

Isolation, identification, and characterization of antibiotic-producing microorganisms from soil



University of Venda

by

Adivhaho Khwathisi

Thesis

Submitted in fulfilment of the requirements for the degree

PHILOSOPHIAE DOCTOR

in

MICROBIOLOGY

In the

Department of Biochemistry and Microbiology

Faculty of Science Engineering and Agriculture

At the

University of the Venda

Promoter: Prof. A Samie

Co-Promoters: Prof. N.E Madala & Prof A.N Traore

March 2025

A special dedication to my late parents:

I dedicate this work to my late parents, Vho-Khwathisi “Nndileleni” Andries and Khwathisi Thifhelimbilu Cecelia. I only wish you were here to witness this achievement. I will forever be grateful for your unwavering support and for giving me the opportunity to pursue my dreams, especially when other parents were encouraging their children to work. I feel truly blessed and humbled to have had both of you as my parents. This is for you!!

Declaration

I, **Adivhaho Khwathisi** [student number: **11625080**], declare that; this thesis titled, “Isolation, identification, and characterization of antibiotic producing microorganisms from soil,” is my original work and that I have not submitted it previously in its entirety or partially for any other qualification. The thesis is being submitted for the degree of PhD in Microbiology at the University of Venda, Thohoyandou, South Africa.



Adivhaho Khwathisi

Preface

This thesis explores the bioprospection of soil-borne microorganisms and chemical dereplication of their antimicrobial-microbial constituents, employing ultrahigh performance liquid chromatography quadrupole time of flight mass spectrometry (UHPLC-q-TOF-MS) and Molecular networking approach; and whole genomic analysis to assess their biosynthetic potential, thereof. Moreover, given that metabolomics studies are primarily data-driven in offering insights into a certain area of study rather than following a predefined hypothesis, novel findings of this study are also included as objectives and experimental chapters in this thesis. This thesis is therefore presented as a compilation of seven chapters in which experimental chapters (chapter 3-6) were written in the journal format in which they were published or submitted. The outline of this thesis is shown below:

Chapter 1: Background, study rationale, and objectives. Despite the advancement in natural products and drug discovery, this **chapter** articulates the foundation for this research by providing an overview of the research, highlighting its significance and relevance. It justifies the need for the study by identifying gaps in the current knowledge and explains why this research is important. The **chapter** clearly states what the study aims to achieve, setting the stage for the rest of the research.

Chapter 2: Literature Review. This **chapter** integrates existing research on the topic, systematically identifies knowledge gaps, and contextualizes the current study within the broader scientific landscape, providing a comprehensive background that informs the research methodology. It emphasizes the importance of bio-guided approaches in the discovery of bioactive natural products and highlights the transformative impact of advancements in analytical techniques and computational tools, including molecular networking, MS2LDA, and *in silico* tools. By integrating these elements, the chapter establishes a robust foundation for the study's objectives and methodological framework.

Chapter 3: Isolation and identification of antibiotic-producing bacteria from soil. This **chapter** outlines the systematic isolation of soil-derived bacteria capable of producing antimicrobial compounds, followed by their characterization and preliminary efficacy testing. Soil samples were collected and subjected to selective culturing to isolate bacterial strains with potential bioactive properties. These isolates are subsequently screened against clinically relevant pathogenic strains to evaluate their antimicrobial activity. The **chapter** employs 16S rRNA

gene sequencing to taxonomically classify promising strains, establishing a molecular foundation for downstream analyses aimed at elucidating their biosynthetic capabilities.

Chapter 4: Assessing the metabolic landscape of soilborne *Bacillus* spp. with an aid of UPLC-qTOF-MS and computational metabolomics. This **chapter** focuses on utilizing digitalized LC-MS/MS and computational tools like molecular networking to analyse the chemical constituents of bacterial methanolic extracts. This chapter aims to understand the metabolic profile of the isolated bacteria, providing insights into their potential to produce valuable compounds. By employing advanced computational metabolomics tools to enhance metabolite identification and annotation, this study adopts modern methodologies that address longstanding challenges in natural product research. The integration of molecular networking, substructure annotation via MS2LDA, and other *in silico* approaches exemplifies contemporary strategies designed to transcend the limitations of traditional NP discovery frameworks.

Chapter 5: Genomic Insights into three soilborne *Bacillus* sp.: Unveiling their Potential as a plant growth promoting and Antibiotic-producing Bacterium. This **chapter** delves into the genomic elements that contribute to the adaptability and biosynthesis of antimicrobial compounds. It provides the precise taxonomic classification of the bacterial strains and explores their potential as antibiotic producers, offering a deeper understanding of their genetic capabilities.

Chapter 6: Multi-parametric optimization of Data Dependent Acquisition towards more holistic bacterial metabolite coverage through Molecular networking. This **chapter** addresses the analytical challenges in metabolite detection by proposing group-based optimization over single-point optimization for more comprehensive metabolite coverage and in-depth exploration. It focuses on enhancing metabolome coverage using molecular networking to improve the detection of bacterial metabolites, ensuring a more comprehensive analysis of the bacterial extracts.

Chapter 7 General Conclusion and recommendations

Abstract

The emergence and spread of antibiotic resistance in pathogenic bacteria have posed a serious concern in global healthcare. Highlighting a need for novel antimicrobial agents with diverse mechanisms of action to combat current pathogenic threats. Soil-borne bacteria are constantly exposed to various environmental stresses, which leads to the production of antimicrobial compounds as a strategy for their survival. In this context, soil-borne bacteria are key producers of antimicrobials with valuable applications in medicine, agriculture, and animal husbandry. The main aim of this study was to isolate, identify, and characterize microorganisms with the potential to produce antimicrobial compounds from soil.

Preliminary screening was conducted using agar well diffusion, revealing that the bacterial isolates exhibited significant antimicrobial activity against the four pathogenic bacterial strains tested (**Chapter 3**). Furthermore, the use of 16S rRNA sequencing aided the identification of the active bacterial isolates at the molecular level, where isolates 1 and 2 were identified as strains of *Bacillus pumilus*, whilst isolate 3 was found to be *Bacillus subtilis* (**Chapter 3**). This study further explored the chemical composition of the bacterial extract using ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry and classical molecular networking. Computational metabolomic analysis revealed that metabolites from these soilborne *Bacillus* species are predominantly composed of peptides, dipeptides, benzenoids, organic acids, and derivatives. Interestingly, the methanolic crude extract of the *Bacillus subtilis* isolate was mainly characterized by the abundance of dipeptides and lipopeptides, such as surfactins, suggesting that the combination of cyclopeptides and lipopeptides together could be the attribute of the enhanced antimicrobial activity observed in agar well diffusion (**Chapter 4**).

The current study further explored the applications of Whole Genome Sequencing (WGS) to enhance understanding of the overall distribution of genomic elements that contribute toward the adaptability and biosynthesis of antimicrobial compounds (**Chapter 5**). The results of the WGS showed that *Bacillus* species presented several genes for adaptability, such as genes that encode for motility, quorum sensing, stress response, desiccation tolerance, heavy metal tolerance, synthesis of siderophores, and most genes involved in antibiotic biosynthesis, such as NRPS, beta lactone, RiPP, and terpene gene clusters. Furthermore, the use of WGS aided a more precise taxonomic classification of the identities of the strains, highlighting the importance of WGS in the identification of bacterial isolates.

Lastly, inspired by the diverse chemical composition and bioactivity of *Bacillus subtilis*, this study further explored the potential to enhance metabolite detection and expand the data coverage in the data-dependent acquisition (DDA) mode of the UHPLC-qTOF-MS technique (**Chapter 6**). Interestingly, using Central composite design (CCD), this study has demonstrated that even though both collision energy and intensity threshold independently enhance metabolite coverage, it is noteworthy that when both parameters are combined, the comprehensiveness in maximizing metabolome coverage is even greater. Therefore, revealing that to enhance exploration and maximize the potential of untargeted metabolomics, this study recommends future studies transitioning from single-point optimization to group-based optimization. Overall, contributes to the growing field of natural product discovery and provides insights into the biosynthetic pathways and insight into the chemical diversity of microbial secondary metabolites. Furthermore, these findings highlight the diverse potential of microbial-derived compounds as potential sources of new bioactive molecules.

Dedications

This work is dedicated to my son Ronewa, you have Daddy's same kind heart and compassionate nature, hold on to that, even when life gets hard. There will be times in your life where you'd rather hide or run or bury your head in the sand than face whatever challenge is in your way. I know because I've been there many times. In those times, I want you to remember to be brave and show courage.

Acknowledgements

I would like to express my sincere appreciation to the following people for their individual contributions to my social and academic life:

- My supervisor, **Prof. Amidou Samie** is thanked for allowing me an opportunity to study and pursue my dream of having a Ph.D. Your patience from my early honours, master's, and now Ph.D., is really appreciated. I thank you for your unique supervision style, guidance, and support. I thank you for all the wonderful, educative, and insightful discussions we had.
- My co-supervisor, **Prof Ntakadzeni Edwin Madala**, I get very emotional thinking of all the effort you have made in my study and personal life in general. I knew the day you became part of this study that my life will change for good. Never in a single day, you failed to impress me with your knowledge. Your knowledge of advance Mass spectrometry is amazing. I thank you for sharing knowledge with me and working with me when I needed supervision the most. Your strong command of basic writing language is the only reason this thesis is even written in the first place. I thank you for being instrumental for my career development.
- My second co-supervisor, **Prof Afsatou Ndama Traore**, you have been a mother and support structure since day one and I just want to thank you for your support and guidance in the journey. I wholeheartedly appreciate all your support and guidance.
- My laboratory friends and colleagues, **Dr. Anza-Tshilidzi Ramabulana-Mhlongo**, words cannot express my gratitude. You came in and changed the game. Thank you for always sharing your knowledge and wisdom. From observing you from a distance, I've learned that true blessings come from those who give. Learning about molecular networking has been one of the best experiences of my life, and I am incredibly grateful for your selflessness. It is because of you that this thesis is what it is today. **Dr. Barbra Moyo**, thank you for always offering a helping hand and for encouraging me when Prof was being tough on me. I am also deeply grateful for your willingness to share your knowledge. **Dr. Ronewa Khumela, Phindulo Mathobo, Nzumbululo Ndou, Dr. Dakalo Ndou, Dakalo Mugwena, Consol Kubayi, Tendani Matamela, Dr. Edgar Musie**, and **Dr. Mpho Magwalivha** thank you for all the advice and insightful discussions you shared throughout this journey. Whenever things weren't going as planned, you were always the first to step in and offer support.
- **Rabelani**, words can't express my sincere gratitude, I wholeheartedly thank you for your love and support.
- **Dr Kedibone Masinye**, Thank you for the genomic analysis.
- University of Venda and the NRF are thanked for the scholarships I received throughout my studies.
- To God Almighty, I thank you for the mercies that a forever new every day and the grace that is always sufficient. I am constantly encouraged by the promise in **Isaiah 43 verse 2**, "When you pass through the waters, I will be with you and when you pass through the rivers, they will not sweep over you".

Conference presentations

1. **Adivhaho Khwathisi**, Afsatou Ndama Traore, Ntakadzeni Edwin Madala, and Amidou Samie. Bioprospecting of soil-borne microorganisms and chemical dereplication of their anti-microbial constituents with the aid of UPLC-QTOF-MS and molecular networking approach. 2024. SASBMB Congress 2024, 7–10 July, The Ranch Hotel, Polokwane.
2. **Adivhaho Khwathisi**, Ntakadzeni Edwin Madala, Afsatou Ndama Traore, and Amidou Samie. Analysis of anti-microbial constituents of soil-borne microorganisms with the aid of UPLC-QTOF-MS and Molecular Networking approach. 2nd annual SAMRC-Univen antimicrobial resistance & global health research unit symposium, 26-27 September 2024, 2ten hotel, Sibasa.

Publications

1. Bioprospecting of soil-borne microorganisms and chemical dereplication of their anti-microbial constituents with the aid of UPLC-QTOF-MS and molecular networking approach. Khwathisi A, Madala NE, Traore AN, Samie A. 2024, <https://doi.org/10.7717/peerj.17364>
2. Multi-parametric optimization of Data Dependent Acquisition towards more holistic bacterial metabolite coverage through Molecular networking. Khwathisi A, Madala NE, Traore AN, Samie A. <https://doi.org/10.1155/ijm/4388417>
3. Genomic Insights into soilborne *Bacillus* sp.: Unveiling their Potential as a plant growth promoting and Antibiotic-producing Bacterium. Khwathisi A, Masenye K, Traore AN, Madala NE, Samie A. To be submitted to *Journal of Biotechnology*

Table of contents

Contents

Declaration.....	ii
Preface.....	iii
Abstract.....	v
Dedications	vii
Acknowledgements.....	viii
Conference presentations	ix
Publications.....	x
Table of contents.....	xi
List of figures.....	xiv
List of tables.....	xvi
List of abbreviations	xvii
CHAPTER 1	1
Study Background, Rationale, And Objectives	1
1.1 Background.....	1
1.2 Study Rationale.....	4
1.3 OBJECTIVES OF THE STUDY	6
1.3.1 MAIN OBJECTIVE.....	6
1.3.2 SECONDARY OBJECTIVES.....	6
1.4 Hypothesis.....	6
1.5 Research questions.....	6
1.6 REFERENCE.....	7
CHAPTER 2	11
LITERATURE REVIEW	11
2.1 Antimicrobial resistance (AMR).....	11
2.2 History of antibiotic and their nomenclature	12
2.3 Mechanisms of AMR and microbes involved	13
2.4 The development of antibiotics.....	15
2.5 Soil as a source of microorganisms	18
2.6 Antibiotic-Producing Microorganisms from the soil.....	18
2.7 Secondary metabolites from soil borne bacteria.....	19
2.7.1 Bacterial secondary metabolites	19
2.7.2 Microbial biosynthesis of secondary metabolites	20
2.8 Analysis of secondary metabolites through Analytical techniques	24
	xi

2.8.1 Isolation and cultivation for the production of secondary metabolites	24
2.8.2 Genome Mining of WGS for secondary metabolites.....	25
2.8.3 Chemistry techniques used to analyse secondary metabolites.....	26
2.8.4 Chromatographic techniques in secondary metabolites analysis (HPLC).....	27
2.8.5 Mass spectrometric characterisation of secondary metabolites	28
2.8.6 Nuclear magnetic resonance spectroscopy (NMR).....	29
2.9 Computational Tools for analysis of these metabolomics data	30
2.9.1 Molecular networking	30
2.9.2 Spec2vec	32
2.9.3 MS2LDA.....	32
2.9.4 GNPS <i>in silico</i> tools.....	33
2.10 References.....	35
Chapter 3: Isolation and identification of antibiotic-producing bacteria from soil.	46
3.1 Summary	46
3.2 Introduction.....	47
3.3 Experimental.....	49
3.3.1 Sample collection.....	49
3.3.2 Isolation of antibiotic producing strains	49
3.3.3 Primary screening of antimicrobial compounds	49
3.3.4 Taxonomic identification of the bacterial isolates	50
3.4 RESULTS	51
3.4.1 Morphological characteristics.....	51
3.4.2 Antimicrobial activity testing	51
3.4.3 16S rDNA Aided Identification of Bacterial Isolates	52
3.5 DISCUSSION	55
3.6 Conclusions.....	56
3.7 Reference	57
Chapter 4: Assessing the metabolic landscape of soilborne <i>Bacillus spp.</i> with an aid of UPLC-qTOF-MS and computational metabolomics	60
4.1 Summary	60
4.2 Introduction.....	61
4.3 EXPERIMENTAL	63
4.3.1 Metabolite extraction	63
4.3.2 Liquid Chromatography Mass Spectrometry analysis.....	63
4.3.3 Metabolite annotation and Molecular Networking	64

