMASH v5.0 (189) to predict biosynthesis gene clusters (BGCs) and corresponding secondary metabolites. A total of 16,718 BGCs were identified, including 614 P450-containing BGCs. These P450-containing BGCs were further analyzed by BiG-SCAPE v.0.0.0r (293) to construct clusters of networks based on their sequence similarity. The raw distance matrices from the BiG-SCAPE analysis were visualized and annotated using Cytoscape (375). Finally, all 614 P450-containing BGCs were classified into seven types based on the structural and chemical properties of predicted secondary metabolites: non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), a hybrid of PKS and NRPS (PKS-NRPS), ribosomally-synthesized and post-translationally modified peptides (RiPP), terpenes, cyclodipeptide synthases (CDPS), and BGCs outside these six classes (other).

4.3.4. *Bacillus* BGC-affiliated P450 analyses

All 614 BGC-affiliated P450 proteins were further grouped into their families and subfamilies based on the rule of the International P450 Nomenclature Committee (222). All P450 protein sequences were aligned using Clustal Omega (376). A phylogenetic tree for these P450 sequences was constructed using MEGA X (35). Finally, the returned distance matrices of the phylogenetic analysis were visualized and annotated on the iTOL (interactive Tree of Life) webpage (https://itol.embl.de/) (377). Further analysis of amino acid sequence conservation of the five most common BGC-affiliated P450 families was performed using PROMALS3D ((378). According to this analysis, the conservation index of each amino acid position of a given sequence within a P450 family was ranked from 0 to 9. A conservation index of 9 indicates the highest conservation (invariantly conserved), whereas 0 indicates no conservation at all. Further, accounting for all P450 sequences in each family, a cumulative index (0 to 9) for each P450 family was calculated as follows:

$$\frac{\sum(\text{conservation index} \times \text{number of amino acids belonged to that index})}{\text{the average number of amino acids of the given P450 family}}$$

The distribution of the conservation index of each P450 family, the number of amino acids that belonged to that index, and derived cumulative index for each P450 family are provided in Table 4.1.

Table 4.1: Amino acid sequence conservation of 614 BGC-affiliated P450s

P450 Family	Conservation Index										Cumulative Score[1]	Cumulative Index[2]
	9	8	7	6	5	4	3	2	1	0		
CYP113	350	0	1	0	0	0	30	0	0	4	3,247	8.4
CYP134	163	0	87	25	62	3	34	19	10	3	2,698	6.6
CYP102	209	0	10	90	5	67	44	28	6	10	2,978	6.3
CYP109	116	0	4	109	1	129	37	8	1	1	2,375	5.8
CYP107	54	0	42	22	45	51	55	46	91	16	1,689	4.0

[1]**Cumulative Score**= $\sum$ (conservation index × amino acid counts at that index).
[2]**Cumulative Index**= Cumulative Score/total amino acid count.

4.3.5. High-resolution homology modeling of five BGC-affiliated P450s

A representative P450 protein sequence from each BGC-affiliated P450 family was selected to construct a high-resolution homology model using RosettaCM (379). Homology models were constructed using the multi-template comparative modeling protocol implemented in RosettaCM described on the rosettacommons.org webpage (380). Briefly, three templates for each target P450 protein sequence were selected based on the BLASTp search at NCBI by specifying "Protein Data Bank proteins (pdb)" as the Database. This allowed the identification of P450s that have crystal structures available in the protein databank. The crystal structure of the top hits from the BLASTp was further compared for higher resolution and percent identity. The top three hits with high-resolution crystal structures were selected to use as templates. Each P450 target sequence and respective three template sequences were aligned using Clustal Omega (376). The template and target sequences were then converted to Grishin format. The fragment files (9-mers and 3-mers) for all target sequences were generated at the Robetta webpage (http://old.robetta.org/fragmentsubmit.jsp) (381). The membrane region of each target P450 sequence was predicted on the Octopus webpage (https://octopus.cbr.su.se/) (382). The target P450

sequence was threaded on each structure of the templates by using Rosetta's partial_thread program. The heme structure was threaded onto each target P450 structure by using the molfile_to_params.py program from the Rosetta suite of programs. Finally, the homology modeling was carried out by running RosettaCM Hybridize function employing rosetta_cm.xml and rosetta_cm.options by providing the requisite information about the target P450 (heme incorporated) and template structures. Nearly 1,000 structures were generated for each target P450 sequence, and the top three structures with the lowest total energy were selected. Subsequently, the structures were optimized using a relaxation step. The optimized structure with the lowest total energy was selected as a candidate structure. Finally, the stereochemistry and enatic quality for candidate structure were further evaluated and refined at the Structural Analysis and Verification (SAVES) and ProsaII webpage (383). Each P450 refined structure was selected for further analysis.

4.3.6. Molecular docking simulations

Each representative modeled structure of five BGC-affiliated P450 families was selected for molecular docking simulation. The secondary metabolite (or associated intermediate compound) predicted to be produced by each P450-containing BGC was used as a putative substrate for the docking simulation. Cyclo (leu-leu), difficidin, dihydroxybenzoate, bacillaene, and plantazolicin were selected as the putative substrates for CYP134, CYP113, CYP109, CYP107, and CYP102, respectively. The 3D structure of each candidate substrate was downloaded from the ChemSpider database. Each substrate structure was optimized using Gaussian 16 package employing the B3LYP level of theory (387). The harmonic frequencies for the ground state of each compound were calculated using the same level of theory. The geometry optimization was

performed and further verified by the absence of imaginary frequencies. This confirmed that the stationary points of the optimized geometry were corresponded to the global minima. The coordinates of the optimized structure of each compound were selected for molecular docking simulation. For input file preparation, the gaussian output file of each optimized substrate was first converted to pdb file using Open Babel, and then converted to pdbqt file using Auto-Dock-Tools-1.5.6 (388, 389). The pdb file of each P450 model structure was also converted to pdbqt file using Auto-Dock-Tools-1.5.6 (388). The commonly known P450 active site at the heme-distal side was selected as the center of gridbox for substrate binding. Finally, the molecular docking simulation was carried out using AutoDock Vina employing the following parameters: exhaustiveness=400, num_modes=10, energy_range =2 (388).

4.3.7. *Ba*CYP102A2 construct for heterologous expression

In order to evaluate the functional role of CYP102 in plantazolicin biosynthetic pathway, the CYP102A2 gene-encoding DNA sequence from *Bacillus amyloliquefaciens* (*Ba*CYP102A2) was selected and optimized for *E. coli* expression at GenScript (Piscataway, NJ). The optimized sequence was then subcloned into a pET21a (+) expression vector using XbaI and XhoI as restriction sites. The transformants were selected based on ampicillin resistance.

4.3.8. Transformation of the *Ba*CYP102A2 construct

The *Ba*CYP102A2 gene-containing plasmid was transferred into *E. coli* XL-1 Blue competent cell and BL21-[DE3] expression system. An aliquot of 500 µL of *E. coli* cells was thawed on ice. Two µL of plasmid-bearing *Ba*CYP102A2 gene was added into 200 µL of thawed cells, mixed gently, and left on the ice for 20 mins. The mixture was then heat shocked at 42 °C for 120 s in a water bath and returned to the ice immediately for 120 s. 800 µL of pre-warmed LB

medium was added to the cells and then incubated for 40 mins at 37 °C. 300-500 µL of the transformed culture was plated onto freshly prepared pre-warmed LB-agar plates supplemented with appropriate antibiotics (depending on the expression system) and incubated overnight at 37 °C.

4.3.9. Chemicals, buffers, and growth conditions

Ampicillin (AMP), tetracycline, phenylmethylsulfonyl fluoride (PMSF), δ-aminolevulinic acid (δ-ALA), and imidazole were purchased from Sigma-Aldrich (St. Louis, MO). Isopropyl-β-d-thiogalactopyranoside (IPTG) was purchased from Thermo Fisher Scientific. Bugbuster, and benzonase nuclease were purchased from Novagen (Madison, WI). The *E. coli* cells (BL21-[DE3] and XL-1 Blue) were purchased from Agilent Technologies. Nickel-nitrilotriacetic acid (Ni-NTA) resin was purchased from Qiagen (Valencia, CA). Desalting 10DG chromatography columns were purchased from Bio-Rad Laboratories (Hercules, CA). Sodium dodecyl sulfate (SDS), reduced nicotinamide adenine dinucleotide phosphate (NADPH), myristic acid, palmitic acid, and oleic acid were purchased from VWR (Radnor, PA). Buffers and media were prepared using the water purified by a Milli-Q purification system (18.2 MΩ/cm). Sterile LB-agar plate was prepared using 25 g/L of LB and 15g/L of agar. The appropriate antibiotic was added once the autoclaved LB and agar medium cooled below 50 °C. Approximately 15 ml of the medium was transferred onto the petri dish and left to solidify in an aseptic condition.

4.3.10. *Ba*CYP102A2 enzyme expression

*Ba*CYP102A2 expression was carried out in *E. coli* BL21-[DE3] expression system using LB broth supplemented with ampicillin (100 µg/mL) and tetracycline (10 µg/mL). Cells were grown at 37 °C with constant agitation (220 rpm) to the mid-log phase (OD_{600}= 0.6-0.8) and then

induced by the addition of 1 mM IPTG. The temperature was lowered to 28 °C after 20 mins of IPTG addition, and the culture continued for 12h post-induction. The cells were harvested by centrifugation at 6,000 × g for 20 mins at 4 °C. The supernatant was discarded, and the cell pellet was stored at -80 °C until use. The expression analysis was carried out using the trichloroacetic acid (TCA) precipitation protocol as described elsewhere (390). A solubility test was carried out to determine the extent of the solubility of desired *Ba*CYP102A2 protein.

4.3.11. *Ba*CYP102A2 enzyme purification

The cell pellet was thawed on ice and resuspended in buffer A (50 mM KPi, pH 7.0). In the presence of 0.1 mM PMSF, the cell suspension was homogenized using a bead-beater. The homogenized cells were then lysed using a Branson Sonifier 250 (Branson Ultrasonics, Corp., Danbury, CT, USA). Sonication was performed with an output control of 3.5 and a duty cycle of 30% for 8 intervals of 42 s with cooling on ice between each sonication. After sonication, cell lysate was incubated for 2 hours at 4 °C with gentle stirring in the presence of benzonase nuclease (250 U). The cell lysate was then centrifuged at 7,500 × g for 35 mins at 4 °C. Pellets were discarded, and the supernatant was collected and transferred into 50 mL Falcon tubes containing pre-washed Ni-NTA resin. The mixture was incubated overnight at 4°C with constant rocking and shaking. The mixture was then loaded into a column and washed with buffer B (50 mM KPi, pH 7.5, 50 mM NaCl). The desired protein was eluted by successive addition of buffer B containing imidazole of 2 mM, 5 mM, 7.5 mM, 10 mM, 50 mM, and 200 mM. The eluted fractions were collected and analyzed using SDS-PAGE. The fractions containing the desired enzyme were pooled and concentrated by ultrafiltration using a 30 kDa molecular cutoff filter and subjected to buffer exchange with buffer A. To remove any fatty acid contaminants, extensive dialysis was

carried out against buffer A at 4 °C for 24 hours. The proteins were then run through a desalting DG10 column for further buffer exchange. Finally, the proteins were purified using an anion-exchange (AEC) column. Fractions collected from the AEC containing the desired protein were pooled, concentrated, and aliquoted as 250 µL stock into 1.5 mL Eppendorf tubes and stored at -80 °C until use.

4.3.12. SDS-PAGE analysis

SDS-PAGE was used to evaluate the purity and relative quantity of the *Ba*CYP102A2 protein based on the size. The protein samples were made up to 15 µL with the sample buffer (4% SDS, 10% β-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 125 mM Tris HCl at pH 6.8), 5 µL protein sample, and H_2O. A broad range pre-stained protein ladder (Bio-Rad, California, USA) was used as a reference. The mixture was boiled for 10 mins at 95 ºC and loaded onto a 7% acrylamide SDS-PAGE gel. The gel was run using the running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) with a voltage of 300 V. The gel was then stained using the staining buffer, followed by de-staining with 7.5% glacial acetic acid. Subsequently, the gel was dried overnight at ambient temperature.

4.3.13. UV-vis spectroscopic analysis

UV-vis absorption spectra of *Ba*CYP102A2 were recorded in a glass cuvette using a UV-vis spectrophotometer (Shimadzu, UV–1601). The absorption spectra were recorded in the range of 200-700 nm for characteristic P450 Soret band and charge-transfer bands of *Ba*CYP102A2 protein. The protein concentration was determined on the basis of the Soret band (419.5 nm), using the estimated absorption coefficient of full-length $P450_{BM3}$ protein ($\varepsilon_{419.5}$= 105 $mM^{-1}cm^{-1}$) as

described elsewhere (391). The purity (Reinheitzahl, Rz) of the protein was determined by comparing the absorbance of the heme Soret band at 419.5 nm and the protein peak at 280 nm.

4.3.14. Substrate-binding titration

Substrate binding affinity of the *Ba*CYP102A2 enzyme was determined for the P450 substrate sodium dodecyl substrate (SDS) and oleic acid (OA). 1 μM of *Ba*CYP102A2 enzyme and 0 to 500 μM of substrate (depending on the substrate's saturation concentration) were used for the assay. The mixture of *Ba*CYP102A2 and a given substrate was incubated at room temperature for 30 mins. Finally, the UV-vis absorption spectra were recorded for the substrate-bound *Ba*CYP102A2 enzyme in the range of 200-700 nm using UV-vis spectrophotometer (Shimadzu, UV–1601).

4.3.15. Steady-state kinetic analysis

Steady-state kinetic analyses were performed on the *Ba*CYP102A2 enzyme to estimate the catalytic rates of NADPH-dependent reactions with typical P450 substrates. 1 mL of sample mixture was prepared using buffer C (50mM KPi, 50 mM KCl, pH 7.20) with the following concentrations: 100 nM *Ba*CYP102A2, 120 μM NADPH, and 0-400 μM of substrates (depending on substrate saturation concentration). After mixing all chemicals into a 1 ml cuvette, the cuvette was promptly transferred into the spectrophotometer. The decrease of absorbance of NADPH at 340 nm due to the oxidation of substrate was recorded for 120 s. At least three traces were recorded for each data point.

4.4. Results and discussion

4.4.1. Diversity of BGC-affiliated cytochromes P450 in *Bacillus* species

Our genome mining effort utilizing sequence from 1,562 *Bacillus* strains identified a total of 5,051 P450 proteins. Annotation of these P450s using the "Cytochromes P450 Homepage" database (374) showed that all of them were distributed across 12 families: CYP107 (1845), CYP102 (1487), CYP109 (749), CYP106 (624), CYP134 (170), CYP113 (101), CYP197 (61), CYP152 (7), CYP1007 (4), CYP120 (1), CYP205 (1) and CYP267 (1) (Fig. 4.1). Interestingly, the majority of *Bacillus* P450s (93%) were distributed across only 4 families: CYP107, CYP102, CYP109, and CYP106. This indicates that P450 proteins are highly conserved within the *Bacillus* species. From the 5,051 P450s, 614 genes were encoded by secondary metabolite-producing biosynthetic gene clusters (BGCs) and annotated as “accessory genes.” These BGC-affiliated P450 families were CYP107 (326), CYP134 (142), CYP113 (84), CYP109 (33), CYP102 (27), and CYP106 (2). The distribution of these BGC-affiliated P450s as compared to all P450s is shown in Fig. 4.1. All 614 P450-containing BGCs were further grouped into seven classes based on the chemical and structural properties of associated secondary metabolites. These were NRPS, PKS, the hybrid PKS-NRPS, RiPPs, CDPS, terpenes, and BGCs outside these six classes (other). According to this classification, most genes from the CYP107 family were encoded in NRPS and PK-NRPS BGCs, whereas all genes from CYP134 and CYP113 families were exclusively found in CDPS and PKS BGCs, respectively (Fig. 4.2). Enzymes from the CYP109 family were observed in both NRPS and terpene BGCs, whereas most genes from the CYP102 family were encoded in RiPP BGCs. The remaining two P450s from the CYP106 family were associated with PKS BGCs.

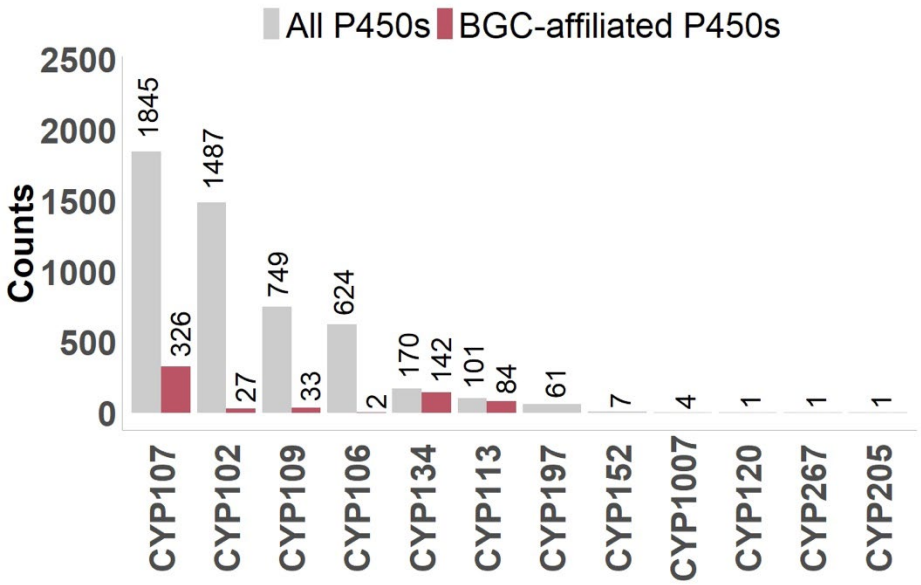


Figure 4.1: Distribution of BGC-affiliated P450s in *Bacillus* species. The distribution of 614 BGC-affiliated P450s as compared to all 5,051 P450s identified in 1,562 *Bacillus* strains. All P454s are distributed across twelve families, while the BGC-affiliated P450s are distributed within only six families.

In order to evaluate the extent of amino acid sequence conservation across the protein sequences in each BGC-affiliated P450 family, a conservation index (0 to 9) was calculated (see *Materials and Methods*); a score of 9 for a give position indicates invariably conserved, whereas 0 indicates no conservation at all. Based on this score, CYP113 (8.4) was the most conserved P450 family, whereas the CYP107 (4.0) was the least conserved (Fig. 4.3). Interestingly, enzymes from the highly conserved CYP113 family were exclusively observed in CDPS BGCs while enzymes from the CYP134 family were only found in PKS BGCs. As a contrast, enzymes from the least conserved CYP107 family were encoded within NRPS and PK-NRPS BGCs, while the enzymes from the CYP109 family were encoded within NRPS and terpene BGCs.

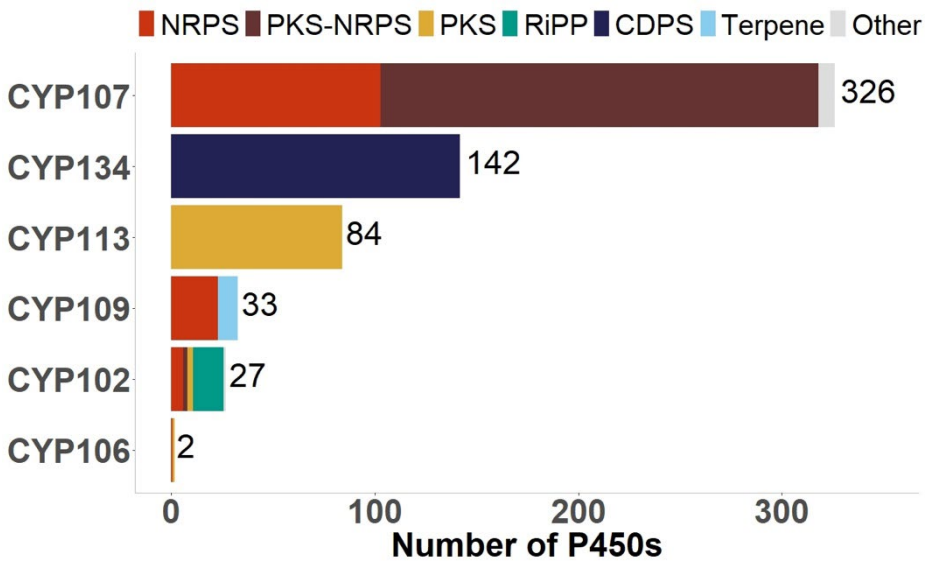


Figure 4.2: Distribution of each BGC-affiliated P450 family among various BGC classes. All P450-containing BGCs are grouped into seven classes based on the chemical and structural properties of predicted secondary metabolites. Each BGC is shown with distinct color in the barplot matching with the associated P450 family (see *Materials and Methods).*

It is notable that highest CYP family conservation (CYP113 and CYP134) corresponded to P450s found exclusively in one type of BGC, and those least conserved (CYP107 and CYP109) were identified in at least two types of BGCs. This suggests that the former have been tuned to serve very specialized roles in secondary metabolite biosynthesis, perhaps acting on a very narrow substrate profile. On the other hand, those least conserved P450 families may fill more generalized roles acting, for example, across multiple substrate types and metabolic pathways. In summary, these data indicate that the level of amino acid sequence conservation across P450 families is linked with substrate specificity associated with respective secondary metabolite biosynthetic pathways.

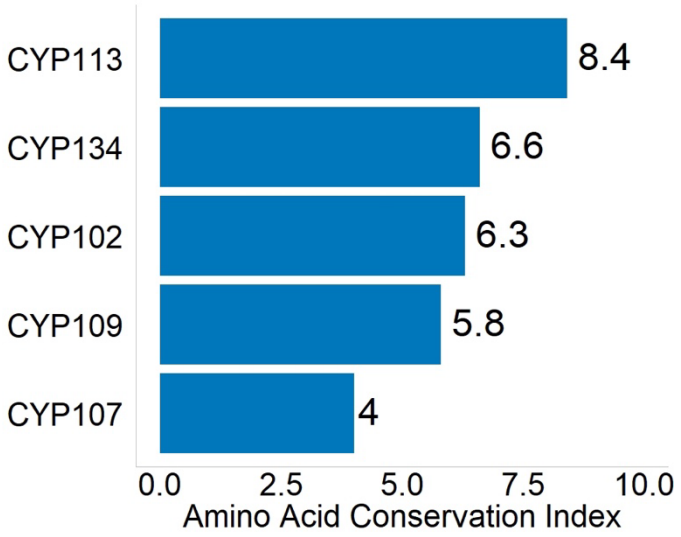


Figure 4.3: Amino acid sequence conservation (score from 0 to 9) of each BGC-affiliated P450 family; score of 9 indicates invariantly conserved, and 0 indicates no conservation at all (see *Materials and Methods*).

In order to visualize whether all proteins from a P450 family may evolve from a common ancestor and may associate with a specific BGC type, a phylogenetic tree was constructed for all BGC-affiliated P450s (Fig. 4.4). All enzymes from each P450 family (see Fig. 4.4 inner circle) were positioned in close proximity to one another, indicating they are phylogenetically more closely related. This also confirms that all protein sequences within a P450 family are highly similar in amino acid sequence. In order to compare how each P450 may associate with the respective BGC type, each BGC was shown as the outer circle next to its associated P450 (inner circle). Interestingly, enzymes from the highly conserved CYP113 family were aligned exclusively with the difficidin PKS BGC, whereas CYP134 aligned with CDPS (cyclodipeptide) BGC. Notably, the CYP113 family is situated next to the CYP109 family in the phylogenetic tree, where both occupy a common clade, indicating that they share more similar sequences and structures, and perhaps, metabolic function. Conversely, the CYP134 family originated from a separate clade

and further divided into three sub-clades. This observation suggests that there may be three types of structurally distinct enzymes within the CYP113 family.

The least conserved CYP107 family occupies two proximally located clades and aligns with two BGCs, a PKS-NRPS (bacillaene) and an NRPS (fengycin). Interestingly, P450s that are linked with bacillaene are clearly separated from those associated with the fengycin BGC within the CYP107 family, indicating each subgroup is likely different from one another. The CYP109 family occupies two sub-clades, where one subgroup is found with bacillibactin (NRPS) BGCs and the other is with a terpene BGC. It is reasonable to suppose that these are also distinct from one another in structure and function. Finally, the enzymes from the CYP102 family occupy a single clade and are aligned primarily with plantazolicin (a RiPP BGC) biosynthesis with a few exceptions. In summary, these data suggest that amino acid sequence conservation of each P450 family is phylogenetically related and associated with a specific biosynthetic pathway. Further, the degree of sequence conservation of each P450 family appears to correlate with substrate specificity linked with their respective biosynthetic pathway(s).

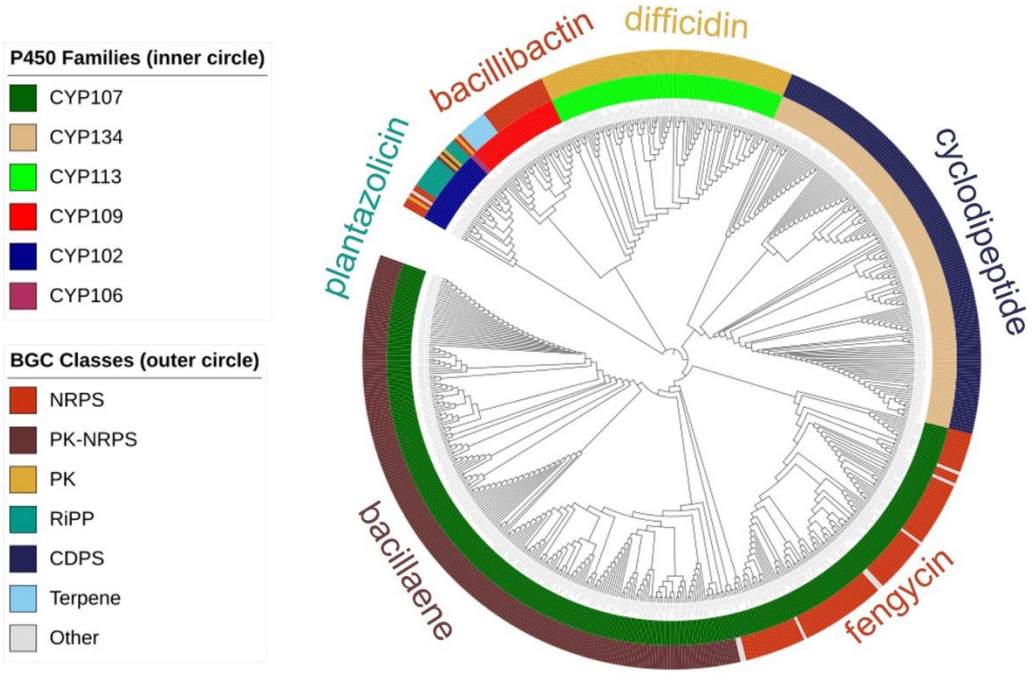


Figure 4.4: Phylogenetic tree showing the distribution of BGC-affiliated P450s among P450 families and BGC types, where families are shown in the inner circle, and BGC types are shown in the outer circle. Prominent secondary metabolites are labeled next to the corresponding BGC.

4.4.2. Each BGC-affiliated P450 family is involved in a specific secondary metabolite biosynthetic pathway

In order to visualize the association of P450 families with respective secondary metabolite biosynthetic pathways, similarity analysis of P450-containing BGCs was carried out (BiG-SCAPE v.0.0.0r) (293). In this analysis, most of the BGCs were distributed into several distinct clusters. Note that all BGCs within a cluster are highly similar and they are likely involved in a common secondary metabolite biosynthetic pathway. In order to highlight the P450 family distribution among various BGC clusters, each BGC was colored based on the distinct P450 family (node color). Interestingly, as shown in Figure 4.5A, each large cluster of BGCs appeared to integrate a specific P450 family. The CYP107 family primarily associated with BGCs for bacillaene and

fengycin biosynthesis. Most CYP109 members occupied bacillibactin BGCs, but there were a handful distributed in unknown BGCs. With the exception of one member, the difficidin cluster was densely packed, and it exclusively contained P450s from the CYP113 family. Similarly, CYP113s were not observed to occupy any other BGC. Though the CYP134 family only observed in CDPS BGCs, these divided into three distinct clusters, one of whose members appeared to be rather loosely associated and contained one member with an alternate P450 gene (a CYP102 member). These data suggest that there is substantial variation of CYP134-containing CDPS BGCs. The CYP102 family was primarily represented in the plantazolicin cluster, with some exceptions occupying an otherwise unknown BGC as well as a handful of pairs and singletons. The two CYP106-containing BGCs appeared as singletons. This clustering analysis clearly shows that proteins from each P450s family are linked with specific biosynthetic pathways, suggesting the conservation of functional roles in their respective biosynthetic assignments. Further, this visualization supports our hypothesis that relatively highly conserved CYP134 and CYP113 families are exclusively associated with only one type of biosynthetic pathway, and perhaps, a single substrate and specific transformation. The least conserved P450 families, CYP107 and CYP109, are associated with more than one biosynthetic pathway. The gene organization of five P450-containing BGCs is shown in Fig. 4.5B, where the relative position of P450 in the BGC is highlighted by an upward (black) arrow.

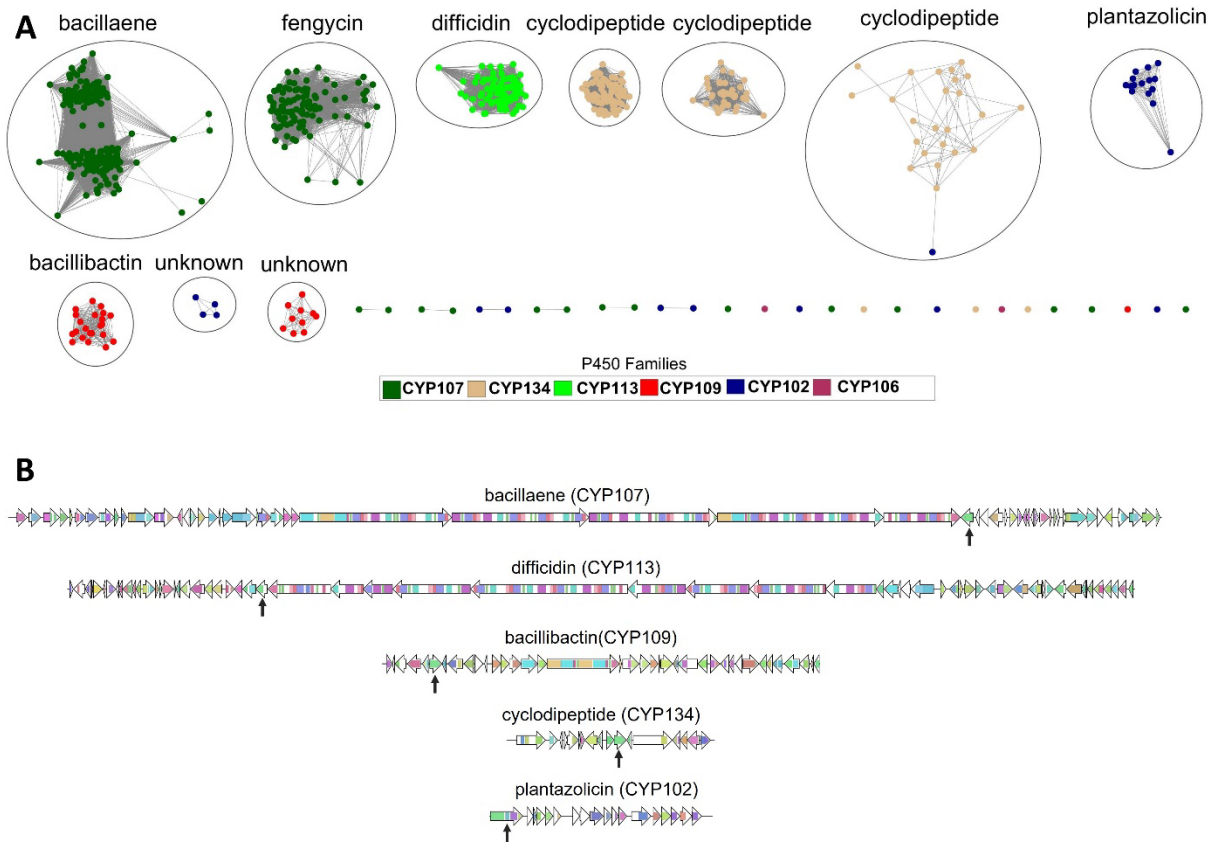


Figure 4.5: Network analysis and visualization of 614 P450-containing BGCs. Out of 614 BGCs, 583 are distributed among ten clusters (A). Each node represents a single BGC. The nodes are colored according to the family of the P450 gene which occupies the BGC. With the exception of pairs and singletons, clusters of BGCs are labeled according to the secondary metabolites they are predicted to produce. The gene organization of five P450-containing BGCs is shown (B). The location of the P450 gene within each BGC is indicated by the black arrow.

4.4.3. High-resolution homology models of representative BGC-affiliated P450s showed distinct structural features

In order to compare the structures of the five most common BGC-affiliated P450 families, high-resolution homology models were constructed using the RosettaCM suite of programs (379). Comparison of structures showed some obvious differences in the relative organization of secondary and tertiary structures, especially α-helices within the tertiary structure among five P450

families (Fig. 4.6). The structure of CYP107 appeared as relatively broad active site pocket where the B/C loop-containing an extended B' helix was in close proximity to the F/G loop. In contrast for CYP134, the B/C loop is longer, and the F/G loop is inclined toward the active site. The secondary structure near the active site of CYP113 is very different from either CYP107 or CYP134. Here, the B´ helix orients orthogonally to the heme plane and well above the prosthetic group's distal face. The F/G loop protrudes out from the active site. The structure of CYP109 is similar to CYP134, however, the F/G loop is more inclined toward the active site and extends toward the A/B loop. The structure of CYP102 has a broad opening near the A/B loop, where the F/G loop protrudes out from the active site. Based on these observations, we surmised that the observed structural difference may be linked with substrate preference for each P450 family in relation to the respective secondary metabolite biosynthesis.

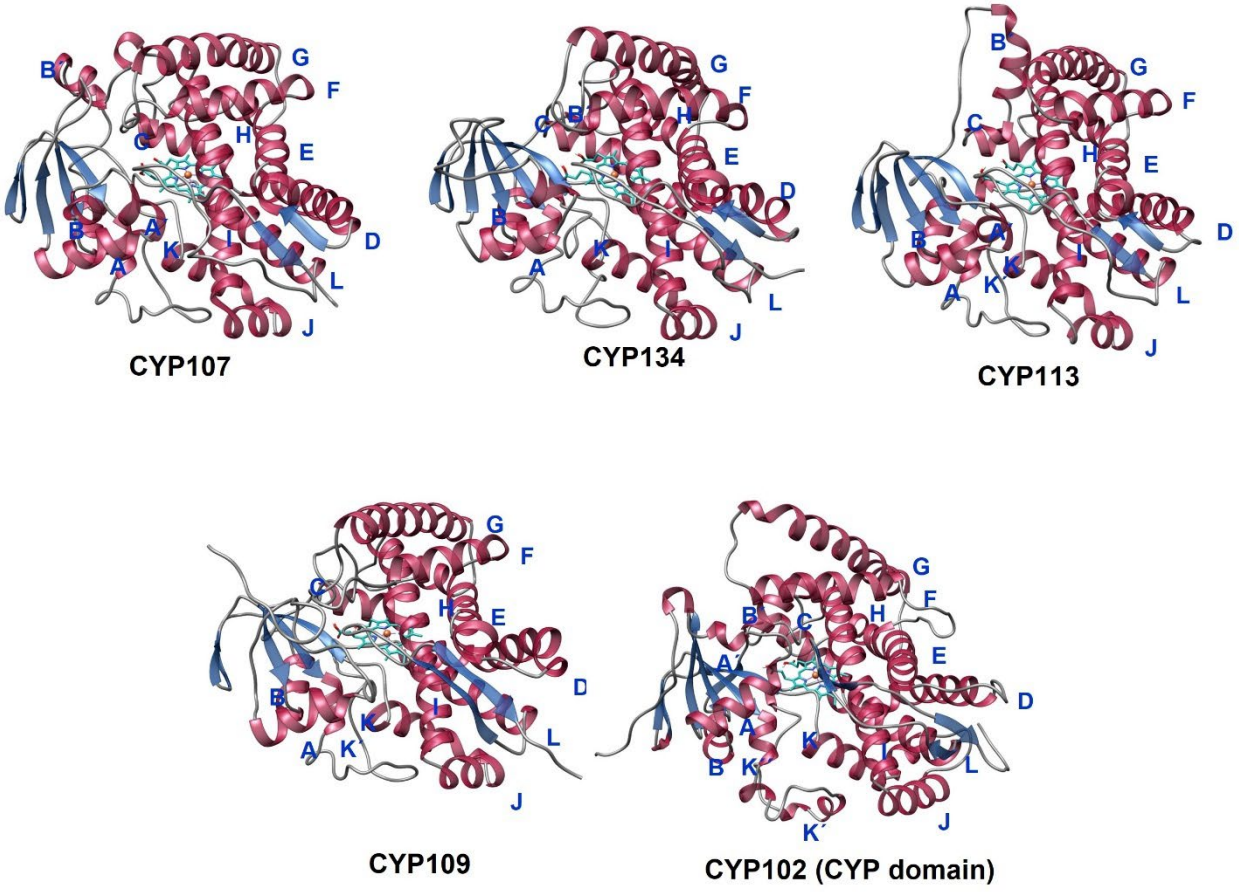
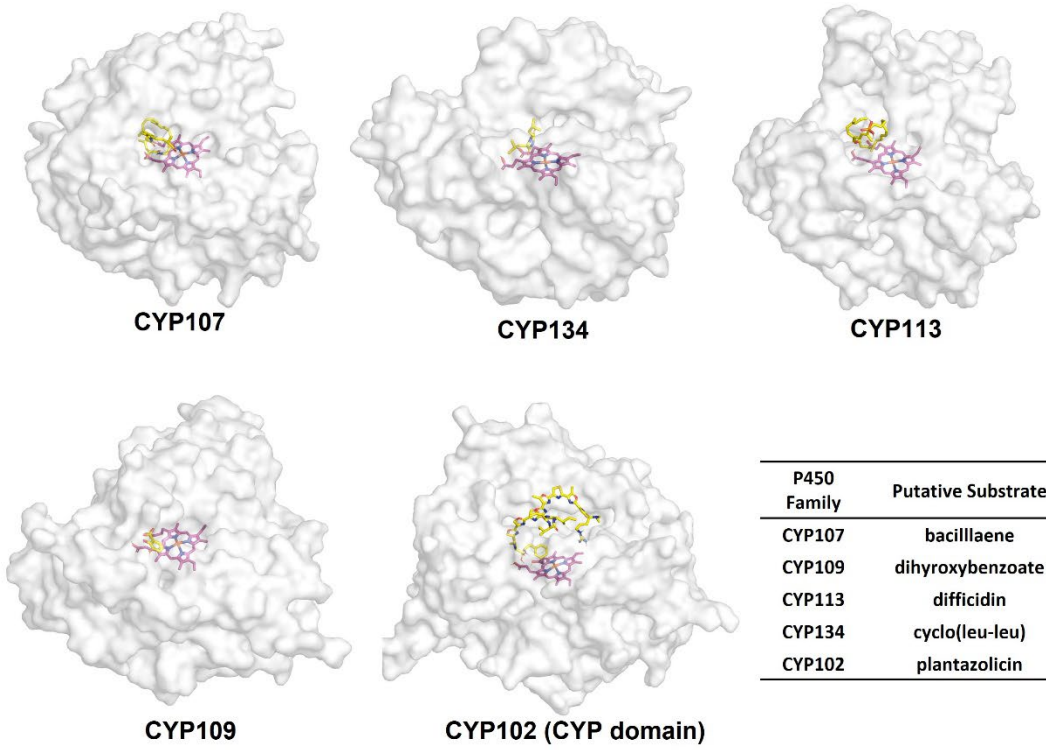


Figure 4.6: Homology models of five BGC-affiliated P450s. Homology model showing the tertiary structure of representative P450 protein from five BGC-affiliated P450 families. Each model structure is selected as the best model with the lowest total energy from nearly 1,000 structures generated for each P450 family using the high-resolution RosettaCM suite of programs (379).