4.4 RESULTS	65
4.4.1 UPLC-QTOF-MS Analyses.....	65
4.5 Discussion	70
4.6 Conclusion	72
4.7 Reference	79
Chapter 5: Genomic Insights into three soilborne <i>Bacillus</i> sp.: Unveiling their Potential as a plant growth promoting and Antibiotic-producing Bacterium	82
5.1 Summary	82
5.2 Introduction.....	83
5.3 Experimental	85
5.3.1 Isolation and culture maintenance	85
5.3.2 DNA extraction and Genome sequencing.....	85
5.3.3 Genome Assembly and bioinformatics analysis	85
5.3.4 Genome Comparison	85
5.4 Results.....	86
5.4.1 Genomic Characteristics and Assembly Metrics	86
5.4.2 Genome annotation of the <i>Bacillus</i> spp strains.....	92
5.4.3 The predictive analysis of secondary metabolites	98
5.4.5 Antibiotic-resistant gene prediction.....	100
5.5 Discussion	103
5.6 Conclusion	107
5.7 Reference	108
Chapter 6: Multi-parametric optimization of Data Dependent Acquisition towards more holistic bacterial metabolite coverage through Molecular networking.....	113
6.1 Summary	113
6.2 Introduction.....	114
6.3 Experimental	116
6.3.1 Bacterial cultivation and Metabolite extraction.....	116
6.3.2 Liquid Chromatography Mass Spectrometry analysis.....	116
6.3.3 Metabolite annotation and Molecular Networking	117
6.3.4 Statistical analysis.....	118
6.4 Results.....	119
6.5 Discussion	130
6.6 Conclusions.....	134
6.7 REFERENCE.....	136
Chapter 7.....	141

7.1 General conclusion.....	141
7.2 Challenges and future studies	142

List of figures

Figure 1: Combination of factors driving the increase of antimicrobial resistance and the future of antimicrobial resistance.....	12
Figure 2: Bacterial mechanisms of resistance to antimicrobial agents include several common strategies	15
Figure 3: Timeline illustrating the discovery of different antibiotic classes used in clinical practice.	17
Figure 4: Applications of Microbial Secondary Metabolites, ...	20
Figure 5: Machine and deep learning algorithms in genomics.	26
Figure 6: Range of Sensitivities in Metabolomic Technologies.....	27
Figure 7: An illustration of the Stable Isotope-Resolved Metabolomic pipeline that integrates NMR and MS to enhance metabolite detection and measure metabolic fluxes.	29
Figure 8: An illustration of the procedure for constructing a classical molecular network within the GNPS ecosystem.....	31
Figure 9: An illustration of the procedure for GNPS in-silico tools within the GNPS ecosystem	34
Figure 10: Showing characteristics of three soil-borne isolates on nutrient medium.....	51
Figure 11: Infographic display of bacterial isolates extracts analyzed by agar diffusion technique against 4 pathogenic bacteria showing antimicrobial activity after 48 hours of incubation.	52
Figure 12: Phylogenetic tree based on 16s RNA gene fragment. The Neighbor-Joining technique was used to generate the tree.....	53
Figure 13: Phylogenetic tree based on the 16s RNA gene fragment. Evolutionary relationship inferred using the Neighbour-Joining method. In this study, the soil isolate is represented by a squared black dot..	54
Figure 14: Molecular network of Bacillus species extracts analysed through liquid chromatography-tandem mass spectrometry using electrospray ionisation (ESI) in Positive mode.....	66
Figure 15: Molecular network of Bacillus species extracts analysed through liquid chromatography-tandem mass spectrometry using electrospray ionisation in Positive mode..	68
Supplementary Figure 16: Representative UHPLC-qTOF-MS/MS chromatogram showing single ion chromatogram of putatively identified Isarrin D	69
Supplementary Figure 17: Representative base peak intensity chromatograms (BPC) indicating the separation of metabolites in methanol extracts of isolate 3	73
Supplementary Figure 18: Chemical classification by MolNetEnhancer of Positive ionization datasets acquired through UPLC-QTOF-MS	74
Supplementary Figure 19: Typical mass spectra of the fragmentation patterns of cyclo (L-Pro-L-Val) (A), cyclo (D-Pro-D-Leu) (B), cyclo(L-Tyr-L-Pro) (D) and cyclo (L-Pro-D-Phe) (C) with their typical structures.	75

Supplementary Figure 20: Molecular network Analysis of antimicrobial compounds from three bacterial isolates.	76
Figure 21: Heat-map, based on ANI values, for <i>Bacillus</i> sp. Strain KH1 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness	90
Figure 22: Heat-map, based on ANI values, for <i>Bacillus</i> sp. Strain KH2 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness	91
Figure 23: Heat-map, based on ANI values, for <i>Bacillus</i> sp. Strain KH2 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness.....	92
Figure 24: RAST annotation results for the genome of <i>Bacillus</i> sp. strain KH1 genome. The annotation categorizes the identified genes into various functional subsystems.....	94
Figure 25: Rast annotation results for the genome <i>Bacillus</i> sp. Strain KH2 genome. The annotation categorizes the identified genes into various functional subsystems.....	94
Figure 26: RAST annotation results for the genome of <i>Bacillus</i> sp. strain KH1 genome. The annotation categorizes the identified genes into various functional subsystems.....	95
Figure 27: A bar graph showing the functional classification of the genome of <i>Bacillus</i> sp. strain KH1 based on the EggNOG.	96
Figure 28: A bar graph showing the functional classification of the genome of <i>Bacillus</i> sp. strain KH2 based on the EggNOG.	97
Figure 29: A bar graph showing the functional classification of the genome of <i>Bacillus</i> sp. strain KH3 based on the EggNOG.	97
Figure 30: Identification of Gene Clusters for Antimicrobial Secondary Metabolites in <i>Bacillus subtilis</i> Bbv57 Using antiSMASH 6.0.....	100
Figure 31: Molecular networks derived from bacterial extracts illustrating the effect of collision energy on MS data acquisition.....	120
Figure 32: Molecular networks derived from bacterial extracts illustrating the effect of threshold intensity on MS data acquisition.	120
Figure 33: Showing the different molecular networks generated from the Design of Experiment (DOE) analysis of threshold and collision energy, visualized in Cytoscape. ...	122
Figure 34: (A) Upset plot illustrating the overlap of different metabolites among the ten experimental conditions. (B) Venn diagram showed the overlapped metabolites between four different conditions with the combined events.	
Figure_35: The molecular network of <i>Bacillus subtilis</i> extracts was analysed using liquid chromatography-tandem mass spectrometry with electrospray ionization in Positive mode..	127
Figure 36: Response surfaces evaluating the Synergistic effect of collision energy and Threshold of intensity on the metabolite coverage.	129

List of tables

Table 1: Biosynthetic gene clusters and their secondary metabolites, including NRPS and PKS pathways.	22
Table 2: Metabolite annotation table. Profiling of metabolites reported to level 2 or 3 of the Metabolomics standards initiative.	77
Table 3: Genome characteristic for <i>Bacillus</i> sp. strain KH1	86
Table 4: Pairwise comparisons of user genomes (KH1) vs. type strain genomes	88
Table 5: Pairwise comparisons of user genomes (KH2) vs. type strain genomes	88
Table 6: Pairwise comparisons of user genomes (KH3) vs. type strain genomes	89
Table 7: Predicted secondary metabolites gene cluster in the <i>Bacillus</i> sp. Strain	98
Table 8: Antimicrobial resistance genes predicted from the genome of <i>Bacillus</i> spp., using CARD	102
Table 9: Parameters that were varied for MS ² scan in a DDA method.....	117
Table 10: Showing the "Model Summary Statistics": Focus on the model maximizing the "Adjusted R-Squared" and the "Predicted R-Squared".	129

List of abbreviations

AMR -	Antimicrobial Resistance
ANI -	Average Nucleotide Identity
ATCC -	American Type Culture Collection
BGCs -	Biosynthetic Gene Clusters
BLAST -	Basic Local Alignment Search Tool
CARD -	Comprehensive Antibiotic Resistance Database
CCD -	Central Composite Design
COGs -	Clusters of Orthologous Groups
DDA -	Data-Dependent Acquisition
dDDH -	Digital DNA-DNA Hybridization
DIA -	Data-Independent Acquisition
DoE -	Design of Experiments
ESI -	Electrospray Ionization
GBDP -	Genome-Based Distance Phylogeny
GGDC -	Genome-to-Genome Distance Calculator
GNPS -	Global Natural Products Social
GO -	Gene Ontology
HMBD -	Human Microbiome Database
HPLC -	High-Performance Liquid Chromatography
IAA -	Indole-3-Acetic Acid
IAR -	Indole-3-Acetamide Receptor
LC-MS -	Liquid Chromatography-Mass Spectrometry
LDL -	Lower Detection Limit
m/z -	Mass-to-Charge Ratio
MN -	Molecular Networking
MRD -	Minimal Residual Disease
MS -	Mass Spectrometry
MS/MS -	Tandem Mass Spectrometry
MS2LDA -	Mass Spectrometry-based LDA (Linear Discriminant Analysis)

MSI -	Metabolomics Standards Initiative
NA -	Nutrient agar
Na -	Sodium
NAP -	Network Annotation Propagation
NCBI -	National Center for Biotechnology Information
NGS -	Next-Generation Sequencing
NIST -	National Institute of Standards and Technology
NMR -	Nuclear Magnetic Resonance
NPs -	Natural Products
NRPS -	Non-Ribosomal Peptide Synthetase
OVAT -	One Variable At a Time
PCR -	Polymerase Chain Reaction
PGPB -	Plant Growth-Promoting Bacteria
pH -	Potential of Hydrogen
PKS -	Polyketide Synthase
QACs -	Quaternary Ammonium Compounds
RSM -	Response Surface Methodology
TYGS -	Type Strain Genome Server
UPLC-q-TOF-MS -	Ultra-Performance Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry

CHAPTER 1

Study Background, Rationale, And Objectives

1.1 Background

Soil contains a high diversity of microorganisms (Nannipieri et al., 2003), some of which are essential for supporting life (Mhlongo et al., 2018). The numbers and kinds of microorganisms present in the soil are dependent on many environmental factors such as soil type, nutrient availability, the degree of aeration, pH, temperature, and soil moisture (Sharma, 2020). However, large variation in biotic and abiotic conditions of the soil make its microbial inhabitants adapt and develop strategies for their survival (Soussi et al., 2016). Production of antimicrobial compounds is one of the fundamental key strategies for their survival (Munita and Arias, 2016), ensuring little competition for nutrients by other co-existing microorganisms (Alves-Barroco et al., 2020).

Recently, researchers have shown increasing interest in underexploited habitats in soil as potential sources of bioactive metabolites for the benefit of society and mankind. Previous studies have isolated different microbial strains from various habitats including farming soil (Oskay et al., 2004), coal mine soil (Gopinath et al., 2013), rhizosphere (Nanjwade et al., 2010), dumping site (Kuta et al., 2009), and mangrove sediment (Naikpatil and Rathod, 2011). These microbial strains from these dynamic environments have been shown to produce novel bioactive compounds (Nanjwade et al., 2010, Sethi et al., 2013, Sekurova et al., 2019, Jamil et al., 2007), most of which aid them to survive these harsh environments. Examples of such compounds include polymyxin (Stansly and Schlosser, 1947), bacitracin (Ansari et al., 2018) and picolinamycin (Maiti et al., 2020).

Several soil-borne microorganisms are known to produce a wide variety of antimicrobial compounds (Caulier et al., 2019). This capacity is driven by the intense pressure and competition they experience from coexisting microorganisms. The largest number which are available commercially are derived from relatively small group of microorganisms belonging

to the genera *Streptomyces*, *Cephalosporium*, *Micromonospora*, *Bacillus* and *Penicillium* (Awais et al., 2007). While many antibiotics are known to exist, efforts to discover new antimicrobial compounds continue. Many species such as *Streptomyces*, *Penicillium* and *Bacillus* have been studied continuously for their ability to produce antibiotics (Procópio et al., 2012). In addition, many organisms belonging to the *Bacillus* species have been isolated and screened from the soil in the past several decades (Sha et al., 2020, Foysal and Lisa, 2018, Saxena et al., 2020).

Soil-borne microorganisms are constantly facing environmental pressures that drive them to develop survival mechanisms, including drug resistance (Yilmaz et al., 2021). This resistance is primarily associated with the overuse and improper use of drugs, threatening the effectiveness of treatments and limiting alternatives for patients. It is particularly prevalent in oncology and infectious diseases. Pathogens evolve to resist various antimicrobials, including antibiotics, antivirals, antifungals, and antiprotozoals. In agriculture, resistance develops from the overuse of herbicides and pesticides. In cancer treatment, resistance often emerges with most therapeutic regimens, while in infectious diseases, it arises as pathogens spread through large populations, exacerbated by the overuse of antibiotics (Llor and Bjerrum., 2014). The emergence and spread of drug resistance across these various diseases have serious implications for public health, threaten millions of lives, and create a significant financial burden, highlighting the urgent need for new strategies to combat and prevent drug resistance.

In the post-metabolomics era, genetic and molecular techniques have been applied to demonstrate conclusively that microorganisms have the capacity to synthesize a variety of bioactive metabolites (Kaur et al., 2014). However, the identification of bioactive compounds was based primarily on their extraction from the culture medium, often through traditional bioassays evaluating their inhibitory effects on other microorganisms (Oberpaul et al., 2022, Sasidharan et al., 2011). This approach is also accompanied or followed by analysis of their chemical composition and structure using various analytical techniques (Berthod et al., 2009, Feng et al., 2019, Granados-Chinchilla and Rodríguez, 2017). Recent advancements in whole genome sequencing have facilitated the exploration of secondary metabolites, while high-resolution analytical techniques such as NMR (Koskela, 2012) and UPLC-QTOF-MS (Ncube et al., 2014) have been used in the dereplication of low-molecular-weight metabolites in biological samples (Gathungu et al., 2020), including bacteria (Nephali et al., 2022). However, bacteria like many other organisms, produce wide spectrum of compounds with diverse physical–chemical properties making their analysis cumbersome.

Despite the solid foundation established by traditional statistical methods, the data generated from these advanced instruments constitutes big data, necessitating the use of multivariate statistical models to effectively analyse and interpret the inherited biological properties of bioactive compounds (Tugizimana et al., 2013). These models offer a crucial step by integrating essential components for data inspection and evaluation, enabling a streamlined and automated process for mining bioactive secondary metabolites. Furthermore, a complementary method for interpreting the chemical data obtained from analytical instruments, especially mass spectrometry, is molecular networking (Aron et al., 2020). This approach allows for the grouping of metabolites based on similar fragmentation patterns resulting from shared structural moieties. Molecular networking facilitates the visualization of complex data and the identification of unknown compounds by linking them to known chemical entities within a given sample (Clement et al., 2021). This technique has been successfully utilized to isolate antimicrobial compounds from bacterial strains (Nephali et al., 2022). Therefore, this study aimed to isolate and identify bacterial strains exhibiting antimicrobial activity using the sprinkling technique and 16S rRNA gene sequencing. Furthermore, methanolic extracts from bacterial strains isolated from agricultural soils were analysed utilizing a combination of UPLC-QTOF-MS and molecular networking techniques to identify potential antimicrobial compounds, thereof. The results highlight the efficacy of LC-MS-based metabolite profiling as a robust approach for identifying bacterial strains with antimicrobial properties, which may hold promise for the development of therapies targeting drug-resistant pathogens. Additionally, whole genome sequencing provided valuable insights into the functional characteristics of the isolates, highlighting its crucial role in the accurate taxonomic classification of closely related bacterial species. The integration of molecular networking facilitated a rapid and insightful visualization of complex LC-MS/MS data, offering a powerful tool for the analysis of microbial metabolomes.

1.2 Study Rationale

The emergence of acquired antibiotic resistance (AMR) in bacteria poses a serious global health threat, compromising the ability to treat not only resistant diseases but also regular non-resistant illnesses (Perez et al., 2012). AMR is a natural process that occurs over time due to genetic changes in pathogens. However, the significant rise in antibiotic resistance is largely driven by various factors such as increased antibiotic use, the growing number of patients with severe illnesses, longer hospital stays, and challenges in healthcare supply chains. To address this issue, studies have suggested minimizing unnecessary antibiotic prescriptions, ensuring the correct use of prescribed antibiotics, and maintaining good hygiene and infection control practices are essential (Llor and Bjerrum., 2014). Given the rapid development of resistance to each new class of antibiotics and the adverse effects of synthetic drugs, there is an urgent need to explore new antibiotics with novel modes of action. However, soil microorganisms, offer a promising avenue for discovering such antibiotics, as natural product screening has the potential to uncover new substances that could help address the growing problem of antibiotic resistance.

Microorganisms have long been recognized as untapped reservoirs for promising bioactive compounds with diverse structural and functional antimicrobial activity. Traditionally, bioprospection from microbial sources has involved the isolation of microorganisms, followed by 'blind' culture optimization and extensive chemical analysis. This approach has proven to be both labour-intensive and costly. Furthermore, determining the optimal conditions for the growth and production of valuable bacterial or microalgal metabolites has remained highly unpredictable, thereby constraining the full exploration of these microorganisms' potential. Additionally, research on the chemical arsenals employed by soil-borne bacteria to defend themselves in their natural environment remains limited in the literature. Therefore, investigating the chemical diversity of these bacteria could provide a deeper understanding of the molecular mechanisms they use to compete and defend against one another in their ecological niches.

Recent advancements have substantially improved the bioprospecting process. The development of -omics technologies, efficient model organisms, coupled with cutting-edge molecular biology techniques, and machine learning, has greatly accelerated the discovery of novel antimicrobial agents and the identification of new drug targets. The most common instruments in bioprospection are next-generation sequencing (NGS) and Mass Spectrometry

(MS). The advent of next-generation sequencing (NGS) technologies has provided a platform for pan-genomic analysis of microbes, ranging from community and strain levels down to the gene level. While the integration of metabolomics in bioprospecting has accelerated drug discovery, and related fields due to its high sensitivity, manual annotation of metabolites is time-consuming and labour-intensive, given that modern MS technologies can generate hundreds and thousands of MS/MS spectra per instrument run.

In recent years, computational tools and Machine Learning algorithms like molecular networking (Penaloza et al. 2021; Farag et al. 2020; Hegazi et al. 2021) and MS2LDA (Wang et al. 2021; Zhou et al. 2022; Qin et al. 2023) have allowed automated annotations of compounds from MS/MS data through library matching to reference spectra that currently exists in various databases. While these tools have significantly improved metabolite coverage in natural product research, compound isolation by fractionation for biological tests is still a time-consuming and laborious task. In this research, bacterial isolates with the potential to produce antimicrobial substances were isolated from soil and their identities were determined using 16S rRNA. Then their metabolic profiles were investigated using UHPLC-QTOF-MS and molecular networking. Moreover, a complementary tool to molecular networking was MS2LDA and spec2vec which were used to uncover the fragmentation patterns of secondary metabolites produced by these isolates and provide a robust network of the metabolites, respectively. Lastly, the whole genome sequences of these isolates were analysed to mine for secondary metabolite potential and to determine the precise taxonomic position of bacterial strains through phylogenetic analysis. This study contributes directly to the global fight against infectious diseases, particularly in the context of rising antibiotic resistance. Moreover, understanding the specific strains capable of antibiotic production could provide insights into the ecological roles these compounds play and how they might be harnessed for medical applications.

1.3 Objectives of the Study

1.3.1 Main Objective

- The main aim of this study was to isolate, identify, and characterize microorganisms from the soil with the potential to produce antibiotics.

1.3.2 Secondary objectives

- Isolate and identify antibiotic-producing microorganisms from soil and test their inhibitory efficacy against selected pathogens
- Characterize bacterial isolates that secrete compounds with antibacterial properties using 16S rRNA sequencing.
- Study the chemistry of bacterial extracts of using UPLC-QTOF-MS and Molecular networking approach
- Determine the full genome sequences of bacterial isolates that were found to produce antimicrobial compounds.
- Develop an analytical approach to enhance the detection of bacterial metabolome.

1.4 Hypothesis

- We hypothesized that soil harbours bacteria that are capable of producing antimicrobial compounds that can inhibit the growth of specific pathogenic bacteria.

1.5 Research questions

- What types of antibiotic-producing microorganisms can be isolated from soil, and how effective are they in inhibiting the growth of selected pathogenic bacteria?
- What bioactive compounds are present in the bacterial extracts from soil-derived microorganisms?

1.6 References

- ALVES-BARROCO, C., RIVAS-GARCÍA, L., FERNANDES, A. R. & BAPTISTA, P. V. 2020. Tackling multidrug resistance in Streptococci—From novel biotherapeutic strategies to nanomedicines. *Frontiers in microbiology*, 11, 579916.
- ANSARI, A., ZOHRA, R. R., TARAR, O. M., QADER, S. A. U. & AMAN, A. 2018. Screening, purification and characterization of thermostable, protease resistant Bacteriocin active against methicillin resistant Staphylococcus aureus (MRSA). *BMC microbiology*, 18, 1-10.
- AWAIS, M., SHAH, A. A., HAMEED, A. & HASAN, F. 2007. Isolation, identification and optimization of bacitracin produced by Bacillus sp. *Pakistan Journal of Botany*, 39, 1303.
- BERTHOD, A., MARYUTINA, T., SPIVAKOV, B., SHPIGUN, O. & SUTHERLAND, I. A. 2009. Countercurrent chromatography in analytical chemistry (IUPAC Technical Report). *Pure and Applied chemistry*, 81, 355-387.
- 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. *Frontiers in microbiology*, 10, 302.
- EL-KHAWAGA, M. & MEGAHED, M. 2012. Antibacterial and insecticidal activity of actinomycetes isolated from sandy soil of (Cairo-Egypt). *Egypt Acad J Biol Sci*, 4, 53-67.
- FENG, W., LI, M., HAO, Z. & ZHANG, J. 2019. Analytical methods of isolation and identification. *Phytochemicals in human health*. IntechOpen.
- FOYSAL, M. J. & LISA, A. K. 2018. Isolation and characterization of Bacillus sp. strain BC01 from soil displaying potent antagonistic activity against plant and fish pathogenic fungi and bacteria. *Journal of Genetic Engineering and Biotechnology*, 16, 387-392.
- GATHUNGU, R. M., KAUTZ, R., KRISTAL, B. S., BIRD, S. S. & VOUIROS, P. 2020. The integration of LC-MS and NMR for the analysis of low molecular weight trace analytes in complex matrices. *Mass spectrometry reviews*, 39, 35-54.
- GOPINATH, B., VOOTLA, P. K., JYOTHI, R. & REDDY, K. S. 2013. Antimicrobial activity of actinomycetes isolated from coal mine soils of godavari belt region, AP, India. *Asian Journal of Experimental Biological Sciences*, 4, 518-523.

- GRANADOS-CHINCHILLA, F. & RODRÍGUEZ, C. 2017. Tetracyclines in food and feedingstuffs: from regulation to analytical methods, bacterial resistance, and environmental and health implications. *Journal of Analytical Methods in Chemistry*, 2017.
- JAMIL, B., HASAN, F., HAMEED, A. & AHMED, S. 2007. Isolation of *Bacillus subtilis* MH-4 from soil and its potential of polypeptidic antibiotic production. *Pak. J. Pharm. Sci*, 20, 26-31.
- KAUR, G., PANDEY, O. P., SINGH, K., HOMA, D., SCOTT, B. & PICKRELL, G. 2014. A review of bioactive glasses: their structure, properties, fabrication and apatite formation. *Journal of Biomedical Materials Research Part A: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials*, 102, 254-274.
- KOSKELA, H. 2012. A set of triple-resonance nuclear magnetic resonance experiments for structural characterization of organophosphorus compounds in mixture samples. *Analytica chimica acta*, 751, 105-111.
- KUTA, F. A., NIMZING, L. & ORKA'A, P. Y. 2009. Screening of *Bacillus* species with potentials of antibiotics production. *Applied Medical Informatics*, 24, 42-46.
- LLOR, C. AND BJERRUM, L., 2014. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic advances in drug safety*, 5(6), pp.229-241.
- MAITI, P. K., DAS, S., SAHOO, P. & MANDAL, S. 2020. *Streptomyces* sp SM01 isolated from Indian soil produces a novel antibiotic picolinamycin effective against multi drug resistant bacterial strains. *Scientific reports*, 10, 1-12.
- MHLONGO, M. I., PIATER, L. A., MADALA, N. E., LABUSCHAGNE, N. & DUBERY, I. A. 2018. The chemistry of plant–microbe interactions in the rhizosphere and the potential for metabolomics to reveal signaling related to defense priming and induced systemic resistance. *Frontiers in plant science*, 9, 112.
- MUNITA, J. M. & ARIAS, C. A. 2016. Mechanisms of antibiotic resistance. *Virulence mechanisms of bacterial pathogens*, 481-511.
- NAIKPATIL, S. V. & RATHOD, J. 2011. Selective isolation and antimicrobial activity of rare actinomycetes from mangrove sediment of Karwar. *Journal of Ecobiotechnology*.
- NANJWADE, B., CHANDRASHEKHARA, S., GOUDANAVAR, P., SHAMAREZ, A. & MANVI, F. 2010. Production of antibiotics from soil-isolated actinomycetes and

- evaluation of their antimicrobial activities. *Tropical Journal of Pharmaceutical Research*, 9.
- NANNIPIERI, P., ASCHER, J., CECCHERINI, M., LANDI, L., PIETRAMELLARA, G. & RENELLA, G. 2003. Microbial diversity and soil functions. *European journal of soil science*, 54, 655-670.
- NCUBE, E. N., MHLONGO, M. I., PIATER, L. A., STEENKAMP, P. A., DUBERY, I. A. & MADALA, N. E. 2014. Analyses of chlorogenic acids and related cinnamic acid derivatives from *Nicotiana tabacum* tissues with the aid of UPLC-QTOF-MS/MS based on the in-source collision-induced dissociation method. *Chemistry Central Journal*, 8, 1-10.
- NEPHALI, L., STEENKAMP, P., BURGESS, K., HUYSER, J., BRAND, M., VAN DER HOOFT, J. J. & TUGIZIMANA, F. 2022. Mass Spectral Molecular Networking to Profile the Metabolome of Biostimulant *Bacillus* Strains. *Frontiers in plant science*, 1916.
- OBERPAUL, M., BRINKMANN, S., MARNER, M., MIHAJLOVIC, S., LEIS, B., PATRAS, M. A., HARTWIG, C., VILCINSKAS, A., HAMMANN, P. E. & SCHÄBERLE, T. F. 2022. Combination of high-throughput microfluidics and FACS technologies to leverage the numbers game in natural product discovery. *Microbial biotechnology*, 15, 415-430.
- OSKAY, A. M., ÜSAME, T. & CEM, A. 2004. Antibacterial activity of some actinomycetes isolated from farming soils of Turkey. *African journal of Biotechnology*, 3, 441-446.
- PROCÓPIO, R. E. D. L., SILVA, I. R. D., MARTINS, M. K., AZEVEDO, J. L. D. & ARAÚJO, J. M. D. 2012. Antibiotics produced by *Streptomyces*. *Brazilian Journal of Infectious Diseases*, 16, 466-471.
- SASIDHARAN, S., CHEN, Y., SARAVANAN, D., SUNDRAM, K. & LATHA, L. Y. 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. *African journal of traditional, complementary and alternative medicines*, 8.
- SAXENA, A. K., KUMAR, M., CHAKDAR, H., ANUROOPA, N. & BAGYARAJ, D. 2020. *Bacillus* species in soil as a natural resource for plant health and nutrition. *Journal of applied microbiology*, 128, 1583-1594.
- SEKUROVA, O. N., SCHNEIDER, O. & ZOTCHEV, S. B. 2019. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microbial biotechnology*, 12, 828-844.

- SETHI, S., KUMAR, R. & GUPTA, S. 2013. Antibiotic production by microbes isolated from soil. *International Journal of Pharmaceutical Sciences and Research*, 4, 2967.
- SHA, Y., ZENG, Q. & SUI, S. 2020. Screening and application of *Bacillus* strains isolated from nonrhizospheric rice soil for the biocontrol of rice blast. *The Plant Pathology Journal*, 36, 231.
- SHARMA, I. 2020. Bioremediation techniques for polluted environment: concept, advantages, limitations, and prospects. *Trace metals in the environment-new approaches and recent advances*. IntechOpen.
- SOUSSI, A., FERJANI, R., MARASCO, R., GUESMI, A., CHERIF, H., ROLLI, E., MAPELLI, F., OUZARI, H. I., DAFFONCHIO, D. & CHERIF, A. 2016. Plant-associated microbiomes in arid lands: diversity, ecology and biotechnological potential. *Plant and Soil*, 405, 357-370.
- STANSLY, P. & SCHLOSSER, M. 1947. Studies on polymyxin: isolation and identification of *Bacillus polymyxa* and differentiation of polymyxin from certain known antibiotics. *Journal of bacteriology*, 54, 549-556.

CHAPTER 2

LITERATURE REVIEW

2.1 Antimicrobial resistance (AMR)

In the 21st century, antimicrobial resistance (AMR) has emerged as one of the most pressing global public health threats (Tang et al., 2023). This phenomenon is a natural process that occurs when microbes (like bacteria, fungi, parasites, and viruses) develop defence mechanisms against antimicrobial treatments like antibiotics, which are commonly used to treat infections (Murugaiyan et al., 2022). The escalation of AMR is largely attributed to intrinsic and extrinsic factors (Figure 1) which include the overuse, misuse, or inappropriate application of antibiotics across various sectors, including healthcare, agriculture, and veterinary medicine (Singh et al., 2024). Currently, there is a growing number of reports highlighting the development of bacterial resistance to almost all antimicrobial agents (Christaki et al., 2020). It is predicted to cause of approximately 10 million deaths annually by the year 2050, with a total cost of 100 trillion USD (Kathir, 2024).