4.4.4. Favorable binding of BGC-affiliated P450s with candidate substrates

In order to evaluate how each BGC-affiliated P450 enzyme interacts with the respective candidate substrate, molecular docking simulations were performed using the P450 homology model and candidate substrate associated with the secondary metabolite biosynthetic pathway. In docking analyses, favorable substrate binding with negative binding energy was observed for each P450 enzyme (Fig. 4.7). Bacillaene is positioned at the active site of CYP107 with a binding energy

of -6.3 kcal/mol. Interestingly, this binding pose is in agreement with the previously proposed position (14′-15') of bacillaene/dihydrobacillaene oxygenation by a P450 enzyme, PKsS (362). The docking pose of CYP134 showed a favorable binding interaction (-7.0 kcal/mol) with a cyclodipeptide (cyclo-leu-leu) at the active site of CYP134. This pose is consistent with the reported site of cyclodipeptide oxygenation. Cryle *et al.* previously reported that a CYP134 homolog (CypX) involves in the oxidative transformation of cyclo-leu-leu to pulcherriminic acid (261). A docking pose of difficidin at the active site of CYP113 showed a favorable binding pose (-6.0 kcal/mol). This particular pose supports the previously indicated hydroxylation position of difficidin, to produce oxydifficidin (338). Molecular docking simulation of CYP109 with a candidate substrate, dihydroxybenzoate (an intermediate of bacillibactin biosynthesis) showed favorable enzyme-substrate interaction (-5.3 kcal/mol). Notably, bacillibactin is a siderophore that is biosynthesized by an NRPS BGC through three repeating units of dihydroxybenzoate (164). It is possible that the dihydroxybenzoate may be further transformed to trihydroxybenzoate by the CYP109 enzyme to produce a novel bacillibactin derivative. At present, several homologs of CYP109 have been reported to be involved in the oxidative transformation of fatty acids, steroids, and α/β-ionone. Due to its broad substrate tolerance, the CYP109 is commonly referred to as the "versatile enzyme" (272, 392).



P450 Family	Putative Substrate	Binding Energy (kcal/mol)
CYP107	bacilllaene	-6.3
CYP109	dihyroxybenzoate	-5.3
CYP113	difficidin	-6.0
CYP134	cyclo(leu-leu)	-7.0
CYP102	plantazolicin	-6.6

Figure 4.7: Favorable docked pose of a putative substrate in each BGC-affiliated P450 model. Secondary metabolite or precursor compound associated with the respective biosynthesis pathway is selected as a putative substrate for the docking analysis. bacillaene, cyclo-(leu-leu), difficidin, dihydroxybenzoate, and plantazolicin are used as a putative substrate for CYP107, CYP134, CYP113, CYP109, and CYP102, respectively. The table on the lower right shows estimated binding energy of each P450 with the respective candidate substrate.

The docking simulation of a representative CYP102 (from *B. amyloliquefaciens*) (*Ba*CYP102A2) with candidate substrate plantazolicin showed favorable enzyme-substrate interactions (-6.6 kcal/mol), and these placed the molecule in the enzyme's active site adjacent to the heme prosthetic group. In this pose, the C-terminal of plantazolicin is positioned close to the heme iron, suggesting this may be the site of post-translational modification of plantazolicin, as reported by a previous study (393). However, they demonstrated that the azole ring of Thr in this site (pointed by a black arrow in Fig. 4.8) undergoes hydrolytic ring opening by the addition of mild acid. Based on our data, we anticipate this transformation could be mediated by the

*Ba*CYP102A2 enzyme, in addition to the comparatively slow chemical hydrolysis as they reported. A similar ring-opening reaction has been reported to be catalyzed by a P450 enzyme CYP1A2 in setileuton, a 5-lipoxygenase inhibitor (394).

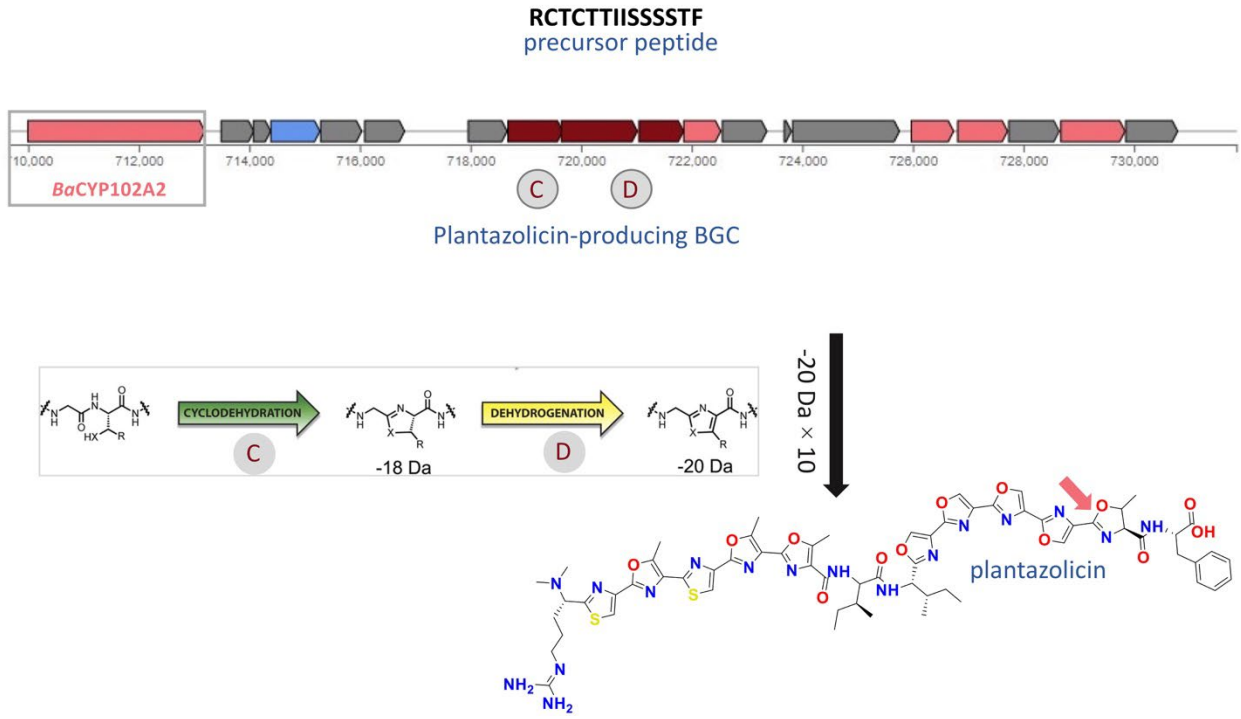


Figure 4.8: Putative role of *Ba*CYP102A2 in plantazolicin biosynthesis. Plantazolicin producing BGC containing *Ba*CYP102A2, cyclodehydratase (tagged as (C), and dehydrogenase (tagged as (D)) enzymes modify the precursor peptide post-translationally to produce plantazolicin. The black arrow pointing to the plantazolicin structure indicates the putative site of modification by *Ba*CYP102A2.

4.4.5. Expression, purification, and identification of *Ba*CYP102A2

To take an additional step toward discerning the catalytic role of *Ba*CYP102A2, the gene encoding the *Ba*CYP102A2 enzyme was synthesized, cloned into a pET21 vector, and used to transform *E. coli* (BL21-[DE3]) for expression of the target enzyme. The expressed enzyme was purified using Ni-NTA affinity, anion-exchange, and size-exclusion chromatographic techniques

followed by extensive dialysis. The identity of the full-length *Ba*CYP102A2 enzyme was confirmed by the presence of a major band at ~120 kDa in SDS-PAGE gel and a heme-based Soret band at 418.5 nm in the UV-vis absorption spectrum. The SDS-PAGE and UV-vis absorbance spectra for the oxidized and reduced form of *Ba*CYP102A2 are shown in Fig. 4.9.

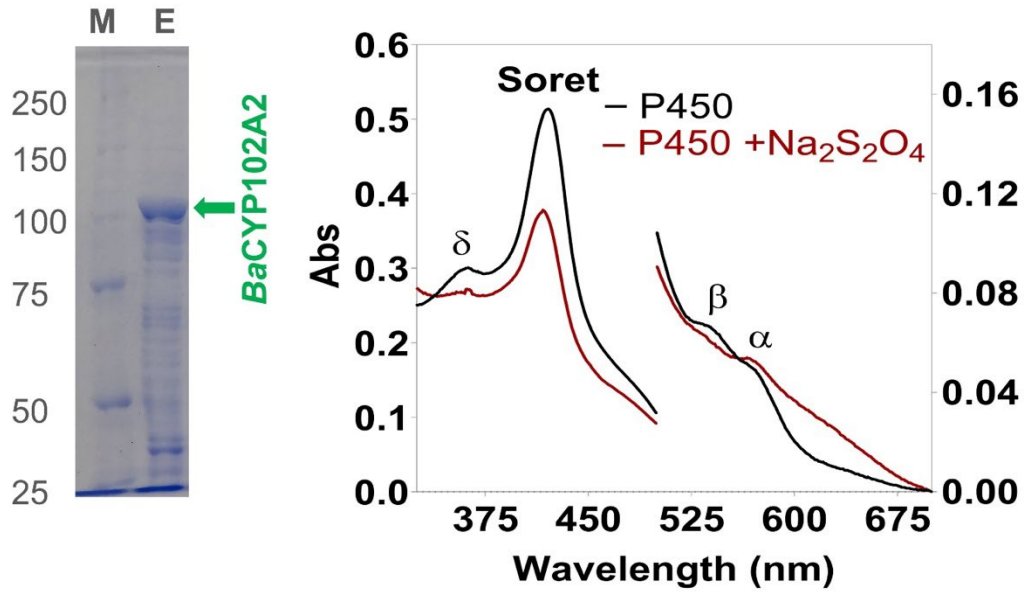


Figure 4.9: SDS-PAGE and UV-vis absorbance of *Ba*CYP102A2. (A) SDS-PAGE for a marker (left lane, M), and *Ba*CYP102A2 (right lane, E) highlighting the band of full-length *Ba*CYP102A2 at ~120 kDa, (B) The characteristic absorption features of *Ba*CYP102A2 in its oxidized and reduced form showing Soret band as well as charge transfer bands.

4.4.6. Substrate binding and NADPH oxidation activity of *Ba*CYP102A2

The CYP102A2 enzyme was subjected to substrate binding titration to observe a typical spectral shift due to the change of spin state of iron of the heme prosthetic group. Type-I spectral shift was observed upon binding of SDS (sodium dodecyl sulfate) and OA (oleic acid) with *Ba*CYP102A2. A type-I spectral shift is indicative of a shift in the spin state of the heme's Fe^{III} ion from low- to high-spin, and is commonly observed in cytochromes P450 upon substrate binding

(395). Importantly, both SDS and OA are known substrates for CYP102. Titration of *Ba*CYP102A2 with SDS produced a sigmoidal response, whereas OA produced a hyperbolic response (Fig. 4.10). The $K_{1/2}$ and K_D estimated by spectral fittings were 174 ± 35 μM and 31 ± 13 μM for SDS and OA, respectively.

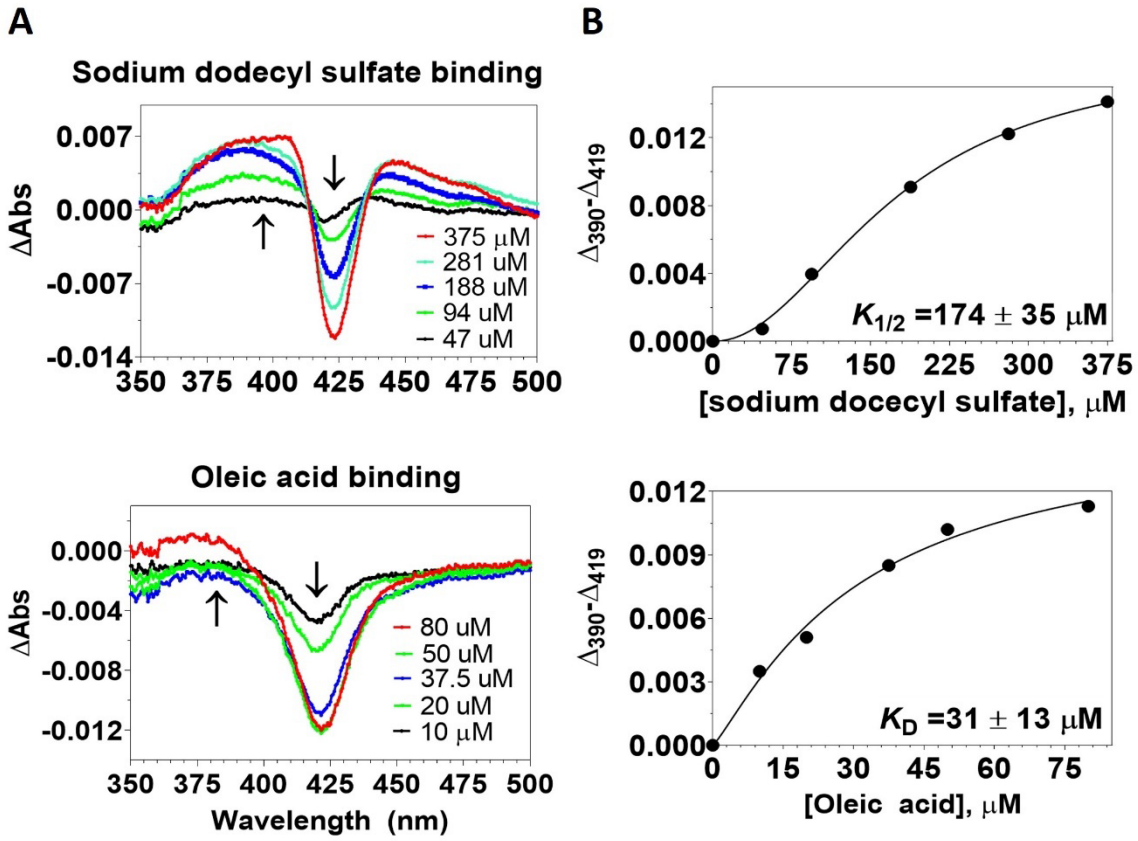


Figure 4.10: UV-vis absorption difference spectra and substrate binding spectral fitting of full-length *Ba*CYP102A2. (A) The absorption difference spectra of *Ba*CYP102A2 for sodium dodecyl substrate (SDS) and oleic acid (OA). (B) The spectral fitting of absorbance difference of *Ba*CYP102A2 for SDS and OA, estimating their $K_{1/2}$ = 174 ± 35 μM and K_D = 31 ± 13 μM, respectively.

Furthermore, oxidation of NADPH by *Ba*CYP102A2 was carried out in the presence of SDS, myristic acid (MA), palmitic acid (PA), and OA. Kinetic fitting of the substrate-dependent

response of SDS, MA, and PA produced a sigmoid response with respective turnover numbers (k_{cat}) of 86 ± 4 min^{-1}, 38 ± 2 min^{-1}, and 19 ± 1 min^{-1} (Fig. 4.11). In contrast, kinetic fitting of OA produced a hyperbolic response with a turnover number of 45 ± 2 min^{-1}. In agreement with the substrate binding titration of SDS and OA, the estimated $K_{1/2}$ and K_M for SDS, MA, PA, and OA were 168 ± 9 μM, 105 ± 3 μM, 35 ± 1 μM, and 23 ± 3 μM, respectively. Overall, these data suggested that the *Ba*CYP102A2 enzyme has a substrate preference for long-chain unsaturated fatty acid, as reported previously for a homolog of the CYP102A2 enzyme (369).

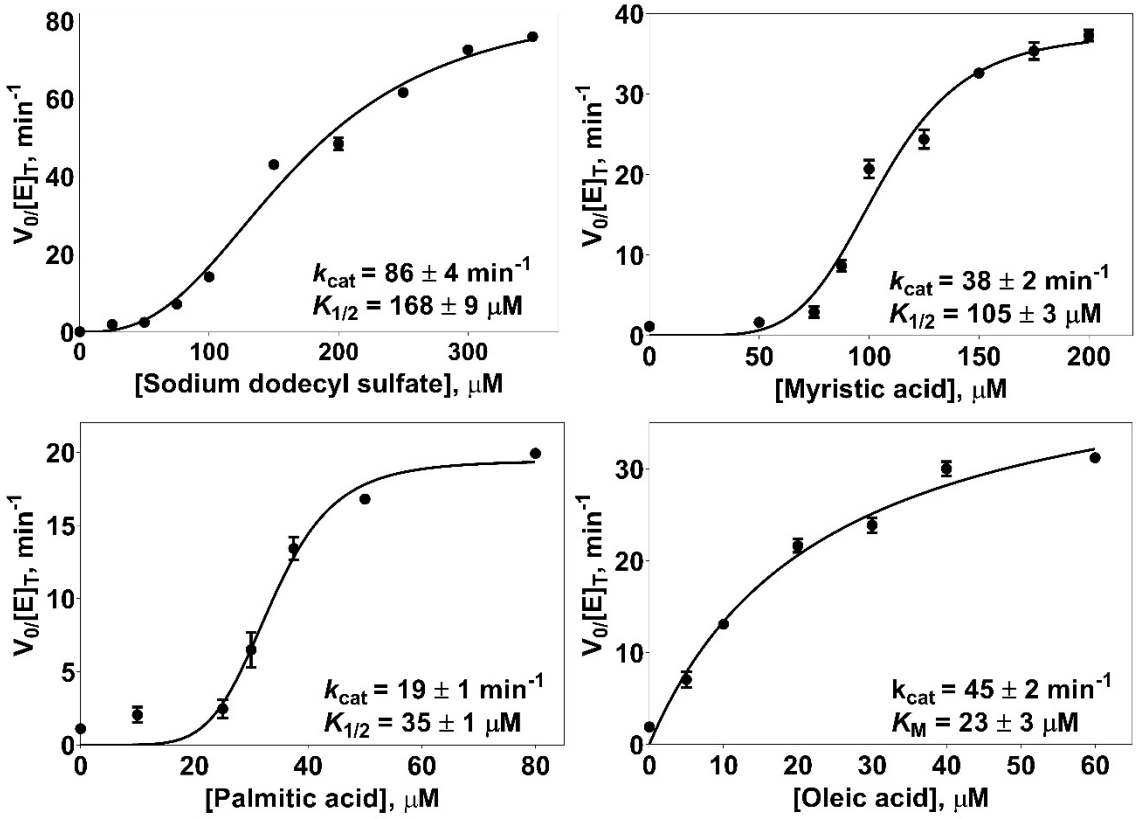


Figure 4.11: The steady-state kinetic analysis. Kinetic fitting of NADPH oxidation by *Ba*CYP102A2 in the presence of sodium dodecyl sulfate (SDS), myristic acid (MA), palmitic acid (PA), and oleic acid (OA), estimating their turnover numbers (k_{cat}) 86 ± 4 min^{-1}, 38 ± 2 min^{-1}, 19

±1 min^{-1}, and 45 ± 2 min^{-1}, and their $K_{1/2}$ = 168 ± 9 μM, 105 ± 3 μM, 35 ±1 μM, and K_M = 23 ± 3 μM, respectively.

In summary, we identified 614 BGC-affiliated P450s distributing across six P450 families in 1,562 *Bacillus* genomes. Comparison of amino acid sequence conservations within each P450 family showed a broad range of sequence conservation, in the order from highest to lowest: CYP113 > CYP134 > CYP102 > CYP109 > CYP107. Enzymes from each P450 family were involved in specific secondary metabolite biosynthetic pathways. The most conserved families (CYP113 and CYP134) were involved in only one biosynthetic pathway, whereas the least conserved families (CYP109 and YCP107) were involved in at least two pathways. Phylogenetic analyses showed that enzymes within each P450 family were highly similar in amino acid sequences and occupied common clade in the phylogenetic tree. High-resolution homology models of representative enzymes from five BGC-affiliated P450s showed distinct structural features at the P450 active site. Molecular docking simulations of the representative enzymes showed favorable binding of CYP113, CYP134, CYP102, CYP109, and CYP107 with the candidate substrates difficidin, cyclodipeptide, plantazolicin, dihydroxybenzoate, and bacillaene. A representative *Ba*CYP102A2 enzyme involved in plantazolicin biosynthesis was heterologuesly expressed and evaluated for characteristic UV-vis absorbance, substrate binding, and steady-state kinetic responses. Substrate binding of *Ba*CYP102A2 with SDS produced a sigmoidal response and with OA produced a hyperbolic response. Steady-state kinetic evaluation of *Ba*CYP102A2 with SDS, MA, and PA produced a sigmoidal response, whereas OA produced a hyperbolic response.

Chapter 5: Summary and future work

5.1. Summary

In the present study, a total of 288 novel PGPR strains across 17 *Bacillus* species were evaluated for their potential as commercially viable biological agents against root-associated plant pathogens. A comprehensive genomic analysis was carried out to predict their biosynthetic gene clusters (BGCs) and secondary metabolites. Concurrently, antibiosis screening of each *Bacillus* strain was carried out against the plant-pathogenic oomycete, *P. nicotianae* to identify candidate *Bacillus* strains with potent biocontrol capacity. The most promising 59 *Bacillus* strains that exhibited strong antibiosis activity against *P. nicotianae* were evaluated against three root-associated plant-pathogenic fungi: *F. oxysporum*, *F. graminearum*, and *R. solani*. Further, secondary metabolites from the culture of a representative strain of five *P. nicotianae*-inhibitory *Bacillus* species were extracted using a semi-targeted acid-methanolic extraction procedure. Chemical characteristic of each extract was performed using UV-vis absorption, LC-elution profile, and high-resolution LC-MS and MS^2 analysis. Bacillomycin L (an iturin), fengycin, and surfactin were purified using a semi-preparative HPLC system and evaluated for antibiosis activity against oomycete/fungi using plate-based and 96-well microtiter-based antibiosis assay. In addition, 18 strains from the most promising *Bacillus* species*, B. velezensis*, were evaluated for quantitative antifungal activity and intraspecies inhibition and resistance activity in order to identify candidate strains with potent antibiotic activity. Furthermore, to study the variation of secondary metabolites scaffolds in relation to one of the commonly observed accessory enzymes, cytochromes P450, 1,562 *Bacillus* genomes were analyzed to identify BGC-affiliated P450s. BGC-affiliated P450s were analyzed for amino acid sequence conservation and structural and

functional features linked with respective biosynthetic pathways. For a representative enzyme from five BGC-affiliated P450 families, a high-resolution homology model was constructed, and molecular docking simulation was performed with candidate substrates linked with respective biosynthetic pathway. Finally, a representative P450 (*Ba*CYP102A2) predicted to involve in plantazolicin biosynthesis was heterologously expressed, purified, and evaluated for substrate binding and steady-state kinetic responses using known P450 substrates.

5.1.1. Broad biological activity of *B. velezensis* and *B. subtilis* is accounted for by conserved ability for secondary metabolite biosynthesis

Our genomic analysis of 288 strains from 17 *Bacillus* species identified a total of 2,442 BGCs (8.5 per species on average), including 1,259 known and 1,183 unknown BGCs. A comparison of BGCs among different species showed a substantial variation in numbers and types of BGCs among different *Bacillus* species. Species containing relatively higher number of BGCs per strain were *B. velezensis*, *B. pumilus*, *B. safensis*, *B. subtilis,* and *B. altitudinis*. *B. velezensis* had the highest number of BGCs per strain (13.4 on average), whereas *B. firmus* had the lowest number of BGCs per strain (4.2 on average). Our antibiosis screening of all 288 *Bacillus* strains against *P. nicotianae* identified a total of 100 (~35%) strains with promising antagonistic activity, where 59 (~21%) and 41 (~14%) strains exhibited strong and weak inhibition, respectively. Interestingly, 56 out of the 59 strongly inhibitory *Bacillus* strains were distributed within only five species: *B. pumilus*, *B. safensis*, *B. altitudinis*, *B. subtilis*, and *B. velezensis*. Strikingly, these five species are phylogenetically closely related than the other species. This result indicated that the observed antibiosis phenotype among these five species may be linked with the expression of phylogenetically-related conserved factors. *Bacillus* species were further classified into strongly,

sparsely, and non-inhibitory species based on their bioactivity index (BI) (0 to 1) by accounting for overall antibiosis activity of strains within each species. The top five species with the highest BI (0.39 to 0.85) were *B. pumilus*, *B. velezensis*, *B. safensis*, *B. subtilis*, and *B. altitudinis*. Interestingly, strains from these five species carried a higher number of NRPS, especially lipopeptide BGCs as compared to the other species. Based on this data, we hypothesized that the expression of these commonly observed lipopeptides may be the conserved factors that contributes to the observed antibiosis response. These findings prompted further investigations on evaluating all 59 *P. nicotianae*-inhibitory *Bacillus* strains against three fungal pathogens.

Evaluation for antifungal activity of 59 *P. nicotianae*-inhibitory *Bacillus* strains against *F. oxysporum*, *F. graminearum*, and *R. solani* showed all strains from *B. velezensis* and all but one from *B. subtilis* exhibited strong inhibition against each fungus being evaluated. In contrast, strains from *B. pumilus*, *B. safensis,* and *B. altitudinis* showed no inhibition against *F. oxysporum*, and *F. graminearum*, and only few strains showed weak inhibition against *R. solani*. Based on this result, we designated *B. velezensis* and *B. subtilis* as *generalists* for the broad-spectrum antibiosis activity against fungi/oomycete. Conversely, *B. pumilus*, *B. safensis,* and *B. altitudinis* were classified as *P. nicotianae specialists* for narrow-spectrum antibiosis activity. Interestingly, the *generalists* two species are phylogenetically more closely related, and the *specialists* three species are also closely related. This data suggests that the expression of phylogenetically conserved common factors among the *generalists* and *specialists* may be linked with the observed breadth of antibiosis activity.

Bacillus BGC analyses showed that several antimicrobial secondary metabolite-producing BGCs were highly conserved within the *generalists* and *specialists*. These were iturin, fengycin,

surfactin, surfactin-like (lichenysin), bacillibactin, and bacilysin. In contrast, these were less commonly observed among *P. nicotianae*-noninhibitory *Bacillus* species. Clustering analysis of all BGCs showed that most commonly observed BGC classes from phylogenetically-related *generalists*, *specialists*, and non-inhibitors were subdivided into separate clusters. This indicates that these commonly observed BGCs within each antibiosis species group are phylogenetically conserved and some highly conserved BGCs from the *generalists* and *specialists* may serve as common factors for the observed antibiosis activity. Interestingly, NRPS, especially lipopeptides, were the most commonly observed BGCs among *generalists* and *specialists* and virtually absent among non-inhibitory species. Notably, all strains from *B. velezensis* carried iturin, fengycin, and surfactin BGCs, while *B. subtilis* carried fengycin and surfactin BGC. As a contrast, each strain from *B. pumilus*, *B. safensis,* and *B. altitudinis* carried a surfactin-like (lichenysin) BGC. The breadth of antibiosis activity of the *generalists* and *specialists* is consistent with the ability to produce numbers and types of lipopeptides; species from the *generalists* produce 2-3 lipopeptides (iturin/bacillomycin, fengycin, and surfactin) while the species from the *specialists* produce only one lipopeptide (surfactin-like).

Furthermore, *Bacillus* secondary metabolites that may be responsible for the observed antibiosis activity were extracted from the culture media of representative *Bacillus* species. Secondary metabolite total extracts were further analyzed using UV-vis absorption, HPLC elution profile, and high-resolution LC-MS identification. UV-vis absorption spectra of the total extract showed characteristic absorbance bands for lipopeptide and polyketide compounds. LC elution of the total extract detected at 220nm showed clusters of peaks in different retention times. High-resolution MS and MS^2 analyses identified six to ten derivatives of lipopeptide with variable length

fatty acid side chain and/or change in amino acid in the core peptide. Bacillomycin L (an iturin), fengycin, and surfactin were identified in extract of *B. velezensis* (JJ334), whereas only the fengycin and surfactin were identified in *B. subtilis* (JM553). Consistent with the BGC analysis, only the surfactin was identified in the total extract of *B. pumilus* (JJ1622), *B. safensis* (JJ1244), and *B. altitudinis* (JJ1138). Each lipopeptide was further purified using a semi-preparative HPLC system and evaluated for antibiosis activity against *P. nicotianae* and fungal pathogens using both plate-based and 96-well microtiter-based antibiosis assay. Each lipopeptide exhibited antibiosis activity against the target pathogens. Quantitative antibiosis evaluation of each lipopeptide against *P. nicotianae* and *F. oxysporum* showed bacillomycin L and fengycin were relatively strong inhibitors than surfactin.

5.1.2. Toward identifying promising *Bacillus* strains for developing commercially viable biological agents

Strong antibiotic activity of *B. velezensis* against *P. nicotianae* and three fungal pathogens and the ability to produce relatively higher number and diverse classes of antimicrobial secondary metabolites prompted further investigation into 18 *B. velezensis* strains. *Bacillus* intraspecies interactions among all 18 strains showed each strain is self-resistant against itself. However, a substantial variation in inhibitory and resistant activity was observed among various strains. Interestingly, strains showing the greatest inhibition across other strains were AB01 and JJ951, and strains with the highest resistance ability were AB01, JJ951, AP46, and JJ747. Strains exhibiting the lowest antibiosis against other strains were JJ213, AP81, JJ1284, and JM204 and with the lowest resistance ability against other strains were JJ1043, AP81, JJ334, and JM204. Further evaluation of antibiotic activity against three fungal pathogens showed a broad range of

antifungal activity among 18 *B. velezensis* strains. Antifungal activity was also dependent on the fungus being evaluated, where *R. solani* was the most sensitive and *F. oxysporum* was the least sensitive. Accounting for overall antibiotic activity against all three fungi, strain AP215 exhibited the most robust antifungal activity than other strains. Further, comparison of the growth phenotype of all 18 *B. velezensis* strains showed all but JM204, JJ907, and JJ1043 produce robust and highly complex biofilm. Finally, a comparison of the production of three polyketides and three lipopeptides showed that strains AB01 and JJ951 produced relatively higher quantity of polyketides+lipopeptides than any other strains.

5.1.3. Structural and functional insights of *Bacillus* cytochromes P450 associated with secondary metabolite biosynthesis

In order to evaluate functional roles of *Bacillus* cytochromes P450 in secondary metabolites biosynthesis, comprehensive genomic analysis, homology modeling, and molecular docking analyses were carried out. A total of 5,051 P450 genes were identified in 1,562 *Bacillus* genomes. All P450s were distributed within twelve P450 families. Among them, 614 P450 genes were identified as accessory genes in various secondary metabolite-producing BGCs. The five most commonly observed P450 families were CYP113, CYP134, CYP102, CYP109, and CYP107. Each family was involved in a specific secondary metabolite biosynthetic pathway: difficidin (CYP113), cyclodipeptide (CYP134), plantazolicin (CYP102), bacillibactin/surfactin (CYP109), and bacillaene/fengycin (CYP107). A representative enzyme from the P450 families was selected to construct homology models and perform molecular docking simulations. A high-resolution homology model was constructed for each P450 protein using RosettaCM employing multiple reference templates. Each modeled P450 structure revealed distinct structural features in

the active site that may be linked with substrate preference for respective secondary metabolite. Each P450 structure was then used for molecular docking simulation with the putative substrate associated with the secondary metabolite biosynthetic pathway. A favorable binding mode with negative binding energy of P450-docked candidate substrate suggests that CYP113, CYP134, CYP109, CYP107, and CYP102 may involve in the oxidative transformation of difficidin, cyclodipeptide, dihydroxybenzoate, bacillaene, plantazolicin, respectively.

Further, a representative CYP102 gene (*Ba*CYP102A2) from *Bacillus amyloliquefaciens* was selected to evaluate its functional role in plantazolicin biosynthetic pathway. The gene was synthesized and cloned in a pET 21 (+) vector at GenScript (Piscataway, NJ) followed by the transformation of *E. coli* BL21 [DE3] for heterologous expression. The expression and purification of the *Ba*CYP102A2 enzyme were optimized and the enzyme was purified employing chromatographic techniques, including affinity (Ni-NTA resin), size exclusion, and anion exchange. The identity of the enzyme was confirmed by SDS-PAGE band at ~120 kDa and a heme-based Soret band (419.5 nm) and charge transfer bands in UV-vis spectra. Substrate binding titrations of *Ba*CYP102A2 with SDS produced a sigmoidal response, whereas OA produced a hyperbolic response. Steady-state kinetic evaluation of *Ba*CYP102A2 with SDS, myristic acid (MA), and palmitic acid (PA) produced a sigmoidal response, whereas OA produced a hyperbolic response.

5.2. Future work

The research of this dissertation has laid out a foundation for three major aspects of future research directions: (a) develop *Bacillus* species as biological agents, (b) discover agriculturally-relevant novel antimicrobial secondary metabolites produced by *Bacillus* species, and (c) elucidate

functional roles of *Bacillus* cytochromes P450 in secondary metabolite biosynthesis. Future works from some of these potential research directions include (i) production, isolation, and characterization of industrially-relevant novel secondary metabolites, (ii) cloning bioactive and/or cryptic secondary metabolites-producing BGCs into heterologous expression system, (iii) evaluation of secondary metabolites as a potential pesticide/fungicide, (iv) discovery of antibiotics for clinically-relevant pathogens, (v) evaluation of plant growth promotion of bioactive *Bacillus* strains in greenhouse and field conditions, and (vi) engineering novel *Bacillus* bioactive strain by combining multiple bioactive strains, BGCs, and/or existing synthetic fungicides.

5.2.1. Identification and characterization of novel secondary metabolites produced by *Bacillus* species

In order to extract, isolate, and identify novel secondary metabolites produced by *Bacillus* species, a multi-dimensional optimization will be carried out in culture medium, extraction solvents, and isolation techniques. Different proportions of carbon and nitrogen sources will be evaluated as nutrient media to express large quantities and more diverse secondary metabolites. Polar, non-polar, and/or semi-polar solvents will be evaluated as extraction solvents for optimum extraction of secondary metabolites. For purification, multi-step separations will be carried out using various chromatographic techniques, including normal-phase silica-gel column, and reverse-phase preparative/semi-preparative HPLC column. Chemical properties of novel compounds will be evaluated by UV-vis absorption, FTIR, HPLC elution, and LC-MS/MS^n analysis. Putative structure of novel compounds informed by MS and MS^n will be unambiguously determined using NMR and/or X-ray diffraction analysis. Finally, bioactivity of a novel compound will be evaluated against plant pathogens, including fungi, oomycetes, bacteria, and nematodes.

Further, a more targeted approach will be considered to increase the production of as yet unknown secondary metabolites. For quantification, secondary metabolites will be produced, extracted, isolated, dried, and weighed. Purified secondary metabolites will be evaluated for antibiotic activity against bacterial and fungal pathogens. The inhibitory effect including IC_{50} and MIC (minimum inhibitory concentration) of the bioactive secondary metabolites will be estimated by monitoring the growth of the target pathogen in the presence of the bioactive compound using microtiter-based assays. For fungal and oomycetes inhibition, quantitative antibiosis assay will be performed in a diluted V8 medium (80 mL/L of V8 juice and 0.5 g/L $CaCO_3$). Similarly, microtiter-based antibacterial activity will be carried out in tryptic soy broth medium. Further, various concentrations of secondary metabolites will be evaluated against plant pathogens and the IC_{50} and MIC will be estimated from the concentration-dependent growth inhibition measurements.

Predicted BGCs that do not produce secondary metabolites in typical laboratory conditions will be cloned into an engineered *Bacillus subtilis* 168 expression system. The regulatory genes and/or the core genes in BGCs will be optimized/modified for optimal expression of the desired secondary metabolites. This approach would allow the production and identification of novel and/or cryptic compounds that are otherwise unidentifiable in typical laboratory conditions. Using an engineered heterologous expression system, isolation of the target secondary metabolite would be much easier by deleting/silencing the BGCs that would otherwise produce unwanted secondary metabolites.

5.2.2. Develop commercially viable *Bacillus* biological agents against plant pathogens

The long-term goal of our research is to exploit promising *Bacillus* strains as sustainable biofertilizers and/or biopesticides as commercially viable biocontrol agents. Toward that direction, the next step would be to evaluate the impact of *Bacillus* strain on plant growth in a greenhouse and field trial experiments. In addition, the direct disease-suppression ability of promising *Bacillus* strains would be evaluated by introducing plant pathogens in the greenhouse and field experiments as mentioned above. Along with traditional foliar treatment, the experiments will also be done by coating *Bacillus* strain on seeds to monitor the effect of plant growth and disease suppression.

Multiple bioactive secondary metabolites producing-BGCs from several promising *Bacillus* strains would be cloned into a single strain in order to produce broad range of bioactive compounds. This may produce a strain with a stronger and broader antibiotic activity against various plant pathogens. Conversely, unnecessary secondary metabolites-producing BGCs could be silenced/deleted from a promising *Bacillus* strain in order to redirect the energy source toward production of target antimicrobial compounds.

Combining a bioactive *Bacillus* strain with existing (or novel) fungicide is another viable strategy to enhance the strength and breadth of bioactivity against wide range of plant pathogens. This strategy has been getting more traction recently since the development of a novel biofungicide *Esendo* by the agrochemical company *AgBiome, LLC*, where they combined their previously EPA-approved biological agent, *Howler* with a synthetic fungicide, *azoxystrobin*.

More potent bioactive *Bacillus* strains with strong and/or broad-spectrum bioactivity may be further engineered by combining multiple *Bacillus* strains with complementary biological activity. One such approach includes combining and shuffling multiple bioactive strains from

different *Bacillus* species to maximize the bioactivity against a wide range of pathogenic microorganisms. For instance, an engineered novel *Bacillus* strain produced by combining one *B. velezensis* with one *B. pumilus* may be a more powerful biological agent than each individual one.

5.2.3. Elucidate the structural and functional properties of *Bacillus* BGC-affiliated P450s

In the future, the role of *Ba*CYP102A2 in the functional modification of plantazolicin will be evaluated *in vitro*. This will be done by identifying the products generated by *Ba*CYP102A2 following reaction with plantazolicin or truncated plantazolicin analogues. To achieve this, plantazolicin and its truncated analogues will be extracted and/or synthesized for assaying with *Ba*CYP102A2 and to produce and identify *Ba*CYP102A2-modified plantazolicin from the *Bacillus* strain. In addition, the plantazolicin products from the culture of a *Ba*CYP102A2-expressing *Bacillus* strain will be isolated and verified the expected modification by *Ba*CYP102A2, *in vivo*. In addition, an attempt will be made to investigate how plantazolicin or its truncated analogues interact with *Ba*CYP102A2 using molecular dynamics simulation.

In the future, an attempt will be made to crystallize the full-length *Ba*CYP102A2 since the crystal structure of a full-length CYP102 ($P450_{BM3}$) is not available as of today. The main reason for not having a full-length X-ray crystal structure of $P450_{BM3}$ is due to challenges in crystallization, partly mediated by the highly mobile linker region (30-40 aa residues) between the CYP-domain and the reductase domain. To assist crystallization and to limit the mobility, the enzyme will be engineered by decreasing size of the linker region and/or modifying amino acid at the linker region while keeping the enzyme fully functional. In the future, conditions for crystallization of full-length *Ba*CYP102A2 enzyme will be optimized using the automated screening setup available at the X-ray diffraction facility in the department of chemistry and

biochemistry at Auburn University. Subsequently, attempt will be made to crystalize *Ba*CYP102A2 using the optimum condition, and high-quality crystals will be sent to the nearest synchrotron radiation facility for recording its X-ray diffractions.