The available evidence suggests a connection between personal and organizational misuse of antibiotics and the rise of resistant bacteria. Eventually, the treatment of emerging diseases has increasingly become more complicated as microorganisms develop resistance to antimicrobial compounds that are available. Over time, the effectiveness of different antibiotics has decreased, resulting in the growth of resistance mechanisms in organisms (figure 1) (Darby et al., 2023). These mechanisms involve the emergence of new penicillin-binding proteins, catalyst-dependent drug switching, altered drug targets, and heightened expression of efflux pumps, all of which contribute to multidrug resistance (MRD) (Ramírez-Rendon et al., 2022). In light of the rise in drug resistance, there is a pressing need for the development of new antimicrobial compounds to address the growing threat of antimicrobial resistance.

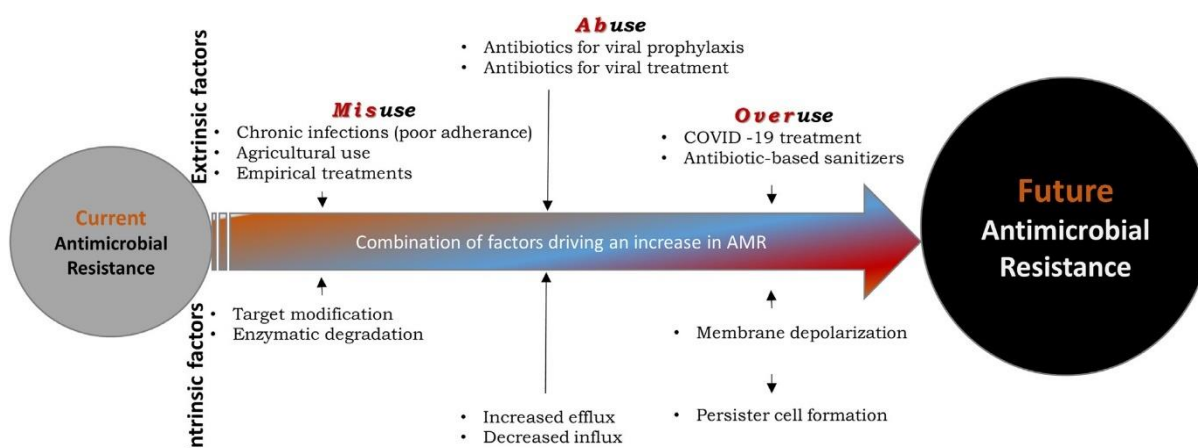


Figure 1: Combination of factors driving the increase of antimicrobial resistance and the future of antimicrobial resistance.

2.2 History of antibiotic and their nomenclature

The derivation of the term antibiotic can be traced back to the 19th century. It was coined or first used as an antonym to symbiosis to describe the antagonistic action between different microorganisms. However, it was later used to describe secondary metabolites which are produced by microorganisms that either has the capability to inhibit or kill other microorganisms. The term “antibiotic” is derived from two words: “anti” means against, and “biota” means life. To date, the term has a broader meaning. Generally, antibiotic refers to antibacterial, the term is loosely defined, and preferable to specify compounds as being antibacterial, antifungal and antiviral (Fadiji et al., 2023). However, this theory now is broadened to consist of any chemical compound of natural origin that has an effect against the growth of other organisms. Scientists distinguish this sort of antimicrobial agents from another group called chemotherapeutic agents which are chemically synthesized (Hobson et al., 2021). Furthermore, a mixture of these two concepts now yields some antimicrobial compounds that are produced by organisms and chemically modified to obtain some desired properties (semi-synthetic antibiotics) (Leisner, 2020).

In classifying antibiotics, various distinct factors are involved, leading to multiple methods of classification being employed. Some classifications are based on the mode of action of the compounds, such as whether they target the cell wall or membrane, inhibit protein synthesis, or interfere with nucleic acid synthesis and cellular metabolism (Ongpipattanakul et al., 2022). Additionally, antibiotics can be classified according to their biosynthetic pathways. However, the presence of commonalities among different biosynthetic routes complicates this

classification. Another approach involves categorizing antibiotics based on their spectrum of activity, such as those that specifically affect bacteria, fungi, or protozoa (Angelini, 2024). Nonetheless, some antibiotics within a particular group may exhibit varying spectra. For instance, paromomycin is typically classified as an antiprotozoal agent, despite having a spectrum of activity and chemical structure nearly identical to other aminoglycosides like neomycin and kanamycin, which are recognized as antibacterial agents (Bellucci and Volonterio, 2020, Haider et al., 2023).

In some cases, antibiotics are classified based on the organisms that produce them. However, certain organisms can synthesize multiple antibiotics; for instance, *Cephalosporium acremonium* is known to produce both penicillin N and cephalosporin C (Bibián et al., 2020, Niu et al., 2021). Conversely, the same antibiotic may be synthesized by several organisms. For example, penicillin has been identified as being produced by a diverse array of organisms, including fungal species such as *Aspergillus*, *Cephalosporium*, and *Trichophyton*, as well as certain actinomycetes like *Streptomyces* spp. (Mishra et al., 2024, Sułkowska-Ziaja et al., 2023). This overlap in antibiotic production complicates classification efforts and underscores the diverse biosynthetic capabilities of these microorganisms. Given the complexity of antibiotic classification, there is a pressing need for more refined and systematic approaches. One effective method involves categorizing antibiotics based on their chemical structure or mode of action, which provides a more consistent framework than relying solely on the producing organism. Another promising approach utilizes genomic and bioinformatics tools to gain a deeper understanding of the biosynthetic pathways of antibiotics (Panter et al., 2021). These advanced technologies enable researchers to identify the specific genes responsible for antibiotic production, thereby clarifying the origins of these compounds and highlighting the potential for discovering novel antibiotics.

2.3 Mechanisms of AMR and microbes involved

In their natural habitats, microorganisms constantly encounter various biotic and abiotic challenges including exposure to antimicrobial compounds (Srikanth et al., 2024). This exposure compels them to develop a range of innovative strategies to survive these environmental pressures (Zhang et al., 2021). These adaptive mechanisms allow microorganisms to withstand attacks from other microorganisms, which may involve antibiotics production and other antimicrobial agents that can impede their growth or lead to their total destruction (Smith et al., 2023). These microorganisms utilize various strategic

metabolic pathways that allow them to evade or neutralize harmful antimicrobial substances. Common resistance mechanisms include enzymatic modification or degradation of antibiotics, limiting their entry into cells, alterations in metabolic pathways, and modifications of binding sites such as ribosomes to reduce drug effectiveness (figure 2) (Varela et al., 2021, Mohanty et al., 2021). Additionally, many microorganisms increase the activity of efflux pumps, which actively expel antibiotics from the cell before they can reach effective concentrations (Henderson et al., 2021).

Bacteria, in particular, can form biofilms, surface-attached communities with varying nutrient availability and limited antibiotic penetration (Abebe, 2020). These biofilms provide added protection to the bacteria. Additionally, bacteria are highly capable of acquiring resistance genes from nearby cells, even from different species through horizontal gene transfer, a process facilitated by plasmids and other mobile genetic elements (Rodríguez-Beltrán et al., 2021). These acquired genes frequently encompass multiple complex resistance mechanisms within a single unit, facilitating the rapid spread of multidrug resistance (MDR) among microbial populations (Jhalora and Bist, 2024). The efficient horizontal transfer of genetic information allows microorganisms to adopt a wide array of resistance strategies, which can be swiftly adapted to ensure survival against the ongoing development and use of antimicrobial treatments in the medical field (Muteeb et al., 2023).

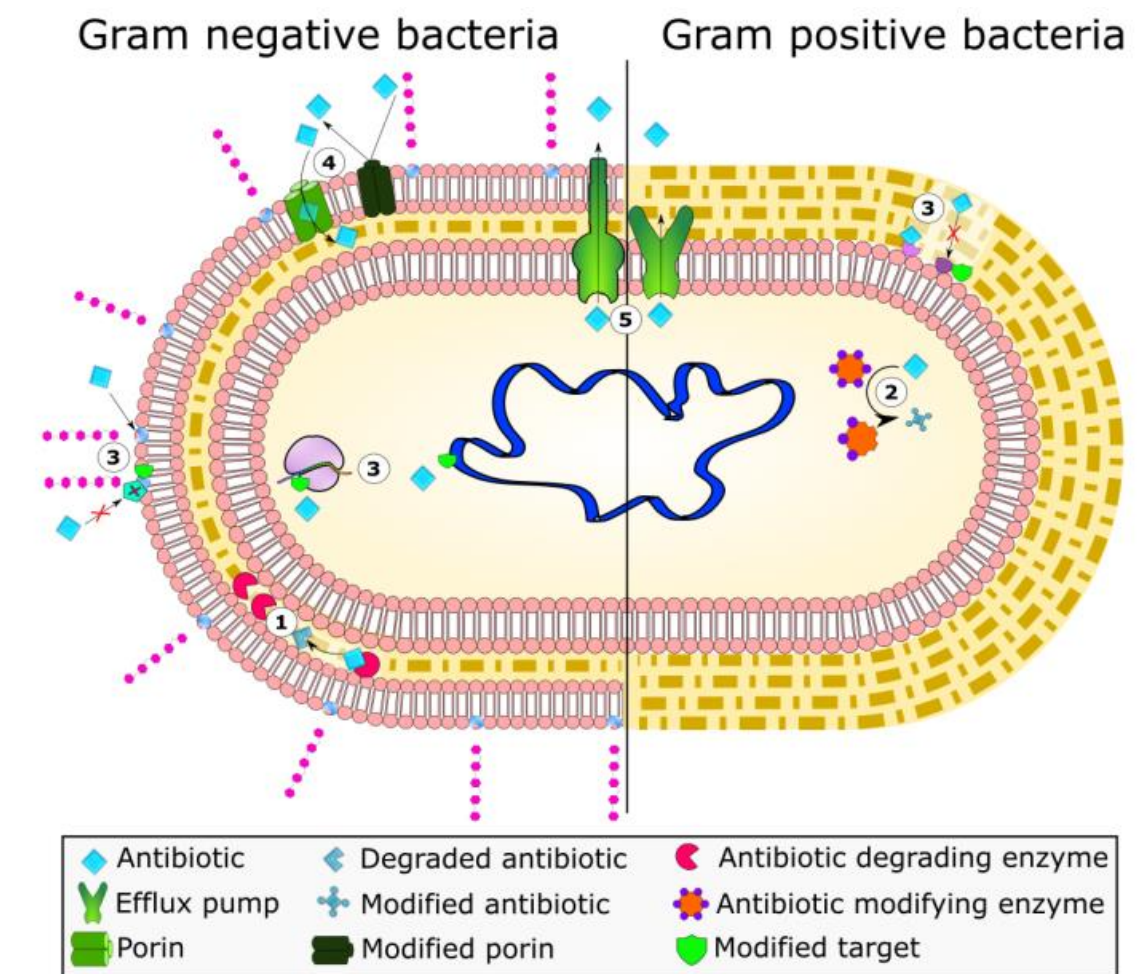


Figure 2: Bacterial mechanisms of resistance to antimicrobial agents include several common strategies: enzymatic hydrolysis (1), enzymatic modification of antibiotics through group transfer and redox processes (2), alteration of antibiotic targets (3), reduced antibiotic permeability due to porin modifications (4), and active extrusion of antibiotics via membrane efflux pumps (5) (Mohanty et al., 2021).

2.4 The development of antibiotics

The discovery of antibiotics and their applications in clinical settings has revolutionized modern medicine, significantly reducing mortality rates and enhancing health outcomes (Jamal, 2023). However, despite this significant achievement, their ongoing use in the 21st century faces challenges that threaten their effectiveness. One major challenge is the gradual development of resistance by the microbes targeted by these drugs (Saha and Sarkar, 2021). Another challenge is the growing cost ineffectiveness of discovering and developing new antibiotics under traditional reimbursement models (Cook and Wright, 2022). As a result, the

potential dangers of entering a post-antibiotic era have compelled policymakers to acknowledge this significant threat to human health and to pledge additional grant funding. This commitment is gradually revitalizing interest in the discovery and development of antibiotics (Figure 3). Given the relatively low number of successes in bringing effective synthetic antibiotics to market, the best opportunity for developing a new generation of anti-infective drugs lies in identifying new microbial natural products (NPs), which are unmatched in their chemical diversity and antibiotic effectiveness.

Since the discovery of penicillin, the concept of drug discovery from microbial natural products has significantly transformed research in this field, establishing it as a crucial avenue for drug development (Tiwari and Bae, 2022). This shift is attributed to the availability, variability, and rich biodiversity of microbes, as well as the chemical diversity and bioactivities they produce (Farda et al., 2022). Today, a wide variety of antibiotics derived from microbial sources are available on the market, with more than 120 of the most essential medicines originating from terrestrial microorganisms (Abdel-Razek et al., 2020, Karthikeyan et al., 2022). Recently, there has been a growing interest in natural products from unexplored microbial sources, including marine environments, extreme habitats such as hot springs and deep-sea ecosystems, and diverse soil microbiomes (Karthikeyan et al., 2022). However, one of the most important challenges has been the difficulty in replicating the natural environments where these microorganisms thrive. The production of secondary metabolites in bacteria is significantly influenced by fermentation parameters, including nutrient availability, pH, aeration, temperature, mineral salts, heavy metals, precursors, inducers, and inhibitors. These factors can vary widely from one organism to another (Pal et al., 2024, Khaochamnan et al., 2024, Huang et al., 2024, Al Daccache et al., 2021).

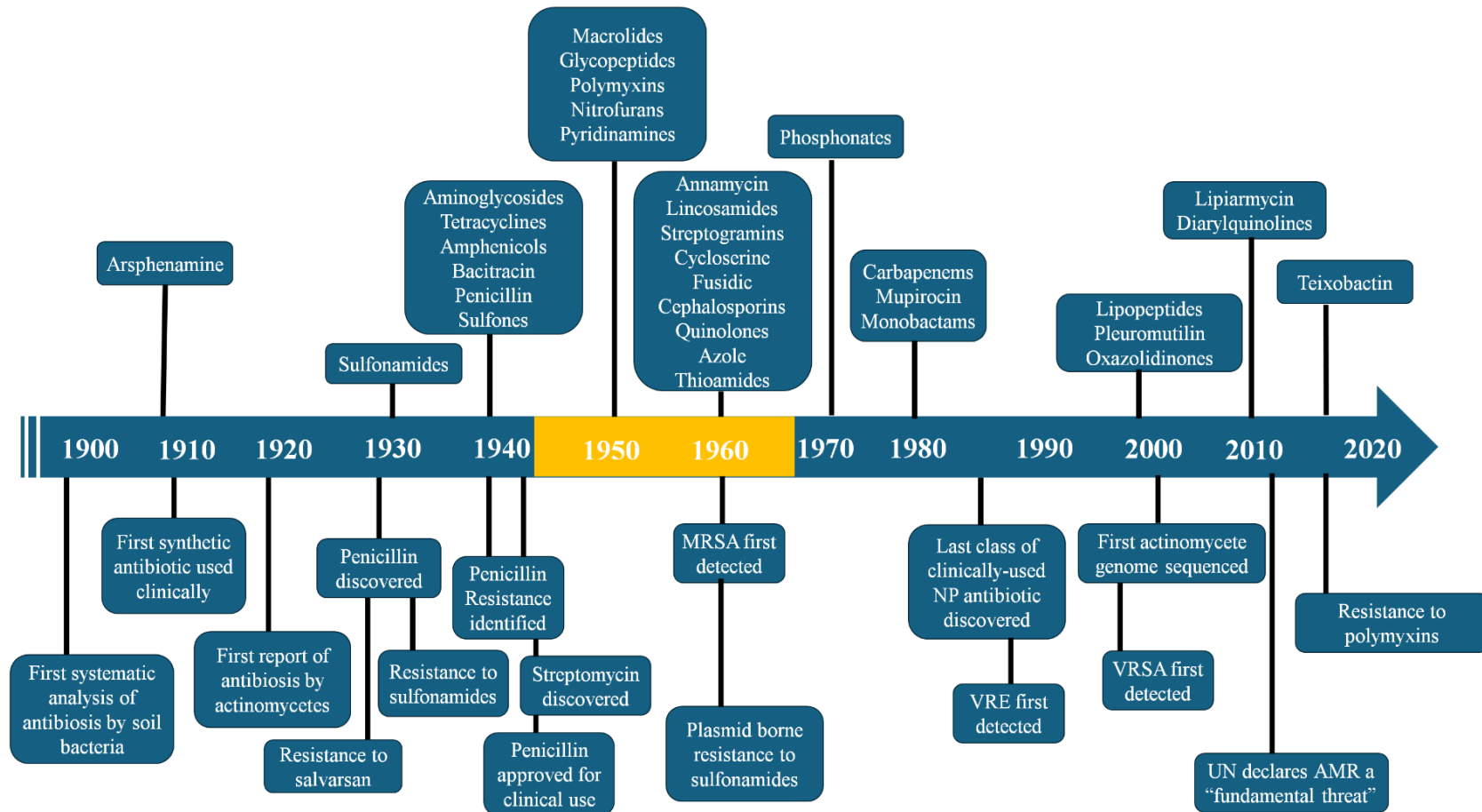


Figure 3: Timeline illustrating the discovery of different antibiotic classes used in clinical practice. The term "discovery void" denotes the period from 1987 to the present, highlighting the gap since the last antibiotic class was successfully introduced for treatment in 1987 (Figure inspired by Hutchings et al., 2019).

2.5 Soil as a source of microorganisms

Soil is a heterogeneous environment composed of organic matter, liquids, and gases, creating a biologically active and porous medium (Lv et al., 2022). Beyond its physical components, soil is home to a diverse array of microorganisms that are vital for sustaining life. These microorganisms are fundamental in nutrient cycling, improving soil structure suppressing diseases, remediating pollutants and fostering beneficial plant-microbe interaction (Maurya, 2024). Moreover, the diversity of soilborne microorganisms, driven by the survival challenges in a complex environment, makes soil an excellent source for isolating organisms with biotechnological potential and the rich microbial landscape offers significant opportunities for innovative advancements in medicine and agriculture (Farda et al., 2022).

2.6 Antibiotic-Producing Microorganisms from the soil

Soil is a rich source of antibiotic-producing microorganisms, making it a vital reservoir for bioactive compounds. The inherent heterogeneity of soil, characterized by varying physical and chemical properties, fosters a diverse array of microorganisms. This diversity arises not only from the multitude of nutrients and substrates available but also from dynamic processes such as decomposition and nutrient cycling occurring within the soil ecosystem (Prasad et al., 2021). The microorganisms that thrive in this environment have evolved various mechanisms to defend themselves against competing organisms, leading to the production of antimicrobial substances that inhibit the growth of nearby microbial populations (McLaren and Callahan, 2020). It is estimated that natural soil can harbor over 10^9 different types of microorganisms per gram, highlighting its significance as a reservoir of potential antibiotic producers. Among these microorganisms, a select group is particularly notable for their ability to produce potent antimicrobial compounds (Erdem Büyükkiraz and Kesmen, 2022). This includes genera such as *Streptomyces*, which produces well-known antibiotics like streptomycin and tetracycline; *Bacillus*, known for producing bacitracin and subtilin; and fungi like *Penicillium*, which is the source of penicillin (Cobongela, 2022, Berillo et al., 2024). Additionally, other important groups include *Cephalosporium*, which produces cephalosporins, and *Micromonospora*, another source of valuable antibiotics (Romero et al., 2024, Roy et al., 2021).

2.7 Secondary metabolites from soil-borne bacteria

2.7.1 Bacterial secondary metabolites

Secondary metabolites, also known as specialized metabolites or natural products, are organic compounds produced by various organisms, including bacteria, fungi, plants, and animals (Bruce, 2022, Oladipo et al., 2022). Unlike primary metabolites, which are essential for growth and development, secondary metabolites are not directly involved in the growth and development or physiological processes of the organism. Instead, they serve critical ecological functions that enhance an organism's survival and reproductive success. These compounds can mediate interactions between organisms, providing advantages such as defence against other predatory organisms (such as Myxobacteria) and competition for resources with the co-existing organisms (Furness et al., 2020, Pérez et al., 2020).

Bacteria are remarkable producers of a diverse array of secondary metabolites, both soluble and volatile, which play crucial ecological roles in their environments (Buddhika and Abeysinghe, 2021). These includes siderophore, antibiotics, plant growth promoting compounds, bioactive compounds and quorum sensing molecules. Some of the applications of microbial secondary metabolites are shown on Figure 4. These metabolites serve as mechanistic links between bacterial interactions and their surroundings, facilitating complex relationships among microbial communities. Among the various functions of these secondary metabolites, antibiosis is particularly significant; it enables bacteria to inhibit the growth of competing microorganisms in their vicinity (Yu et al., 2024). Through this competitive advantage, bacteria can thrive in nutrient-rich yet competitive habitats such as soil.

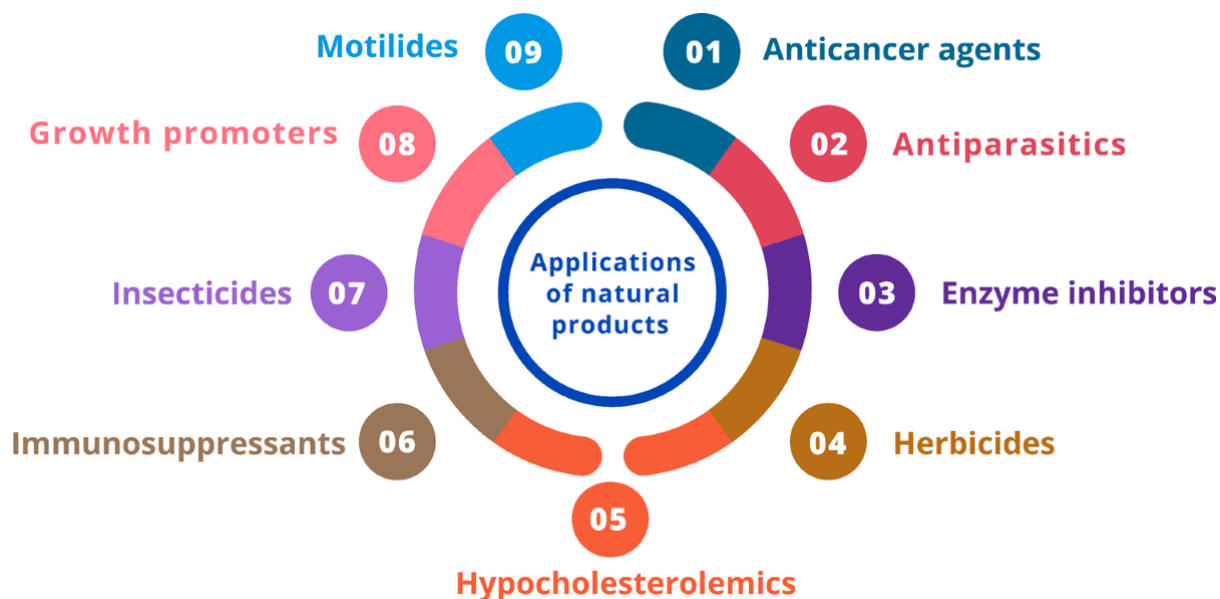


Figure 4: Applications of Microbial Secondary Metabolites, including Anticancer, Antiparasitic, Plant Growth-Promoting Compounds, Bioactive Compounds, and immunosuppression molecules.

Most secondary metabolites produced by bacteria are vital for antibiosis, competition, and parasitism. They contribute to the antagonistic activity of bacteria against coexisting species by producing a variety of antibiotics and toxins. Additionally, these metabolites play an essential role in inducing defence responses in plants, particularly in those associated with plant growth-promoting bacteria (PGPB) (Mhlongo et al., 2018). These interactions highlight the ecological importance of bacterial metabolites by enhancing plant resilience against pathogens while promoting growth.

The production of secondary metabolites typically occurs during the stationary phase of bacterial growth and is often triggered by nutrient depletion or environmental stressors. This strategic response allows bacteria to adapt to changing conditions while engaging in ecological interactions that influence biogeochemical cycles. Notably, the genetic potential for producing these metabolites varies across different bacterial species and is influenced by factors such as soil depth and vegetation type.

2.7.2 Microbial biosynthesis of secondary metabolites

These compounds are synthesized at the of the exponential growth phase and their formation is highly influenced by the growth conditions, especially by the composition of the culture medium. In adapting to new environments, such as nutrient depletion or environmental stress,

bacteria enter a phase where their metabolism shifts from growth to the production of compounds that enhance survival and ecological interactions (Zhang et al., 2021). This synthesis of secondary metabolites is primarily regulated by biosynthetic gene clusters (BGCs), these are groups of two or more genes that collectively encode the biosynthetic pathways in the genome, working together to produce specific molecules (Chakraborty, 2022).

There are different classes of biosynthetic gene clusters, that are known to exist including non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), terpenes, and bacteriocins (Dilshad et al., 2022, Dilshad et al., 2021). However, the key classes of BGCs are non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), which are responsible for the majority of known antibiotics and antifungals (Salam et al., 2021, Robertsen and Musiol-Kroll, 2019, Mohan et al., 2024). The genes within these clusters encode various enzymes that facilitate the assembly of complex molecules, with NRPSs characterized by condensation and adenylation domains, while PKSs contain ketosynthase domains along with other enzymatic functionalities (Table 1).

A comprehensive analysis of secondary metabolite production in bacteria has provided valuable insights into the sophisticated secondary metabolome of various species (Dilshad et al., 2021, Dilshad et al., 2022). For example, *Bacillus* species exhibit a diverse array of biosynthetic gene clusters (BGCs), including polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), siderophores, phosphonates, ectoines, terpenes, thiopeptides, lanthipeptides, bacteriocins, and other non-traditional hybrid BGCs (Yin et al., 2023, Othoum et al., 2018). Recently, studies have shown that these biosynthetic gene clusters (BGCs) encode megasynthases, which are large multi-enzymatic proteins essential for the synthesis of various secondary metabolites (Othoum et al., 2019, Dilshad et al., 2022). These megasynthases utilize precursor units to carry out condensation and modification reactions, resulting in a diverse array of secondary metabolites.

Table 1: Biosynthetic gene clusters and their secondary metabolites, including NRPS and PKS pathways. NRPSs, characterized by condensation and adenylation domains, and PKSs, containing ketosynthase domains along with other enzymatic functionalities, are involved in the production of complex secondary metabolites by their respective microbial producers

Secondary Metabolite	Key Genes	Biosynthetic Class	Organism	Reference
Fimsbactin A	<i>fmsA, fmsB, fmsC, fmsD</i>	NRP	<i>Acinetobacter sp. ADP1</i>	Bohac et al., 2019
Maduropeptin	<i>madA, madB, madC, madD</i>	Enediyne type I polyketide	<i>Actinomadura madurae</i>	Van Lanen et al., 2007
Labyrinthopeptin A1, labyrinthopeptin A2,	<i>labA, labB, labC, labD</i>	Lanthipeptide	<i>Actinomadura namibiensis</i>	Krawczyk et al., 2013
Cetoniacytone A	<i>cetA, cetB, cetC</i>	Other	<i>Actinomyces sp. Lu 9419</i>	Wu et al., 2009
Teicoplanin	<i>tehA, tehB, tehC</i>	NRP	<i>Actinoplanes teichomyceticus</i>	Somma et al., 1984
Marneral	<i>marA, marB, marC</i>	Terpene	<i>Arabidopsis thaliana</i>	Tholl et al., 2011
Thaliandiol, thalianol	<i>thaA, thaB, thaC</i>	Terpene	<i>Arabidopsis thaliana</i>	Tholl et al., 2011
Petrobactin	<i>petA, petB, petC</i>	Other	<i>Bacillus anthracis str. Ames</i>	Hagan et al., 2017
Cereulide	<i>cerA, cerB, cerC</i>	NRP	<i>Bacillus cereus</i>	Toh et al., 2004
Thiocillin I	<i>thiA, thiB, thiC</i>	Thiopeptide	<i>Bacillus cereus ATCC 14579</i>	Müller et al., 2023
Haloduracin α , haloduracin β	<i>halA, halB, halC</i>	Lanthipeptide	<i>Bacillus halodurans C-125</i>	McClerren et al., 2006
Plantazolicin	<i>plzA, plzB, plzC</i>	LAP	<i>Bacillus pumilus ATCC 7061</i>	Molohon et al., 2011
Ericin S	<i>eriA, eriB, eriC</i>	Lanthipeptide	<i>Bacillus subtilis</i>	Stein et al., 2002
Subtilomycin	<i>subA, subB, subC</i>	Lanthipeptide	<i>Bacillus subtilis</i>	Phelan et al., 2013
Entianin	<i>entA, entB, entC</i>	Lanthipeptide	<i>Bacillus subtilis subsp. spizizenii</i>	Fuchs et al., 2011
Subtilin	<i>subA, subB, subC</i>	Lanthipeptide	<i>Bacillus subtilis subsp. spizizenii</i>	Zhang et al., 2022
Sporulation killing factor	<i>skfA, skfB</i>	Sactipeptide	<i>Bacillus subtilis</i>	Allenby et al., 2006
Andalusicin A, andalusicin B	<i>andA, andB, andC</i>	Lanthipeptide	<i>Bacillus thuringiensis</i>	Grigoreva et al., 2021
Difficidin	<i>difA, difB, difC</i>	Polyketide	<i>Bacillus velezensis FZB42</i>	Fan et al., 2018
Macrolactin H	<i>macA, macB, macC</i>	Polyketide	<i>Bacillus velezensis FZB42</i>	Wang et al., 2023
Surfactin	<i>srfA, srfB, srfC</i>	NRP	<i>Bacillus velezensis FZB42</i>	Fan et al., 2018
Amylocyclicin	<i>bacA, bacB, bacC</i>	Sactipeptide	<i>Bacillus velezensis FZB42</i>	Scholz et al., 2014