Azole drugs are of great interest to treat fungal pathogens. Recently, Jeffreys *et al*. reported that the CYP102A1 ($P450_{BM3}$) can bind with azole drugs in its wild-type and double mutant variant (396). However, their data indicated that the variant binds more tightly than the wild-type. This result suggests that various CYP102 homologs or potential variants may catalyze the oxidative transformations of azole drugs. In the future, attempt will be made to produce a library of variants of *Ba*CYP102A2 enzyme to screen for oxidative transformation of azole drugs by employing directed evolution and/or rational design techniques (397).

In the near future, attempt will be made to produce a construct-containing a CYP113 gene (*Bv*CYP113) from *Bacillus velezensis* that is predicted to be involved in difficidin biosynthesis. The *Bv*CYP113-containing plasmid will be transferred into *E. coli* XL-1 Blue competent cell and BL21-[DE3] expression system. The genes for P450 redox partners (flavodoxin, and flavodoxin reductase) will be obtained from *B. velezensis* or *B. subtilis* and cloned into the same or a separate plasmid. The plasmid will be transferred into the *E. coli* expression system for the heterologous expression of target enzyme. Both *Bv*CYP113 and its redox partners enzymes will be expressed and purified to perform substrate-dependent and steady-state kinetic assays. Further, difficidin, the putative substrate of *Bv*CYP113, will be extracted and purified from the *B. velezensis* strain (JJ951 or AP215) and used as a substrate for *Bv*CYP113. In addition, commonly known P450 substrates (e.g., SDS and fatty acids) will be evaluated as substrates for the *Bv*CYP113 enzyme.

In order to investigate the functional role of CYP107, CYP109, and CYP134 in respective biosynthetic pathways of bacillaene, bacillibactin, and cyclodipeptide, genes for each enzyme will be synthesized and cloned into plasmids for heterologous expression. The appropriate construct for the redox partners (flavodoxin/ferredoxin reductase and flavodoxin) will also be cloned for heterologous expressions. Subsequently, each P450 and the redox partners will be expressed and purified to perform substrate dependent and steady-state kinetic assay using the respective candidate substrate. Each substrate will be either isolated from the culture of *Bacillus* species or purchased/synthesized, depending on the availability.

To understand the enzyme-substrate interaction at the active site and the catalytic mechanism of the P450 enzyme, molecular dynamics simulations will be carried out in both substrate-free and substrate-bound forms. Simulation for the substrate-free form will help understand P450's structure in its relaxed, open conformation state. Conversely, simulation for the substrate-bound form would provide information about non-covalent interactions between enzyme and substrate in P450's closed conformation state. In addition, relatively long (50-100 μs) simulation with substrate-bound forms will help understand changes in P450's secondary and tertiary structures due to the effect of substrate binding.

Due to the recent genomic revolution, numbers of genomic sequences and bioinformatics algorithms have continued to increase at an astounding rate. These advances have produced large numbers of publicly available genome sequences from all kingdoms of life and made them accessible to virtually everybody. There is a great potential for the discovery of P450s with novel structure and catalytic function utilizing large-scale genomic, bioinformatics, and computational analysis. In the future, a more accurate evaluation of P450's structure and function will be possible

by employing advanced predictive computation and novel algorithms such as machine learning, alpha fold, Rosetta, and Foldit (380, 398, 399).

References

1. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327:812–818.

2. Lamichhane JR, Debaeke P, Steinberg C, You MP, Barbetti MJ, Aubertot J-N. 2018. Abiotic and biotic factors affecting crop seed germination and seedling emergence: a conceptual framework. Plant Soil 432:1–28.

3. Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. 2014. Abiotic and biotic stress combinations. New Phytol 203:32–43.

4. Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. 2020. Threats to global food security from emerging fungal and oomycete crop pathogens. Nat Food 1:332–342.

5. Ma Z, Michailides TJ. 2005. Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. Crop Prot 24:853–863.

6. Hollomon DW. 2015. Fungicide resistance: facing the challenge - a review. Plant Prot Sci 51 (2015):170–176.

7. Brent K, Hollomon D. 2007. Fungicide Resistance In Crop Pathogens How Can It Be Managed. https://www.semanticscholar.org/paper/Fungicide-Resistance-In-Crop-Pathogens-How-Can-It-Brent-Hollomon/0c840bbfcda7afb121aa9d774b4951810b154035. Retrieved 20 June 2022.

8. KLOEPPER JW. 1978. Plant growth-promoting rhizobacteria on radishes. Proc 4th Internet Conf Plant Pathog Bacter Stn Pathol Veg Phytobacteriologie INRA Angers Fr 1978 2:879–882.

9. Bhattacharyya PN, Jha DK. 2012. Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. World J Microbiol Biotechnol 28:1327–1350.

10. Backer R, Rokem JS, Ilangumaran G, Lamont J, Praslickova D, Ricci E, Subramanian S, Smith DL. 2018. Plant Growth-Promoting Rhizobacteria: Context, Mechanisms of Action, and Roadmap to Commercialization of Biostimulants for Sustainable Agriculture. Front Plant Sci 9:1473.

11. Ali S, Hameed S, Shahid M, Iqbal M, Lazarovits G, Imran A. 2020. Functional characterization of potential PGPR exhibiting broad-spectrum antifungal activity. Microbiol Res 232:126389.

12. Fravel DR. 2005. Commercialization and implementation of biocontrol. Annu Rev Phytopathol 43:337–359.

13. Villarreal-Delgado MF, Villa-Rodríguez ED, Cira-Chávez LA, Estrada-Alvarado MI, Parra-Cota FI, Santos-Villalobos S de los, Villarreal-Delgado MF, Villa-Rodríguez ED, Cira-Chávez LA, Estrada-Alvarado MI, Parra-Cota FI, Santos-Villalobos S de los. 2018. The genus Bacillus as a biological control agent and its implications in the agricultural biosecurity. Rev Mex Fitopatol 36:95–130.

14. Shafi J, Tian H, Ji M. 2017. Bacillus species as versatile weapons for plant pathogens: a review. Biotechnol Biotechnol Equip 31:446–459.

15. Fira D, Dimkić I, Berić T, Lozo J, Stanković S. 2018. Biological control of plant pathogens by Bacillus species. J Biotechnol 285:44–55.

16. Radhakrishnan R, Hashem A, Abd_Allah EF. 2017. Bacillus: A Biological Tool for Crop Improvement through Bio-Molecular Changes in Adverse Environments. Front Physiol 8:667.

17. Kaspar F, Neubauer P, Gimpel M. 2019. Bioactive Secondary Metabolites from Bacillus subtilis: A Comprehensive Review. J Nat Prod 82:2038–2053.

18. Lewis K. 2020. The Science of Antibiotic Discovery. Cell 181:29–45.

19. Hutchings MI, Truman AW, Wilkinson B. 2019. Antibiotics: past, present and future. Curr Opin Microbiol 51:72–80.

20. Hoffman PS. 2020. Antibacterial Discovery: 21st Century Challenges. 5. Antibiotics 9:213.

21. Crüsemann M. 2021. Coupling Mass Spectral and Genomic Information to Improve Bacterial Natural Product Discovery Workflows. Mar Drugs 19.

22. Zerikly M, Challis GL. 2009. Strategies for the Discovery of New Natural Products by Genome Mining. ChemBioChem 10:625–633.

23. Medema MH, Fischbach MA. 2015. Computational approaches to natural product discovery. Nat Chem Biol 11:639–648.

24. Ren H, Shi C, Zhao H. 2020. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. iScience 23:100795.

25. Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH. 2009. Automated genome mining for natural products. BMC Bioinformatics 10:185.

26. Medema MH. 2021. The year 2020 in natural product bioinformatics: an overview of the latest tools and databases. Nat Prod Rep 38:301–306.

27. Kim H, Wang M, Leber C, Nothias L-F, Reher R, Kang KB, Hooft JJJ van der, Dorrestein P, Gerwick W, Cottrell G. 2020. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products https://doi.org/10.26434/chemrxiv.12885494.v1.

28. Zin PPK, Williams GJ, Ekins S. 2020. Cheminformatics Analysis and Modeling with MacrolactoneDB. 1. Sci Rep 10:6284.

29. Skinnider MA, Johnston CW, Gunabalasingam M, Merwin NJ, Kieliszek AM, MacLellan RJ, Li H, Ranieri MRM, Webster ALH, Cao MPT, Pfeifle A, Spencer N, To QH, Wallace DP, Dejong CA, Magarvey NA. 2020. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. 1. Nat Commun 11:6058.

30. Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, Zechel DL, Ma B, Magarvey NA. 2015. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. 1. Nat Commun 6:1–11.

31. Hu D, Gao C, Sun C, Jin T, Fan G, Mok KM, Lee SM-Y. 2019. Genome-guided and mass spectrometry investigation of natural products produced by a potential new actinobacterial strain isolated from a mangrove ecosystem in Futian, Shenzhen, China. Sci Rep 9:823.

32. Latijnhouwers M, de Wit PJGM, Govers F. 2003. Oomycetes and fungi: similar weaponry to attack plants. Trends Microbiol 11:462–469.

33. Rossman AY. 2009. The impact of invasive fungi on agricultural ecosystems in the United States, p. 97–107. *In* Langor, DW, Sweeney, J (eds.), Ecological Impacts of Non-Native Invertebrates and Fungi on Terrestrial Ecosystems. Springer Netherlands, Dordrecht.

34. Woodham-smith C. 1962. The Great Hunger: Ireland 1845-9. Gt Hunger Irel 1845-9.

35. Padmanabhan SY. 1973. The Great Bengal Famine. Annu Rev Phytopathol 11:11–24.

36. Dean R, Van Kan J a. L, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430.

37. Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, Roy SG, Schena L, Zambounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen X-R, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BM, Grünwald NJ, Mukhtar MS, Tomé DFA, Tör M, Van Den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuze H, Meijer HJG, Petre B, Ristaino J, Yoshida K, Birch PRJ, Govers F. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol Plant Pathol 16:413–434.

38. Casadevall A. 2018. Fungal Diseases in the 21st Century: The Near and Far Horizons. Pathog Immun 3:183–196.

39. Khush GS. 2005. What it will take to Feed 5.0 Billion Rice consumers in 2030. Plant Mol Biol 59:1–6.

40. Poveda J, Barquero M, González-Andrés F. 2020. Insight into the Microbiological Control Strategies against Botrytis cinerea Using Systemic Plant Resistance Activation. 11. Agronomy 10:1822.

41. Dean R, Van Kan J a. L, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430.

42. Edel-Hermann V, Lecomte C. 2019. Current Status of Fusarium oxysporum Formae Speciales and Races. Phytopathology® 109:512–530.

43. Srinivas C, Nirmala Devi D, Narasimha Murthy K, Mohan CD, Lakshmeesha TR, Singh B, Kalagatur NK, Niranjana SR, Hashem A, Alqarawi AA, Tabassum B, Abd_Allah EF, Chandra

Nayaka S, Srivastava RK. 2019. Fusarium oxysporum f. sp. lycopersici causal agent of vascular wilt disease of tomato: Biology to diversity– A review. Saudi J Biol Sci 26:1315–1324.

44. Ajayi-Oyetunde OO, Bradley CA. 2018. Rhizoctonia solani: taxonomy, population biology and management of rhizoctonia seedling disease of soybean. Plant Pathol 67:3–17.

45. Moni ZR, Ali MA, Alam MS, Rahman MA, Bhuiyan MR, Mian MS, Iftekharuddaula KM, Latif MA, Khan MAI. 2016. Morphological and Genetical Variability among Rhizoctonia solani Isolates Causing Sheath Blight Disease of Rice. Rice Sci 23:42–50.

46. Atkinson D, Thornton MK, Miller JS. 2010. Development of Rhizoctonia solani on Stems, Stolons and Tubers of Potatoes I. Effect of Inoculum Source. Am J Potato Res 87:374–381.

47. Meng S, Torto-Alalibo T, Chibucos MC, Tyler BM, Dean RA. 2009. Common processes in pathogenesis by fungal and oomycete plant pathogens, described with Gene Ontology terms. BMC Microbiol 9:S7.

48. Kamoun S, Huitema E, Vleeshouwers VGAA. 1999. Resistance to oomycetes: a general role for the hypersensitive response? Trends Plant Sci 4:196–200.

49. Van de Peer Y, De Wachter R. 1997. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. J Mol Evol 45:619–630.

50. Williamson-Benavides BA, Dhingra A. 2021. Understanding Root Rot Disease in Agricultural Crops. 2. Horticulturae 7:33.

51. Jiang RHY, Tyler BM. 2012. Mechanisms and Evolution of Virulence in Oomycetes. Annu Rev Phytopathol 50:295–318.

52. Cooke DEL, Drenth A, Duncan JM, Wagels G, Brasier CM. 2000. A Molecular Phylogeny of Phytophthora and Related Oomycetes. Fungal Genet Biol 30:17–32.

53. Zhang BQ, Yang XB. 2000. Pathogenicity of Pythium Populations from Corn-Soybean Rotation Fields. Plant Dis 84:94–99.

54. Goss EM, Tabima JF, Cooke DEL, Restrepo S, Fry WE, Forbes GA, Fieland VJ, Cardenas M, Grünwald NJ. 2014. The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. Proc Natl Acad Sci 111:8791–8796.

55. Tyler BM. 2007. Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol Plant Pathol 8:1–8.

56. Lamour KH, Stam R, Jupe J, Huitema E. 2012. The oomycete broad-host-range pathogen Phytophthora capsici. Mol Plant Pathol 13:329–337.

57. Bulone V, Chanzy H, Gay L, Girard V, Fe`vre M. 1992. Characterization of chitin and chitin synthase from the cellulosic cell wall fungusSaprolegnia monoi¨ca. Exp Mycol 16:8–21.

58. Werner S, Steiner U, Becher R, Kortekamp A, Zyprian E, Deising HB. 2002. Chitin synthesis during in planta growth and asexual propagation of the cellulosic oomycete and obligate biotrophic grapevine pathogen Plasmopara viticola. FEMS Microbiol Lett 208:169–173.

59. ten Have A, Tenberge KB, Benen JAE, Tudzynski P, Visser J, van Kan JAL. 2002. The Contribution of Cell Wall Degrading Enzymes to Pathogenesis of Fungal Plant Pathogens, p. 341–358. *In* Kempken, F (ed.), Agricultural Applications. Springer, Berlin, Heidelberg.

60. JARVIS MC, THRELFALL DR, FRIEND J. 1981. Potato Cell Wall Polysaccharides: Degradation with Enzymes from Phytophthora infestans. J Exp Bot 32:1309–1319.

61. Tyler B. 2002. Molecular basis of recognition between Phytophthora pathogens and their hosts [Review]. Annu Rev Phytopathol 40:137–67.

62. Dangl JL, Jones JDG. 2001. Plant pathogens and integrated defence responses to infection. 6839. Nature 411:826–833.

63. Shahbazi H, Aminian H, Sahebani N, Halterman DA. 2011. Activity of ß-1,3-glucanase and ß-1,4-glucanase in two potato cultivars following challenge by the fungal pathogen Alternaria solani. Phytoparasitica 39:455–460.

64. Spanu P, Kämper J. 2010. Genomics of biotrophy in fungi and oomycetes—emerging patterns. Curr Opin Plant Biol 13:409–414.

65. Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev 35:57–65.

66. Bebber DP, Gurr SJ. 2015. Crop-destroying fungal and oomycete pathogens challenge food security. Fungal Genet Biol 74:62–64.

67. Lucas JA, Hawkins NJ, Fraaije BA. 2015. Chapter Two - The Evolution of Fungicide Resistance, p. 29–92. *In* Sariaslani, S, Gadd, GM (eds.), Advances in Applied Microbiology. Academic Press.

68. Deliopoulos T, Kettlewell PS, Hare MC. 2010. Fungal disease suppression by inorganic salts: A review. Crop Prot 29:1059–1075.

69. Hahn M. 2014. The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J Chem Biol 7:133–141.

70. Deising HB, Reimann S, Pascholati SF. 2008. Mechanisms and significance of fungicide resistance†. Braz J Microbiol 39:286–295.

71. McKay GJ, Egan D, Morris E, Brown AE. 1998. Identification of benzimidazole resistance in Cladobotryum dendroides using a PCR-based method. Mycol Res 102:671–676.

72. Hahn M. 2014. The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J Chem Biol 7:133–141.

73. Handelsman J, Stabb E. 1996. Biocontrol of Soilborne Plant Pathogens. Plant Cell 8:1855–1869.

74. Baker BP, Green TA, Loker AJ. 2020. Biological control and integrated pest management in organic and conventional systems. Biol Control 140:104095.

75. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16:115–125.

76. Holt RD, Hochberg ME. 1997. When Is Biological Control Evolutionarily Stable (or Is It)? Ecology 78:1673–1683.

77. Ons L, Bylemans D, Thevissen K, Cammue BPA. 2020. Combining Biocontrol Agents with Chemical Fungicides for Integrated Plant Fungal Disease Control. Microorganisms 8:1930.

78. Akinrinlola RJ, Yuen GY, Drijber RA, Adesemoye AO. 2018. Evaluation of Bacillus Strains for Plant Growth Promotion and Predictability of Efficacy by In Vitro Physiological Traits. Int J Microbiol 2018:5686874.

79. Walker TS, Bais HP, Grotewold E, Vivanco JM. 2003. Root exudation and rhizosphere biology. Plant Physiol 132:44–51.

80. Lwin KM, Myint MM, Tar T, Aung WZM. 2012. Isolation of Plant Hormone (Indole-3-Acetic Acid - IAA) Producing Rhizobacteria and Study on Their Effects on Maize Seedling. Eng J 16:137–144.

81. Sengin EL, Rosmana A. Screening Of Free-Living Indole Acetic Acid Producing Rhizobacteria From Shallot Rhizospheres In The Island Of Sulawesi.

82. Joo G-J, Kim Y-M, Kim J-T, Rhee I-K, Kim J-H, Lee I-J. 2005. Gibberellins-producing rhizobacteria increase endogenous gibberellins content and promote growth of red peppers. J Microbiol Seoul Korea 43:510–515.

83. Beneduzi A, Peres D, Vargas LK, Bodanese-Zanettini MH, Passaglia LMP. 2008. Evaluation of genetic diversity and plant growth promoting activities of nitrogen-fixing bacilli isolated from rice fields in South Brazil. Appl Soil Ecol 39:311–320.

84. Zablotowicz RM, Tipping EM, Lifshitz R, Kloepper JW. 1991. Plant growth promotion mediated by bacterial rhizosphere colonizers, p. 315–326. *In* Keister, DL, Cregan, PB (eds.), The Rhizosphere and Plant Growth: Papers presented at a Symposium held May 8–11, 1989, at the Beltsville Agricultural Research Center (BARC), Beltsville, Maryland. Springer Netherlands, Dordrecht.

85. Lugtenberg B, Kamilova F. 2009. Plant-growth-promoting rhizobacteria. Annu Rev Microbiol 63:541–556.

86. Glick BR. 2012. Plant growth-promoting bacteria: mechanisms and applications. Scientifica 2012:963401.

87. Majeed A, Muhammad Z, Ahmad H. 2018. Plant growth promoting bacteria: role in soil improvement, abiotic and biotic stress management of crops. Plant Cell Rep 37:1599–1609.

88. Hider RC, Kong X. 2010. Chemistry and biology of siderophores. Nat Prod Rep 27:637–657.

89. Iqbal A, Hasnain S. 2013. Aeromonas punctata PNS-1: a promising candidate to change the root morphogenesis of Arabidopsis thaliana in MS and sand system. Acta Physiol Plant 35:657–665.

90. Galland M, Gamet L, Varoquaux F, Touraine B, Touraine B, Desbrosses G. 2012. The ethylene pathway contributes to root hair elongation induced by the beneficial bacteria Phyllobacterium brassicacearum STM196. Plant Sci 190:74–81.

91. Lara-Chavez A, Lowman S, Kim S, Tang Y, Zhang J, Udvardi M, Nowak J, Flinn B, Mei C. 2015. Global gene expression profiling of two switchgrass cultivars following inoculation with Burkholderia phytofirmans strain PsJN. J Exp Bot 66:4337–4350.

92. Tahir HAS, Gu Q, Wu H, Raza W, Hanif A, Wu L, Colman MV, Gao X. 2017. Plant Growth Promotion by Volatile Organic Compounds Produced by Bacillus subtilis SYST2. Front Microbiol 8.

93. Spaepen S, Vanderleyden J. 2011. Auxin and Plant-Microbe Interactions. Cold Spring Harb Perspect Biol 3:a001438.

94. Tsavkelova EA, Klimova SYu, Cherdyntseva TA, Netrusov AI. 2006. Microbial producers of plant growth stimulators and their practical use: A review. Appl Biochem Microbiol 42:117–126.

95. Kloepper JW, Leong J, Teintze M, Schroth MN. 1980. Enhanced plant growth by siderophores produced by plant growth-promoting rhizobacteria. 5776. Nature 286:885–886.

96. Kramer J, Özkaya Ö, Kümmerli R. 2020. Bacterial siderophores in community and host interactions. 3. Nat Rev Microbiol 18:152–163.

97. Page MGP. 2019. The Role of Iron and Siderophores in Infection, and the Development of Siderophore Antibiotics. Clin Infect Dis Off Publ Infect Dis Soc Am 69:S529–S537.

98. Wang W, Qiu Z, Tan H, Cao L. 2014. Siderophore production by actinobacteria. BioMetals 27:623–631.

99. Fernando WGD, Nakkeeran S, Zhang Y. 2006. Biosynthesis of Antibiotics by PGPR and its Relation in Biocontrol of Plant Diseases, p. 67–109. *In* Siddiqui, ZA (ed.), PGPR: Biocontrol and Biofertilization. Springer Netherlands, Dordrecht.

100. Gomes RC, Semêdo LT a. S, Soares RMA, Alviano CS, And LFL, Coelho RRR. 2000. Chitinolytic activity of actinomycetes from a cerrado soil and their potential in biocontrol. Lett Appl Microbiol 30:146–150.

101. Chigaleĭchik AG. 1976. [Chitinase of Bacillus thuringiensis]. Mikrobiologiia 45:966–972.

102. Sindhu SS, Dadarwal KR. 2001. Chitinolytic and cellulolytic Pseudomonas sp. antagonistic to fungal pathogens enhances nodulation by Mesorhizobium sp. Cicer in chickpea. Microbiol Res 156:353–358.

103. Picard K, Tirilly Y, Benhamou N. 2000. Cytological Effects of Cellulases in the Parasitism of Phytophthora parasitica by Pythium oligandrum. Appl Environ Microbiol 66:4305–4314.

104. De Vrieze M, Pandey P, Bucheli TD, Varadarajan AR, Ahrens CH, Weisskopf L, Bailly A. 2015. Volatile Organic Compounds from Native Potato-associated Pseudomonas as Potential Anti-oomycete Agents. Front Microbiol 6.

105. Ossowicki A, Jafra S, Garbeva P. 2017. The antimicrobial volatile power of the rhizospheric isolate Pseudomonas donghuensis P482. PLOS ONE 12:e0174362.

106. Powell KA, Jutsum AR. 1993. Technical and commercial aspects of biocontrol products. Pestic Sci 37:315–321.

107. Borriss R. 2011. Use of Plant-Associated Bacillus Strains as Biofertilizers and Biocontrol Agents in Agriculture, p. 41–76. *In* Maheshwari, DK (ed.), Bacteria in Agrobiology: Plant Growth Responses. Springer, Berlin, Heidelberg.

108. Basu A, Prasad P, Das SN, Kalam S, Sayyed RZ, Reddy MS, El Enshasy H. 2021. Plant Growth Promoting Rhizobacteria (PGPR) as Green Bioinoculants: Recent Developments, Constraints, and Prospects. 3. Sustainability 13:1140.

109. Syed Ab Rahman SF, Singh E, Pieterse CMJ, Schenk PM. 2018. Emerging microbial biocontrol strategies for plant pathogens. Plant Sci 267:102–111.

110. Logan NA, Vos PD. 2015. Bacillus, p. 1–163. *In* Bergey's Manual of Systematics of Archaea and Bacteria. American Cancer Society.

111. Hanna PC, Ireland JAW, Hanna PC, Ireland JAW. 1999. Understanding Bacillus anthracis pathogenesis. Trends Microbiol 7:180–182.

112. Messelhäußer U, Ehling-Schulz M. 2018. Bacillus cereus—a Multifaceted Opportunistic Pathogen. Curr Clin Microbiol Rep 5:120–125.

113. Kovács ÁT. 2019. Bacillus subtilis. Trends Microbiol 27:724–725.

114. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi S-K, Codani J-J, Connerton IF, Cummings NJ, Daniel RA, Denizot F, Devine KM, Düsterhöft A, Ehrlich SD, Emmerson PT, Entian KD, Errington J, Fabret C, Ferrari E, Foulger D, Fritz C, Fujita M, Fujita Y, Fuma S, Galizzi A, Galleron N, Ghim S-Y, Glaser P, Goffeau A, Golightly EJ, Grandi G, Guiseppi G, Guy BJ, Haga K, Haiech J, Harwood CR, Hénaut A, Hilbert H, Holsappel S, Hosono S, Hullo M-F, Itaya M, Jones L, Joris B, Karamata D, Kasahara Y, Klaerr-Blanchard M, Klein C, Kobayashi Y, Koetter P, Koningstein G, Krogh S, Kumano M, Kurita K, Lapidus A, Lardinois S, Lauber J, Lazarevic V, Lee S-M, Levine A, Liu H, Masuda S, Mauël C, Médigue C, Medina N, Mellado RP, Mizuno M, Moestl D, Nakai S, Noback M, Noone D, O'Reilly M, Ogawa K, Ogiwara A, Oudega B, Park S-H, Parro V, Pohl TM, Portetelle D, Porwollik S, Prescott AM, Presecan E, Pujic P, Purnelle B, Rapoport G, Rey M, Reynolds S, Rieger M, Rivolta C, Rocha E, Roche B, Rose M, Sadaie Y, Sato T, Scanlan E, Schleich S, Schroeter R, Scoffone F, Sekiguchi J, Sekowska A, Seror SJ, Serror P, Shin B-S, Soldo B, Sorokin A, Tacconi E, Takagi T, Takahashi H, Takemaru K, Takeuchi M, Tamakoshi A, Tanaka T, Terpstra P, Tognoni A, Tosato V, Uchiyama S, Vandenbol M, Vannier F, Vassarotti A, Viari A, Wambutt R, Wedler E, Wedler H, Weitzenegger T, Winters P, Wipat A, Yamamoto H, Yamane K, Yasumoto K,

Yata K, Yoshida K, Yoshikawa H-F, Zumstein E, Yoshikawa H, Danchin A. 1997. The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. Nature 390:249–256.

115. Lateef A, Adelere IA, Gueguim-Kana EB. 2015. The biology and potential biotechnological applications of Bacillus safensis. Biologia (Bratisl) 70:411–419.

116. Contesini FJ, Melo RR de, Sato HH. 2018. An overview of Bacillus proteases: from production to application. Crit Rev Biotechnol 38:321–334.

117. Schallmey M, Singh A, Ward OP. 2011. Developments in the use of Bacillus species for industrial production. Can J Microbiol https://doi.org/10.1139/w03-076.

118. Pérez-García A, Romero D, de Vicente A. 2011. Plant protection and growth stimulation by microorganisms: biotechnological applications of Bacilli in agriculture. Curr Opin Biotechnol 22:187–193.

119. Roh J-Y, Choi J-Y, Li M-S, Jin B-R, Je Y-H. 2007. Bacillus thuringiensis as a Specific, Safe, and Effective Tool for Insect Pest Control. J Microbiol Biotechnol 17:547–559.

120. Nester EW, Thomashow LS, Metz M, Gordon M. 2002. 100 Years of Bacillus thuringiensis: A Critical Scientific Assessment: This report is based on a colloquium, "100 Years of Bacillis thuringiensis, a Paradigm for Producing Transgenic Organisms: A Critical Scientific Assessment," sponsored by the American Academy of Microbiology and held November 16–18, in Ithaca, New York. American Society for Microbiology, Washington (DC). http://www.ncbi.nlm.nih.gov/books/NBK559445/. Retrieved 23 June 2022.

121. Jouzani GS, Valijanian E, Sharafi R. 2017. Bacillus thuringiensis: a successful insecticide with new environmental features and tidings. Appl Microbiol Biotechnol 101:2691–2711.

122. McKenney PT, Driks A, Eichenberger P. 2013. The Bacillus subtilis endospore: assembly and functions of the multilayered coat. Nat Rev Microbiol 11:33–44.

123. Arnaouteli S, Bamford NC, Stanley-Wall NR, Kovács ÁT. 2021. Bacillus subtilis biofilm formation and social interactions. Nat Rev Microbiol 1–15.

124. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Microbiol 2:95–108.

125. Outtrup H, Jørgensen ST. 2002. The Importance of Bacillus Species in the Production of Industrial Enzymes, p. 206–218. *In* Applications and Systematics of Bacillus and Relatives. John Wiley & Sons, Ltd.

126. Danilova I, Sharipova M. 2020. The Practical Potential of Bacilli and Their Enzymes for Industrial Production. Front Microbiol 11:1782.

127. Su Y, Liu C, Fang H, Zhang D. 2020. Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. Microb Cell Factories 19:173.

128. Stein T. 2005. Bacillus subtilis antibiotics: structures, syntheses and specific functions. Mol Microbiol 56:845–857.

129. Mohsin MZ, Omer R, Huang J, Mohsin A, Guo M, Qian J, Zhuang Y. 2021. Advances in engineered Bacillus subtilis biofilms and spores, and their applications in bioremediation, biocatalysis, and biomaterials. Synth Syst Biotechnol 6:180–191.

130. Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R. 2013. Sticking together: building a biofilm the Bacillus subtilis way. Nat Rev Microbiol 11:157–168.

131. Zhang N, Yang D, Wang D, Miao Y, Shao J, Zhou X, Xu Z, Li Q, Feng H, Li S, Shen Q, Zhang R. 2015. Whole transcriptomic analysis of the plant-beneficial rhizobacterium Bacillus amyloliquefaciens SQR9 during enhanced biofilm formation regulated by maize root exudates. BMC Genomics 16:685.

132. Rabbee MF, Ali MS, Choi J, Hwang BS, Jeong SC, Baek K. 2019. Bacillus velezensis: A Valuable Member of Bioactive Molecules within Plant Microbiomes. 6. Molecules 24:1046.

133. Kilian M, Steiner U, Krebs B, Junge H, Schmiedeknecht G, Hain R. 2000. FZB24 Bacillus subtilis - mode of action of a microbial agent enhancing plant vitality. Pflanzenschutz-Nachrichten Bayer 1:72–93.

134. Hashem A, Tabassum B, Fathi Abd_Allah E. 2019. Bacillus subtilis: A plant-growth promoting rhizobacterium that also impacts biotic stress. Saudi J Biol Sci 26:1291–1297.

135. Kumar A, Prakash A, Johri BN. 2011. Bacillus as PGPR in Crop Ecosystem, p. 37–59. *In* Maheshwari, DK (ed.), Bacteria in Agrobiology: Crop Ecosystems. Springer, Berlin, Heidelberg.

136. Kloepper JW, Ryu C-M, Zhang S. 2004. Induced Systemic Resistance and Promotion of Plant Growth by Bacillus spp. Phytopathology® 94:1259–1266.

137. Mondol MAM, Shin HJ, Islam MT. 2013. Diversity of secondary metabolites from marine Bacillus species: chemistry and biological activity. Mar Drugs 11:2846–2872.

138. Carrillo C, Teruel JA, Aranda FJ, Ortiz A. 2003. Molecular mechanism of membrane permeabilization by the peptide antibiotic surfactin. Biochim Biophys Acta BBA - Biomembr 1611:91–97.

139. Aranda FJ, Teruel JA, Ortiz A. 2005. Further aspects on the hemolytic activity of the antibiotic lipopeptide iturin A. Biochim Biophys Acta BBA - Biomembr 1713:51–56.

140. Deleu M, Paquot M, Nylander T. 2008. Effect of Fengycin, a Lipopeptide Produced by Bacillus subtilis, on Model Biomembranes. Biophys J 94:2667–2679.

141. Abriouel H, Franz CMAP, Omar NB, Gálvez A. 2011. Diversity and applications of Bacillus bacteriocins. FEMS Microbiol Rev 35:201–232.

142. Liu Y, Ding S, Shen J, Zhu K. 2019. Nonribosomal antibacterial peptides that target multidrug-resistant bacteria. Nat Prod Rep 36:573–592.

143. Dunlap CA, Bowman MJ, Rooney AP. 2019. Iturinic Lipopeptide Diversity in the Bacillus subtilis Species Group – Important Antifungals for Plant Disease Biocontrol Applications. Front Microbiol 10.

144. Zeriouh H, Romero D, Garcia-Gutierrez L, Cazorla FM, de Vicente A, Perez-Garcia A. 2011. The iturin-like lipopeptides are essential components in the biological control arsenal of Bacillus subtilis against bacterial diseases of cucurbits. Mol Plant-Microbe Interact MPMI 24:1540–1552.

145. de Faria AF, Stéfani D, Vaz BG, Silva ÍS, Garcia JS, Eberlin MN, Grossman MJ, Alves OL, Durrant LR. 2011. Purification and structural characterization of fengycin homologues produced by Bacillus subtilis LSFM-05 grown on raw glycerol. J Ind Microbiol Biotechnol 38:863–871.

146. Ait Kaki A, Smargiasso N, Ongena M, Kara Ali M, Moula N, De Pauw E, Kacem Chaouche N. 2020. Characterization of New Fengycin Cyclic Lipopeptide Variants Produced by Bacillus amyloliquefaciens (ET) Originating from a Salt Lake of Eastern Algeria. Curr Microbiol 77:443–451.

147. Zhang L, Sun C. 2018. Fengycins, Cyclic Lipopeptides from Marine Bacillus subtilis Strains, Kill the Plant-Pathogenic Fungus Magnaporthe grisea by Inducing Reactive Oxygen Species Production and Chromatin Condensation. Appl Environ Microbiol 84:e00445-18.

148. Kang BR, Park JS, Jung W-J. 2020. Antifungal evaluation of fengycin isoforms isolated from Bacillus amyloliquefaciens PPL against Fusarium oxysporum f. sp. lycopersici. Microb Pathog 149:104509.

149. Kang BR, Park JS, Jung W-J. 2021. Antiviral activity by lecithin-induced fengycin lipopeptides as a potent key substrate against Cucumber mosaic virus. Microb Pathog 155:104910.

150. Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT. 2002. Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE. Structure 10:301–310.

151. Théatre A, Cano-Prieto C, Bartolini M, Laurin Y, Deleu M, Niehren J, Fida T, Gerbinet S, Alanjary M, Medema MH, Léonard A, Lins L, Arabolaza A, Gramajo H, Gross H, Jacques P. 2021. The Surfactin-Like Lipopeptides From Bacillus spp.: Natural Biodiversity and Synthetic Biology for a Broader Application Range. Front Bioeng Biotechnol 9.

152. Meena KR, Dhiman R, Singh K, Kumar S, Sharma A, Kanwar SS, Mondal R, Das S, Franco OL, Mandal AK. 2021. Purification and identification of a surfactin biosurfactant and engine oil degradation by Bacillus velezensis KLP2016. Microb Cell Factories 20:26.

153. Bonmatin J-M, Laprévote O, Peypoux F. 2003. Diversity among microbial cyclic lipopeptides: iturins and surfactins. Activity-structure relationships to design new bioactive agents. Comb Chem High Throughput Screen 6:541–556.

154. Li Y, Héloir M, Zhang X, Geissler M, Trouvelot S, Jacquens L, Henkel M, Su X, Fang X, Wang Q, Adrian M. 2019. Surfactin and fengycin contribute to the protection of a Bacillus subtilis strain against grape downy mildew by both direct effect and defence stimulation. Mol Plant Pathol 20:1037–1050.

155. Horng Y-B, Yu Y-H, Dybus A, Hsiao FS-H, Cheng Y-H. 2019. Antibacterial activity of Bacillus species-derived surfactin on Brachyspira hyodysenteriae and Clostridium perfringens. AMB Express 9:188.

156. Wang Y, Zhang C, Liang J, Wang L, Gao W, Jiang J, Chang R. 2020. Surfactin and fengycin B extracted from Bacillus pumilus W-7 provide protection against potato late blight via distinct and synergistic mechanisms. Appl Microbiol Biotechnol 104:7467–7481.

157. Wang X, Zhou H, Chen H, Jing X, Zheng W, Li R, Sun T, Liu J, Fu J, Huo L, Li Y, Shen Y, Ding X, Müller R, Bian X, Zhang Y. 2018. Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. Proc Natl Acad Sci 115:E4255–E4263.

158. Gavriilidou A, Mackenzie TA, Sánchez P, Tormo JR, Ingham C, Smidt H, Sipkema D. 2021. Bioactivity Screening and Gene-Trait Matching across Marine Sponge-Associated Bacteria. 2. Mar Drugs 19:75.

159. Luo Chuping, Liu Xuehui, Zhou Xian, Guo Junyao, Truong John, Wang Xiaoyu, Zhou Huafei, Li Xiangqian, Chen Zhiyi, Elliot M. A. 2015. Unusual Biosynthesis and Structure of Locillomycins from Bacillus subtilis 916. Appl Environ Microbiol 81:6601–6609.

160. Caulier S, Nannan C, Gillis A, Licciardi F, Bragard C, Mahillon J. 2019. Overview of the Antimicrobial Compounds Produced by Members of the Bacillus subtilis Group. Front Microbiol 10.

161. Luo C, Chen Y, Liu X, Wang X, Wang X, Li X, Zhao Y, Wei L. 2019. Engineered biosynthesis of cyclic lipopeptide locillomycins in surrogate host Bacillus velezensis FZB42 and derivative strains enhance antibacterial activity. Appl Microbiol Biotechnol 103:4467–4481.

162. Shabeer Ali H, Ajesh K, Dileep KV, Prajosh P, Sreejith K. 2020. Structural characterization of Kannurin isoforms and evaluation of the role of β-hydroxy fatty acid tail length in functional specificity. 1. Sci Rep 10:2839.

163. Ajesh K, Sudarslal S, Arunan C, Sreejith K. 2013. Kannurin, a novel lipopeptide from Bacillus cereus strain AK1: isolation, structural evaluation and antifungal activities. J Appl Microbiol 115:1287–1296.

164. May JJ, Wendrich TM, Marahiel MA. 2001. The dhb Operon of Bacillus subtilisEncodes the Biosynthetic Template for the Catecholic Siderophore 2,3-Dihydroxybenzoate-Glycine-Threonine Trimeric Ester Bacillibactin *. J Biol Chem 276:7209–7217.

165. Dimopoulou A, Theologidis I, Benaki D, Koukounia M, Zervakou A, Tzima A, Diallinas G, Hatzinikolaou DG, Skandalis N. Direct Antibiotic Activity of Bacillibactin Broadens the Biocontrol Range of Bacillus amyloliquefaciens MBI600. mSphere 6:e00376-21.

166. Scholz R, Molohon KJ, Nachtigall J, Vater J, Markley AL, Süssmuth RD, Mitchell DA, Borriss R. 2011. Plantazolicin, a Novel Microcin B17/Streptolysin S-Like Natural Product from Bacillus amyloliquefaciens FZB42. J Bacteriol 193:215–224.