Bacillaene	<i>bmyA, bmyB, bmyC</i>	NRP	<i>Bacillus velezensis FZB42</i>	Fazle Rabbee et al., 2020
Bacillomycin D	<i>fenA, fenB, fenC</i>	NRP	<i>Bacillus velezensis FZB42</i>	Liu et al., 2022
Fengycin	<i>bacA, bacB, bacC</i>	NRP	<i>Bacillus subtilis</i>	Harwood et al., 2018
Sch 18640	<i>evnA, evnB, evnC</i>	Thiopeptide	<i>Micromonospora arborensis</i>	Schwalen et al., 2018
Everninomicin D, everninomicin E,	<i>rosA, rosB, rosC</i>	Saccharide	<i>Micromonospora carbonacea</i>	Limbrick et al., 2020
Rosamicin	<i>rakA, rakB, rakC</i>	Polyketide	<i>Micromonospora carbonacea</i>	Waitz et al., 1972
Kosinostatin	<i>dazA, dazB, dazC</i>	NRP	<i>Micromonospora sp. 28ISP2-46</i>	Back et al., 2021
Diazaquinomycin H, diazaquinomycin J	<i>thcA, thcB, thcC</i>	Other	<i>Micromonospora sp. B006</i>	Braesel et al., 2019
Rakicidin A, rakicidin B	<i>beA, beB, beC</i>	NRP	<i>Micromonospora sp. M42</i>	Oku et al., 2014
Macedocin	<i>mutA, mutB, mutC</i>	Lanthipeptide	<i>Streptococcus macedonicus</i>	Georgalaki et al., 2002
Mutanocyclin	<i>slsA, slsB, slsC</i>	NRP	<i>Streptococcus mutans B04Sm5</i>	Baker et al., 2022
Streptolysin S	<i>strA, strB, strC</i>	LAP	<i>Streptococcus pyogenes M1 GAS</i>	Sumitomo et al., 2011
Streptomonicin	<i>perA, perB, perC</i>	Lasso peptide	<i>Streptomonospora alba</i>	Metelev et al., 2015
Persiamycin A	<i>salA, salB, salC</i>	Type II polyketide	<i>Streptomonospora sp. PA3</i>	Matroodi et al., 2020
Streptamidine	<i>pamA, pamB, pamC</i>	Other	<i>Streptomyces albidoflavus J1074</i>	Russell et al., 2021
Pamamycin-607	<i>salA, salB, salC</i>	Type III polyketide	<i>Streptomyces alboniger</i>	Hashimoto et al., 2005
Salinomycin	<i>staA, staB, staC</i>	Modular type I polyketide	<i>Streptomyces albus</i>	Izumikawa et al., 2003
Stambomycin A, stambomycin B, stambomycin C	<i>isoA, isoB, isoC</i>	Modular type I polyketide	<i>Streptomyces ambofaciens</i>	Laureti et al., 2011
17-hydroxy-8-desmethoxy- isomigrastatin, 8,9- dihydrolactimidomycin,	<i>chlA, chlB, chlC</i>	Modular type I polyketide	<i>Streptomyces amphibiosporus</i>	Ma et al., 2013
Chlorothricin, deschlorothricin	<i>oviA, oviB, oviC</i>	Iterative type I polyketide	<i>Streptomyces antibioticus</i>	Li et al., 2020
Oviedomycin	<i>simA, simB, simC</i>	Type II polyketide	<i>Streptomyces antibioticus</i>	Lombó et al., 2004
Simocyclinone D8	<i>endA, endB, endC</i>	Other	<i>Streptomyces antibioticus</i>	Flatman et al., 2005
Endophenazine A, endophenazine B	<i>fogA, fogB, fogC</i>	Other	<i>Streptomyces anulatus</i>	Krastel et al., 2002
Mithramycin	<i>argA, argB, argC</i>	Type II polyketide	<i>Streptomyces argillaceus</i>	Lombó et al., 1999
Argimycin PI, argimycin PII, argimycin PIV	<i>larA, larB, larC</i>	Modular type I polyketide	<i>Streptomyces argillaceus</i>	Ye et al., 2018

2.8 Analysis of secondary metabolites through analytical techniques

2.8.1 Isolation and cultivation for the production of secondary metabolites

For decades, the identification of bacterial secondary metabolites has primarily relied on culturing methods (Sharrar et al., 2020). These techniques involve isolating bacteria from environmental samples and cultivating them under specific conditions to produce and extract metabolites. Although this approach has provided insight into bacterial secondary metabolites of many bacteria, it has notable limitations. For example, many bacteria cannot be cultured in laboratory settings, resulting in gaps in our understanding of their metabolic profiles. Furthermore, this method may overlook metabolites that are expressed only under certain environmental conditions, thereby restricting the discovery of potentially significant compounds. To address these challenges, several innovative approaches have been developed to comprehensively identify secondary metabolites from bacteria. These include enrichment cultures (Rosero-Chasoy et al., 2021) and co-cultivation techniques (Boruta, 2023), which enhance the chances of discovering a broader range of metabolites by simulating natural environments and interactions (Boruta, 2023).

Traditionally, culturing microorganisms involved establishing laboratory conditions with specific nutrient compositions to support bacterial growth. However, recent innovations derived from these traditional methods have greatly enhanced our understanding of microbial ecology and the functional roles of various organisms within their environment (Boruta, 2023). One notable example is the application of enrichment cultures in the production of secondary metabolites (Rosero-Chasoy et al., 2021). This approach involves creating favourable conditions that promote the growth of target organisms. By simulating natural environments, enrichment cultures can stimulate the growth of microorganisms that express specific metabolic pathways or produce unique metabolites under particular conditions. This adaptability not only facilitates the discovery of novel compounds but also highlights their potential applications in medicine and biotechnology.

Additionally, another approach for enhancing the production of secondary metabolites is co-culturing. Co-culturing microorganisms involves growing different species together to enhance the production of secondary metabolites. Recently, studies have shown that co-culturing is an effective strategy for production of secondary metabolites by inducing production of defensive

secondary metabolites against the co-cultured microorganism (Selegato and Castro-Gamboa, 2023, Li et al., 2023). These secondary metabolites include antibiotics that function as competitive weapons, acting as repellents and toxins, while also serving as metal transport agents and supporting symbiotic relationships between microorganisms (Harwani et al., 2018). These interactions foster an interactive environment that provides essential signals for organization, differentiation, and homeostasis, ultimately promoting the action of these metabolites by mimicking the microbial community.

2.8.2 Genome Mining of WGS for secondary metabolites

Genome mining is a powerful method for identifying secondary metabolites through the analysis of bacterial genomes. This approach enables researchers to uncover the genetic potential of microorganisms, revealing previously uncharacterized biosynthetic pathways responsible for producing bioactive compounds. Tools such as antiSMASH (Blin et al., 2023) and MIBiG (Kautsar et al., 2020) aid in the identification and annotation of biosynthetic gene clusters (BGCs), facilitating the exploration of the metabolic capabilities of various bacterial species. However, it is essential to recognize that genome mining primarily provides an overview of an organism's metabolic potential rather than a direct indication of physiological activity. While the presence of specific BGCs in a genome suggests that an organism can produce certain secondary metabolites, it does not ensure that these compounds are synthesized under all conditions (Dilshad et al., 2021). Environmental factors, regulatory mechanisms, and growth conditions significantly influence metabolite production (Ghosh et al., 2021), meaning that some compounds may only be expressed under specific circumstances or may remain inactive despite the presence of their corresponding genetic pathways (Kumar et al., 2021).

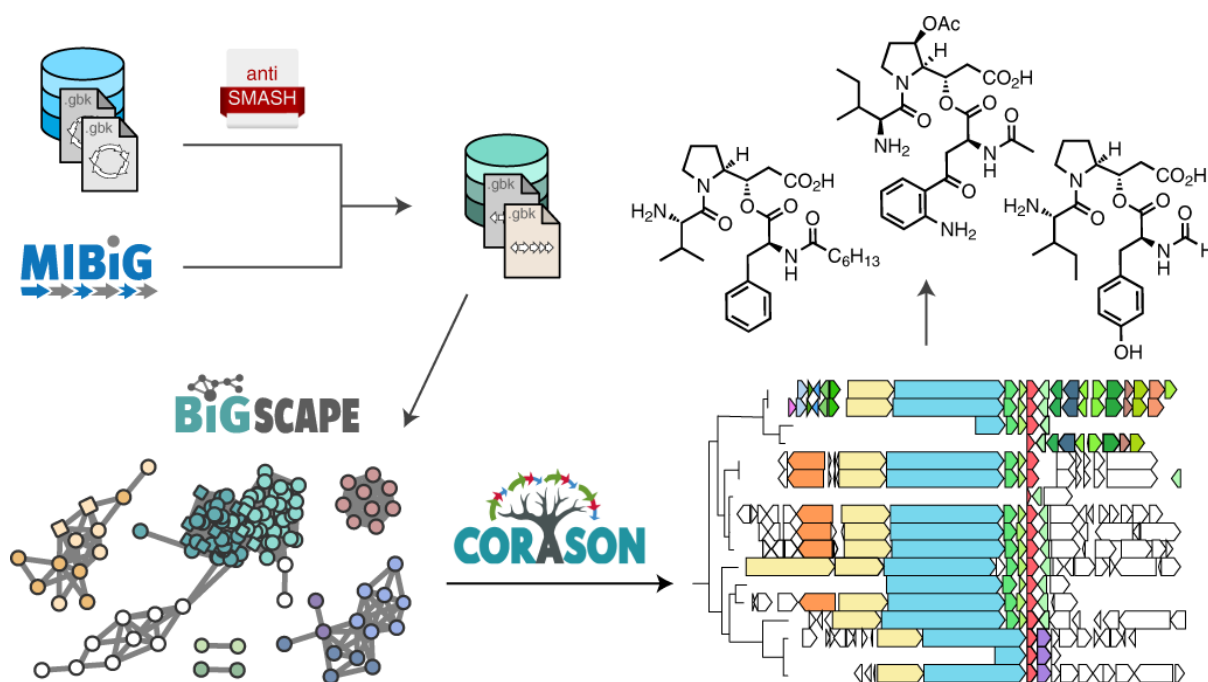


Figure 5: Machine and deep learning algorithms in genomics. The tools antiSMASH and MiBiG are increasingly utilized in genome mining for secondary metabolites, facilitating the acquisition of essential insights into a strain's capacity to produce specific secondary metabolites.

2.8.3 Chemistry techniques used to analyse secondary metabolites

To further study secondary metabolites, various analytical techniques are employed, either individually or in combination. These techniques include High-Performance Liquid Chromatography (HPLC), Mass Spectrometry (MS), Thin Layer Chromatography (TLC), Nuclear Magnetic Resonance (NMR) and combinations like HPLC-MS. However, although these techniques are highly sensitive in detecting of small compounds, they cannot detect all kinds of metabolites (See Figure 4). Some techniques are specific, detecting only particular metabolites, while others, such as MS and NMR, provide valuable structural information about the isolated compounds. However, recent technological advancements have shown that MS-based methods offer greater sensitivity compared to NMR (Labine and Simpson, 2020). As a result, MS has become the preferred choice for metabolomics. Additionally, among MS technologies, the most widely used chromatographic methods for metabolite separation prior to MS analysis are Liquid chromatography and Gas chromatography (Segers et al., 2019). LC is particularly favoured as the chromatography-MS strategy due to its broad range of metabolite coverage (Harrieder et al., 2022).

Technology & Sensitivity

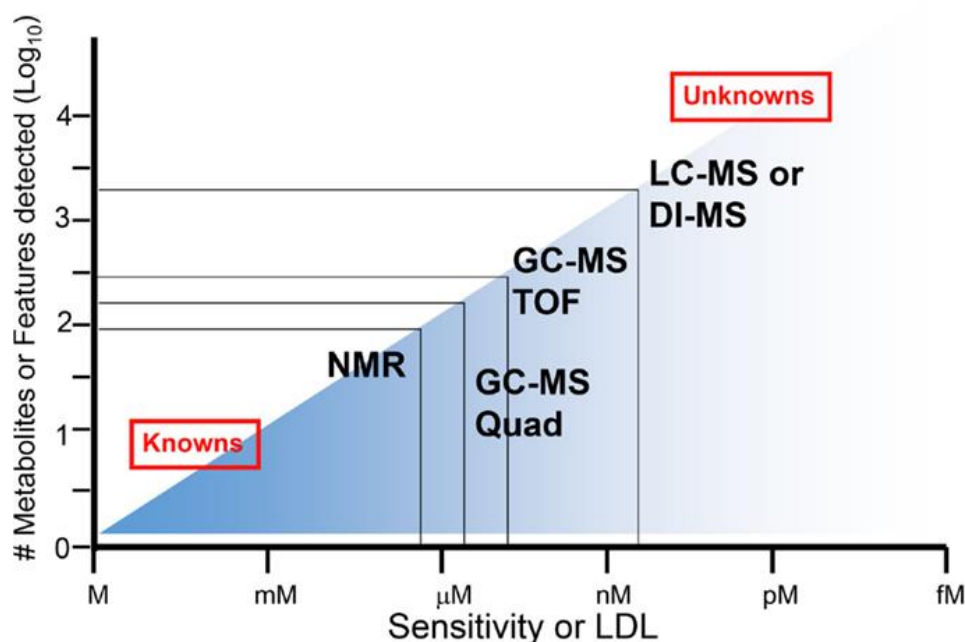


Figure 6: Range of Sensitivities in Metabolomic Technologies: At lower sensitivity or the lower detection limit (LDL), NMR is effective for detecting a smaller number of known metabolites. At higher sensitivity, MS-based technologies are superior for detecting both known and unknown metabolites. Adapted from Stringer et al 2016.

2.8.4 Chromatographic techniques in secondary metabolites analysis (HPLC)

Given the vast diversity of microbial secondary metabolites, chemistry-based techniques like HPLC are essential for analysing the compounds produced by these microorganisms. HPLC is a preferred choice in microbial secondary metabolites because it's high sensitivity and resolution (Kumar et al., 2021). HPLC is routinely employed to purify and analyse bioactive secondary metabolites from bacterial cultures, particularly because it provides excellent separation of complex mixtures allowing for the precise identification and quantification of secondary metabolites even at low concentration (Wolfender et al., 2018, Roopashree and Naik, 2019). Studies have shown that HPLC is a powerful tool for analysing bacterial secondary metabolites due to its high sensitivity, rapid analysis capabilities, versatility, and ability to provide detailed structural insights when combined with other techniques like mass spectrometry (Wolfender et al., 2018). However, despite its advantages, it has limitations, including sensitivity to sample quality and the ability to detect low concentrations.

2.8.5 Mass spectrometric characterisation of secondary metabolites

Generally, Secondary metabolites are the closest reflection of microbial physiological status than gene expression, which tends to provide insights into potential metabolic capabilities without necessarily capturing real-time physiological conditions due to regulatory complexities. Analysing secondary metabolites necessitates the use of high-resolution analytical techniques, such as mass spectrometry. In metabolomics studies, mass spectrometry holds significant promise for investigating microbial physiological states by examining the metabolomes of various microbial species and their responses to different environmental conditions (Kellogg and Kang, 2020, Domenick et al., 2020). Exploring these microbial secondary metabolites using mass spectrometry can lead to the discovery of novel compounds with pharmaceutical significance (Xu et al., 2019, Sigrist et al., 2020). Numerous studies have utilized metabolomics to profile the metabolomic composition of a wide range of microbial species, from bacteria to fungi (Chu et al., 2020, Hautbergue et al., 2018). However, the ecological roles of these metabolites within their environments remain not fully understood.