167. Molohon KJ, Saint-Vincent PMB, Park S, Doroghazi JR, Maxson T, Hershfield JR, Flatt KM, Schroeder NE, Ha T, Mitchell DA. 2016. Plantazolicin Is an Ultranarrow-Spectrum Antibiotic That Targets the Bacillus anthracis Membrane. ACS Infect Dis 2:207–220.

168. Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J. 2007. The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. Proc Natl Acad Sci 104:1506–1509.

169. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, Dean L, Meyers E, Fernandes P, Mayerl F. 1995. Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by Bacillus subtilis: production, taxonomy, isolation, physico-chemical characterization and biological activity. J Antibiot (Tokyo) 48:997–1003.

170. Wilson KE, Flor JE, Schwartz RE, Joshua H, Smith JL, Pelak BA, Liesch JM, Hensens OD. 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by Bacillus subtilis. II. Isolation and physico-chemical characterization. J Antibiot (Tokyo) 40:1682–1691.

171. Ortiz A, Sansinenea E. 2019. Macrolactin Antibiotics: Amazing Natural Products. Mini Rev Med Chem https://doi.org/10.2174/1389557519666191205124050.

172. Kevany BM, Rasko DA, Thomas MG. 2009. Characterization of the Complete Zwittermicin A Biosynthesis Gene Cluster from Bacillus cereus. Appl Environ Microbiol 75:1144–1155.

173. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421.

174. Large-Scale Bioinformatics Analysis of Bacillus Genomes Uncovers Conserved Roles of Natural Products in Bacterial Physiology | mSystems. https://msystems.asm.org/content/2/6/e00040-17. Retrieved 14 February 2020.

175. Fazle Rabbee M, Baek K-H. 2020. Antimicrobial Activities of Lipopeptides and Polyketides of Bacillus velezensis for Agricultural Applications. Mol Basel Switz 25:E4973.

176. Smith S, Tsai S-C. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat Prod Rep 24:1041–1072.

177. Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition | Wiley. Wiley.com. https://www.wiley.com/en-us/Fundamentals+of+Biochemistry%3A+Life+at+the+Molecular+Level%2C+5th+Edition-p-9781118918401. Retrieved 13 January 2022.

178. Chen X-H, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süssmuth RD, Borriss R. 2006. Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in Bacillus amyloliquefaciens FZB 42. J Bacteriol 188:4024–4036.

179. Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC. 2006. Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis. Proc Natl Acad Sci 103:8977–8982.

180. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30:108–160.

181. Song Y, Guan Z, van Merkerk R, Pramastya H, Abdallah II, Setroikromo R, Quax WJ. 2020. Production of Squalene in Bacillus subtilis by Squalene Synthase Screening and Metabolic Engineering. J Agric Food Chem 68:4447–4455.

182. Pramastya H, Song Y, Elfahmi EY, Sukrasno S, Quax WJ. 2021. Positioning Bacillus subtilis as terpenoid cell factory. J Appl Microbiol 130:1839–1856.

183. Zhang K, El Damaty S, Fasan R. 2011. P450 fingerprinting method for rapid discovery of terpene hydroxylating P450 catalysts with diversified regioselectivity. J Am Chem Soc 133:3242–3245.

184. Lewis K. 2012. Recover the lost art of drug discovery. Nature 485:439–440.

185. Morens DM, Folkers GK, Fauci AS. 2004. The challenge of emerging and re-emerging infectious diseases. 6996. Nature 430:242–249.

186. Vouga M, Greub G. 2016. Emerging bacterial pathogens: the past and beyond. Clin Microbiol Infect 22:12–21.

187. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. 7393. Nature 484:186–194.

188. Nathan C, Cars O. 2014. Antibiotic Resistance — Problems, Progress, and Prospects. N Engl J Med 371:1761–1763.

189. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87.

190. Yadav G, Gokhale RS, Mohanty D. 2003. Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. J Mol Biol 328:335–363.

191. Tripathi A, Vázquez-Baeza Y, Gauglitz JM, Wang M, Dührkop K, Nothias-Esposito M, Acharya DD, Ernst M, van der Hooft JJJ, Zhu Q, McDonald D, Brejnrod AD, Gonzalez A, Handelsman J, Fleischauer M, Ludwig M, Böcker S, Nothias L-F, Knight R, Dorrestein PC. 2021. Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. 2. Nat Chem Biol 17:146–151.

192. Du C, van Wezel GP. 2018. Mining for Microbial Gems: Integrating Proteomics in the Postgenomic Natural Product Discovery Pipeline. PROTEOMICS 18:1700332.

193. Kersten RD, Yang Y-L, Xu Y, Cimermancic P, Nam S-J, Fenical W, Fischbach MA, Moore BS, Dorrestein PC. 2011. A mass spectrometry–guided genome mining approach for natural product peptidogenomics. 11. Nat Chem Biol 7:794–802.

194. Kurita KL, Glassey E, Linington RG. 2015. Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. Proc Natl Acad Sci 112:11999–12004.

195. Rakshith D, Santosh P, Pradeep TP, Gurudatt DM, Baker S, Yashavantha Rao HC, Pasha A, Satish S. 2016. Application of Bioassay-Guided Fractionation Coupled with a Molecular Approach for the Dereplication of Antimicrobial Metabolites. Chromatographia 79:1625–1642.

196. Nothias L-F, Nothias-Esposito M, da Silva R, Wang M, Protsyuk I, Zhang Z, Sarvepalli A, Leyssen P, Touboul D, Costa J, Paolini J, Alexandrov T, Litaudon M, Dorrestein PC. 2018. Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. J Nat Prod 81:758–767.

197. Li X, Munir S, Xu Y, Wang Y, He Y. 2021. Combined mass spectrometry-guided genome mining and virtual screening for acaricidal activity in secondary metabolites of Bacillus velezensis W1. RSC Adv 11:25441–25449.

198. Pareek CS, Smoczynski R, Tretyn A. 2011. Sequencing technologies and genome sequencing. J Appl Genet 52:413–435.

199. Bachmann BO, Van Lanen SG, Baltz RH. 2014. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? J Ind Microbiol Biotechnol 41:175–184.

200. Liu X, Ser Z, Locasale JW. 2014. Development and Quantitative Evaluation of a High-Resolution Metabolomics Technology. Anal Chem 86:2175–2184.

201. Issaq HJ, Abbott E, Veenstra TD. 2008. Utility of separation science in metabolomic studies. J Sep Sci 31:1936–1947.

202. Ludwig C, Günther UL. 2011. MetaboLab - advanced NMR data processing and analysis for metabolomics. BMC Bioinformatics 12:366.

203. Bingol K. 2018. Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. 2. High-Throughput 7:9.

204. Liu X, Zhou L, Shi X, Xu G. 2019. New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. TrAC Trends Anal Chem 121:115665.

205. Freiberg C, Brötz-Oesterhelt H. 2005. Functional genomics in antibacterial drug discovery. Drug Discov Today 10:927–935.

206. Xu M, Wang Y, Zhao Z, Gao G, Huang S-X, Kang Q, He X, Lin S, Pang X, Deng Z, Tao M. 2016. Functional Genome Mining for Metabolites Encoded by Large Gene Clusters through Heterologous

Expression of a Whole-Genome Bacterial Artificial Chromosome Library in Streptomyces spp. Appl Environ Microbiol 82:5795–5805.

207. Ho CH, Piotrowski J, Dixon SJ, Baryshnikova A, Costanzo M, Boone C. 2011. Combining functional genomics and chemical biology to identify targets of bioactive compounds. Curr Opin Chem Biol 15:66–78.

208. Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264.

209. Przybyla L, Gilbert LA. 2021. A new era in functional genomics screens. Nat Rev Genet 1–15.

210. Montague MG, Lartigue C, Vashee S. 2012. Synthetic genomics: potential and limitations. Curr Opin Biotechnol 23:659–665.

211. Wang L, Jiang S, Chen C, He W, Wu X, Wang F, Tong T, Zou X, Li Z, Luo J, Deng Z, Chen S. 2018. Synthetic Genomics: From DNA Synthesis to Genome Design. Angew Chem Int Ed 57:1748–1756.

212. Handelsman J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. Microbiol Mol Biol Rev 68:669–685.

213. Culligan EP, Sleator RD, Marchesi JR, Hill C. 2014. Metagenomics and novel gene discovery. Virulence 5:399–412.

214. Hugenholtz P, Tyson GW. 2008. Metagenomics. Nature 455:481–483.

215. Kumar PS, Dabdoub SM, Ganesan SM. 2021. Probing periodontal microbial dark matter using metataxonomics and metagenomics. Periodontol 2000 85:12–27.

216. Hedlund BP, Dodsworth JA, Murugapiran SK, Rinke C, Woyke T. 2014. Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." Extremophiles 18:865–875.

217. Cook DJ, Finnigan JD, Cook K, Black GW, Charnock SJ. 2016. Cytochromes P450: History, Classes, Catalytic Mechanism, and Industrial Application. Adv Protein Chem Struct Biol 105:105–126.

218. Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. Genome Biol 1:reviews3003.1.

219. Guengerich FP, Martin MV, Sohl CD, Cheng Q. 2009. Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. Nat Protoc 4:1245–1251.

220. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187–D191.

221. Lamb DC, Lei L, Warrilow AGS, Lepesheva GI, Mullins JGL, Waterman MR, Kelly SL. 2009. The First Virally Encoded Cytochrome P450. J Virol 83:8266–8269.

222. Nelson DR. 2006. Cytochrome P450 Nomenclature, 2004, p. 1–10. *In* Phillips, IR, Shephard, EA (eds.), Cytochrome P450 Protocols. Humana Press, Totowa, NJ.

223. Nebert DW, Wikvall K, Miller WL. 2013. Human cytochromes P450 in health and disease. Philos Trans R Soc B Biol Sci 368:20120431.

224. Lamb DC, Follmer AH, Goldstone JV, Nelson DR, Warrilow AG, Price CL, True MY, Kelly SL, Poulos TL, Stegeman JJ. 2019. On the occurrence of cytochrome P450 in viruses. Proc Natl Acad Sci 116:12343–12352.

225. Degtyarenko KN. 1995. Structural domains of P450-containing monooxygenase systems. Protein Eng 8:737–747.

226. Guengerich FP. 2018. Mechanisms of Cytochrome P450-Catalyzed Oxidations. ACS Catal 8:10964–10976.

227. Jung ST, Lauchli R, Arnold FH. 2011. Cytochrome P450: taming a wild type enzyme. Curr Opin Biotechnol 22:809–817.

228. Thistlethwaite S, Jeffreys LN, Girvan HM, McLean KJ, Munro AW. 2021. A Promiscuous Bacterial P450: The Unparalleled Diversity of BM3 in Pharmaceutical Metabolism. 21. Int J Mol Sci 22:11380.

229. Kelly SL, Kelly DE. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? Philos Trans R Soc B Biol Sci 368:20120476.

230. Li R, Khaleeli N, Townsend CA. 2000. Expansion of the Clavulanic Acid Gene Cluster: Identification and In Vivo Functional Analysis of Three New Genes Required for Biosynthesis of Clavulanic Acid by Streptomyces clavuligerus. J Bacteriol 182:4087–4095.

231. Reddick JJ, Antolak SA, Raner GM. 2007. PksS from Bacillus subtilis is a cytochrome P450 involved in bacillaene metabolism. Biochem Biophys Res Commun 358:363–367.

232. MENDES MV, ANTÓN N, MARTÍN JF, APARICIO JF. 2005. Characterization of the polyene macrolide P450 epoxidase from Streptomyces natalensis that converts de-epoxypimaricin into pimaricin. Biochem J 386:57–62.

233. Stassi D, Donadio S, Staver MJ, Katz L. 1993. Identification of a Saccharopolyspora erythraea gene required for the final hydroxylation step in erythromycin biosynthesis. J Bacteriol 175:182–189.

234. Chen S, Mao X, Shen Y, Zhou Y, Li J, Wang L, Tao X, Yang L, Wang Y, Zhou X, Deng Z, Wei D. 2009. Tailoring the P450 Monooxygenase Gene for FR-008/Candicidin Biosynthesis. Appl Environ Microbiol 75:1778–1781.

235. Xie Y, Wang B, Liu J, Zhou J, Ma J, Huang H, Ju J. 2012. Identification of the Biosynthetic Gene Cluster and Regulatory Cascade for the Synergistic Antibacterial Antibiotics Griseoviridin and Viridogrisein in Streptomyces griseoviridis. ChemBioChem 13:2745–2757.

236. Schmartz PC, Wölfel K, Zerbe K, Gad E, El Tamany ES, Ibrahim HK, Abou-Hadeed K, Robinson JA. 2012. Substituent Effects on the Phenol Coupling Reaction Catalyzed by the Vancomycin Biosynthetic P450 Enzyme OxyB. Angew Chem Int Ed 51:11468–11472.

237. Peschke M, Gonsior M, Süssmuth RD, Cryle MJ. 2016. Understanding the crucial interactions between Cytochrome P450s and non-ribosomal peptide synthetases during glycopeptide antibiotic biosynthesis. Curr Opin Struct Biol 41:46–53.

238. Barry SM, Kers JA, Johnson EG, Song L, Aston PR, Patel B, Krasnoff SB, Crane BR, Gibson DM, Loria R, Challis GL. 2012. Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. 10. Nat Chem Biol 8:814–816.

239. Tomita H, Katsuyama Y, Minami H, Ohnishi Y. 2017. Identification and characterization of a bacterial cytochrome P450 monooxygenase catalyzing the 3-nitration of tyrosine in rufomycin biosynthesis. J Biol Chem 292:15859–15869.

240. Kelly WL, Townsend CA. 2002. Role of the cytochrome P450 NocL in nocardicin A biosynthesis. J Am Chem Soc 124:8186–8187.

241. Foulston LC, Bibb MJ. 2010. Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. Proc Natl Acad Sci 107:13461–13466.

242. 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.

243. Zheng Q, Wang S, Liao R, Liu W. 2016. Precursor-Directed Mutational Biosynthesis Facilitates the Functional Assignment of Two Cytochromes P450 in Thiostrepton Biosynthesis. ACS Chem Biol 11:2673–2678.

244. Zhang Z, Hudson GA, Mahanta N, Tietz JI, van der Donk WA, Mitchell DA. 2016. Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin. J Am Chem Soc 138:15511–15514.

245. Poulos TL, Finzel BC, Gunsalus IC, Wagner GC, Kraut J. 1985. The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450. J Biol Chem 260:16122–16130.

246. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, Benson DE, Sweet RM, Ringe D, Petsko GA, Sligar SG. 2000. The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. Science 287:1615–1622.

247. M. Podust L, H. Sherman D. 2012. Diversity of P450 enzymes in the biosynthesis of natural products. Nat Prod Rep 29:1251–1266.

248. Poulos TL. 2003. Cytochrome P450 flexibility. Proc Natl Acad Sci 100:13121–13122.

249. Isin EM, Guengerich FP. 2007. Complex reactions catalyzed by cytochrome P450 enzymes. Biochim Biophys Acta BBA - Gen Subj 1770:314–329.

250. Denisov IG, Makris TM, Sligar SG, Schlichting I. 2005. Structure and Chemistry of Cytochrome P450. Chem Rev 105:2253–2278.

251. Sligar SG, Cinti DL, Gibson GG, Schenkman JB. 1979. Spin state control of the hepatic cytochrome P450 redox potential. Biochem Biophys Res Commun 90:925–932.

252. Li Z, Jiang Y, Guengerich FP, Ma L, Li S, Zhang W. 2020. Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. J Biol Chem 295:833–849.

253. Black SD, Martin ST. 1994. Evidence for conformational dynamics and molecular aggregation in cytochrome P450 102 (BM-3). Biochemistry 33:12056–12062.

254. Ruettinger RT, Fulco AJ. 1981. Epoxidation of unsaturated fatty acids by a soluble cytochrome P-450-dependent system from Bacillus megaterium. J Biol Chem 256:5728–5734.

255. Fulco AJ, Kim BH, Matson RS, Owers Narhi L, Ruettinger RT. 1983. Nonsubstrate induction of a soluble bacterial cytochrome P-450 monooxygenase by phenobarbital and its analogs. Mol Cell Biochem 53:155–161.

256. Ariyasu S, Stanfield JK, Aiba Y, Shoji O. 2020. Expanding the applicability of cytochrome P450s and other haemoproteins. Curr Opin Chem Biol 59:155–163.

257. Whitehouse CJC, Bell SG, Wong L-L. 2012. P450BM3 (CYP102A1): connecting the dots. Chem Soc Rev 41:1218–1260.

258. Zuo R, Zhang Y, Jiang C, Hackett JC, Loria R, Bruner SD, Ding Y. 2017. Engineered P450 biocatalysts show improved activity and regio-promiscuity in aromatic nitration. Sci Rep 7:842.

259. Acevedo-Rocha CG, Hoebenreich S, Reetz MT. 2014. Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. Methods Mol Biol Clifton NJ 1179:103–128.

260. Mthethwa BC, Chen W, Ngwenya ML, Kappo AP, Syed PR, Karpoormath R, Yu J-H, Nelson DR, Syed K. 2018. Comparative Analyses of Cytochrome P450s and Those Associated with Secondary Metabolism in Bacillus Species. Int J Mol Sci 19:3623.

261. Cryle MJ, Bell SG, Schlichting I. 2010. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from Bacillus subtilis: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. Biochemistry 49:7282–7296.

262. Girhard M, Klaus T, Khatri Y, Bernhardt R, Urlacher VB. 2010. Characterization of the versatile monooxygenase CYP109B1 from Bacillus subtilis. Appl Microbiol Biotechnol 87:595–607.

263. Geronimo I, Denning CA, Rogers WE, Othman T, Huxford T, Heidary DK, Glazer EC, Payne CM. 2016. Effect of Mutation and Substrate Binding on the Stability of Cytochrome P450BM3 Variants. Biochemistry 55:3594–3606.

264. Bank RPD. RCSB PDB - 7OW9: Crystal structure of a staphylococcal orthologue of CYP134A1 (CYPX) in complex with Cyclo-L-leucyl-L-leucine. https://www.rcsb.org/structure/7OW9. Retrieved 21 December 2021.

265. Bank RPD. RCSB PDB - 4YZR: Bacillus subtilis 168 Bacillaene Polyketide Synthase (PKS) Cytochrome P450 PksS. https://www.rcsb.org/structure/4YZR. Retrieved 21 December 2021.

266. Jóźwik IK, Kiss FM, Gricman Ł, Abdulmughni A, Brill E, Zapp J, Pleiss J, Bernhardt R, Thunnissen AWH. 2016. Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from Bacillus megaterium. Febs J 283:4128–4148.

267. Scotti C, Piatti M, Cuzzoni A, Perani P, Tognoni A, Grandi G, Galizzi A, Albertini AM. 1993. A Bacillus subtilis large ORF coding for a polypeptide highly similar to polyketide synthases. Gene 130:65–71.

268. Warman AJ, Roitel O, Neeli R, Girvan HM, Seward HE, Murray SA, McLean KJ, Joyce MG, Toogood H, Holt RA, Leys D, Scrutton NS, Munro AW. 2005. Flavocytochrome P450 BM3: an update on structure and mechanism of a biotechnologically important enzyme. Biochem Soc Trans 33:747–753.

269. Omura K, Aiba Y, Onoda H, Stanfield JK, Ariyasu S, Sugimoto H, Shiro Y, Shoji O, Watanabe Y. 2018. Reconstitution of full-length P450BM3 with an artificial metal complex by utilising the transpeptidase Sortase A. Chem Commun 54:7892–7895.

270. Furuya T, Nishi T, Shibata D, Suzuki H, Ohta D, Kino K. 2008. Characterization of Orphan Monooxygenases by Rapid Substrate Screening Using FT-ICR Mass Spectrometry. Chem Biol 15:563–572.

271. Abdulmughni A, Jóźwik IK, Brill E, Hannemann F, Thunnissen A-MWH, Bernhardt R. 2017. Biochemical and structural characterization of CYP109A2, a vitamin D3 25-hydroxylase from Bacillus megaterium. FEBS J 284:3881–3894.

272. Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong L-L, Zhou W, Bell SG. 2015. The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from Bacillus subtilis. Mol Biosyst 11:869–881.

273. Dean R, Van Kan J a. L, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J, Foster GD. 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol 13:414–430.

274. Almeida F, Rodrigues ML, Coelho C. 2019. The Still Underestimated Problem of Fungal Diseases Worldwide. Front Microbiol 10:214.

275. Hahn M. 2014. The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J Chem Biol 7:133–141.

276. Gerbore J, Benhamou N, Vallance J, Le Floch G, Grizard D, Regnault-Roger C, Rey P. 2014. Biological control of plant pathogens: advantages and limitations seen through the case study of Pythium oligandrum. Environ Sci Pollut Res 21:4847–4860.

277. Köhl J, Kolnaar R, Ravensberg WJ. 2019. Mode of Action of Microbial Biological Control Agents Against Plant Diseases: Relevance Beyond Efficacy. Front Plant Sci 10:845.

278. Syed Ab Rahman SF, Singh E, Pieterse CMJ, Schenk PM. 2018. Emerging microbial biocontrol strategies for plant pathogens. Plant Sci 267:102–111.

279. Jiang C-H, Liao M-J, Wang H-K, Zheng M-Z, Xu J-J, Guo J-H. 2018. Bacillus velezensis, a potential and efficient biocontrol agent in control of pepper gray mold caused by Botrytis cinerea. Biol Control 126:147–157.

280. Saleh AE, Ul-Hassan Z, Zeidan R, Al-Shamary N, Al-Yafei T, Alnaimi H, Higazy NS, Migheli Q, Jaoua S. 2021. Biocontrol Activity of Bacillus megaterium BM344-1 against Toxigenic Fungi. ACS Omega 6:10984–10990.

281. Fira D, Dimkić I, Berić T, Lozo J, Stanković S. 2018. Biological control of plant pathogens by Bacillus species. J Biotechnol 285:44–55.

282. Marahiel MA, Essen L -O. 2009. Chapter 13 Nonribosomal Peptide Synthetases: Mechanistic and Structural Aspects of Essential Domains, p. 337–351. *In* Methods in Enzymology. Academic Press.

283. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16:115–125.

284. Kiesewalter HT, Lozano-Andrade CN, Wibowo M, Strube ML, Maróti G, Snyder D, Jørgensen TS, Larsen TO, Cooper VS, Weber T, Kovács ÁT. Genomic and Chemical Diversity of Bacillus subtilis Secondary Metabolites against Plant Pathogenic Fungi. mSystems 6:e00770-20.

285. Moreno-Velandia CA, Ongena M, Cotes AM. 2021. Effects of fengycins and iturins on Fusarium oxysporum f. sp. physali and root colonization by Bacillus velezensis Bs006 protect golden berry against vascular wilt. Phytopathology https://doi.org/10.1094/PHYTO-01-21-0001-R.

286. Gu Q, Yang Y, Yuan Q, Shi G, Wu L, Lou Z, Huo R, Wu H, Borriss R, Gao X. Bacillomycin D Produced by Bacillus amyloliquefaciens Is Involved in the Antagonistic Interaction with the Plant-Pathogenic Fungus Fusarium graminearum. Appl Environ Microbiol 83:e01075-17.

287. Xie Y, Peng Q, Ji Y, Xie A, Yang L, Mu S, Li Z, He T, Xiao Y, Zhao J, Zhang Q. 2021. Isolation and Identification of Antibacterial Bioactive Compounds From Bacillus megaterium L2. Front Microbiol 12:662.

288. Wang Y, Liang J, Zhang C, Wang L, Gao W, Jiang J. 2020. Bacillus megaterium WL-3 Lipopeptides Collaborate Against Phytophthora infestans to Control Potato Late Blight and Promote Potato Plant Growth. Front Microbiol 11.

289. QIAGEN CLC Genomics Workbench | QIAGEN Digital Insights.

290. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

291. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288.

292. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617.

293. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale biosynthetic diversity. 1. Nat Chem Biol 16:60–68.

294. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432.

295. Volpon L, Besson F, Lancelin J-M. 1999. NMR structure of active and inactive forms of the sterol-dependent antifungal antibiotic bacillomycin L. Eur J Biochem 264:200–210.

296. Pathak KV, Keharia H, Gupta K, Thakur SS, Balaram P. 2012. Lipopeptides from the banyan endophyte, Bacillus subtilis K1: mass spectrometric characterization of a library of fengycins. J Am Soc Mass Spectrom 23:1716–1728.

297. Savadogo A, Tapi A, Chollet M, Wathelet B, Traoré AS, Jacques Ph. 2011. Identification of surfactin producing strains in Soumbala and Bikalga fermented condiments using Polymerase Chain Reaction and Matrix Assisted Laser Desorption/Ionization-Mass Spectrometry methods. Int J Food Microbiol 151:299–306.

298. Ma Y, Kong Q, Qin C, Chen Y, Chen Y, Lv R, Zhou G. 2016. Identification of lipopeptides in Bacillus megaterium by two-step ultrafiltration and LC–ESI–MS/MS. AMB Express 6:79.

299. Chen Y, Liu SA, Mou H, Ma Y, Li M, Hu X. 2017. Characterization of Lipopeptide Biosurfactants Produced by Bacillus licheniformis MB01 from Marine Sediments. Front Microbiol 8.

300. Romano A, Vitullo D, Senatore M, Lima G, Lanzotti V. 2013. Antifungal Cyclic Lipopeptides from Bacillus amyloliquefaciens Strain BO5A. J Nat Prod 76:2019–2025.

301. Khan N, Maymon M, Hirsch AM. 2017. Combating Fusarium Infection Using Bacillus-Based Antimicrobials. 4. Microorganisms 5:75.

302. Bais HP, Fall R, Vivanco JM. 2004. Biocontrol of Bacillus subtilis against Infection of Arabidopsis Roots by Pseudomonas syringae Is Facilitated by Biofilm Formation and Surfactin Production. Plant Physiol 134:307–319.

303. Jin P, Wang H, Tan Z, Xuan Z, Dahar GY, Li QX, Miao W, Liu W. 2020. Antifungal mechanism of bacillomycin D from Bacillus velezensis HN-2 against Colletotrichum gloeosporioides Penz. Pestic Biochem Physiol 163:102–107.

304. Luo C, Zhou H, Zou J, Wang X, Zhang R, Xiang Y, Chen Z. 2015. Bacillomycin L and surfactin contribute synergistically to the phenotypic features of Bacillus subtilis 916 and the biocontrol of rice sheath blight induced by Rhizoctonia solani. Appl Microbiol Biotechnol 99:1897–1910.

305. Zhao H, Shao D, Jiang C, Shi J, Li Q, Huang Q, Rajoka MSR, Yang H, Jin M. 2017. Biological activity of lipopeptides from Bacillus. Appl Microbiol Biotechnol 101:5951–5960.

306. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16:115–125.

307. Luo Chuping, Liu Xuehui, Zhou Huafei, Wang Xiaoyu, Chen Zhiyi, Kelly R. M. 2015. Nonribosomal Peptide Synthase Gene Clusters for Lipopeptide Biosynthesis in Bacillus subtilis 916 and Their Phenotypic Functions. Appl Environ Microbiol 81:422–431.

308. Jasim B, Sreelakshmi KS, Mathew J, Radhakrishnan EK. 2016. Surfactin, Iturin, and Fengycin Biosynthesis by Endophytic Bacillus sp. from Bacopa monnieri. Microb Ecol 72:106–119.

309. Hu LB, Shi ZQ, Zhang T, Yang ZM. 2007. Fengycin antibiotics isolated from B-FS01 culture inhibit the growth of Fusarium moniliforme Sheldon ATCC 38932. FEMS Microbiol Lett 272:91–98.

310. Romero D, de Vicente A, Rakotoaly RH, Dufour SE, Veening J-W, Arrebola E, Cazorla FM, Kuipers OP, Paquot M, Pérez-García A. 2007. The Iturin and Fengycin Families of Lipopeptides Are Key Factors in Antagonism of Bacillus subtilis Toward Podosphaera fusca. Mol Plant-Microbe Interactions® 20:430–440.

311. Bebber DP, Gurr SJ. 2015. Crop-destroying fungal and oomycete pathogens challenge food security. Fungal Genet Biol 74:62–64.

312. Doehlemann G, Ökmen B, Zhu W, Sharon A. 2017. Plant Pathogenic Fungi. Microbiol Spectr 5:5.1.14.

313. de Chaves MA, Reginatto P, da Costa BS, de Paschoal RI, Teixeira ML, Fuentefria AM. 2022. Fungicide Resistance in Fusarium graminearum Species Complex. Curr Microbiol 79:62.

314. Ogoshi A. 1987. Ecology and Pathogenicity of Anastomosis and Intraspecific Groups of Rhizoctonia Solani Kuhn. Annu Rev Phytopathol 25:125–143.

315. Chung WH, Chung WC, Ting PF, Ru CC, Huang HC, Huang JW. 2009. Nature of Resistance to Methyl Benzimidazole Carbamate Fungicides in Fusarium oxysporum f.sp. lilii and F. oxysporum f.sp. gladioli in Taiwan. J Phytopathol 157:742–747.

316. Thind TS. 2012. Fungicide Resistance in Crop Protection: Risk and Management. CABI.

317. Brent K, Hollomon D. 2007. FUNGICIDE RESISTANCE : THE ASSESSMENT OF RISK. https://www.semanticscholar.org/paper/FUNGICIDE-RESISTANCE-%3A-THE-ASSESSMENT-OF-RISK-Brent-Hollomon/b9b7aa046d5f3437b898c855bbb9a0c8addb81b1. Retrieved 25 November 2021.

318. Sm Z, M J, N X, Rn M. 2020. Antibiotics and antibiotic resistant genes (ARGs) in groundwater: A global review on dissemination, sources, interactions, environmental and human health risks. Water Res 187:116455–116455.

319. Hahn M. 2014. The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J Chem Biol 7:133–141.

320. Becher R, Hettwer U, Karlovsky P, Deising HB, Wirsel SGR. 2010. Adaptation of Fusarium graminearum to tebuconazole yielded descendants diverging for levels of fitness, fungicide resistance, virulence, and mycotoxin production. Phytopathology 100:444–453.

321. Zhao C, Li Y, Liang Z, Gao L, Han C, Wu X. 2022. Molecular Mechanisms Associated with the Resistance of Rhizoctonia solani AG-4 Isolates to the Succinate Dehydrogenase Inhibitor Thifluzamide. Phytopathology® 112:567–578.

322. Duval RE, Grare M, Demoré B. 2019. Fight Against Antimicrobial Resistance: We Always Need New Antibacterials but for Right Bacteria. Molecules 24.

323. Syed Ab Rahman SF, Singh E, Pieterse CMJ, Schenk PM. 2018. Emerging microbial biocontrol strategies for plant pathogens. Plant Sci 267:102–111.

324. Kettles GJ, Luna E. 2019. Food security in 2044: How do we control the fungal threat? Fungal Biol 123:558–564.

325. Tyc Olaf, Song C, Dickschat JS, Vos M, Garbeva P. 2017. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. Trends Microbiol 25:280–292.

326. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16:115–125.

327. Ongena M, Jacques P. 2008. Bacillus lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol 16:115–125.

328. Khalid A, Arshad M, Zahir ZA, Khaliq A. 1997. Potential of plant growth promoting Rhizobacteria for enhancing wheat (Triticum aestivum L.) yield. J Anim Plant Sci Pak.

329. Xu D, Côté J-C. 2003. Phylogenetic relationships between Bacillus species and related genera inferred from comparison of 3’ end 16S rDNA and 5’ end 16S-23S ITS nucleotide sequences. Int J Syst Evol Microbiol 53:695–704.

330. Çakmakçi R, Dönmez F, Aydın A, Şahin F. 2006. Growth promotion of plants by plant growth-promoting rhizobacteria under greenhouse and two different field soil conditions. Soil Biol Biochem 38:1482–1487.

331. Amaya-Gómez CV, Porcel M, Mesa-Garriga L, Gómez-Álvarez MI. 2020. A Framework for the Selection of Plant Growth-Promoting Rhizobacteria Based on Bacterial Competence Mechanisms. Appl Environ Microbiol https://doi.org/10.1128/AEM.00760-20.

332. Bach E, Seger GD dos S, Fernandes G de C, Lisboa BB, Passaglia LMP. 2016. Evaluation of biological control and rhizosphere competence of plant growth promoting bacteria. Appl Soil Ecol 99:141–149.

333. Seemann T. 2022. Barrnap. https://github.com/tseemann/barrnap. Retrieved 9 June 2022.

334. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549.

335. Berlanga-Clavero MV, Molina-Santiago C, Caraballo-Rodríguez AM, Petras D, Díaz-Martínez L, Pérez-García A, de Vicente A, Carrión VJ, Dorrestein PC, Romero D. 2022. Bacillus subtilis biofilm matrix components target seed oil bodies to promote growth and anti-fungal resistance in melon. Nat Microbiol 1–15.

336. Zeng Q, Xie J, Li Y, Chen X, Gu X, Yang P, Hu G, Wang Q. 2021. Organization, evolution and function of fengycin biosynthesis gene clusters in the Bacillus amyloliquefaciens group. Phytopathol Res 3:26.

337. Li H, Han X, Dong Y, Xu S, Chen C, Feng Y, Cui Q, Li W. 2021. Bacillaenes: Decomposition Trigger Point and Biofilm Enhancement in Bacillus. ACS Omega 6:1093–1098.

338. Zimmerman SB, Schwartz CD, Monaghan RL, Pelak BA, Weissberger B, Gilfillan EC, Mochales S, Hernandez S, Currie SA, Tejera E. 1987. Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by Bacillus subtilis. I. Production, taxonomy and antibacterial activity. J Antibiot (Tokyo) 40:1677–1681.

339. Weller DM. 1988. Biological Control of Soilborne Plant Pathogens in the Rhizosphere with Bacteria. Annu Rev Phytopathol 26:379–407.

340. Mazzola M, Freilich S. 2017. Prospects for Biological Soilborne Disease Control: Application of Indigenous Versus Synthetic Microbiomes. Phytopathology® 107:256–263.

341. Miljaković D, Marinković J, Balešević-Tubić S. 2020. The Significance of Bacillus spp. in Disease Suppression and Growth Promotion of Field and Vegetable Crops. 7. Microorganisms 8:1037.

342. Dadrasnia A, Usman MM, Omar R, Ismail S, Abdullah R. 2020. Potential use of Bacillus genus to control of bananas diseases: Approaches toward high yield production and sustainable management. J King Saud Univ - Sci 32:2336–2342.

343. Akila R, Rajendran L, Harish S, Saveetha K, Raguchander T, Samiyappan R. 2011. Combined application of botanical formulations and biocontrol agents for the management of Fusarium oxysporum f. sp. cubense (Foc) causing Fusarium wilt in banana. Biol Control 57:175–183.

344. Zhang N, Wu K, He X, Li S, Zhang Z, Shen B, Yang X, Zhang R, Huang Q, Shen Q. 2011. A new bioorganic fertilizer can effectively control banana wilt by strong colonization with Bacillus subtilis N11. Plant Soil 344:87–97.

345. Ye M, Tang X, Yang R, Zhang H, Li F, Tao F, Li F, Wang Z. 2018. Characteristics and Application of a Novel Species of Bacillus: Bacillus velezensis. ACS Chem Biol 13:500–505.

346. Yan H, Qiu Y, Yang S, Wang Y, Wang K, Jiang L, Wang H. 2020. Antagonistic Activity of Bacillus velezensis SDTB038 against Phytophthora infestans in Potato. Plant Dis PDIS-08-20-1666-RE.

347. Fan B, Wang C, Song X, Ding X, Wu L, Wu H, Gao X, Borriss R. 2018. Bacillus velezensis FZB42 in 2018: The Gram-Positive Model Strain for Plant Growth Promotion and Biocontrol. Front Microbiol 9:2491.

348. Allard-Massicotte R, Tessier L, Lécuyer F, Lakshmanan V, Lucier J-F, Garneau D, Caudwell L, Vlamakis H, Bais HP, Beauregard PB. 2016. Bacillus subtilis Early Colonization of Arabidopsis thaliana Roots Involves Multiple Chemotaxis Receptors. mBio 7:e01664-16.

349. Chu J, Wang Y, Zhao B, Zhang X, Liu K, Mao L, Kalamiyets E. 2019. Isolation and identification of new antibacterial compounds from Bacillus pumilus. Appl Microbiol Biotechnol 103:8375–8381.

350. Ren J-H, Li H, Wang Y-F, Ye J-R, Yan A-Q, Wu X-Q. 2013. Biocontrol potential of an endophytic Bacillus pumilus JK-SX001 against poplar canker. Biol Control 67:421–430.

351. Moreno-Velandia CA, Ongena M, Cotes AM. 2021. Effects of Fengycins and Iturins on Fusarium oxysporum f. sp. physali and Root Colonization by Bacillus velezensis Bs006 Protect Golden Berry Against Vascular Wilt. Phytopathology PHYTO01210001R.

352. Zeriouh H, de Vicente A, Pérez-García A, Romero D. 2014. Surfactin triggers biofilm formation of Bacillus subtilis in melon phylloplane and contributes to the biocontrol activity. Environ Microbiol 16:2196–2211.

353. Omura T, Sato R. 1964. THE CARBON MONOXIDE-BINDING PIGMENT OF LIVER MICROSOMES. I. EVIDENCE FOR ITS HEMOPROTEIN NATURE. J Biol Chem 239:2370–2378.

354. Jung ST, Lauchli R, Arnold FH. 2011. Cytochrome P450: taming a wild type enzyme. Curr Opin Biotechnol 22:809–817.

355. Meunier B, de Visser SP, Shaik S. 2004. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. Chem Rev 104:3947–3980.

356. Bernhardt R. 2006. Cytochromes P450 as versatile biocatalysts. J Biotechnol 124:128–145.

357. Omura T. 1999. Forty Years of Cytochrome P450. Biochem Biophys Res Commun 266:690–698.

358. Munro AW, Girvan HM, McLean KJ. 2007. Variations on a (t)heme—novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. Nat Prod Rep 24:585–609.

359. Zanger UM, Schwab M. 2013. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther 138:103–141.

360. Anzenbacher P, Anzenbacherová E. 2001. Cytochromes P450 and metabolism of xenobiotics. Cell Mol Life Sci CMLS 58:737–747.

361. Guengerich FP, Munro AW. 2013. Unusual Cytochrome P450 Enzymes and Reactions. J Biol Chem 288:17065–17073.

362. Reddick JJ, Antolak SA, Raner GM. 2007. PksS from Bacillus subtilis is a cytochrome P450 involved in bacillaene metabolism. Biochem Biophys Res Commun 358:363–367.

363. Green AJ, Rivers SL, Cheesman M, Reid GA, Quaroni LG, Macdonald IDG, Chapman SK, Munro AW. 2001. Expression, purification and characterization of cytochrome P450 Biol: a novel P450 involved in biotin synthesis in Bacillus subtilis. JBIC J Biol Inorg Chem 6:523–533.

364. Lo N, Aj F. 1986. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. J Biol Chem 261:7160–7169.

365. Cryle MJ, Bell SG, Schlichting I. 2010. Structural and Biochemical Characterization of the Cytochrome P450 CypX (CYP134A1) from Bacillus subtilis: A Cyclo-l-leucyl-l-leucyl Dipeptide Oxidase. research-article. American Chemical Society. https://pubs.acs.org/doi/abs/10.1021/bi100910y. Retrieved 23 August 2020.

366. Budde M, Maurer SC, Schmid RD, Urlacher VB. 2004. Cloning, expression and characterisation of CYP102A2, a self-sufficient P450 monooxygenase from Bacillus subtilis. Appl Microbiol Biotechnol 66:180–186.

367. Dietrich M, Eiben S, Asta C, Do TA, Pleiss J, Urlacher VB. 2008. Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from Bacillus licheniformis. Appl Microbiol Biotechnol 79:931–940.

368. Matsunaga I, Ueda A, Fujiwara N, Sumimoto T, Ichihara K. 1999. Characterization of the ybdT gene product of Bacillus subtilis: Novel fatty acid β-hydroxylating cytochrome P450. Lipids 34:841–846.

369. Gustafsson MCU, Roitel O, Marshall KR, Noble MA, Chapman SK, Pessegueiro A, Fulco AJ, Cheesman MR, von Wachenfeldt C, Munro AW. 2004. Expression, Purification, and Characterization of Bacillus subtilis Cytochromes P450 CYP102A2 and CYP102A3: Flavocytochrome Homologues of P450 BM3 from Bacillus megaterium. Biochemistry 43:5474–5487.

370. Mthethwa BC, Chen W, Ngwenya ML, Kappo AP, Syed PR, Karpoormath R, Yu J-H, Nelson DR, Syed K. 2018. Comparative Analyses of Cytochrome P450s and Those Associated with Secondary Metabolism in Bacillus Species. 11. Int J Mol Sci 19:3623.

371. Grubbs KJ, Bleich RM, Maria KCS, Allen SE, Farag S, Team A, Shank EA, Bowers AA. 2017. Large-Scale Bioinformatics Analysis of Bacillus Genomes Uncovers Conserved Roles of Natural Products in Bacterial Physiology. mSystems 2.

372. Palaniappan K, Chen I-MA, Chu K, Ratner A, Seshadri R, Kyrpides NC, Ivanova NN, Mouncey NJ. 2020. IMG-ABC v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. Nucleic Acids Res 48:D422–D430.

373. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222-226.

374. Nelson DR. 2009. The Cytochrome P450 Homepage. Hum Genomics 4:59.

375. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13:2498–2504.

376. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

377. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245.

378. Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295–2300.

379. Song Y, DiMaio F, Wang RY-R, Kim D, Miles C, Brunette T, Thompson J, Baker D. 2013. High resolution comparative modeling with RosettaCM. Struct Lond Engl 1993 21.

380. Bender BJ, Cisneros A, Duran AM, Finn JA, Fu D, Lokits AD, Mueller BK, Sangha AK, Sauer MF, Sevy AM, Sliwoski G, Sheehan JH, DiMaio F, Meiler J, Moretti R. 2016. Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. Biochemistry 55:4748–4763.

381. Kim DE, Chivian D, Baker D. 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32:W526–W531.

382. Viklund H, Elofsson A. 2008. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinforma Oxf Engl 24:1662–1668.

383. Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407-410.

384. Eisenberg D, Lüthy R, Bowie JU. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404.

385. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. 2. J Appl Crystallogr 26:283–291.

386. Colovos C, Yeates TO. 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519.

387. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam

JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ. 2016. Gaussian˜16 Revision C.01.

388. Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461.

389. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. 2011. Open Babel: An open chemical toolbox. J Cheminformatics 3:33.

390. Varnado CL, Hertwig KM, Thomas R, Roberts JK, Goodwin DC. 2004. Properties of a novel periplasmic catalase-peroxidase from Escherichia coli O157:H7. Arch Biochem Biophys 421:166–174.

391. Quaroni LG, Seward HE, McLean KJ, Girvan HM, Ost TWB, Noble MA, Kelly SM, Price NC, Cheesman MR, Smith WE, Munro AW. 2004. Interaction of nitric oxide with cytochrome P450 BM3. Biochemistry 43:16416–16431.

392. Szaleniec M, Wojtkiewicz AM, Bernhardt R, Borowski T, Donova M. 2018. Bacterial steroid hydroxylases: enzyme classes, their functions and comparison of their catalytic mechanisms. Appl Microbiol Biotechnol 102:8153–8171.

393. Molohon KJ, Melby JO, Lee J, Evans BS, Dunbar KL, Bumpus SB, Kelleher NL, Mitchell DA. 2011. Structure Determination and Interception of Biosynthetic Intermediates for the Plantazolicin Class of Highly Discriminating Antibiotics. ACS Chem Biol 6:1307–1313.

394. Maciolek CM, Ma B, Menzel K, Laliberte S, Bateman K, Krolikowski P, Gibson CR. 2011. Novel Cytochrome P450-Mediated Ring Opening of the 1,3,4-Oxadiazole in Setileuton, a 5-Lipoxygenase Inhibitor. Drug Metab Dispos 39:763–770.

395. C. Whitehouse CJ, G. Bell S, Wong L-L. 2012. P450 BM3 (CYP102A1): connecting the dots. Chem Soc Rev 41:1218–1260.

396. Jeffreys LN, Poddar H, Golovanova M, Levy CW, Girvan HM, McLean KJ, Voice MW, Leys D, Munro AW. 2019. Novel insights into P450 BM3 interactions with FDA-approved antifungal azole drugs. 1. Sci Rep 9:1577.

397. Packer MS, Liu DR. 2015. Methods for the directed evolution of proteins. 7. Nat Rev Genet 16:379–394.

398. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. 7873. Nature 596:583–589.

399. Curtis V. 2015. Motivation to Participate in an Online Citizen Science Game: A Study of Foldit. Sci Commun 37:723–746.

Appendices

Appendix A: Workflow for evaluating *Bacillus* strains to identify candidate strains for biological agents. Figure 1.1 shows our comprehensive pipeline for genomic, antibiosis, and chemical analyses. Figure 1.2 shows steps toward evaluating selected *Bacillus* strains for the identification of candidate strains to develop commercially viable biological agents. Figure 1.3 shows a workflow for the investigation of structural and functional properties of *Bacillus* cytochromes P450 that are involved in secondary metabolite biosynthesis pathways.

Appendix B: MS^2 spectra of lipopeptides, fengycin, surfactin, bacillomycin L, and iturin produced by *Bacillus velezensis* species. MS^2 spectra were produced by a multi-stage MS^2 analyzer from the precursor ions generated by a positive-mode electrospray ionization source. Shown here are MS^2 spectra of five bacillomycin L (an iturin) derivatives, nine fengycin, and five surfactin derivatives that were produced by *B. velezensis* JJ334, whereas four iturin derivatives were produced by *B. velezensis* JJ951. MS^2 spectra of one bacillomycin L, one fengycin, and one surfactin derivative with the highest m/z were not generated even though they were identified in MS spectra.

Appendix C: Extracted ion chromatograms (EIC), MS spectra, MS^2 spectra, and elemental composition analysis of three polyketides with antibacterial activity were produced by *Bacillus velezensis* strain JJ334. These include bacillaene, difficidin, and macrolactin W. In addition, four derivatives of bacillaene and difficidin were identified based on the MS^2 and elemental composition analysis however their exact structure is yet to be determined. These were denoted as putative oxybacillaene, putative oxydifficidin, putative difficidin 1, and putative difficidin 2.

Appendix D: Quantitative antibiosis evaluation of eighteen *Bacillus velezensis* strains for intraspecies inhibition/resistance and antifungal activity. Antibiosis scores (0-4) derived from the observed zone of inhibition (ZOI) were shown as percent inhibition for both intraspecies inhibition/resistance and antifungal activity.

Appendix E: The extent of antimicrobial secondary metabolites production that is derived from the peak area of extracted ion chromatogram (EIC). Peak area for prominent lipopeptides iturin (or bacillomycin L), fengycin, surfactin, and polyketides, bacillaene, difficidin, and macrolactin W are compared.

Appendix F: Chemical properties of secondary metabolites produced by 18 *Bacillus velezensis* strains. UV-vis absorption, absorption-based HPLC elution profile, and total ion current (TIC)-based LC profile are shown.

Appendix G: Amino acid sequences and homology model of representative BGC-affiliated cytochromes P450. Selected protein sequences of five different BGC-affiliated P450 families that

were used to construct high-resolution homology model. A high-resolution homology model of a full-length *Ba*CYP102A2. A schematic of our *Ba*CYP102A2-gene containing construct produced for heterologous expression of *Ba*CYP102A2-encoded enzyme.

Appendix A

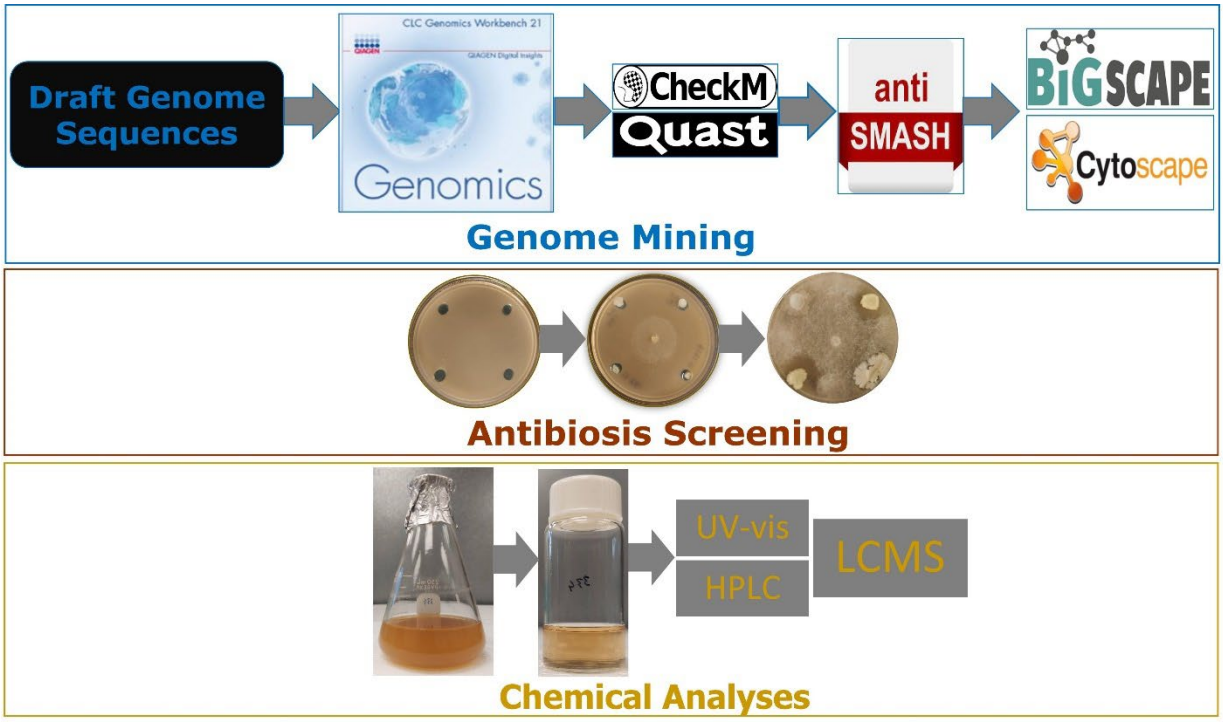


Figure 1.1: Workflow for genomic, antibiosis, and chemical analyses performed for evaluating antibiosis potential of PGPR *Bacillus* species. Genomic analyses are performed to predict biosynthetic gene clusters and secondary metabolites produced by *Bacillus* species. Bioinformatics analyses employing CLC genomics, Quast, CheckM, antiSMASH, BiG-SCAPE, and Cytoscape are carried out to predict the ability to produce antimicrobial secondary metabolites. Antibiosis screenings are performed to identify the most promising strains with strong antibiosis ability against plant pathogens. Chemical analyses are carried out on extracted secondary metabolites from the culture of *Bacillus* species using UV-vis absorption, HPLC elution, and high-resolution LC-MS identification.

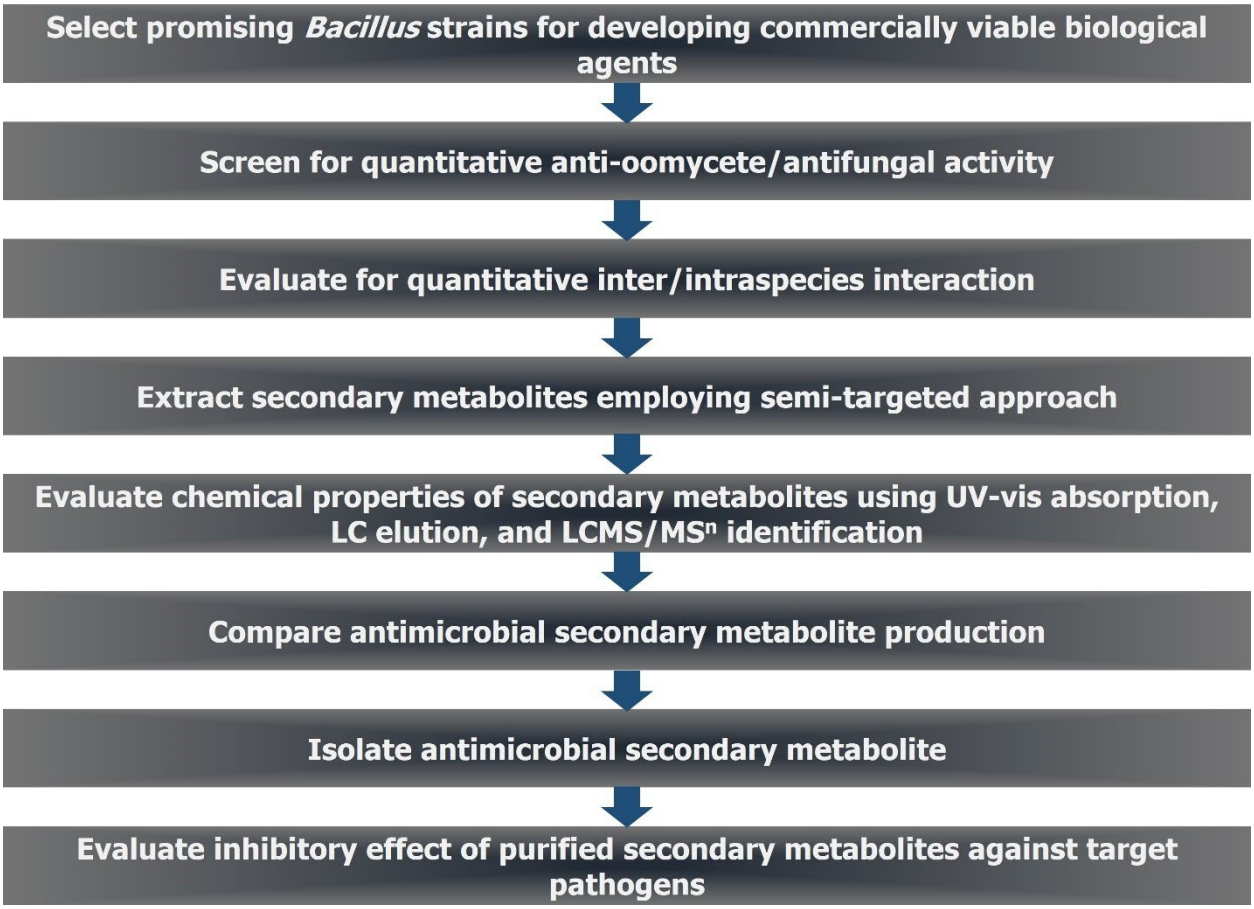


Figure 1.2: Workflow for identifying most promising *Bacillus strains* for developing commercially viable biological agents. Candidate PGPR *Bacillus* strains with antibiosis activity against plant pathogens are evaluated for quantitative antibiosis activity against oomycetes/fungi, inter/intraspecies inhibition/resistance activity, and ability to produce antimicrobial secondary metabolites.

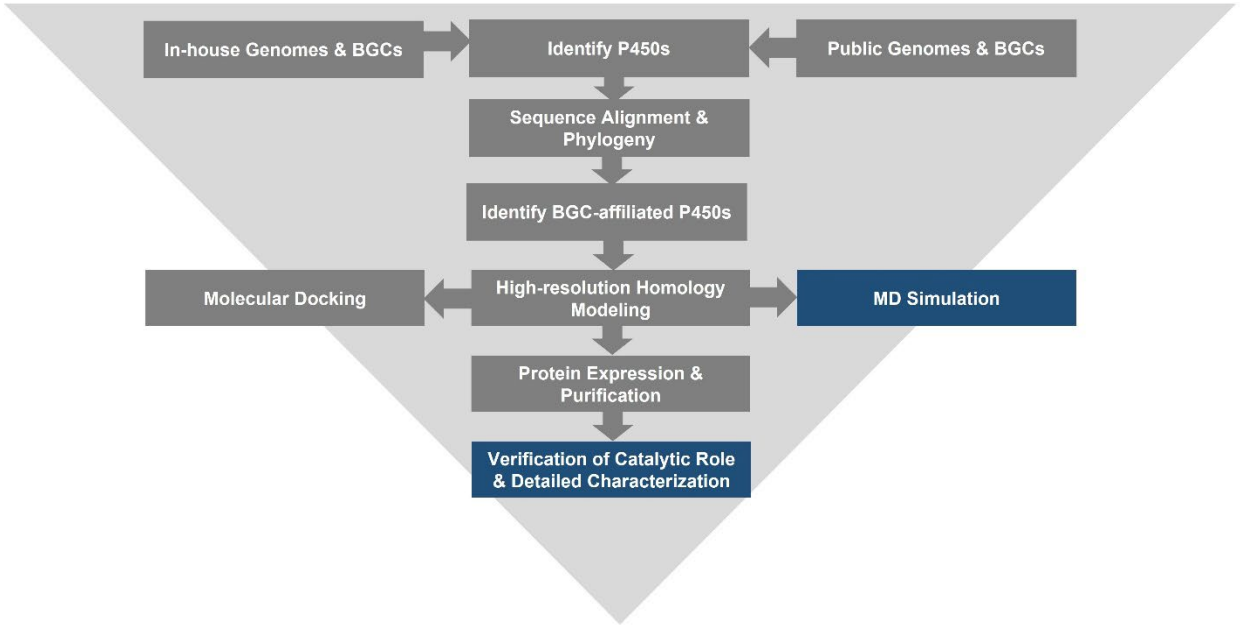


Figure 1.3: Workflow for our approach to investigating BGC-affiliated P450s. Amino acid sequences of BGC-affiliated P450s are collected from either in-house or publicly available genome sequences. Subsequently, P450 annotation, phylogenic analysis, and BGC clustering are performed to gain insight into the diversity of P450s and respective secondary metabolite biosynthetic pathways. Representative P450s are selected in order to construct high-resolution homology models followed by molecular dynamics simulation to investigate structural and functional features. To understand the functional role of BGC-affiliated P450s, representative P450s are further investigated for heterologous expression in order to evaluate substrate binding and oxidative transformation of P450-catalyzed reactions with the respective substrate. Further, our workflow includes molecular dynamics simulation to understand changes in P450's active site over time in the presence and absence of candidate substrate.

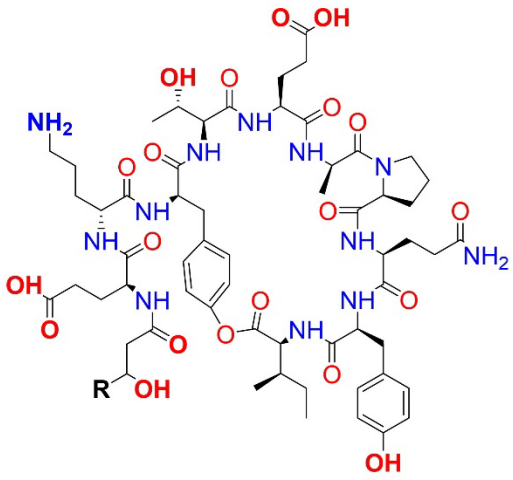
Appendix B

Table 2.1. Fengycin ions generated by MS and MS^2 were used for unequivocal assignment of fengycin derivatives.

Precursor ion (MH^+)	Diagnostic ion (A/B)	Diagnostic Ion (C)	Fengycin Derivative	aa position 6	aa position 10
1435.7730	966.4499/1080.5294	356.2407	C14 fengycin A	Ala	Ile
1449.7878	966.4497/1080.5299	370.2564	C15 fengycin A	Ala	Ile
	952.4338/1066.5136	384.2717	C16 fengycin A2	Ala	Val
1463.8028	966.4498/1080.5306	384.2700	C16 fengycin A	Ala	Ile
	952.4378/1066.5146	398.2852	C17 fengycin A2	Ala	Val
1477.8185	966.4502/1080.5254	398.2860	C17 fengycin A	Ala	Ile
1491.8332	994.4809/1108.5593	384.2726	C16 fengycin B	Val	Ile
	980.4677/1094.5444	398.2726	C17 fengycin B2	Val	Val
1505.8471	994.4833/1108.5606	398.2873	C17 fengycin B	Val	Ile
1519.8666	994.4780/1108.5611	412.3039	C18 fengycin B	Val	Ile
1533.8819	980.4987/1094.5770	ND[1]	C20 fengycin B2	Val	Val
1547.8966	980.4988/1094.5783	ND[1]	C21 fengycin B2	Val	Val
1561.9106	ND[1]				

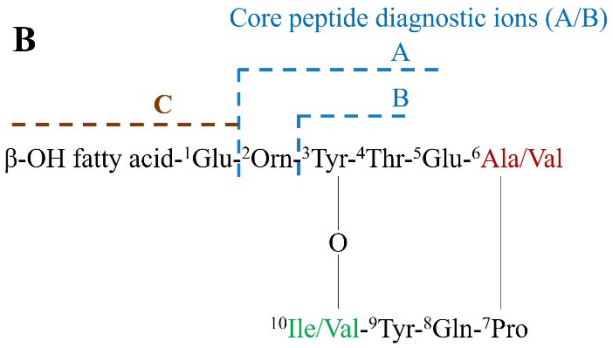
[1]ND=Not Detected.

A



Fengycin

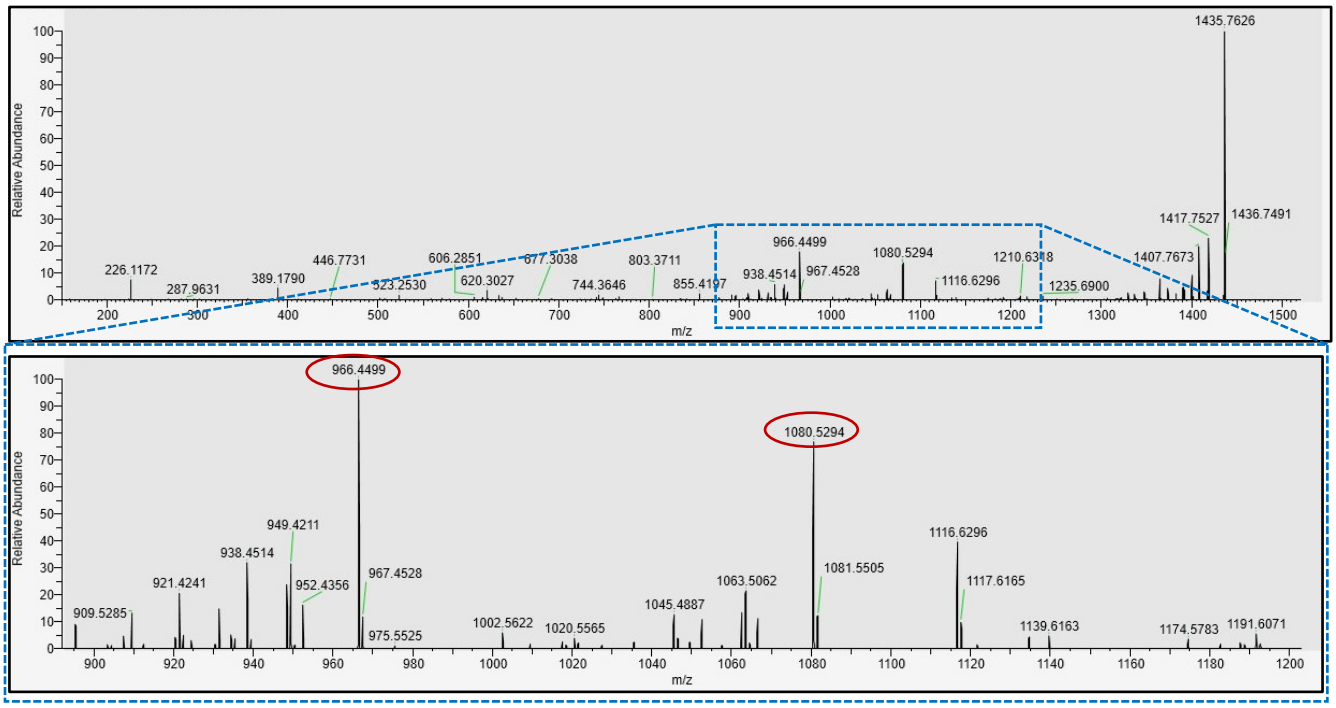
B



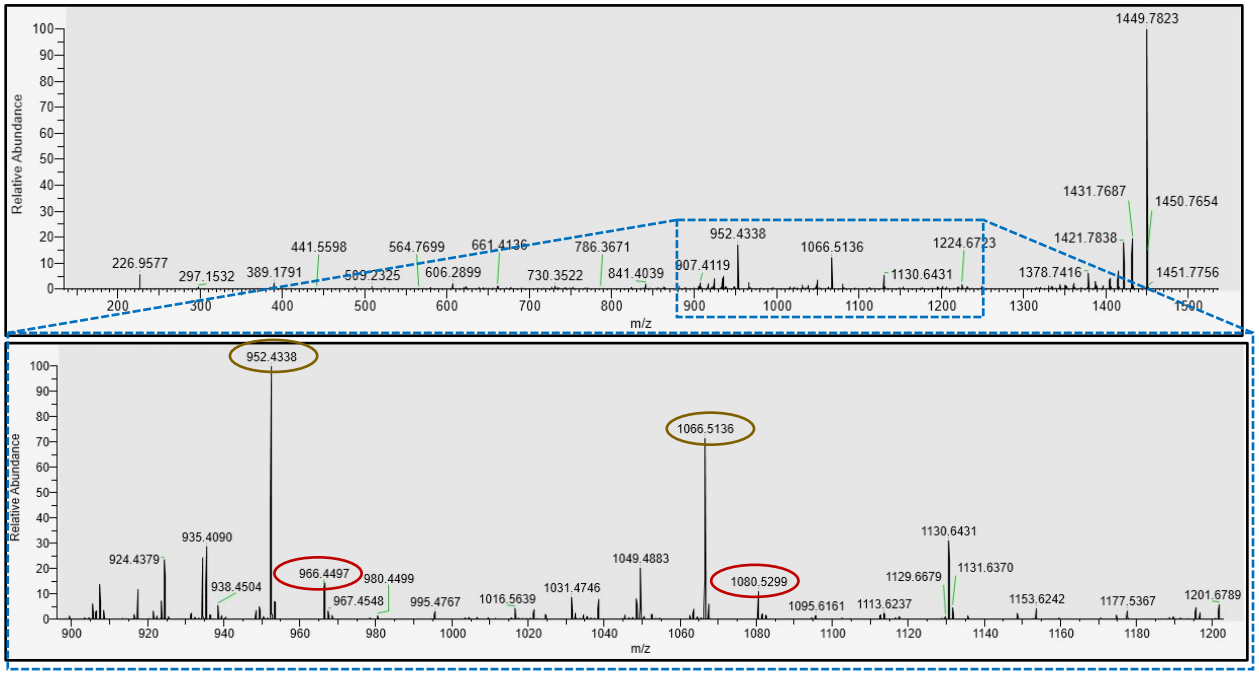
Fengycin derivatives	aa position 6	aa position 10
fengycin A	Ala	Ile
fengycin A2	Ala	Val
fengycin B	Val	Ile
fengycin B2	Val	Val

C

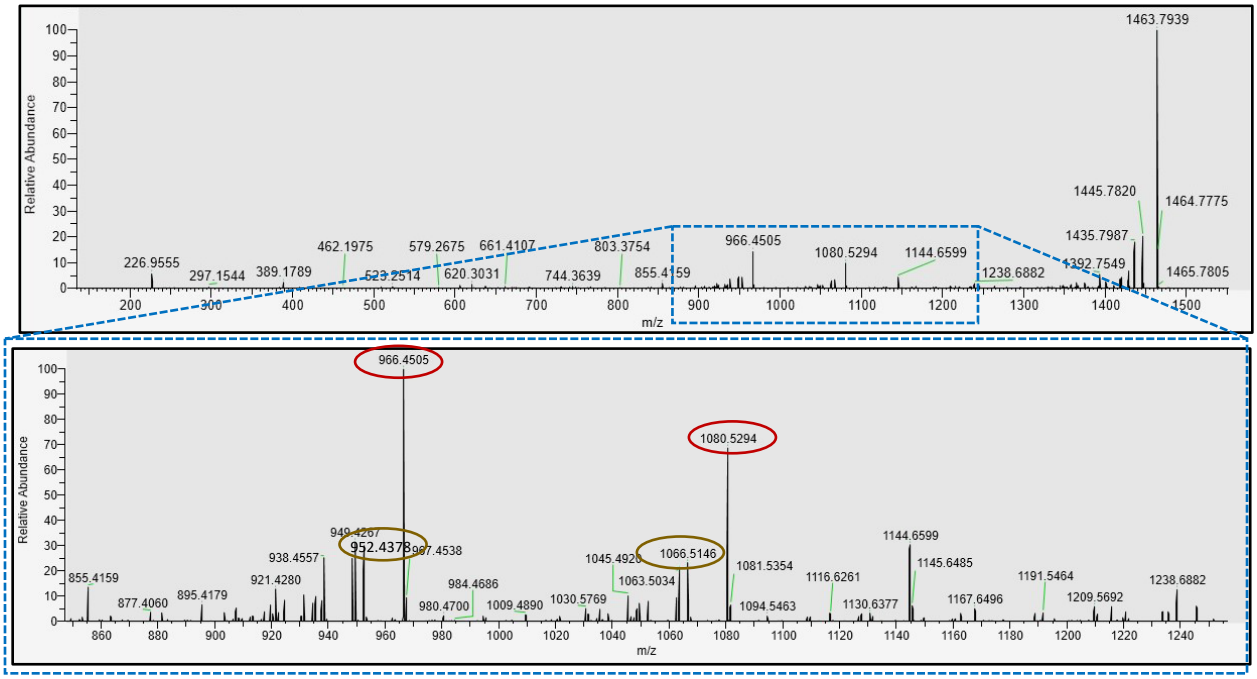
MS² of C14 fengycin A ([M+ H]⁺ = 1435.7730) (*B. velezensis* JJ334)



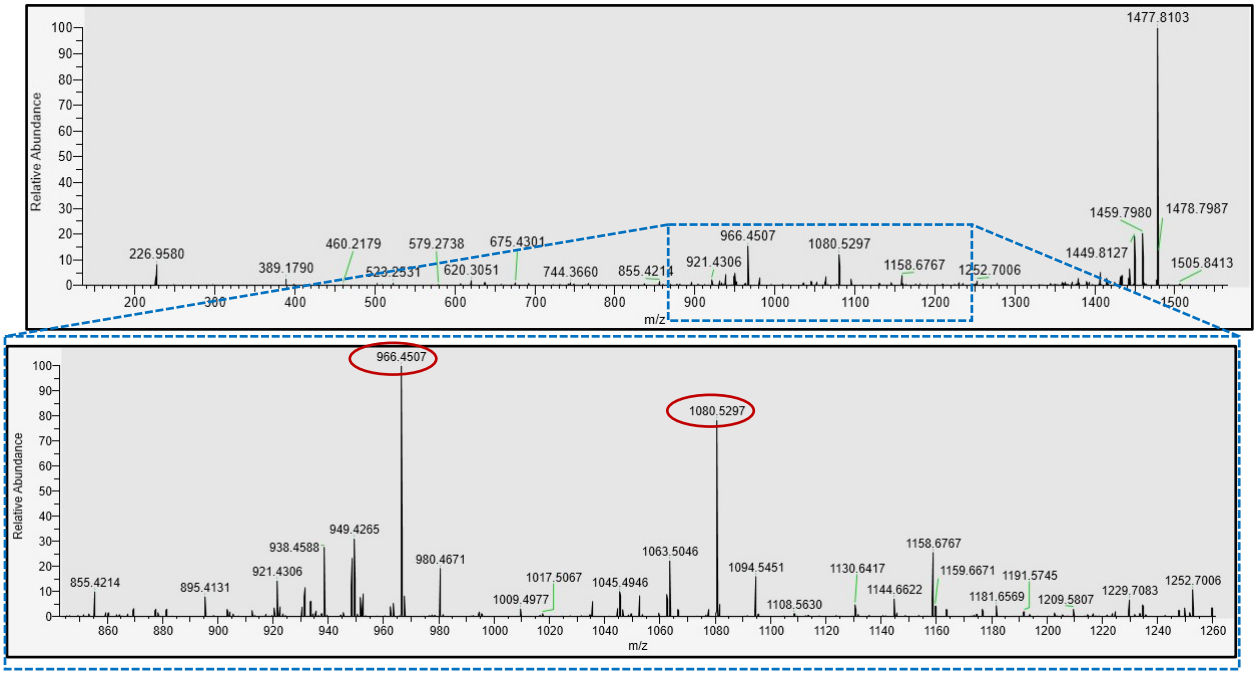
MS^2 of C15 fengycin A/ C16 of fengycin A2 ([M+ H]$^+$ = 1449.5539) (*B. velezensis* JJ334)



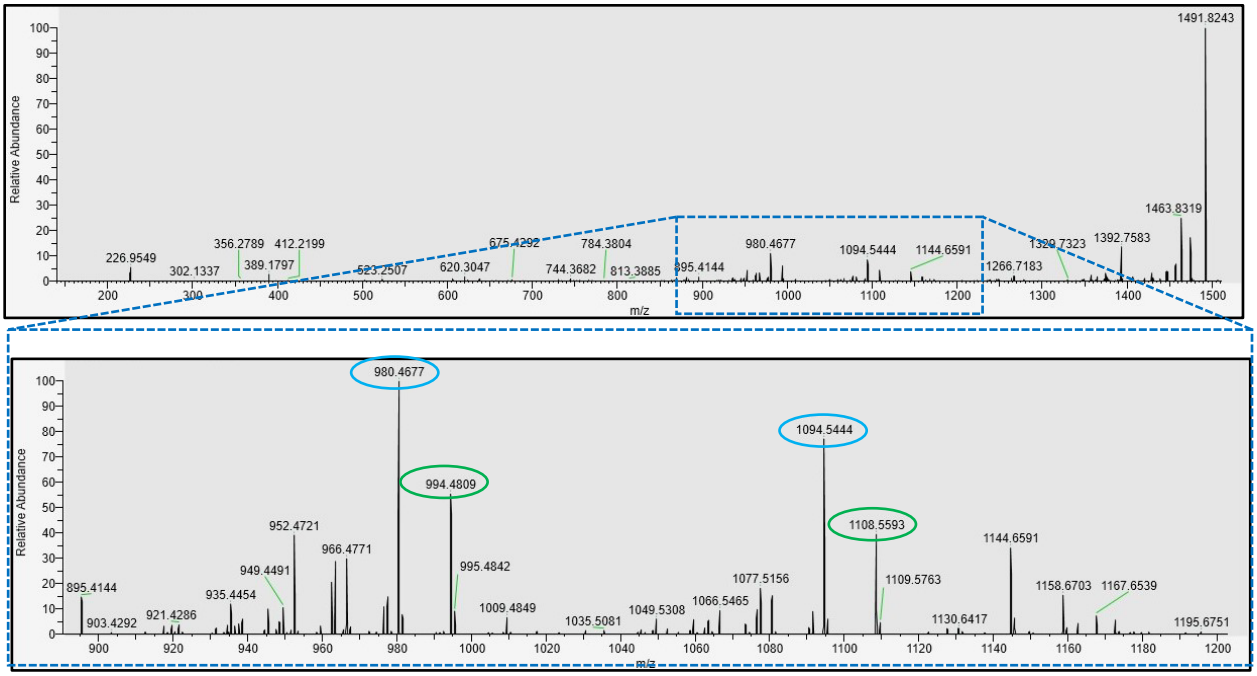
MS^2 of C16 fengycin A/ C17 of fengycin A2 ([M+ H]$^+$ = 1463.8028) (*B. velezensis* JJ334)



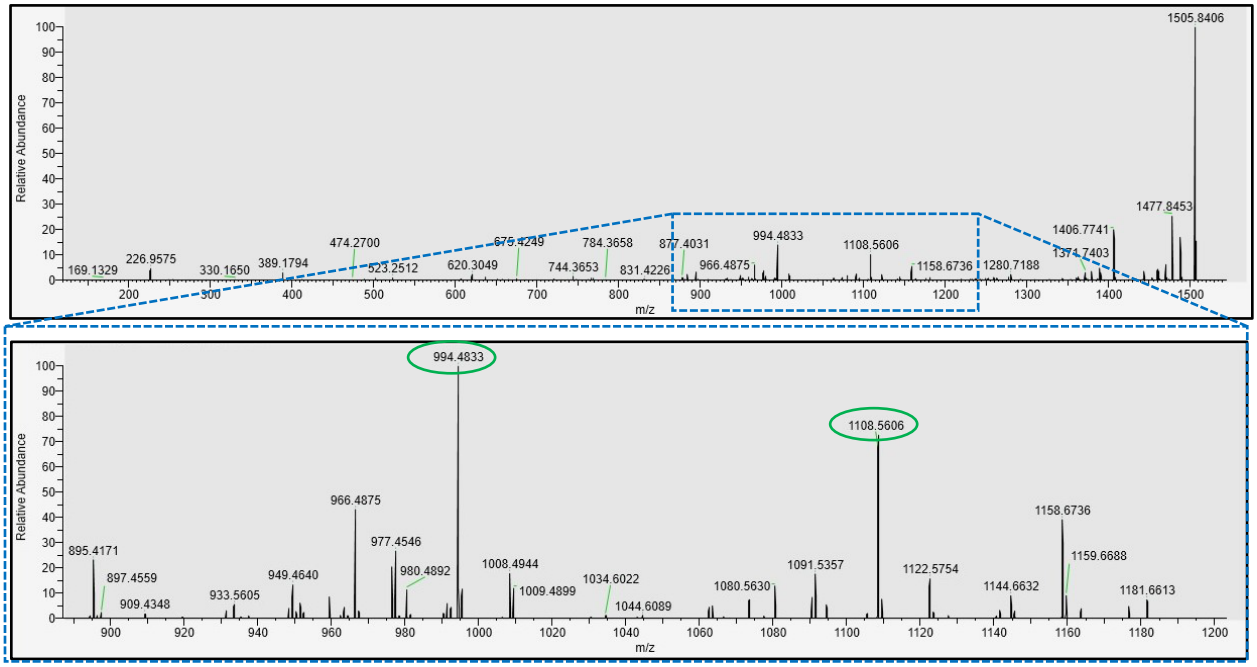
MS^2 of C17 fengycin A ([M+ H]$^+$ = 1477.8185) (*B. velezensis* JJ334)