In addition, Liquid chromatography coupled with tandem mass spectrometry (LC-MSⁿ) has proven to be a powerful tool for the characterization and identification of natural substances (Sarkar et al., 2024). Its effectiveness stems from its versatility in accommodating different ionization modes and employing advanced data acquisition techniques, which expand the range of compounds that can be analyzed. This flexibility is especially beneficial given the diverse physicochemical properties of secondary metabolites. However, quadrupole time-of-flight mass spectrometry (qTOF-MS) instruments are frequently favored in natural product research due to their high accuracy in mass measurement (Shao et al., 2020). MS generally uses data dependent acquisition (DDA) or data independent acquisition (DIA) modes for acquiring tandem mass spectrometry (MS/MS) data. DDA involves selecting specific ions for fragmentation based on their intensity or abundance (Defossez et al., 2023), while DIA involves fragmenting all ions within a predetermined m/z range, regardless of their intensity or abundance (Guo et al., 2020). However, DDA is a more widely used technique and preferred for its high spectral quality and time efficiency, especially for metabolite profiling and annotation in untargeted metabolomics. Despite its limitations in capturing the complete spectrum of metabolites (metabolome), eight guidelines have been introduced for selecting technical parameters to enhance the effectiveness of DDA analyses (Defossez et al., 2023). Given all these improvements, the chemical heterogeneity of metabolites poses a significant challenge in achieving holistic metabolite coverage even with an optimal LC-MS conditions,

which therefore requires the continuous development of new technologies and standards to expand the coverage of metabolites and improve the quality and robustness of the data.

2.8.6 Nuclear magnetic resonance spectroscopy (NMR)

Nuclear magnetic resonance spectroscopy (NMR) is the mostly widely used and powerful technique for structural identification of unknown complex compounds (Sugiki et al., 2018). Although this technique is less sensitive compared to mass spectrometry, it is a major analytical technique in the growing field of metabolomics (Marcarino et al., 2020). This is due to its high reproducibility, quantitative and the ability to identify unknown compound from complex mixtures (Garcia-Perez et al., 2020). Previous studies have employed NMR-based techniques to correlate antibacterial activities by profiling bioactive secondary metabolites from bacteria (Betancur et al., 2020). Recently, researchers have increasingly integrated NMR with mass spectrometry in metabolomics to enhance the detection of metabolites and measure metabolic fluxes using stable isotope-labeled precursors (Letertre et al., 2020). This combination allows for a more comprehensive analysis of metabolic pathways and their dynamics (figure 7).

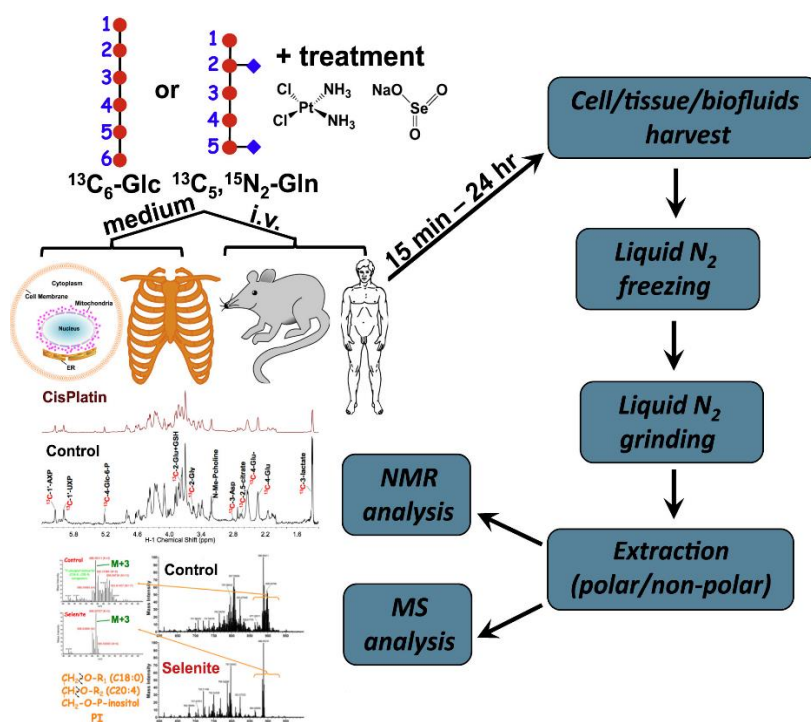


Figure 7: An illustration of the Stable Isotope-Resolved Metabolomic pipeline that integrates NMR and MS to enhance metabolite detection and measure metabolic fluxes. Stable isotope-enriched precursors are introduced into the biological system for a defined period, followed by

the collection of tissue, biofluid, or medium, metabolite extraction, and subsequent analysis using MS and NMR. This diagram was taken from (Fan et al., 2012)

2.9 Computational tools for analysis of these metabolomics data

Many advanced analytical instruments such as MS and NMR generate large datasets that are difficult to analyse using traditional statistical methods. Consequently, machine learning models are increasingly utilized to interpret the complex chemistry within biological systems (Keith et al., 2021). In metabolomics, both targeted and untargeted approaches are employed to explore biological chemistry; however, the processes of isolation, purification, dereplication, and metabolite annotation often create bottlenecks, particularly in untargeted studies (Dong et al., 2024). These traditional characterization methods can be resource-intensive and time-consuming, making large-scale analysis of specialized metabolites challenging. Moreover, the complexity of biological matrices can hinder metabolite identification, complicating the dereplication process. As a result, there is a growing need for automated and high-throughput techniques that can streamline these processes and enhance the efficiency of metabolomic studies (Miggiels et al., 2019). The integration of advanced computational tools with experimental workflows holds promise for overcoming these challenges and facilitating a more comprehensive understanding of specialized metabolites and their roles in biological systems.

2.9.1 Molecular networking

Molecular networking (MN), which is housed in the Global Natural Product Social Molecular Networking (GNPS) ecosystem has emerged as a transformative approach in the field of metabolomics (Yang et al., 2013). This innovative technique leverages tandem mass spectrometry (MS/MS) data to create visual representations of chemical space, enabling the exploration of complex datasets in a systematic manner (Aron et al., 2020). MN facilitates the identification and characterization of metabolites by grouping compounds based on the similarity of their mass spectrometry (MS) spectra, even when specific compounds are not previously known (Ramabulana et al., 2021). The GNPS platform is a comprehensive tool for curating and analysing MS/MS data, enabling putative annotation and interpretation. However, one of the key advantages of GNPS is its ability to handle large volumes of untargeted mass spectrometry data, which traditional analytical methods often struggle to process effectively. It allows users to create molecular networks that reveal relationships between compounds based on spectral similarity, uncovering hidden patterns and connections. MN enhances metabolite

identification and helps discover novel compounds and metabolic pathways, particularly in natural product research where chemical diversity complicates analysis. It enables researchers to visualize complex datasets and gain valuable insights into biochemical interactions within ecosystems (Ramabulana et al., 2023). As a result, molecular networking represents a significant advancement in our ability to explore and understand the vast chemical diversity present in biological systems.

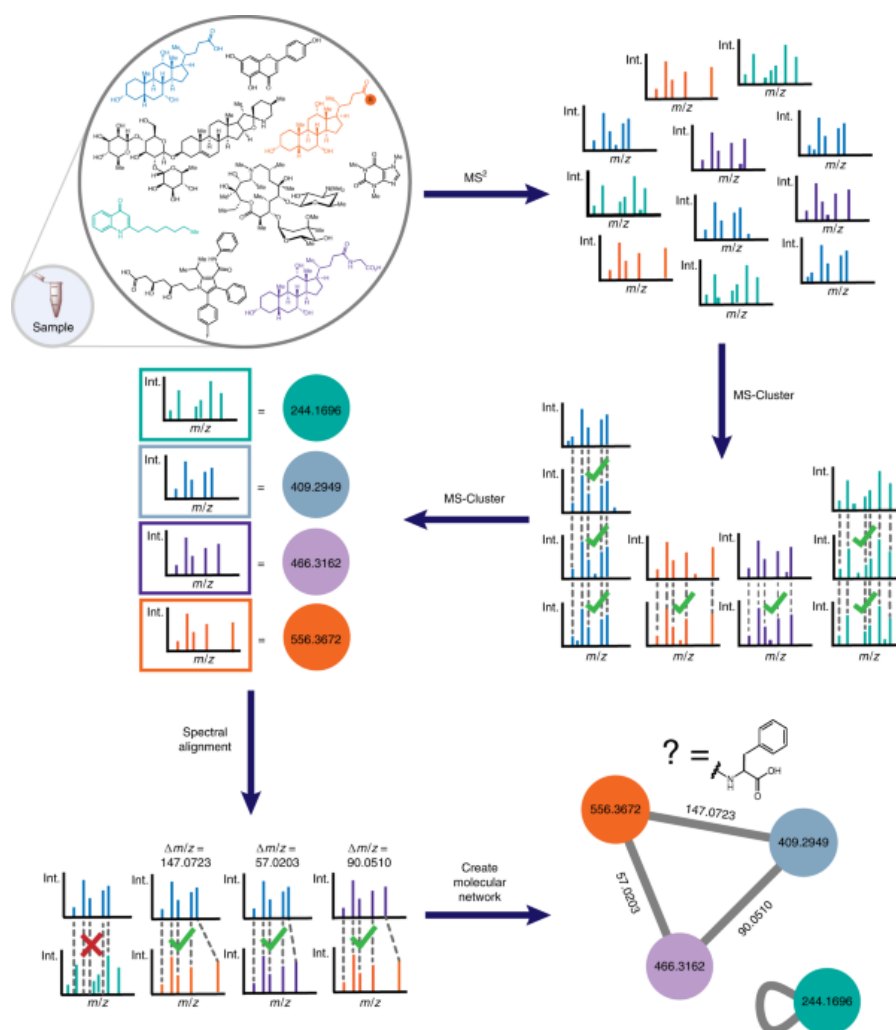


Figure 8: An illustration of the procedure for constructing a classical molecular network within the GNPS ecosystem (Aron et al., 2020). This process involves three primary steps: first, calculating the cosine score to assess similarities between related compounds by comparing all spectra within the sample. Next, identical masses are merged into a consensus spectrum using the MS-Cluster algorithm. Finally, each spectrum in the sample is compared against those in the GNPS databases for annotation, identifying matches with available library spectra.

2.9.2 Spec2vec

Spec2Vec is an innovative tool for analysing mass spectrometry (MS) data, particularly within the Global Natural Product Social Molecular Networking (GNPS) ecosystem. It offers an alternative approach to traditional analysis methods, leveraging machine learning techniques to transform MS data into feature vectors for improved interpretation and comparison (Huber et al., 2021). This tool enhances spectral similarity scoring through learning fragmental relationships from extensive datasets derived from tandem mass spectrometry. Unlike traditional methods that often rely on cosine-based similarity measures, Spec2Vec generates abstract spectral embeddings that capture nuanced relationships between mass spectral peaks. This approach improves the identification of structural analogues and related compounds, making it a valuable tool for library matching and molecular networking tasks (Beniddir et al., 2021). Its computational scalability allows it to handle extensive datasets, making it suitable for high-throughput applications in metabolomics. By bridging traditional mass spectrometry analysis with modern computational techniques, Spec2Vec significantly advances our ability to explore chemical diversity in biological systems, contributing to progress in areas such as drug discovery and environmental science.

2.9.3 MS2LDA

MS2LDA is a powerful tool designed for the decomposition of molecular fragment data derived from large-scale metabolomics experiments. MS2LDA employs a method inspired by Latent Dirichlet Allocation (LDA) to identify co-occurring mass fragments and neutral losses, known as Mass2Motifs, which often correspond to shared chemical substructures across various compounds (Wandy et al., 2018). This unsupervised approach allows researchers to explore complex metabolite mixtures and facilitates the identification of structural families without relying solely on existing spectral databases. The integration of MS2LDA with additional computational tools, such as MAGMa for automated annotation and ClassyFire for structural classification, further enriches its capabilities (Rogers et al., 2019). These enhancements streamline the annotation process of Mass2Motifs, allowing for semi-automated identification of molecular substructures across diverse datasets. MS2LDA fosters a collaborative environment that allows researchers to share and compare their findings, thereby accelerating the analysis of complex mixtures and expanding the chemical space that can be characterized.

2.9.4 GNPS *in silico* tools

The Global Natural Product Social Molecular Networking (GNPS) platform offers a suite of advanced *in silico* annotation tools that enhance the analysis of mass spectrometry (MS) data, particularly for natural product research. Key tools within the GNPS ecosystem include Network Annotation Propagation (NAP), Sirius, Deplicator, and MolNetEnhancer. NAP refines structure annotations by leveraging molecular network topology and structural similarities (Figure 9), improving predictions even without direct spectral library matches (da Silva et al., 2018). Sirius employs machine learning to predict molecular structures from MS/MS fragmentation patterns (Dührkop et al., 2019), while Deplicator reduces redundancy by removing duplicate entries from molecular networks (Mohimani et al., 2018). MolNetEnhancer integrates molecular networks with MS2LDA substructure data and GNPS library matches, offering a comprehensive view of molecular families and their chemical characteristics (Ernst et al., 2019). Together, these tools improve metabolite identification, streamline analysis, and support the discovery of novel compounds in complex biological samples.

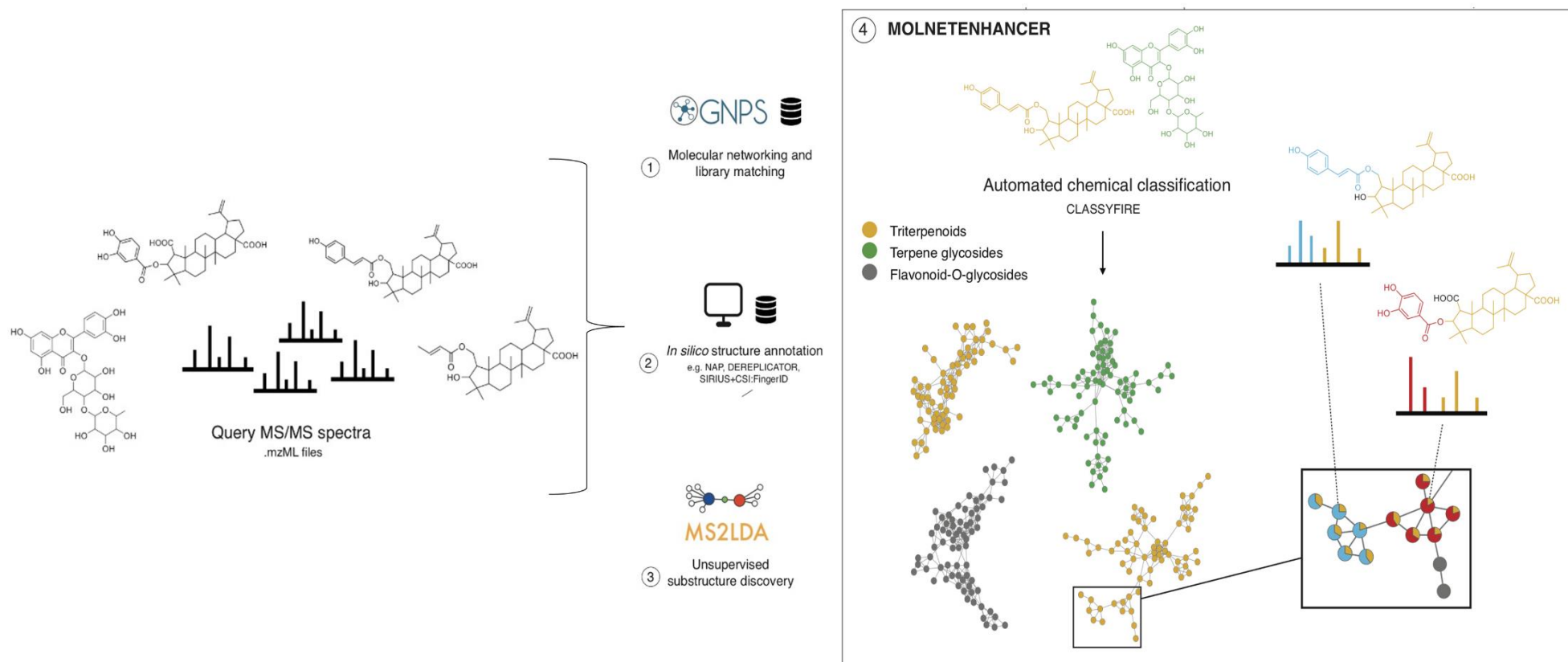


Figure 9: An illustration of the procedure for GNPS *in silico* tools within the GNPS ecosystem (Ernst et al., 2019). This process involves two primary steps: first, integrate MS2LDA substructure information with mass spectral molecular networks. Finally, integrate chemical structural information from GNPS library matching, and *in silico* structure prediction with mass spectral molecular networks to retrieve most abundant chemical classes per molecular family.