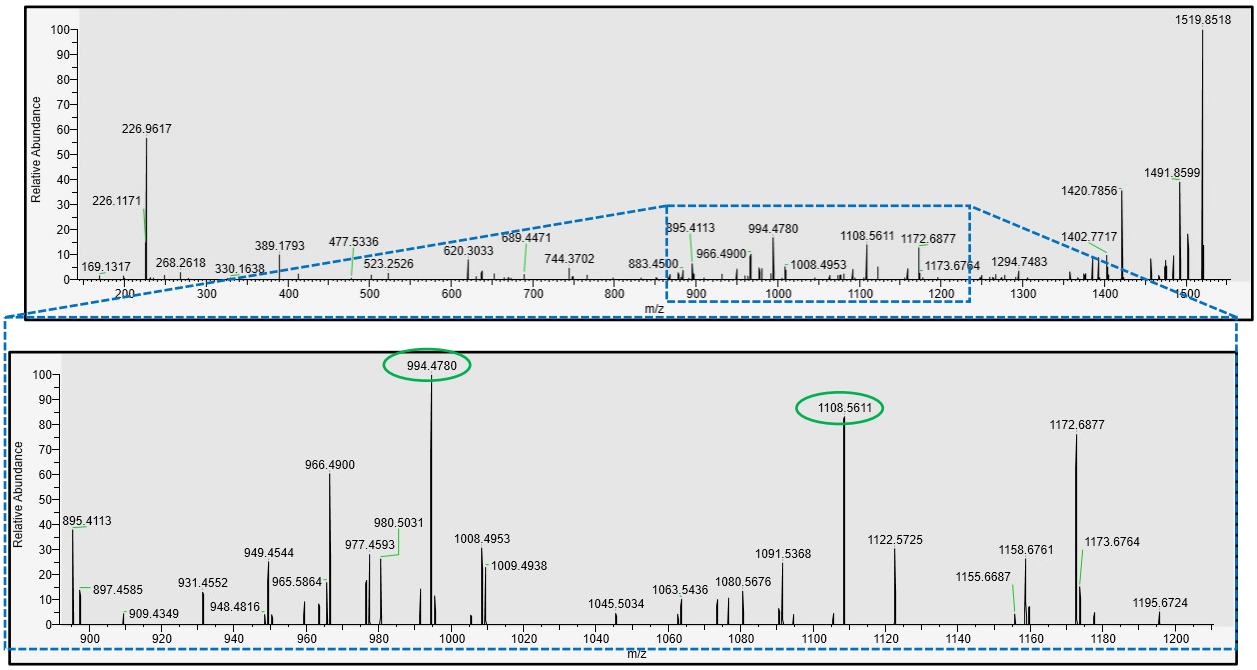
MS^2 of C16 fengycin B/ C17 of fengycin B2 ([M+ H]$^+$ = 1491.8332) (*B. velezensis* JJ334)



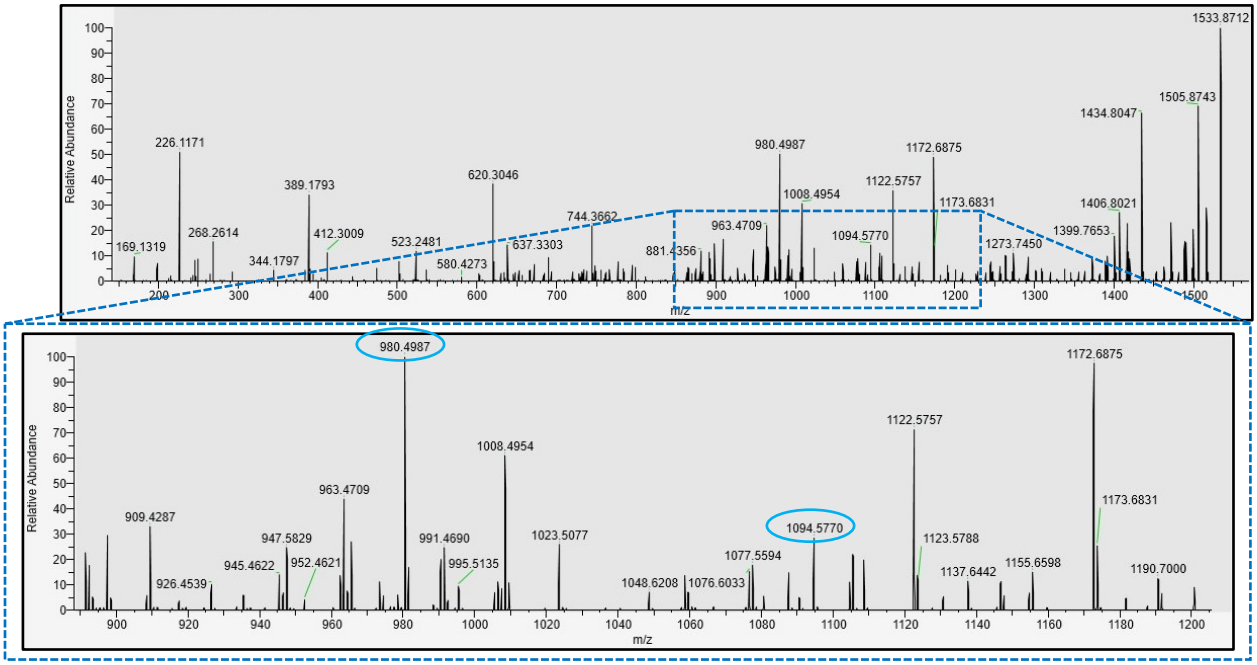
MS^2 of C17 fengycin B ([M+ H]$^+$ = 1505.8471) (*B. velezensis* JJ334)



MS^2 of fengycin B C18 ([M+ H]$^+$ = 1519.8666) (*B. velezensis* JJ334)



MS^2 of C20 fengycin B2 ([M+ H]$^+$ = 1533.8819) (*B. velezensis* JJ334)



MS^2 of C21 ([M+ H]$^+$ = 1547.8966) fengycin B2 (*B. velezensis* JJ334)

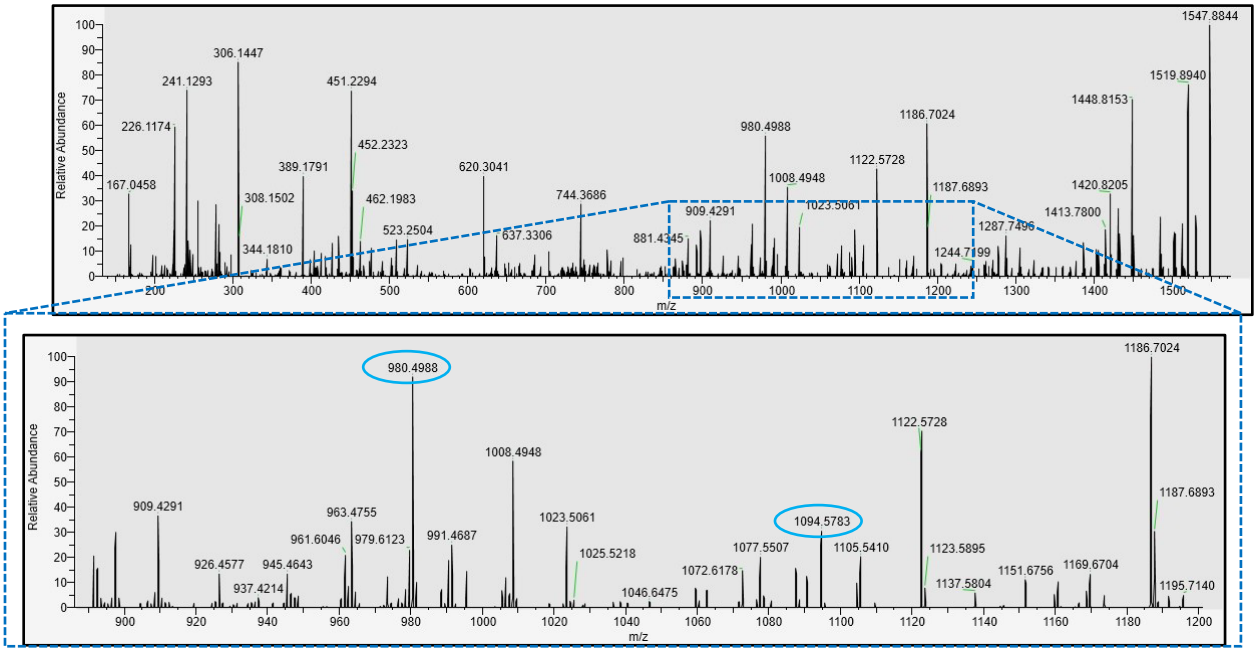


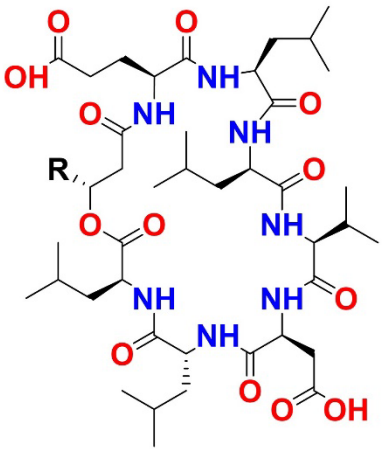
Figure 2.1. The structure of fengycin (A), ions generated from diagnostic fragmentation of the core peptide (B), and MS^2 spectra corresponding to identified fengycin derivatives (C). MS^2 spectra of nine prominent fengycin derivatives are shown. These were obtained from total extracts of *B. velezensis* JJ334. These spectra were generated by a multi-stage MS^n analyzer from respective precursor ions (positive ionization mode). Note that MS^2 is not available for the fengycin derivative 1561.9106. Assignment of fengycin derivatives based on change of aa at position 6^{th} and 10^{th} in the core peptide. Commonly observed diagnostic ions generated by MS^2 of precursor ion were used for assigning the fengycin derivatives. The table shows the assignment of fengycin derivatives based on the variation of aa in the core peptide.

Table 2.2. Ions from MS and MS^2 spectra used for assignment of surfactin derivatives.

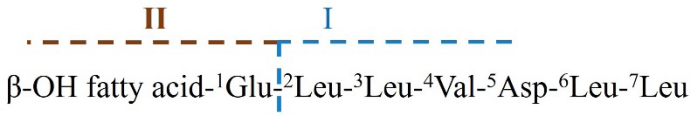
Precursor ion (MH^+)	**Fragment ion (I)**	**Fragment Ion (II)**	**Surfactin derivative**
994.6428	685.4447	310.1996	C12
1008.6587	685.4458	324.2162	C13
1022.6740	685.4451	338.2312	C14
1036.6898	685.4443	352.2463	C15
1050.7053	685.4447	366.2615	C16
1064.7199	685.4449	380.2765	C17
1078.7360	685.4509	394.2923	C18
1092.7519	ND[1]	ND[1]	C19

[1]ND=Not Detected.

A

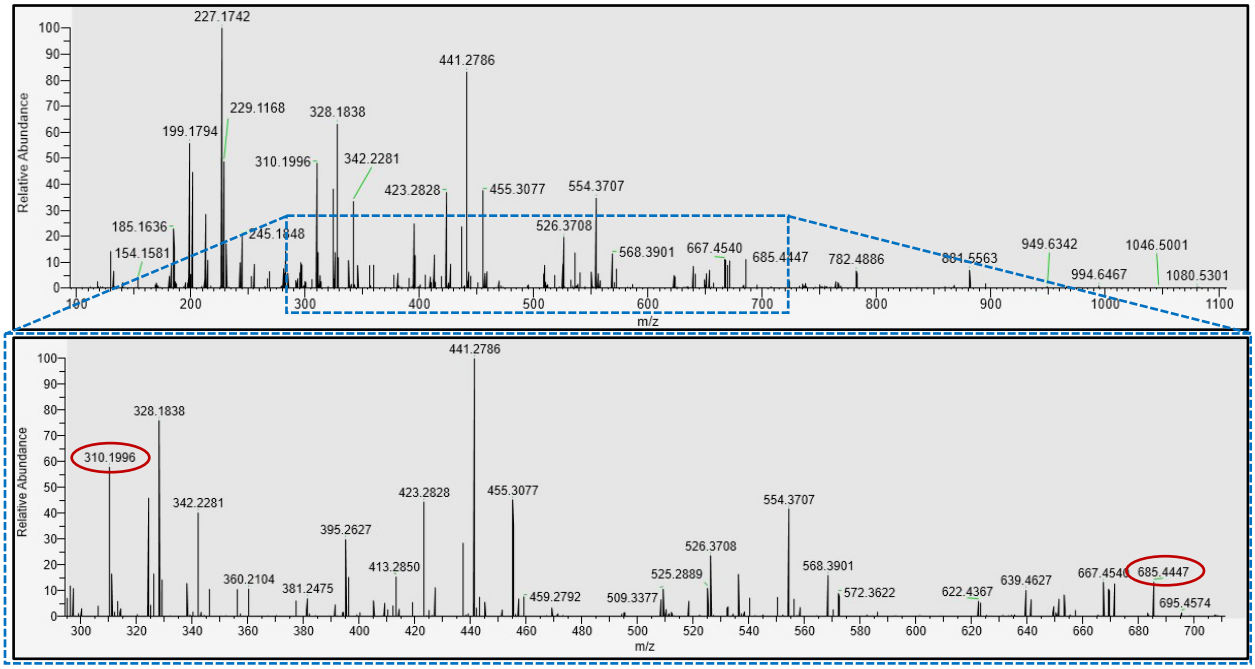


B

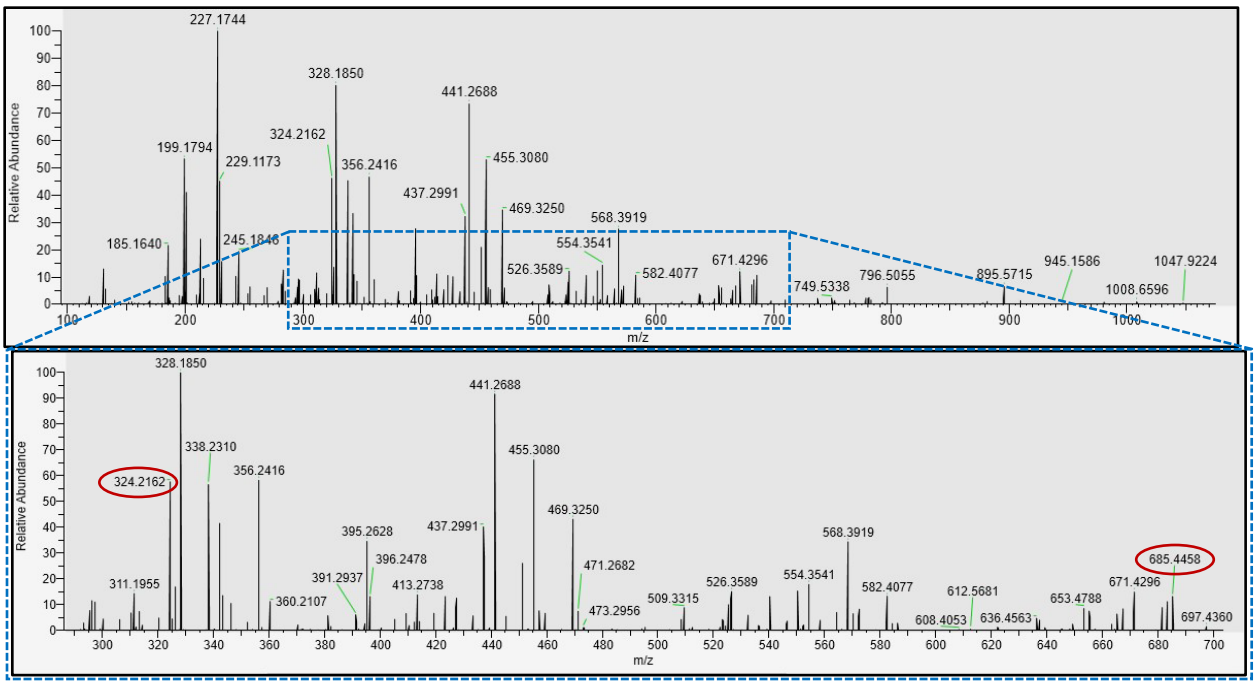


C

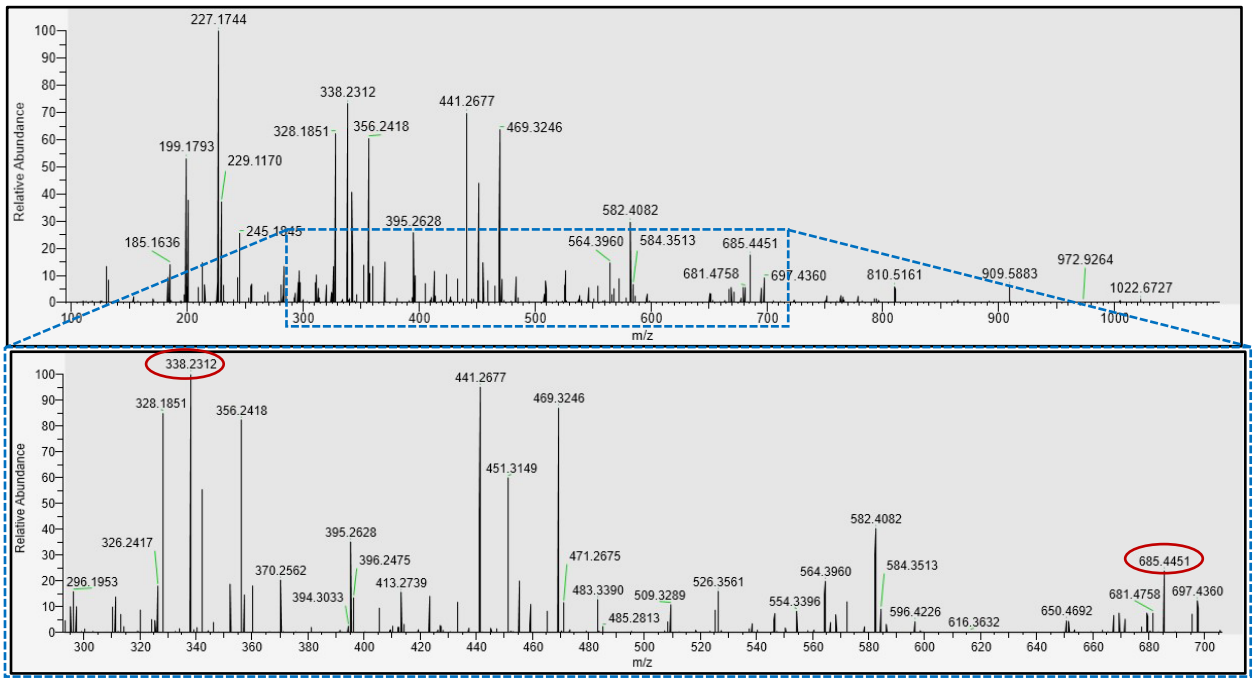
MS^2 of C12 surfactin ([M+ H]$^+$ = 994.6428) (*B. velezensis* JJ334)



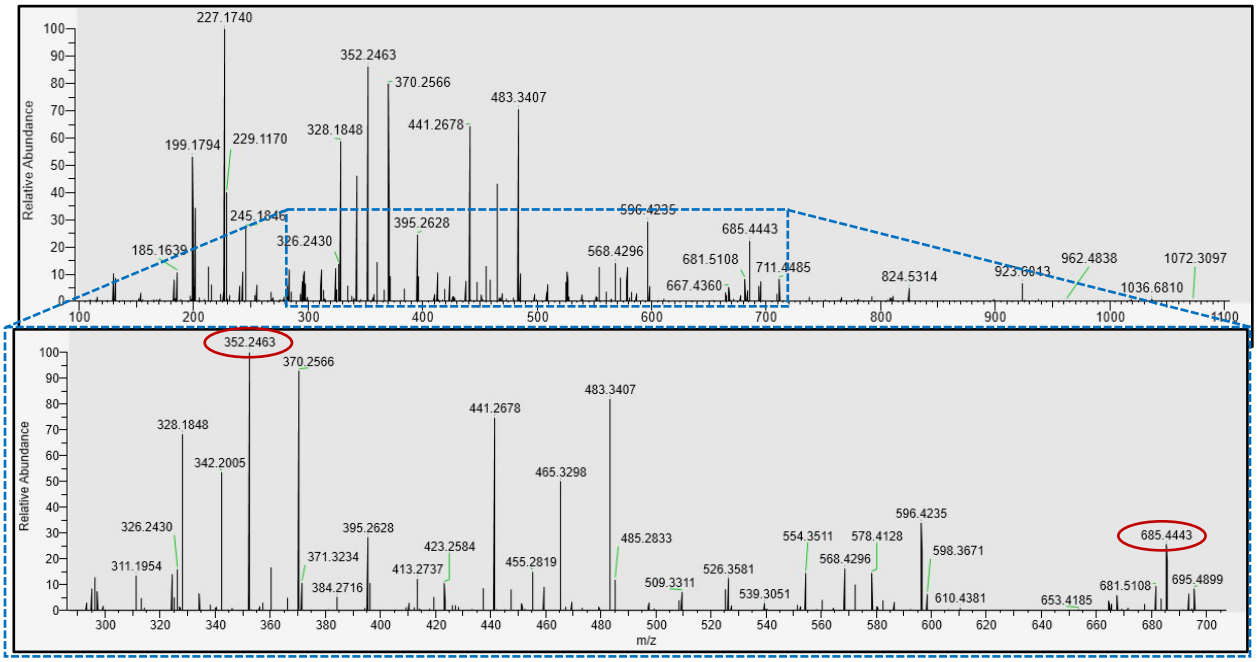
MS^2 of C13 surfactin ([M+ H]$^+$ = 1008.6587) (*B. velezensis* JJ334)



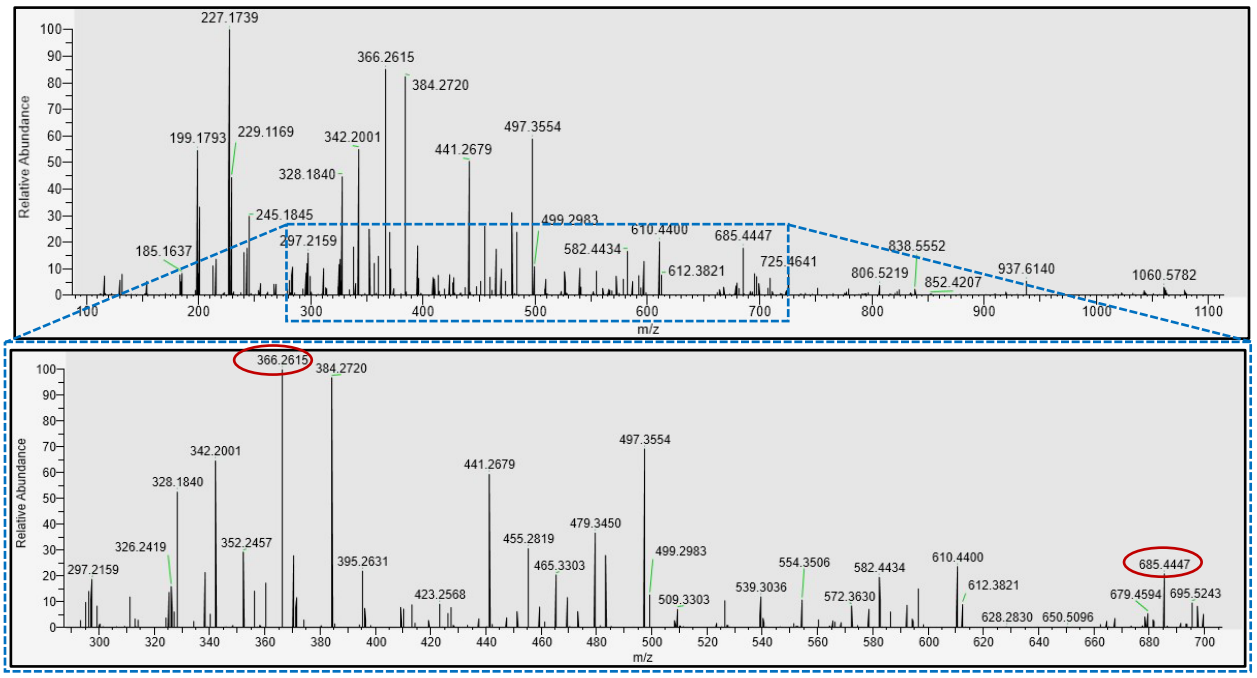
MS² of C14 surfactin ([M+ H]⁺ = 1022.6740) (*B. velezensis* JJ334)



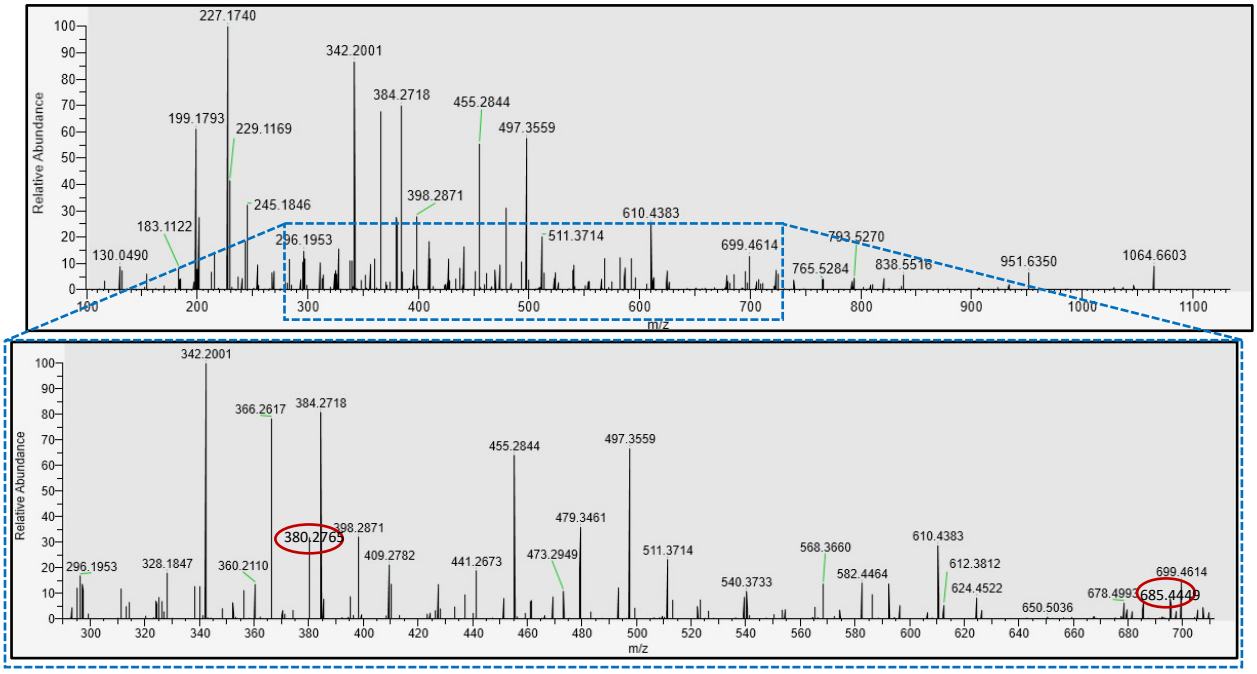
MS^2 of C15 surfactin ($[M+H]^+$ = 1036.6898) (*B. velezensis* JJ334)



MS^2 of C16 surfactin ($[M+H]^+$ = 1050.7053) (*B. velezensis* JJ334)



MS^2 of C17 surfactin ([M+ H]$^+$ = 1064.7199) (*B. velezensis* JJ334)



MS^2 of C18 surfactin ([M+ H]$^+$ = 1078.7360) (*B. velezensis* JJ334)

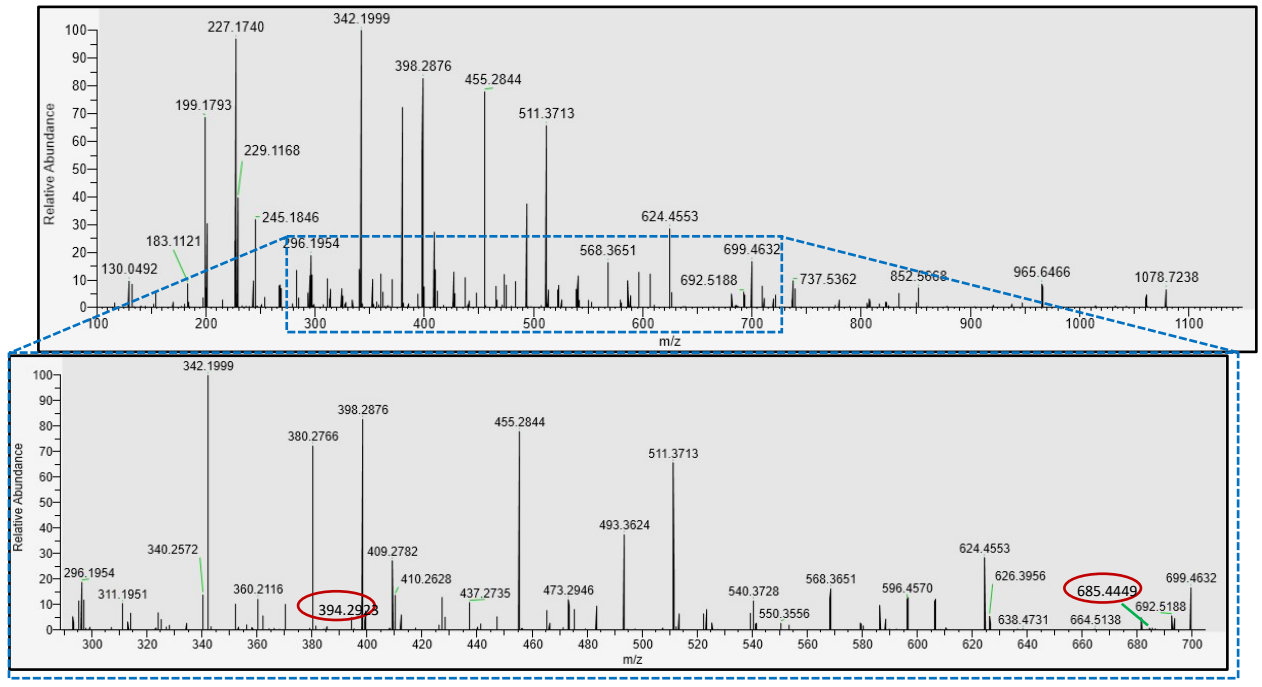


Figure 2.2. The structure of surfactin (A), ions generated from diagnostic fragmentation of the core peptide (B), and MS^2 spectra corresponding to identified surfactin derivatives (C). MS^2

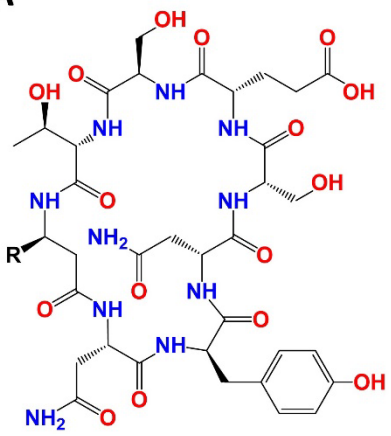
spectra of seven prominent surfactin derivatives were identified from the total extract of *B. velezensis* JJ334. Assignment of surfactin derivatives was based on the cleavage of the Glu1-(Leu/Ile)2 bond, generating the bulk of the core peptide (Leu/Ile2 – Leu7) and the remaining Glu1-FA tail. These spectra were generated by a multi-stage MS^n analyzer from respective precursor ions (positive ionization mode). Note that MS^2 data are not available for the surfactin derivative 1092.9519.

Table 2.3. Ions from MS and MS^2 spectra used for assignment of bacillomycin L derivatives.

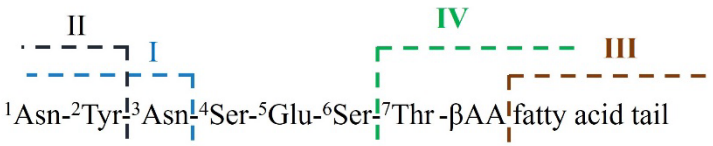
Precursor ion (MH^+)	**Fragment ion (I)**	**Fragment ion (II)**	**Fragment ion (III)**	**Fragment ion (IV)**	**Bacillomycin Derivative**
1007.5044	392.1530	278.1116	170.1891	313.2468	C13
1021.5198	392.1542	278.1115	184.2046	327.2621	C14
1035.5354	392.1525	278.1119	198.2205	341.2778	C15
1049.5539	392.1549	278.1120	212.2361	355.2938	C16
1063.5666	392.1544	278.1117	226.2516	369.3085	C17
1077.5820	ND[1]				

[1]ND=Not Detected.

A

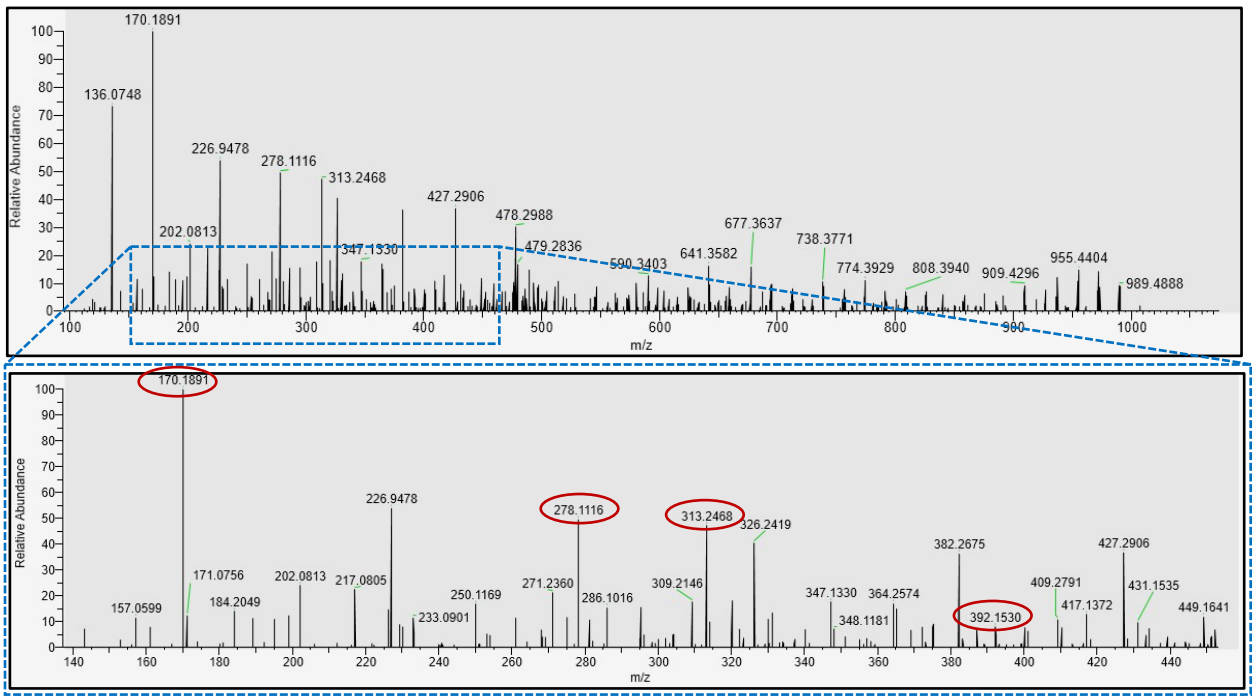


B

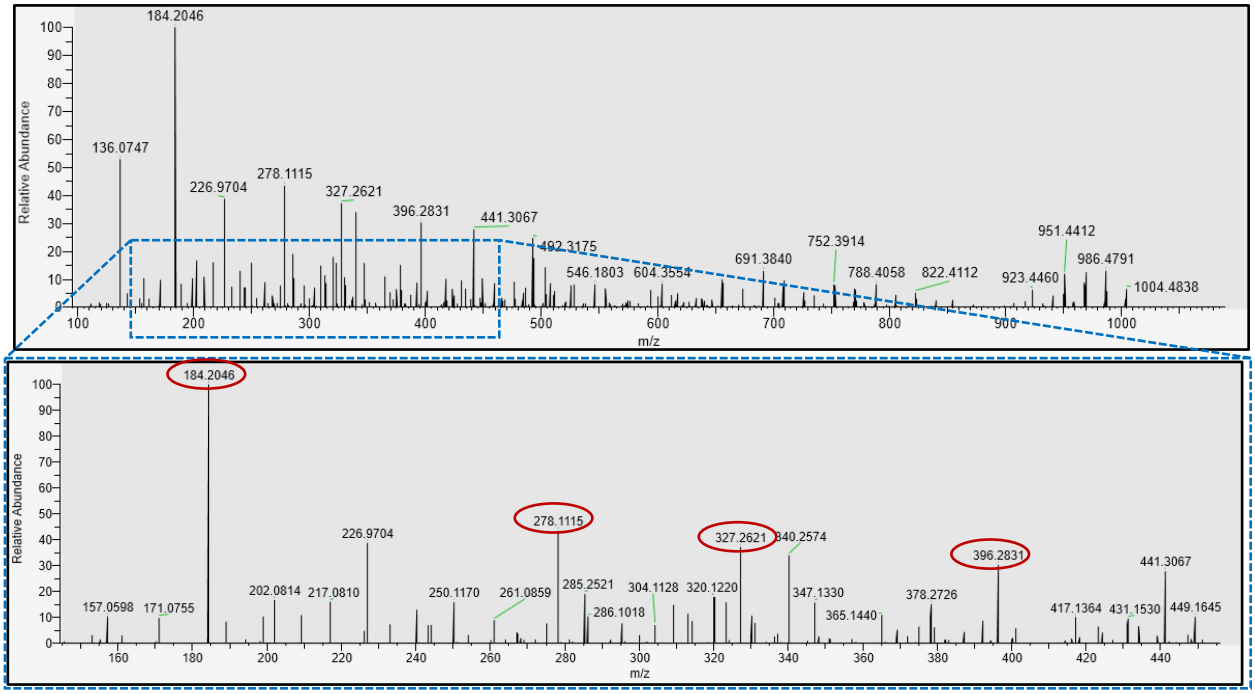


C

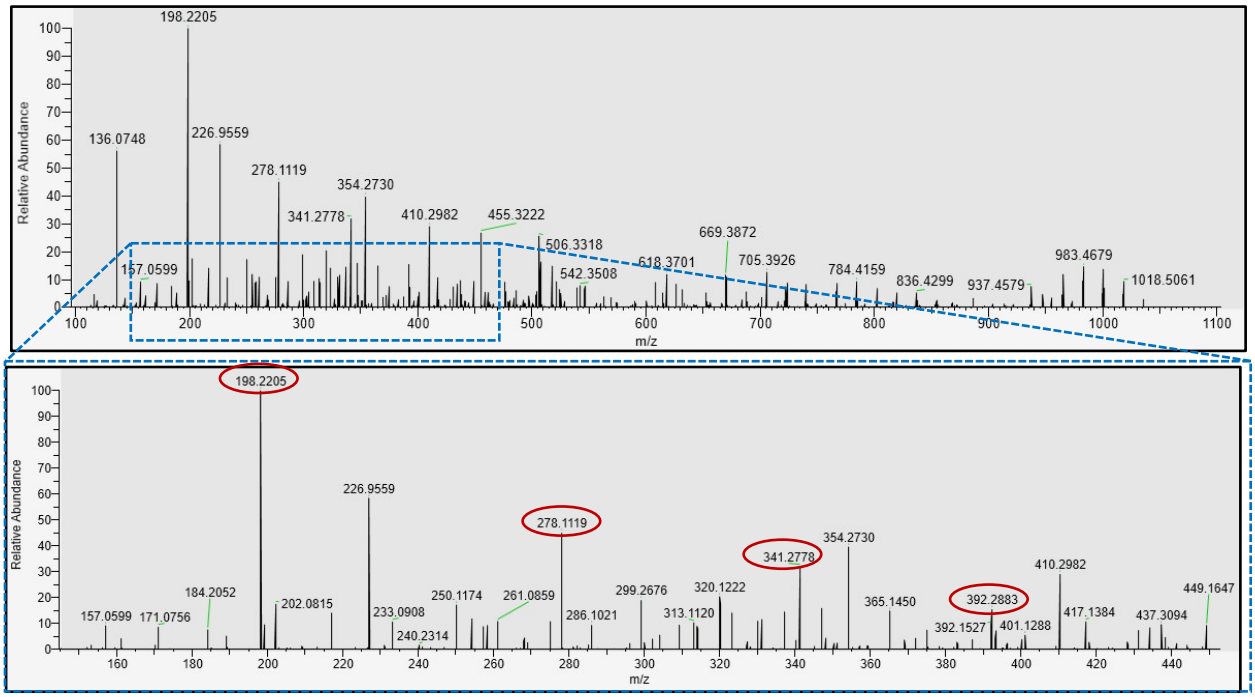
MS^2 of C13 bacillomycin L ([M+ H]$^+$ = 1007.5044) (*B. velezensis* JJ334)



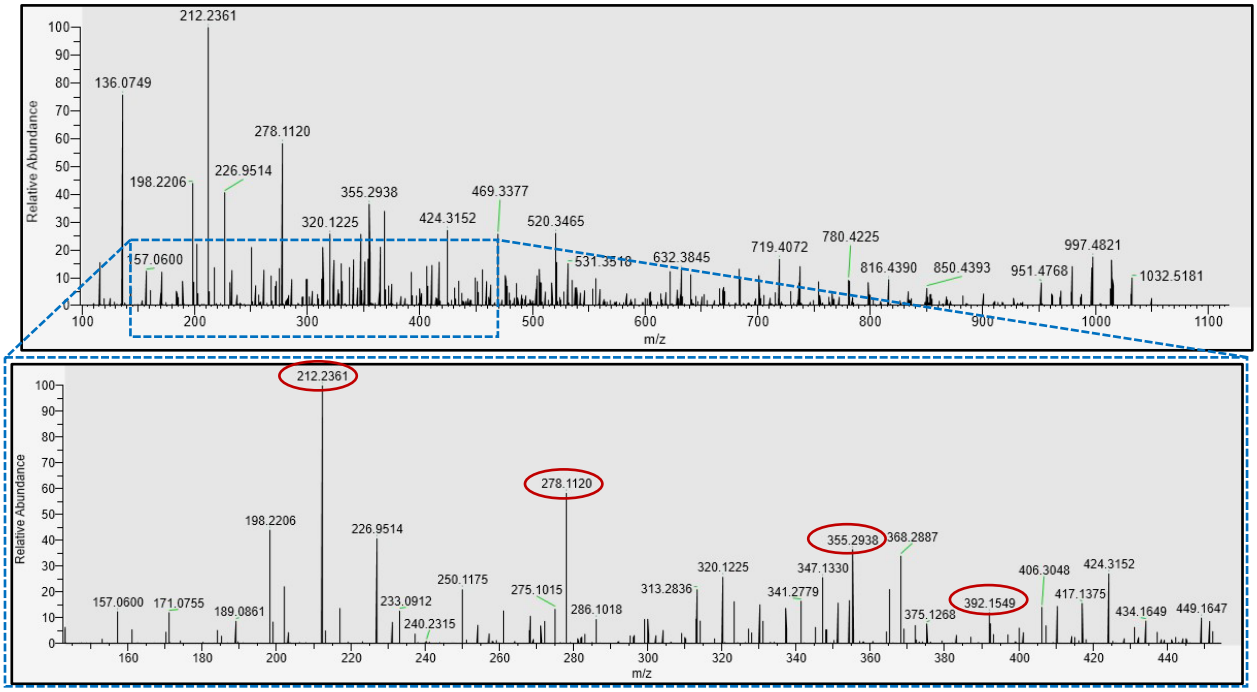
MS^2 of C14 bacillomycin L ([M+ H]$^+$ = 1021.5188) (*B. velezensis* JJ334)



MS^2 of C15 bacillomycin L ([M+ H]$^+$ = 1035.7730) (*B. velezensis* JJ334)



MS^2 of C16 bacillomycin L ([M+ H]$^+$ = 1049.5539) (*B. velezensis* JJ334)



MS^2 of C17 bacillomycin L ([M+ H]$^+$ = 1063.5666) (*B. velezensis* JJ334)

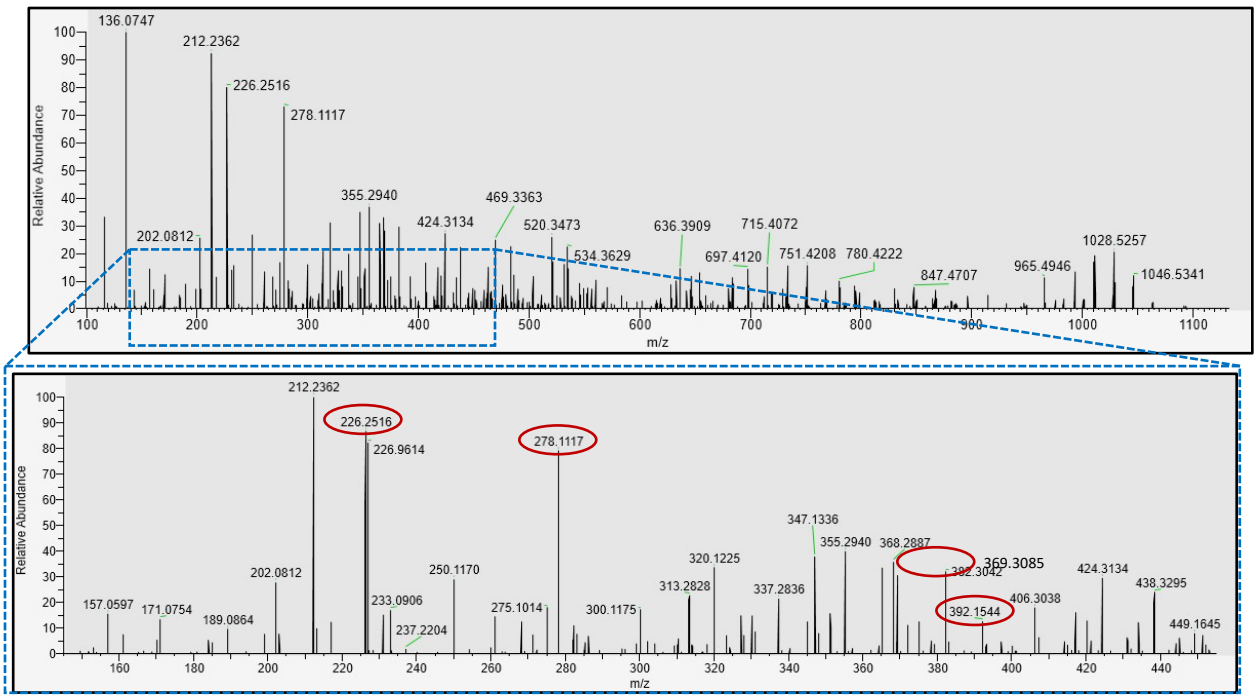


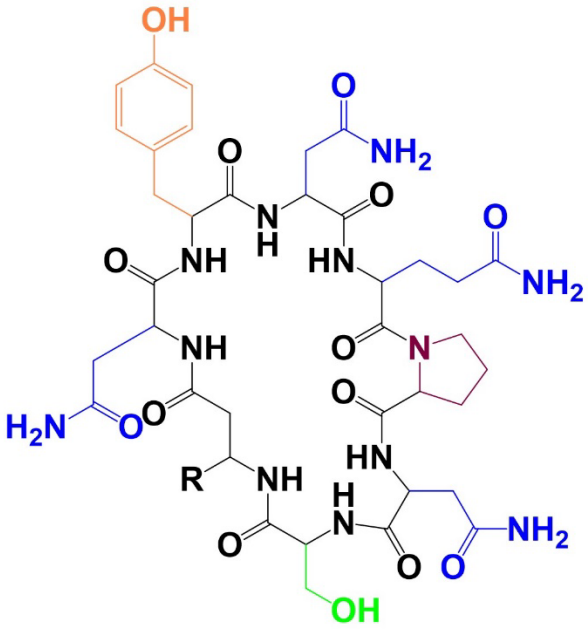
Figure 2.3. The structure of bacillomycin L (A), ions generated from diagnostic fragmentation of the core peptide (B), and MS^2 spectra corresponding to identify bacillomycin L derivatives (C). MS^2 spectra of six prominent bacillomycin L derivatives were identified from total extracts of *B. velezensis* JJ334. Assignment of bacillomycin L derivatives was based on the fragmentation of the core peptide to generate Asn-Tyr-Asn (I), Asn-Tyr (II) ions as well as the cleavage of the β-hydroxy fatty acid tail. Note that MS^2 is not available for the bacillomycin L derivative with a *m/z* of 1077.5820.

Table 2.4 Ions from MS and MS^2 spectra used for assignment of iturin derivatives.

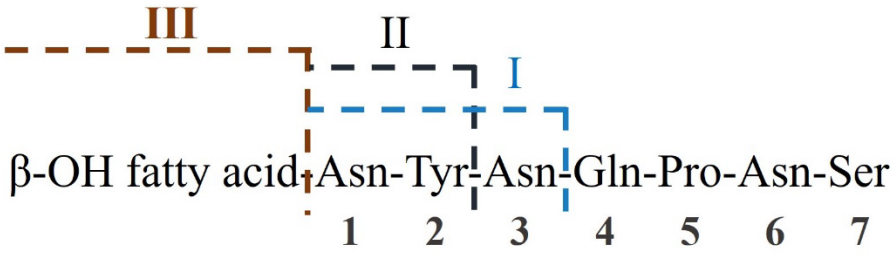
Precursor ion (MH^+)	Fragment ions (I)	Fragment ions (II)	Fragment ions (III)	Bacillomycin Derivatives
1043.5516	392.1527	278.1117	184.2050	C14
1057.5667	392.1536	278.1117	198.2205	C15
1071.5828	392.1540	278.1119	212.2360	C16
1085.5981	392.1538	278.1118	226.2515	C17

B

A

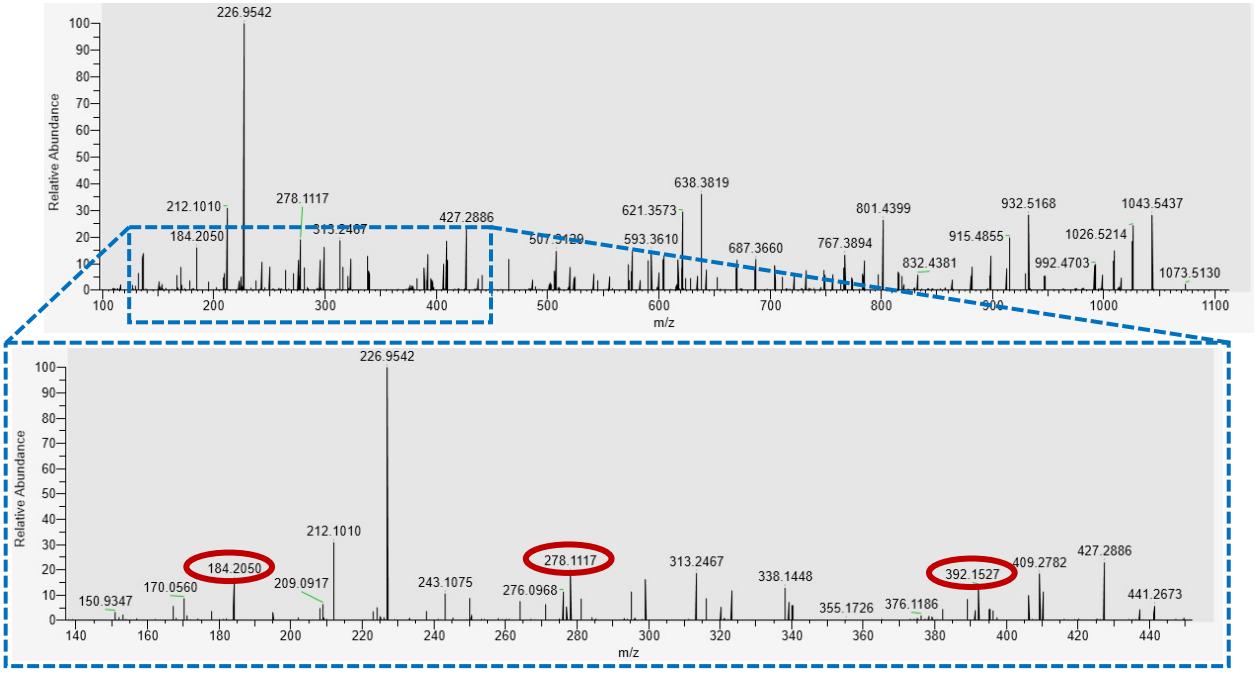

OH
O
NH2
O
N
H
O
NH
HN
O
NH2
NH
O
N
O
H2N
O
O
HN
R
NH
H
N
NH2
O
O
O
OH

B

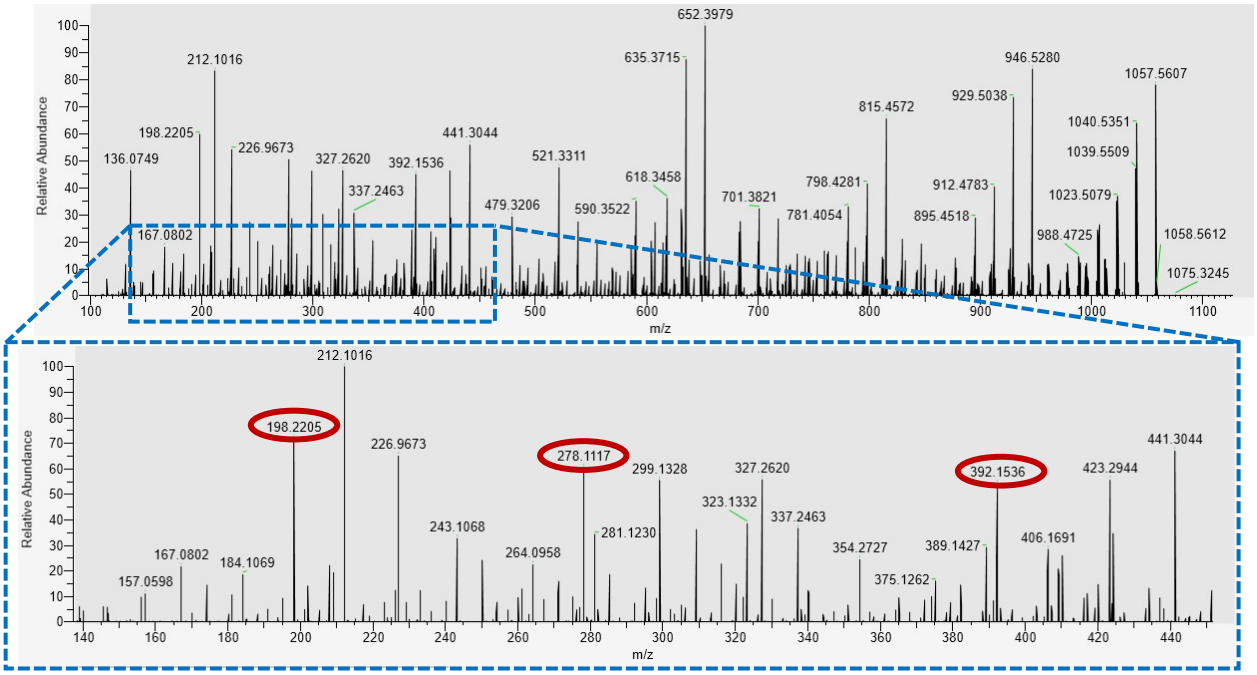

III
II
I
β-OH fatty acid-Asn-Tyr-Asn-Gln-Pro-Asn-Ser
1 2 3 4 5 6 7

C

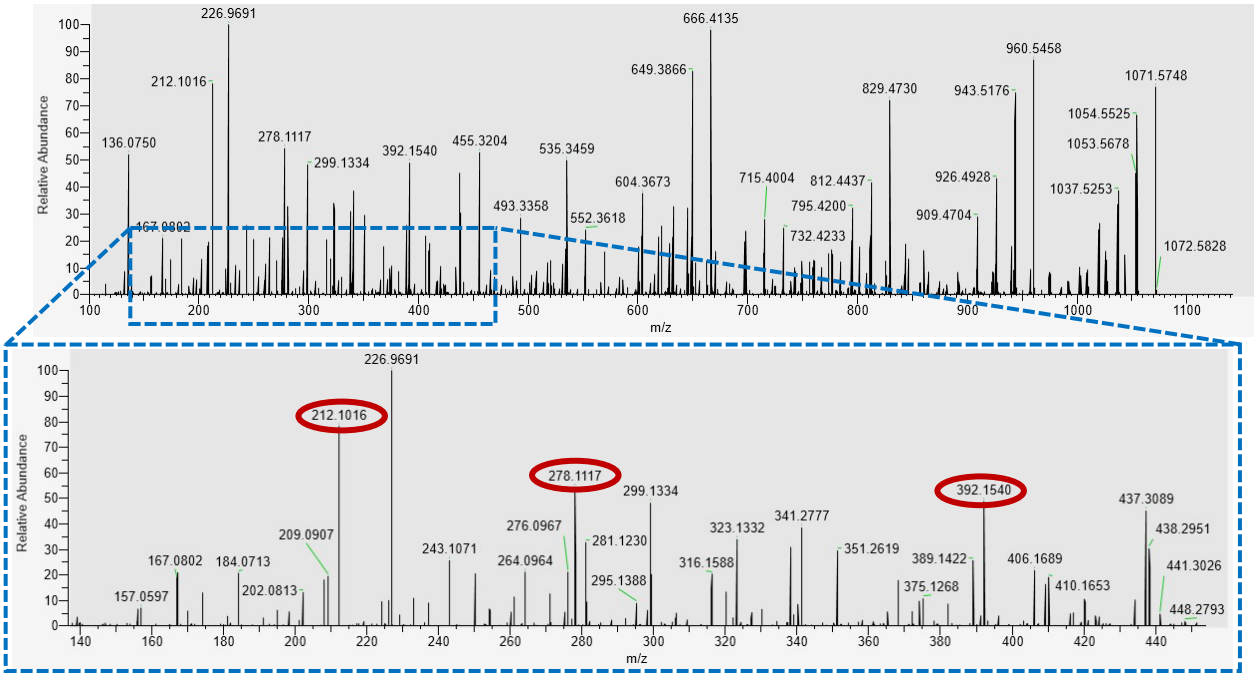
MS² of C14 iturin ([M+ H]⁺ = 1043.5517) (*B. velezensis* JJ951)



MS² of C15 iturin ([M+ H]⁺ = 1057.5675) (*B. velezensis* JJ951)



MS^2 of C16 iturin ([M+ H]$^+$ = 1071.5833) (*B. velezensis* JJ951)



MS^2 of C17 iturin ([M+ H]$^+$ = 1085.5980) (*B. velezensis* JJ951)

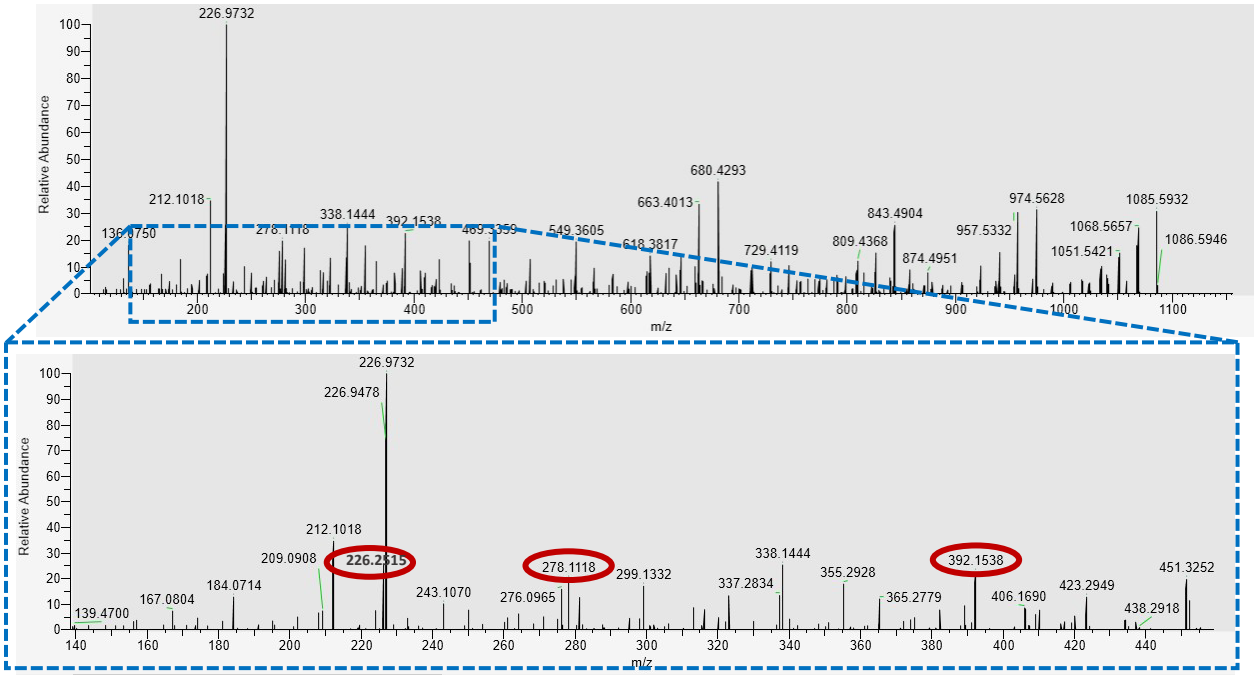
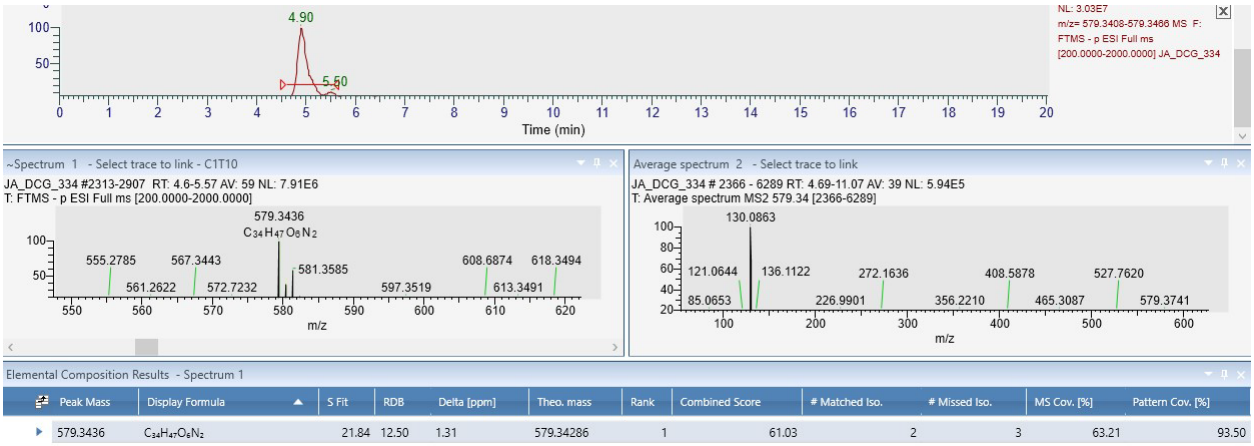


Figure 2.4. The structure of iturin A (A), ions generated from diagnostic fragmentation of the core peptide (B), and MS^2 spectra corresponding to identify iturin A derivatives (C). MS^2 spectra of four prominent iturin A derivatives were identified from total extracts of *B. velezensis* JJ951. Assignment of iturin A derivatives was based on the fragmentation of the core peptide to generate Asn-Tyr-Asn (I), Asn-Tyr (II) ions as well as the cleavage of the β-hydroxy fatty acid tail.

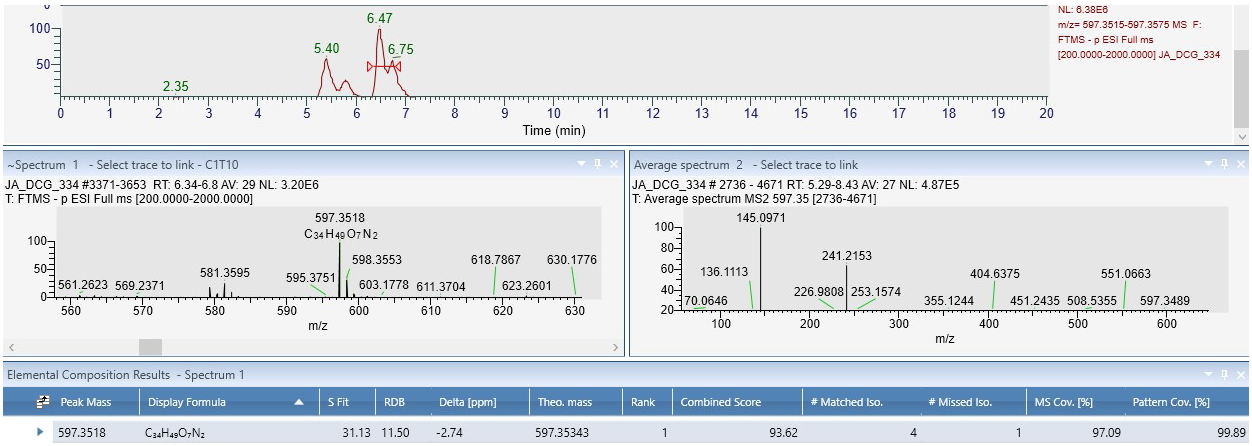
Appendix C

Extracted ion chromatogram from the LC elution, MS/MS^2 spectra, and elemental composition prediction of prominent antimicrobial polyketide compounds identified in the total extract of *B. velezensis* JJ334. Antimicrobial polyketides (bacillaene, putative bacillaene, difficidin, oxydifficidin, and macrolactin W) were detected using the negative ionization mode. Plantazolicin (ribosomal peptide) was detected in positive ionization mode. MS^2 spectra were generated by a multi-stage MS^n analyzer from respective precursor ions.

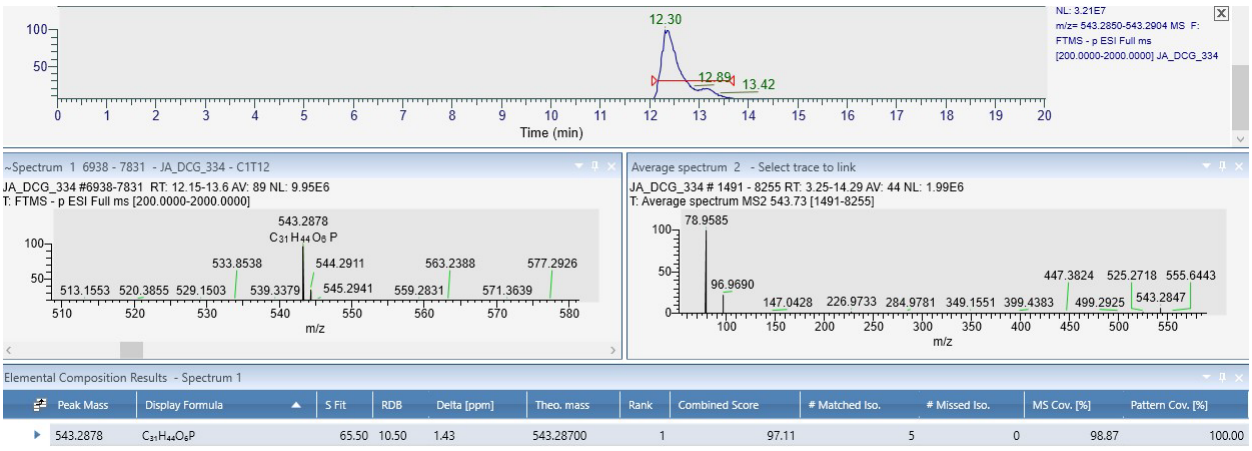
Bacillaene ([M-H]$^-$ = 579.3436): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel)



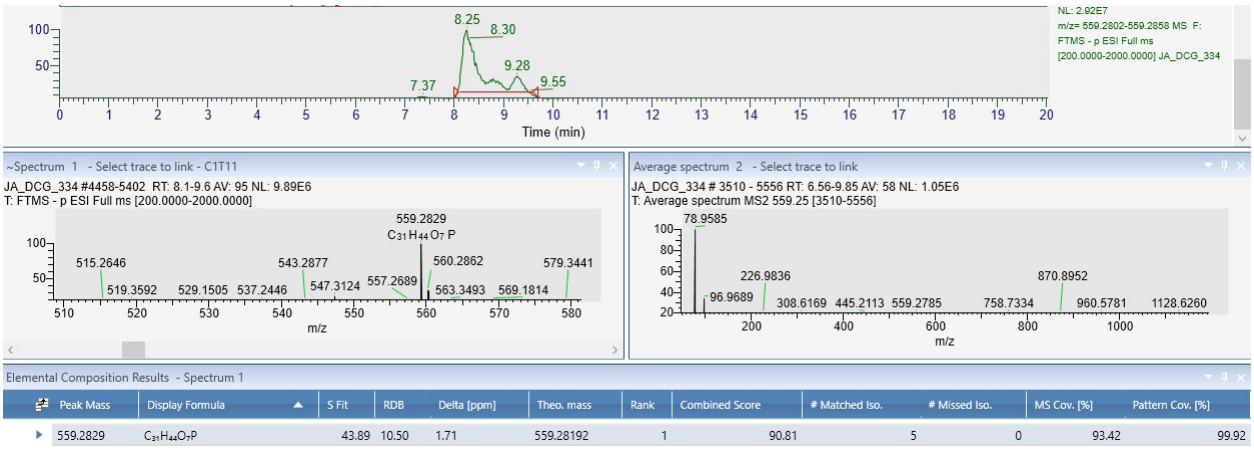
Putative oxybacillaene ([M-H]$^-$ = 597.3518): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel)



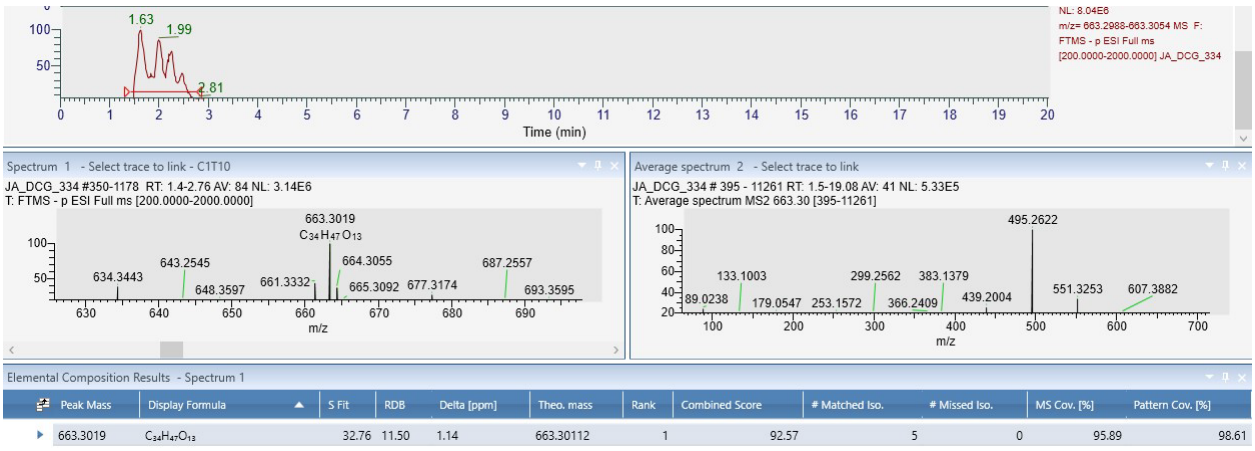
Difficidin ([M-H]$^-$ = 543.2878): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel).



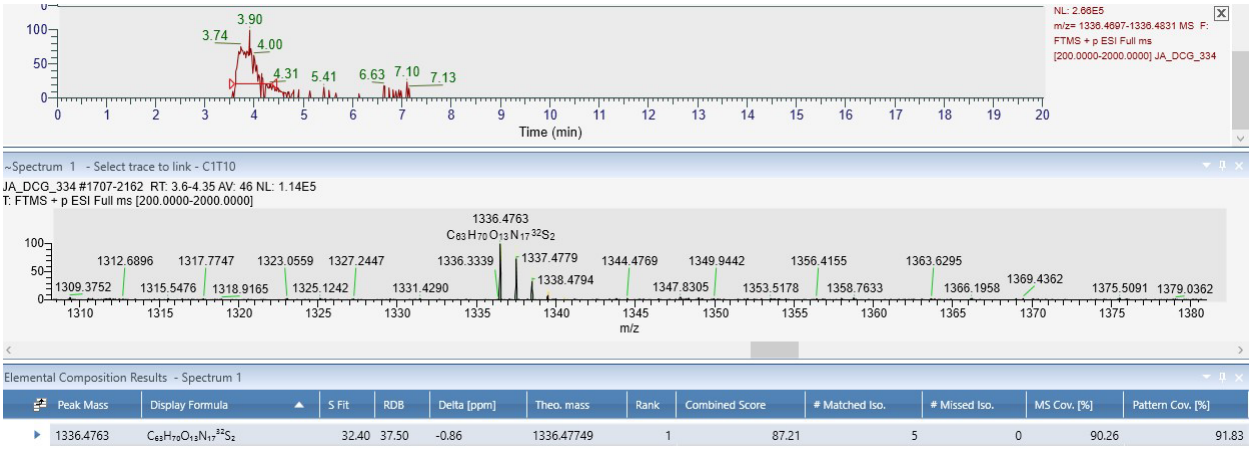
Putative oxydifficidin ([M-H]$^-$ =559.2829): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel).



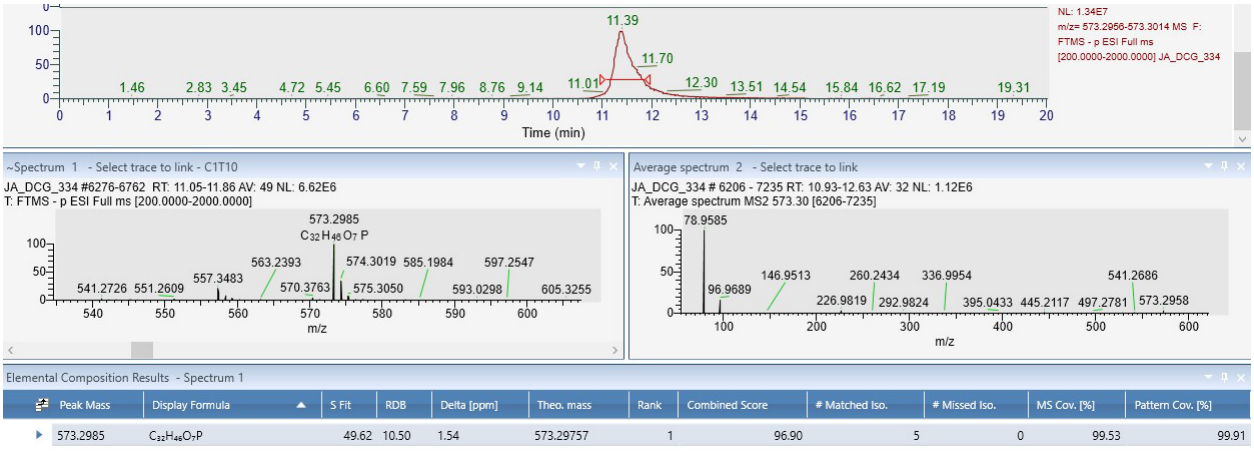
Macrolactin W ([M-H]$^-$ = 663.3019): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel).



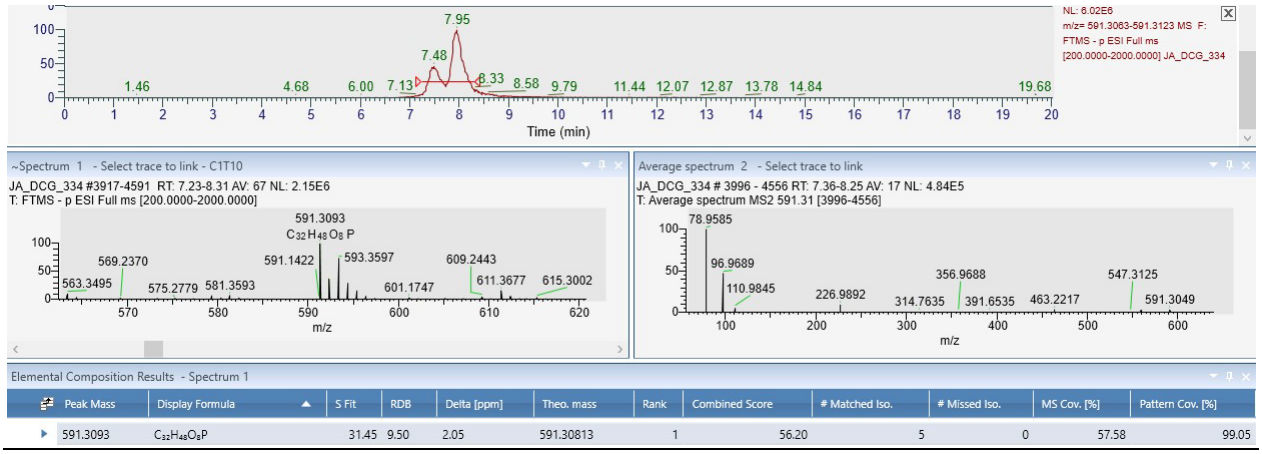
Plantazolicin([M+H]$^+$=543.1336.4763) : LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), and elemental composition (bottom panel).



Putative difficidin derivative 1 ([M-H]$^-$ = 573.2985): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel).



Putative difficidin derivative 2 ([M-H]$^-$ =591.3093): LC elution from the extracted ion chromatogram (EIC) (top panel), MS spectra (middle left panel), MS2 spectra (middle right panel), and elemental composition (bottom panel).



Appendix D

Table 4.1: Quantitative intraspecies inhibition score (0-4) derived from the measured zone of inhibition (ZOI) and morphological changes of both focal and target strains. Inhibition score shown here for each strain is the average score from three replicates.

Heatmap showing average inhibition score (0-4) of each focal strain against each target strain																		
Target Strains	Focal Strains																	
	AB01	AP202	AP215	AP45	AP46	AP52	AP81	JJ1043	JJ1284	JJ213	JJ334	JJ747	JJ947	JJ951	JM199	JM204	JM236	JM907
AB01	0	1	2	1	1	0	1	1	1	0	1	1	3	0	4	1	1	2
AP202	3	0	1	2	3	2	0	0	1	1	1	2	4	3	4	0	2	1
AP215	4	0	0	1	2	2	1	1	0	1	1	1	2	4	4	1	1	1
AP45	4	3	3	0	1	1	0	0	3	1	1	2	3	4	2	1	3	2
AP46	4	0	0	0	0	3	0	0	0	0	2	1	1	4	4	0	0	0
AP52	3	2	3	0	2	0	1	0	0	1	1	2	1	4	4	1	2	1
AP81	3	4	4	0	4	2	0	0	2	1	1	1	4	4	2	2	4	4
JJ1043	4	2	4	2	4	1	0	0	4	1	1	2	4	4	4	4	4	4
JJ1284	4	2	2	4	3	1	2	0	0	1	1	2	4	4	2	1	3	1
JJ213	3	4	4	1	3	1	0	0	1	0	1	1	2	4	1	1	3	3
JJ334	4	1	2	1	3	1	0	0	2	1	0	2	4	4	4	2	4	3
JJ747	4	0	0	2	2	2	0	0	0	1	2	0	0	4	3	0	0	0
JJ947	4	2	3	1	1	1	0	0	2	1	0	1	0	4	2	2	4	3
JJ951	0	1	2	1	4	0	0	0	1	0	1	3	3	0	4	0	1	1
JM199	4	4	4	0	2	3	0	1	0	0	1	2	2	4	0	1	3	1
JM204	4	3	3	2	4	1	1	3	1	1	2	3	2	4	3	0	3	0
JM236	4	3	1	4	3	1	4	0	1	1	2	1	0	4	3	1	0	2
JM907	4	3	3	1	3	1	1	0	0	1	2	2	3	4	3	0	3	0

Table 4.2: Quantitative intraspecies percent inhibition derived from the measured zone of inhibition (ZOI) and morphological changes of both focal and target strains. Percent inhibition score shown here for each strain is the average score from three replicates.

Heatmap showing average percent inhibition of each focal strain against each target strain																		
Target Strains	Focal Strains																	
	AB01	AP202	AP215	AP45	AP46	AP52	AP81	JJ1043	JJ1284	JJ213	JJ334	JJ747	JJ947	JJ951	JM199	JM204	JM236	JM907
AB01	0	33	58	33	17	0	25	17	17	8	25	33	67	0	100	25	33	58
AP202	83	8	17	50	67	42	8	0	25	17	17	58	92	83	100	8	50	17
AP215	100	0	0	33	50	58	25	17	8	25	25	25	58	92	92	17	25	25
AP45	100	67	83	0	33	33	0	8	67	25	25	50	75	92	50	25	75	58
AP46	100	0	0	0	0	67	0	0	0	0	50	17	25	100	100	8	0	0
AP52	75	58	75	0	42	8	25	8	0	17	17	50	33	100	100	25	58	33
AP81	67	100	92	0	92	50	0	8	50	25	25	25	100	100	42	50	100	92
JJ1043	92	58	92	58	100	25	8	0	92	17	25	50	100	100	92	100	100	100
JJ1284	92	58	50	92	83	25	42	8	0	17	25	42	92	100	58	25	67	33
JJ213	83	92	100	17	83	33	0	8	25	0	17	33	50	100	25	25	75	67
JJ334	92	33	58	25	75	17	8	0	58	25	0	50	100	100	92	58	100	83
JJ747	100	8	8	42	58	42	0	0	0	17	42	0	0	100	83	0	0	0
JJ947	92	58	67	33	33	25	0	0	42	25	0	25	0	92	42	50	92	83
JJ951	0	17	50	17	92	8	0	8	17	0	17	75	75	0	100	8	25	17
JM199	100	100	100	0	58	67	8	25	8	0	25	50	50	100	0	33	67	33
JM204	92	75	67	50	100	33	25	75	17	25	50	75	58	92	83	0	75	0
JM236	92	75	25	100	75	25	100	8	17	25	42	33	8	100	83	25	0	58
JM907	92	75	75	25	67	25	25	8	0	25	50	42	83	92	83	0	83	0

Table 4.3: Cumulative % inhibition and % resistance of each strain against all strains.

Strain ID	*Bacillus velezensis* intraspecies inhibition and resistance across all 18 strains							
	Cumulative % inhibition				Cumulative % resistance			
	rep. 1	rep. 2	rep. 3	average	rep. 1	rep. 2	rep. 3	average
AB01	82	79	81	74	69	65	74	69
AP215	54	61	54	60	63	63	63	63
JJ951	85	83	89	77	68	78	67	71
JJ1284	25	24	25	39	54	49	46	50
AP52	29	36	32	48	63	57	60	60
JM204	29	28	24	37	46	44	44	45
JJ213	11	19	18	38	58	54	49	54
JM236	56	58	57	53	53	51	47	50
JJ334	25	28	26	38	47	43	47	46
JM907	40	46	40	48	53	53	53	53
AP45	31	33	32	43	54	50	51	52
AP81	17	18	15	32	42	46	43	44
AP46	58	57	72	69	74	74	75	74
AP202	49	53	51	55	58	58	60	59
JM199	72	72	76	63	57	51	54	54
JJ747	39	43	40	59	78	68	71	72
JJ947	60	57	61	58	57	58	58	58
JJ1043	11	11	11	24	35	31	33	33

Table 4.4: Derived antifungal inhibition score from the measured ZOI. Antibiosis inhibition score was calculated as follows: for ZOI<=2, score =1; ZOI<=3, score = 1.5; ZOI<=4, score= 2, ZOI<=5, score=2.5, ZOI<=6, score=3, ZOI<=7, score =3.5, ZOI>7, score =4

Strains	Antifungal activity: measured zone of inhibition (ZOI), mm									Antibiosis inhibition score derived from the ZOI								
	F. oxysporum			*F. graminearum*			*R. solani*			*F. oxysporum*			*F. graminearum*			*R. solani*		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
AB01	6	5	5	5.5	5.5	5.5	6	6.5	6	3	2.5	2.5	3	3	3	3	3.5	3
AP215	8	7	7	6	7	6	8	8	8	4	3.5	3.5	3	3.5	3	4	4	4
JJ951	4.5	5	5	4.5	5	5	5	5.5	5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	3	2.5
JJ1284	4	4	5	6.5	6.5	6.5	6	6.6	5.5	2	2	2.5	3.5	3.5	3.5	3	3.5	3
AP52	6	6.5	4.5	7	6.5	7	5.5	6	6	3	3.5	2.5	3.5	3.5	3.5	3	3	3
JM204	4	5	5.5	6.5	7	7	4.5	6	6	2	2.5	3	3.5	3.5	3.5	2.5	3	3
JJ213	5	4	5	6	6	5	6	6	6	2.5	2	2.5	3	3	2.5	3	3	3
JM236	1.5	2	1.5	3	2	3	4	3	4	1	1	1	1.5	1	1.5	2	1.5	2
JJ334	4	3.5	3	5	5.5	6	6	5	6	2	2	1.5	2.5	3	3	3	2.5	3
JM907	2	3	3	3	4	3.5	7	7.5	8	1	1.5	1.5	1.5	2	2	3.5	4	4
AP45	5	6	4	5.5	5.5	5.5	4.5	5.5	6	2.5	3	2	3	3	3	2.5	3	3
AP81	4.5	4.5	4.5	4.5	6	6	8	7	8	2.5	2.5	2.5	2.5	3	3	4	3.5	4
AP46	2.3	3	3	5.5	5	5	5	6	7	1.5	1.5	1.5	3	2.5	2.5	2.5	3	3.5
AP202	7	6.5	7	5.5	6	5.5	4.5	4	5	3.5	3.5	3.5	3	3	3	2.5	2	2.5
JM199	5	5	6	6	6	6	6	7	7	2.5	2.5	3	3	3	3	3	3.5	3.5
JJ747	6	5	4.5	5.5	7	6	6	7	7	3	2.5	2.5	3	3.5	3	3	3.5	3.5
JJ947	2	2	2	4	3.5	4	4	5	6	1	1	1	2	2	2	2	2.5	3
JJ1043	4	4	4	6	6	6	6	6	6	2	2	2	3	3	3	3	3	3
Min	1.5	2	1.5	3	2	3	4	3	4	1	1	1	1.5	1	1.5	2	1.5	2
Max	8	7	7	7	7	7	8	8	8	4	3.5	3.5	3.5	3.5	3.5	4	4	4

Table 4.5: Quantitative antifungal response of 18 *B. velezensis* strains against three fungal pathogens.

Strains	Percent inhibition against three fungal pathogen											
	F. oxysporum				*F. graminearum*				*R. solani*			
	rep. 1	rep. 2	rep. 3	average	rep. 1	rep. 2	rep. 3	average	rep. 1	rep. 2	rep. 3	average
AB01	75	63	63	67	75	75	75	75	75	88	75	79
AP215	100	88	88	92	75	88	75	79	100	100	100	100
JJ951	63	63	63	63	63	63	63	63	63	75	63	67
JJ1284	50	50	63	54	88	88	88	88	75	88	75	79
AP52	75	88	63	75	88	88	88	88	75	75	75	75
JM204	50	63	75	63	88	88	88	88	63	75	75	71
JJ213	63	50	63	58	75	75	63	71	75	75	75	75
JM236	25	25	25	25	38	25	38	33	50	38	50	46
JJ334	50	50	38	46	63	75	75	71	75	63	75	71
JM907	25	38	38	33	38	50	50	46	88	100	100	96
AP45	63	75	50	63	75	75	75	75	63	75	75	71
AP81	63	63	63	63	63	75	75	71	100	88	100	96
AP46	38	38	38	38	75	63	63	67	63	75	88	75
AP202	88	88	88	88	75	75	75	75	63	50	63	58
JM199	63	63	75	67	75	75	75	75	75	88	88	83
JJ747	75	63	63	67	75	88	75	79	75	88	88	83
JJ947	25	25	25	25	50	50	50	50	50	63	75	63
JJ1043	50	50	50	50	75	75	75	75	75	75	75	75

Table 4.6: Cumulative antifungal score of each *B. velezensis* strain across all fungi.

Strains	Cumulative % inhibition across three fungi			
	rep. 1	rep. 2	rep 3.	average
AB01	75	75	71	74
AP215	92	92	88	90
JJ951	63	67	63	64
JJ1284	71	75	75	74
AP52	79	83	75	79
JM204	67	75	79	74
JJ213	71	67	67	68
JM236	38	29	38	35
JJ334	63	63	63	63
JM907	50	63	63	58
AP45	67	75	67	69
AP81	75	75	79	76
AP46	58	58	63	60
AP202	75	71	75	74
JM199	71	75	79	75
JJ747	75	79	75	76
JJ947	42	46	50	46
JJ1043	67	67	67	67

Appendix E

Table 4.7: Comparison of production of antimicrobial secondary metabolites produced by 18 B. *velezensis* strains. Peak areas of extracted ion chromatogram of selective compounds are shown. Note that peak area of four derivatives of iturin (or bacillomycin L for JJ334, JJ947, and JJ1043), five derivatives of fengycin, and four derivatives of surfactin were combined and shown here. (itu. = iturins, fen. = fengycin, sur. =surfactin, dif. = difficidin, bac.=bacillaene, mac.=macrolactin W)

Strains	Peak area of extracted ion chromatogram (EIC) of selective antimicrobial secondary metabolites								
	iturins	fengycins	surfactins	itu.+fen.+sur.	difficidin	bacillaene	macrolactin W	dif.+bac.+mac.	itu.+fen.+sur.+dif.+bac.+mac.
AB01	2.61E+09	7.72E+09	8.80E+10	9.83E+10	4.76E+08	7.07E+07	3.54E+07	5.82E+08	9.89E+10
AP202	1.88E+09	6.54E+09	4.70E+10	5.54E+10	1.48E+09	9.27E+07	1.87E+07	1.59E+09	5.70E+10
AP215	1.73E+09	7.96E+09	5.90E+10	6.87E+10	3.44E+09	1.71E+08	2.75E+07	3.64E+09	7.23E+10
AP45	6.92E+08	5.27E+09	6.87E+10	7.47E+10	1.51E+09	5.22E+07	3.10E+07	1.59E+09	7.63E+10
AP46	4.25E+09	9.23E+09	5.35E+10	6.70E+10	8.30E+08	7.58E+07	1.03E+08	1.01E+09	6.80E+10
AP52	7.33E+08	5.09E+09	4.37E+10	4.95E+10	1.54E+09	1.65E+08	2.58E+07	1.73E+09	5.12E+10
AP81	1.62E+09	8.51E+09	7.39E+10	8.41E+10	9.83E+08	8.21E+07	3.14E+07	1.10E+09	8.52E+10
JJ1043	2.41E+09	8.26E+08	4.58E+10	4.91E+10	3.49E+08	2.73E+08	4.10E+07	6.63E+08	4.97E+10
JJ1284	1.61E+09	5.65E+09	7.44E+10	8.17E+10	1.54E+09	4.42E+07	4.38E+07	1.63E+09	8.33E+10
JJ213	5.23E+09	9.40E+09	6.94E+10	8.41E+10	1.36E+09	1.85E+08	1.27E+08	1.67E+09	8.57E+10
JJ334	8.02E+09	5.14E+09	4.22E+10	5.54E+10	6.19E+08	4.37E+08	9.04E+07	1.15E+09	5.65E+10
JJ747	5.25E+08	5.80E+09	6.01E+10	6.64E+10	1.18E+09	4.28E+07	1.36E+07	1.24E+09	6.77E+10
JJ947	4.09E+09	5.51E+09	6.19E+10	7.15E+10	1.37E+09	6.71E+07	1.21E+08	1.56E+09	7.31E+10
JJ951	2.70E+09	9.52E+09	8.81E+10	1.00E+11	4.78E+09	5.44E+07	5.83E+07	4.89E+09	1.05E+11
JM199	1.20E+09	6.09E+09	4.72E+10	5.45E+10	1.17E+09	3.93E+07	1.21E+08	1.33E+09	5.58E+10
JM204	1.61E+09	7.28E+09	6.78E+10	7.67E+10	2.18E+09	7.10E+07	4.36E+07	2.29E+09	7.90E+10
JM236	7.84E+08	3.65E+09	4.74E+10	5.18E+10	5.75E+08	1.92E+08	3.07E+07	7.98E+08	5.26E+10
JM907	1.26E+09	4.89E+09	1.99E+08	6.35E+09	1.37E+09	3.58E+08	1.67E+08	1.90E+09	8.24E+09

Table 5.1: Diversity of iturnic secondary metabolites produced by 18 *B. velezensis* strains

***B. velezensis* strain IDs**	**Iturinic secondary metabolites**
AB01	iturin
AP45	iturin
AP46	iturin
AP52	iturin
AP81	iturin
JM199	iturin
AP202	iturin
JM204	iturin
JJ213	iturin
AP215	iturin
JM236	iturin
JJ334	bacillomycin L
JJ747	iturin
JM907	iturin
JJ947	bacillomycin L
JJ951	iturin
JJ1043	bacillomycin L
JJ1284	iturin

Appendix F

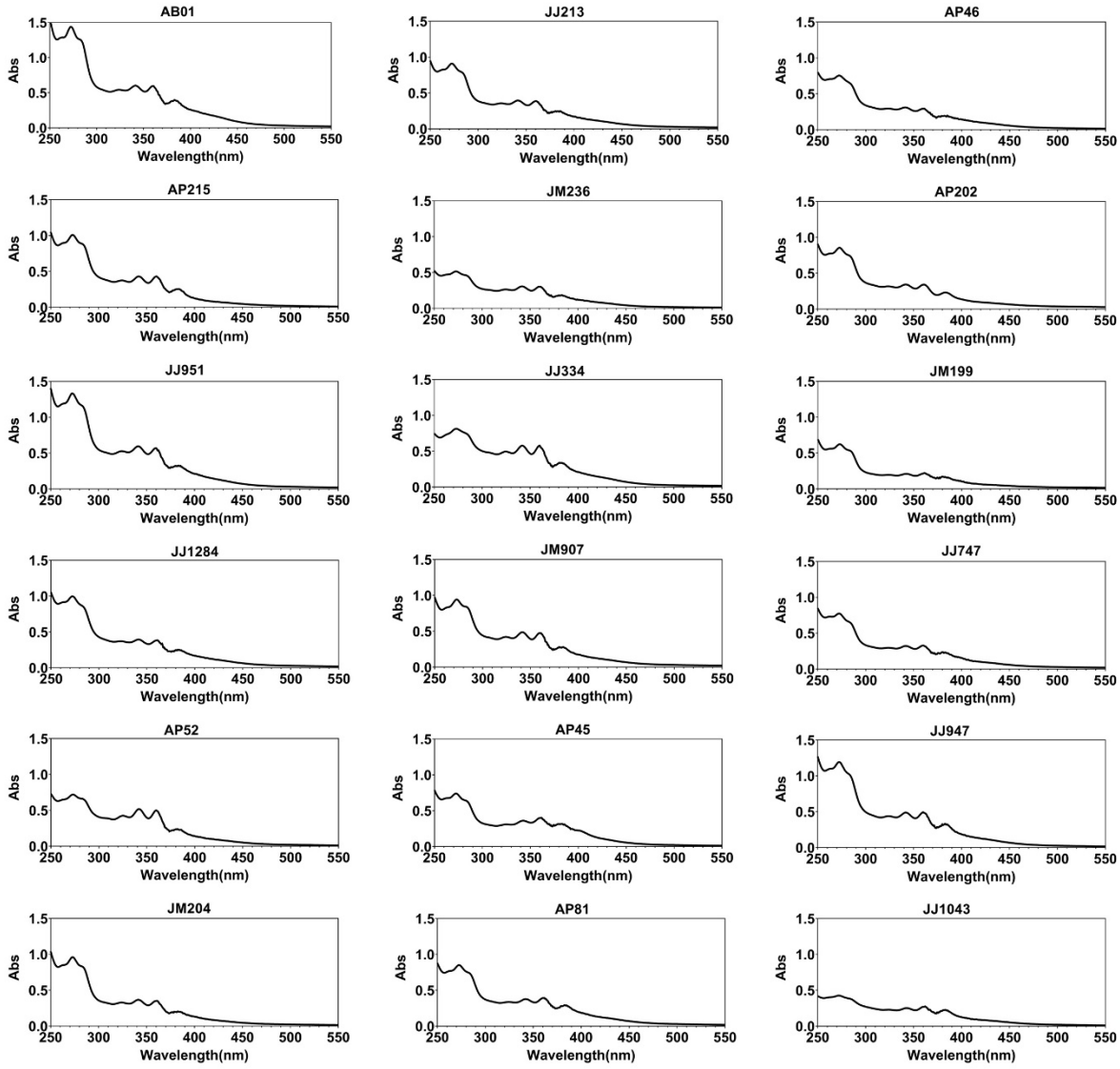


Figure 6.1: UV-vis absorption spectra of the total extract from the culture of 18 *Bacillus velezensis* strains.

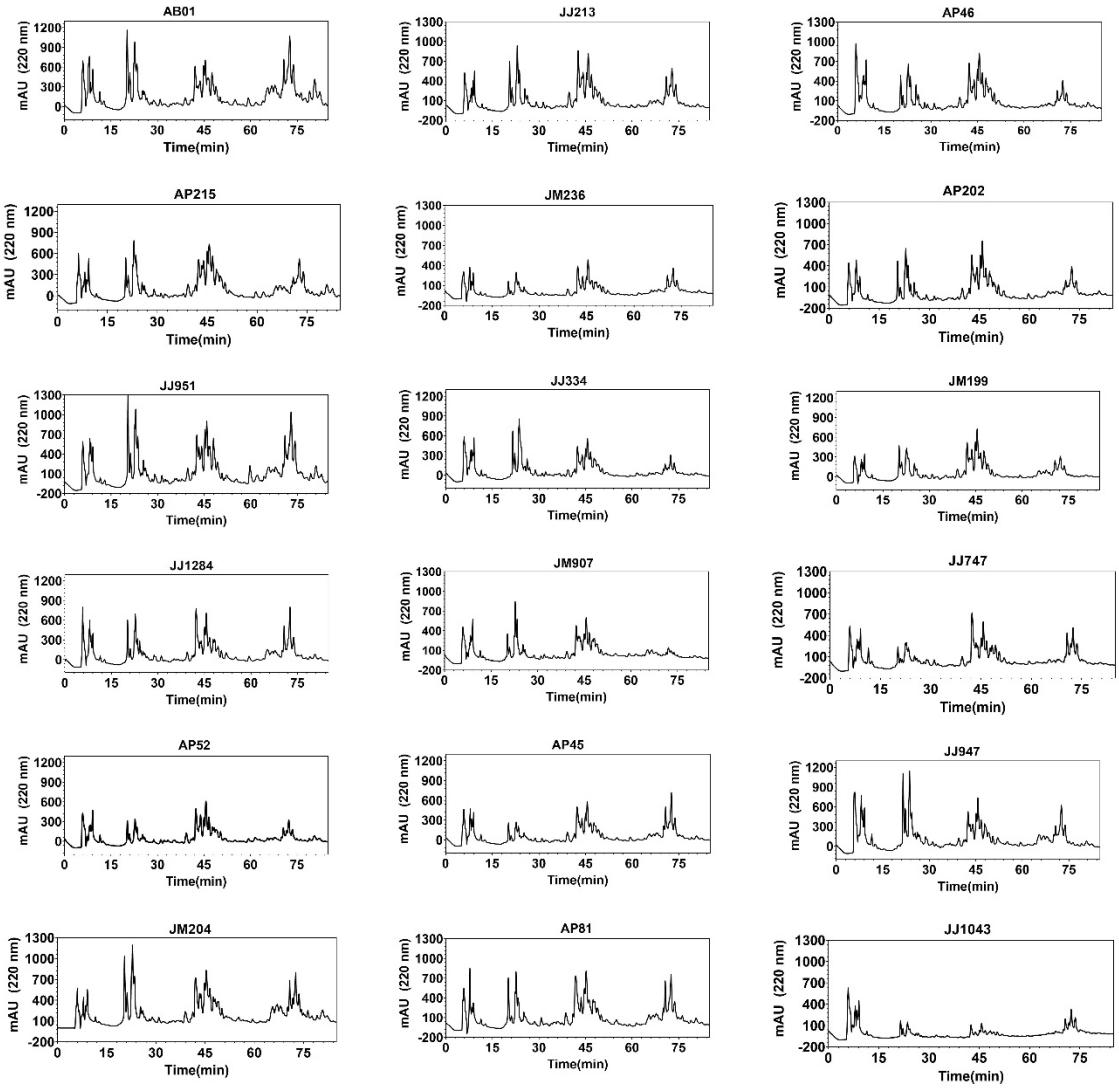


Figure 6.2: HPLC elution profile of the total extract of 18 *B. velezensis* strains.

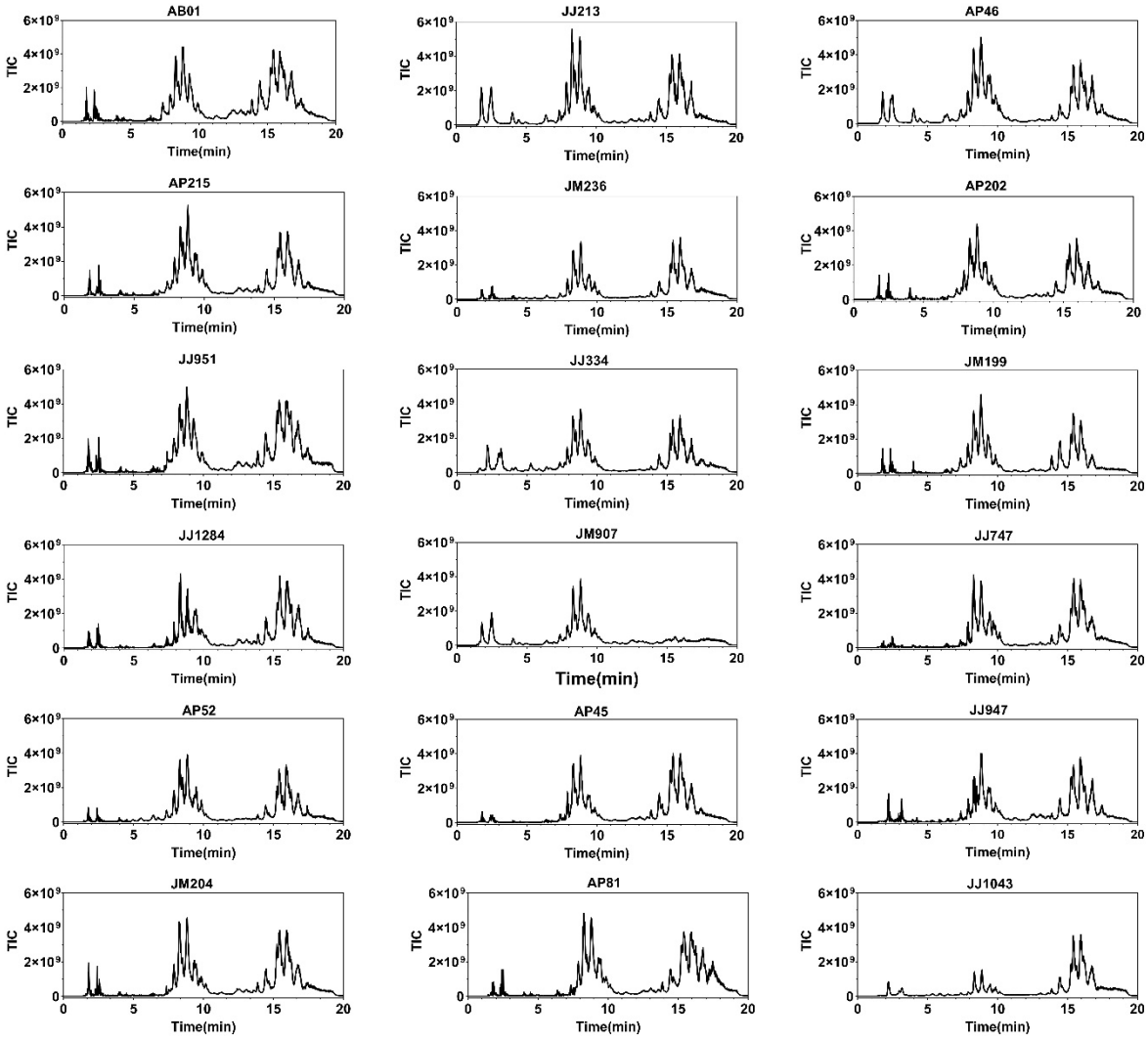


Figure 6.3: LC elution profile based on total ion count (TIC) of the total extract of 18 *B. velezensis* strains.

Appendix G

Amino acid sequences of five BGC-affiliated P450s that were used to construct high-resolution homology models are shown below:

CYP134 (accessory P450 encoded in cyclodipeptide-producing BGC)

MNQSIKLFSVLSDQFQENPYAYFSQLREEDPVHYEESIDSYFISRYHDVRYILQHPDIFTTKSLVQRAEP
VMRGPVLAQMHGKEHSAKRRIVVRSFVGDALDHLSPLIKQNAENLLAPYLERGRIDLVNDFGKTFAV
CVTMDMLGLDKRDHEKIAEWHSGVADFITSISQTPEARAHSLWCSEQLSQYLMPVIEERRVNPGSDLI
SILCTSEYEGMAMSDKDILALILNVLLAATEPADKTLALMIYHLLNNPEQMNDVLADRSLVPRAIAET
LRYKPPVQLIPRQLSQDTVVGGMEIKKDTIVFCMIGAANRDPEAFERPDVFHIHREDLGIKSAFSGAAR
HLAFGSGIHNCVGAAFAKTEIEIVANIVLDKMRNIRLEEGFRYAESGLYTRGPVSLHVAFDRA

CYP113 (accessory P450 encoded in difficidin-producing BGC)

MTSLTKIRQQQPYKWYQTMRETSPVHYNEKEDCWEIFTYDEVKRVISDYSHFSSDHKYLSADKQEK
MIRHINKDSLLKMDPPEHTVFRKLVNQPFMPKSVESLAPRIAAIADDLLQAVRSKGRMDIIEDYAFPLP
IIVIAELLGFPPKDRDIFKSWVDQSQNVKDEKKMNEVQKQMIGYFMQFILQRRKQPQNDLISHLISADL
DGEPLSDKQLIGFCGLLIVAGHVTTENVIGNSFLSLKEFPHILPRLLENKALLPDFIEEVIRLRPSIQRVTR
YTAVESEIGGKTIPAGEKVYAWIGSANRDEKKFENADQIDLGRKPNQHLSFGQGSHYCLGAPLARLE
AKIALSHFFEQMPAWRFTEDQEPNLVPSPVFHGVDRLLVEF

CYP109 (accessory P450 encoded in bacillibactin-producing BGC)

METTSPSAVQKTLLRGKNKQDPYHPFDWYANMRQTSPVHFDEASQTWSVFTYEEAKRVTIDKDTFSS
QPPKNQRKHSLMKTMVMMDPPNHTRVRSIVSKAFTPRVMKLWEPRIYELMDELMAQLEGKKEIDLV
QDISYPLPVIVIAELLGVPSEHKQSFKEWSDILVSMPKSENEEDVAEWQKTRDKGEADMMAFFADTIE
KKRHNLGDDLISLIIQAEENGDKLAADELIPFCNLLLLAGNETTTNLISNMIFSLLEQPGAYEALAQSPE
LIPRAVEEAVRFRAPAPTIVRYVTKDTELGGKVLKKGDNVIVFLASANRDERQFSNAHEFDIHRHPNP
HIGFGHGIHFCLGAPLARLEACTAIKILIERYEALELLSYVPMTSSSMYGLKELKLCVTPRS

CYP107 (accessory P450 encoded in bacillaene-producing BGC)

MEKLMFHPHGKEFHHNPFSVLGRFREEEPIHRFELKRFGATYPAWLITRYDDCMAFLKDNRITRDVK
NVMNQEQIKMLNVSEDIDFVSDHMLAKDTPDHTRLRSLVHQAFTPRTIENLRGSIEQIAEQLLDEMEK
ENKADIMKSFASPLPFIVISELMGIPKEDRSQFQIWTNAMVDTSEGNRELTNQALREFKDYIAKLIHDR
RIKPKDDLISKLVHAEENGSKLSEKELYSMLFLLIVAGLETTVNLLGSGTLALLQHMKECEKLKQHPE
MIATAVEELLRYTSPVVMMANRWAIEDFTYKGHSIKRGDMIFIGIGSANRDPNFFENPEILNINRSPNR
HISFGFGIHFCLGAPLARLEGHIAFNALLKRFPDIELAVAPDDIQWRKNVFLRGLESLPVSLSK

CYP102A2 (CYP domain) (accessory P450 encoded in plantazolicin-producing BGC)

MKETGPIPQPKTFGPLGNLPLLDKDKPTMSLIKLANEQGPIFQLHTPAGAIIVVSGHELVKEVCDEERF
DKSIEGALEKVRAFSGDGLFTSWTHEPNWRKAHHILMPTFSQRAMKDYHSMMTDIAVQLIQKWARL
NPDEAVDVPADMTRLTLDTIGLCGFNYRFNSYYRETPHPFINSMVRALDEAMHQMQRLDVQDKLMI
RTKRQFHHDIQAMFSLVDSIIAERRSGGRDEKDLLARMLNVEDPETGEKLDDENIRFQIITFLIAGHETT
SGLLSFAIYFLLKHPRVLEKAYEEADRVLTDPVPSYKQVLDLTYIRMILQESLRLWPTAPAFSLYAKED
TVIGGKYPITPKDRISVLIPQLHRDKDAWGDNAEEFYPERFEHPDRVPHHAYKPFGNGQRACIGMQFA
LHEATLVLGMILQHFTFIDHTDYELDIKQTLTIKPGDFHIRVRPRNKEDVA

Amino acid sequence of full-length ***B*aCYP102A2** with His_6-tag at the N-terminal

MHHHHHHKETGPIPQPKTFGPLGNLPLLDKDKPTMSLIKLANEQGPIFQLHTPAGAIIVVSGHELVKEV
CDEERFDKSIEGALEKVRAFSGDGLFTSWTHEPNWRKAHHILMPTFSQRAMKDYHSMMTDIAVQLIQ
KWARLNPDEAVDVPADMTRLTLDTIGLCGFNYRFNSYYRETPHPFINSMVRALDEAMHQMQRLDVQ
DKLMIRTKRQFHHDIQAMFSLVDSIIAERRSGGRDEKDLLARMLNVEDPETGEKLDDENIRFQIITFLIA
GHETTSGLLSFAIYFLLKHPRVLEKAYEEADRVLTDPVPSYKQVLDLTYIRMILQESLRLWPTAPAFSL
YAKEDTVIGGKYPITPKDRISVLIPQLHRDKDAWGDNAEEFYPERFEHPDRVPHHAYKPFGNGQRACI
GMQFALHEATLVLGMILQHFTFIDHTDYELDIKQTLTIKPGDFHIRVRPRNKEDVAAALPAAEKAAED
VGKEKRETKGASIIGLDNRPLLILYGSDTGTAEGVARELADTAGMHGVRTETAPLNDRIGKLPKEGAL
LIITSSYNGKPPSNAGQFVQWLEEVKPGELEGVRYAVFGCGDHNWAATYQAVPRLIDEKLAEKGAER
FSSRGEGDVSGDFEGKLDEWKKSMWTDAMKAFGLKLNENAEKERSALGLQFVSGLGGSPLAQTYEA
VYASVAENRELQAPESGRSTRHIEITLPKEAAYHEGDHLGVLPVNSKEQVSRVLRRFNLNGNDQVLLT
ASGQSAAHLPLDRPVRLHDLLSSCVELQEAASRAQIREMAAYTVCPPHKRELEDFLEEGVYQEQILTS
RVSMLDLLEKYEACELPFERFLELLRPLKPRYYSISSSPRKHPGQASITVGVVRGPARSGLGEYRGVAS
NYLADRGPEDGIVMFVRTPETRFRLPEDPEKPIIMVGPGTGVAPFRGFLQARAALKKEGKELGEAHLY
FGCRNDHDFIYRDELEAYEKDGIVTLHTAFSRKEGVPKTYVQHLMAKDAGALISILGRGGHLYVCGD
GSKMAPDVEATLQKAYQSVRETDERQAQEWLLDLQTKGIYAKDVWAGI

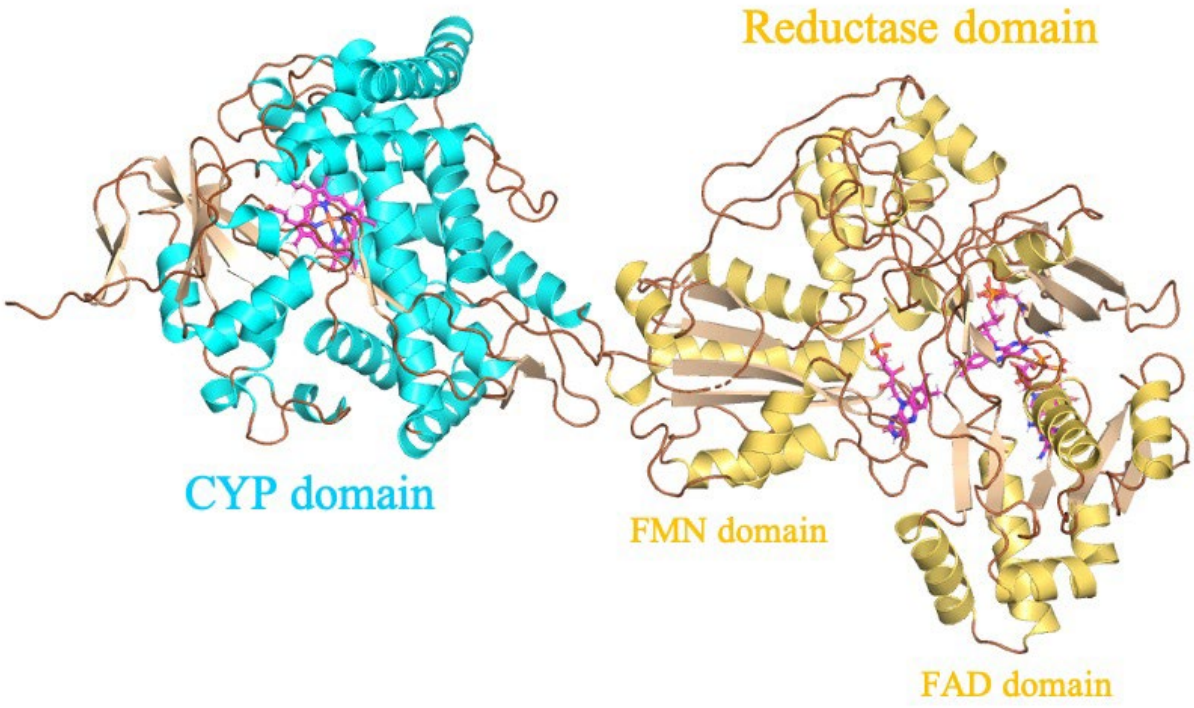


Figure 7.1: High-resolution homology model of a full-length *Ba*CYP102A2 enzyme constructed by RosettaCM. Due to the lack of a full-length crystal structure, a continuous sequence connecting both CYP and reductase domain is not available. Therefore, a model for each domain was constructed separately using the RosettaCM suite of programs and then placed in their linking position.

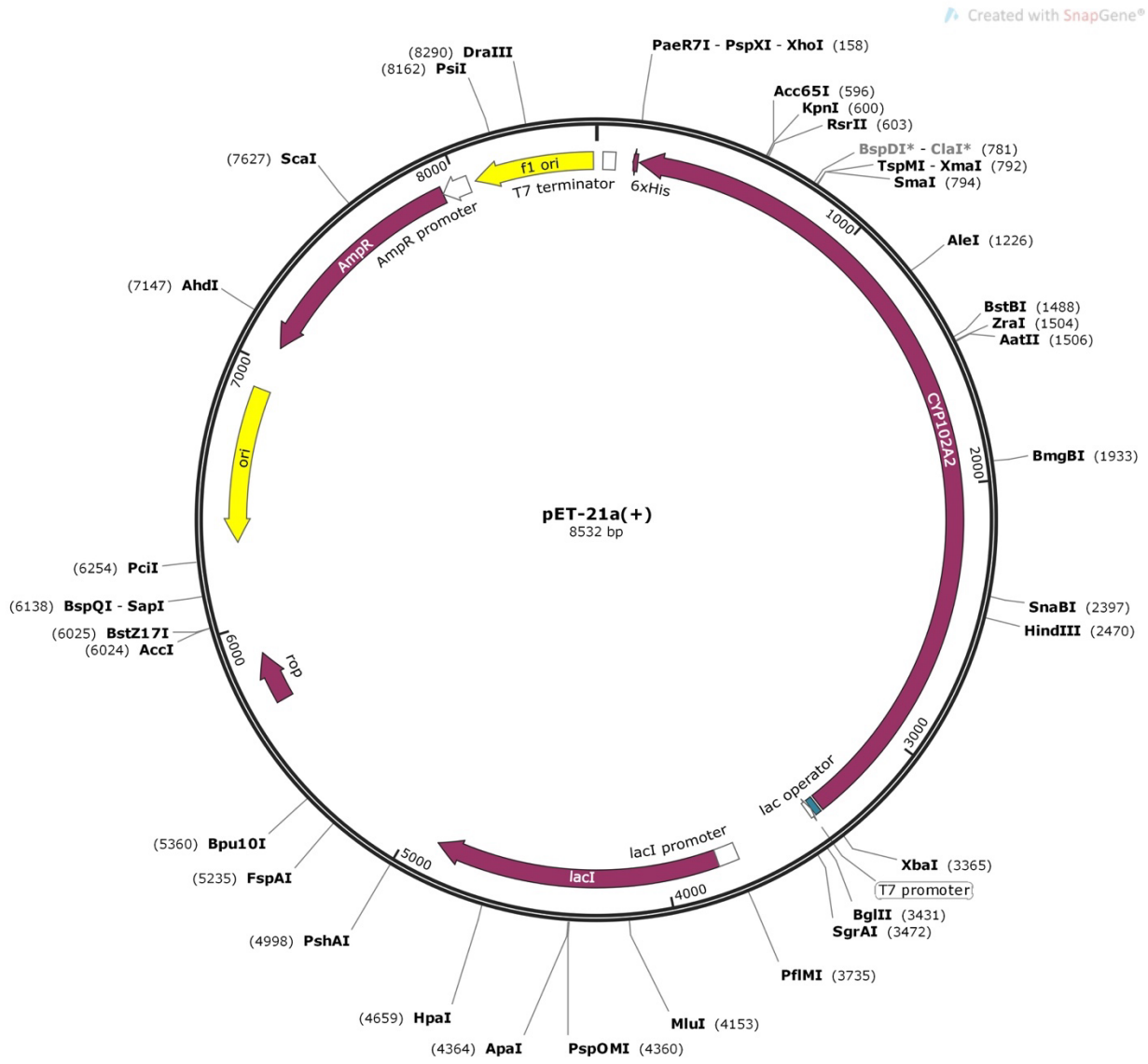


Figure 7.1: Schematic of our pET21a (+) construct-containing the full-length *Ba*CYP102A2 gene.