2.10 References

- ABDEL-RAZEK, A. S., EL-NAGGAR, M. E., ALLAM, A., MORSY, O. M. & OTHMAN, S. I. 2020. Microbial natural products in drug discovery. *Processes*, 8, 470.
- ABEBE, G. M. 2020. The role of bacterial biofilm in antibiotic resistance and food contamination. *International journal of microbiology*, 2020, 1705814.
- AL DACCACHE, M., LOUKA, N., MAROUN, R. G. & SALAMEH, D. 2021. Culture Condition Changes for Enhancing Fermentation Processes. *Fermentation Processes: Emerging and Conventional Technologies*, 97-116.
- ANGELINI, P. 2024. Plant-Derived Antimicrobials and Their Crucial Role in Combating Antimicrobial Resistance. *Antibiotics*, 13, 746.
- ARON, A. T., GENTRY, E. C., MCPHAIL, K. L., NOTHIAS, L.-F., NOTHIAS-ESPOSITO, M., BOUSLIMANI, A., PETRAS, D., GAUGLITZ, J. M., SIKORA, N. & VARGAS, F. 2020. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature protocols*, 15, 1954-1991.
- BELLUCCI, M. C. & VOLONTERIO, A. 2020. Aminoglycosides: from antibiotics to building blocks for the synthesis and development of gene delivery vehicles. *Antibiotics*, 9, 504.
- BENIDDIR, M. A., KANG, K. B., GENTA-JOUVE, G., HUBER, F., ROGERS, S. & VAN DER HOOFT, J. J. 2021. Advances in decomposing complex metabolite mixtures using substructure-and network-based computational metabolomics approaches. *Natural product reports*, 38, 1967-1993.
- BERILLO, D., MALIKA, T., BAIMAKHANOVA, B. B., SADANOV, A. K., BEREZIN, V. E., TRENOZHNIKOVA, L. P., BAIMAKHANOVA, G. B., AMANGELDI, A. A. & KERIMZHANOVA, B. 2024. An Overview of Microorganisms Immobilized in a Gel Structure for the Production of Precursors, Antibiotics, and Valuable Products. *Gels*, 10, 646.
- BETANCUR, L. A., FORERO, A. M., VINCHIRA-VILLARRAGA, D. M., CÁRDENAS, J. D., ROMERO-OTERO, A., CHAGAS, F. O., PUPO, M. T., CASTELLANOS, L. & RAMOS, F. A. 2020. NMR-based metabolic profiling to follow the production of anti-phytopathogenic compounds in the culture of the marine strain *Streptomyces* sp. PNM-9. *Microbiological research*, 239, 126507.
- BIBIÁN, M. E., PÉREZ-SÁNCHEZ, A., MEJÍA, A. & BARRIOS-GONZÁLEZ, J. 2020. Penicillin and cephalosporin biosyntheses are also regulated by reactive oxygen species. *Applied microbiology and biotechnology*, 104, 1773-1783.

- BLIN, K., SHAW, S., AUGUSTIJN, H. E., REITZ, Z. L., BIERMANN, F., ALANJARY, M., FETTER, A., TERLOUW, B. R., METCALF, W. W. & HELFRICH, E. J. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic acids research*, 51, W46-W50.
- BORUTA, T. 2023. Computation-aided studies related to the induction of specialized metabolite biosynthesis in microbial co-cultures: An introductory overview. *Computational and Structural Biotechnology Journal*.
- BRUCE, S. O. 2022. Secondary metabolites from natural products. *Secondary Metabolites: Trends and Reviews*, 51.
- BUDDHIKA, U. & ABEYSINGHE, S. 2021. Secondary metabolites from microbes for plant disease management. *Emerging Trends in Plant Pathology*, 331-342.
- CHAKRABORTY, P. 2022. Gene cluster from plant to microbes: their role in genome architecture, organism's development, specialized metabolism and drug discovery. *Biochimie*, 193, 1-15.
- CHRISTAKI, E., MARCOU, M. & TOFARIDES, A. 2020. Antimicrobial resistance in bacteria: mechanisms, evolution, and persistence. *Journal of molecular evolution*, 88, 26-40.
- CHU, L., HUANG, J., MUHAMMAD, M., DENG, Z. & GAO, J. 2020. Genome mining as a biotechnological tool for the discovery of novel marine natural products. *Critical reviews in biotechnology*, 40, 571-589.
- COBONGELA, S. Z. Z. 2022. Applications of Microbes in Antibiotics. *Application of Microbes in Environmental and Microbial Biotechnology*, 693-710.
- COOK, M. A. & WRIGHT, G. D. 2022. The past, present, and future of antibiotics. *Science translational medicine*, 14, eabo7793.
- DA SILVA, R. R., WANG, M., NOTHIAS, L.-F., VAN DER HOOFT, J. J., CARABALLO-RODRÍGUEZ, A. M., FOX, E., BALUNAS, M. J., KLASSEN, J. L., LOPES, N. P. & DORRESTEIN, P. C. 2018. Propagating annotations of molecular networks using in silico fragmentation. *PLoS computational biology*, 14, e1006089.
- DARBY, E. M., TRAMPARI, E., SIASAT, P., GAYA, M. S., ALAV, I., WEBBER, M. A. & BLAIR, J. M. 2023. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology*, 21, 280-295.
- DEFOSSEZ, E., BOURQUIN, J., VON REUSS, S., RASMANN, S. & GLAUSER, G. 2023. Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 42, 131-143.

- DILSHAD, R., JAMIL, N. & BATOOL, R. 2021. Biosynthetic gene clusters in bacteria: a review: bacterial secondary metabolites producing genes. *Proceedings of the Pakistan Academy of Sciences: B. Life and Environmental Sciences*, 58, 29-42.
- DILSHAD, R., JAMIL, N. & BATOOL, R. 2022. Biosynthetic Gene Clusters in Bacteria: A Review. *Proceedings of the Pakistan Academy of Sciences: Part B*, 58.
- DOMENICK, T. M., GILL, E. L., VEDAM-MAI, V. & YOST, R. A. 2020. Mass spectrometry-based cellular metabolomics: current approaches, applications, and future directions. *Analytical Chemistry*, 93, 546-566.
- DONG, S.-H., DUAN, Z.-K., BAI, M., HUANG, X.-X. & SONG, S.-J. 2024. Advanced technologies targeting isolation and characterization of natural products. *TrAC Trends in Analytical Chemistry*, 117711.
- DÜHRKOP, K., FLEISCHAUER, M., LUDWIG, M., AKSENOV, A. A., MELNIK, A. V., MEUSEL, M., DORRESTEIN, P. C., ROUSU, J. & BÖCKER, S. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature methods*, 16, 299-302.
- ERDEM BÜYÜKKIRAZ, M. & KESMEN, Z. 2022. Antimicrobial peptides (AMPs): A promising class of antimicrobial compounds. *Journal of applied microbiology*, 132, 1573-1596.
- ERNST, M., KANG, K. B., CARABALLO-RODRÍGUEZ, A. M., NOTHIAS, L.-F., WANDY, J., CHEN, C., WANG, M., ROGERS, S., MEDEMA, M. H. & DORRESTEIN, P. C. 2019. MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites*, 9, 144.
- FADIJI, T., RASHVAND, M., DARAMOLA, M. O. & IWARERE, S. A. 2023. A review on antimicrobial packaging for extending the shelf life of food. *Processes*, 11, 590.
- FAN, T. W.-M., LORKIEWICZ, P. K., SELLERS, K., MOSELEY, H. N., HIGASHI, R. M. & LANE, A. N. 2012. Stable isotope-resolved metabolomics and applications for drug development. *Pharmacology & therapeutics*, 133, 366-391.
- FARDA, B., DJEBAILI, R., VACCARELLI, I., DEL GALLO, M. & PELLEGRINI, M. 2022. Actinomycetes from caves: an overview of their diversity, biotechnological properties, and insights for their use in soil environments. *Microorganisms*, 10, 453.
- FURNESS, E., WHITWORTH, D. E. & ZWARYCZ, A. 2020. Predatory interactions between myxobacteria and their prey. *The ecology of predation at the microscale*, 1-36.
- GARCIA-PEREZ, I., POSMA, J. M., SERRANO-CONTRERAS, J. I., BOULANGE, C. L., CHAN, Q., FROST, G., STAMLER, J., ELLIOTT, P., LINDON, J. C. & HOLMES, E.

2020. Identifying unknown metabolites using NMR-based metabolic profiling techniques. *Nature Protocols*, 1-30.
- GHOSH, S., WHITLEY, C. S., HARIBABU, B. & JALA, V. R. 2021. Regulation of intestinal barrier function by microbial metabolites. *Cellular and molecular gastroenterology and hepatology*, 11, 1463-1482.
- GUO, Z., HUANG, S., WANG, J. & FENG, Y.-L. 2020. Recent advances in non-targeted screening analysis using liquid chromatography-high resolution mass spectrometry to explore new biomarkers for human exposure. *Talanta*, 219, 121339.
- HAIDER, A., IKRAM, M., SHAHZADI, I. & ASIF RAZA, M. 2023. Antibiotic Drug Resistance. *Polymeric Nanoparticles for Bovine Mastitis Treatment*. Springer.
- HARRIEDER, E.-M., KRETSCHMER, F., BÖCKER, S. & WITTING, M. 2022. Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics. *Journal of Chromatography B*, 1188, 123069.
- HARWANI, D., BEGANI, J. & LAKHANI, J. 2018. Co-cultivation strategies to induce de novo synthesis of novel chemical scaffolds from cryptic secondary metabolite gene clusters. *Fungi and their Role in Sustainable Development: Current Perspectives*, 617-631.
- HAUTBERGUE, T., JAMIN, E. L., DEBRAUWER, L., PUEL, O. & OSWALD, I. P. 2018. From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites. *Natural Product Reports*, 35, 147-173.
- HENDERSON, P. J., MAHER, C., ELBOURNE, L. D., EIJKELKAMP, B. A., PAULSEN, I. T. & HASSAN, K. A. 2021. Physiological functions of bacterial “multidrug” efflux pumps. *Chemical reviews*, 121, 5417-5478.
- HOBSON, C., CHAN, A. N. & WRIGHT, G. D. 2021. The antibiotic resistome: a guide for the discovery of natural products as antimicrobial agents. *Chemical Reviews*, 121, 3464-3494.
- HUANG, X.-Y., YE, X.-P., HU, Y.-Y., TANG, Z.-X., ZHANG, T., ZHOU, H., ZHOU, T., BAI, X.-L., PI, E.-X. & XIE, B.-H. 2024. Exopolysaccharides of *Paenibacillus polymyxa*: A review. *International Journal of Biological Macromolecules*, 129663.
- HUBER, F., RIDDER, L., VERHOEVEN, S., SPAAKS, J. H., DIBLEN, F., ROGERS, S. & VAN DER HOOFT, J. J. 2021. Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS computational biology*, 17, e1008724.

- HUTCHINGS, M. I., TRUMAN, A. W. & WILKINSON, B. 2019. Antibiotics: past, present and future. *Current opinion in microbiology*, 51, 72-80.
- JAMAL, A. 2023. Antibiotics in Contemporary Medicine: Advances, Obstacles, and the Future. *BULLET: Jurnal Multidisiplin Ilmu*, 2, 548-557.
- JHALORA, V. & BIST, R. 2024. A Comprehensive Review of Molecular Mechanisms Leading to the Emergence of Multidrug Resistance in Bacteria. *Indian Journal of Microbiology*, 1-22.
- KARTHIKEYAN, A., JOSEPH, A. & NAIR, B. G. 2022. Promising bioactive compounds from the marine environment and their potential effects on various diseases. *Journal of Genetic Engineering and Biotechnology*, 20, 14.
- KATHIR, S. 2024. Antibiotic Resistance-The Silent Pandemic. *University of Ottawa Science Undergraduate Research Journal*, 3.
- KAUTSAR, S. A., BLIN, K., SHAW, S., NAVARRO-MUÑOZ, J. C., TERLOUW, B. R., VAN DER HOOFT, J. J., VAN SANTEN, J. A., TRACANNA, V., SUAREZ DURAN, H. G. & PASCAL ANDREU, V. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic acids research*, 48, D454-D458.
- KEITH, J. A., VASSILEV-GALINDO, V., CHENG, B., CHMIELA, S., GASTEGGER, M., MULLER, K.-R. & TKATCHENKO, A. 2021. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical reviews*, 121, 9816-9872.
- KELLOGG, J. & KANG, S. 2020. Metabolomics, an essential tool in exploring and harnessing microbial chemical ecology. *Phytobiomes Journal*, 4, 195-210.
- KHAOCHAMNAN, R., SUANYUK, N., LERTCANAWANICHAKUL, M. & PEDPRADAB, P. 2024. Biological characteristics of marine *Streptomyces* SK3 and optimization of cultivation conditions for production of compounds against Vibriosis pathogen isolated from cultured white shrimp (*Litopenaeus vannamei*). *PeerJ*, 12, e18053.
- KUMAR, V., AHLUWALIA, V., SARAN, S., KUMAR, J., PATEL, A. K. & SINGHANIA, R. R. 2021. Recent developments on solid-state fermentation for production of microbial secondary metabolites: Challenges and solutions. *Bioresource Technology*, 323, 124566.
- LABINE, L. M. & SIMPSON, M. J. 2020. The use of nuclear magnetic resonance (NMR) and mass spectrometry (MS)–based metabolomics in environmental exposure assessment. *Current Opinion in Environmental Science & Health*, 15, 7-15.

- LEISNER, J. J. 2020. The diverse search for synthetic, semisynthetic and natural product antibiotics from the 1940s and up to 1960 exemplified by a small pharmaceutical player. *Frontiers in Microbiology*, 11, 976.
- LETERTRE, M. P., DERVILLY, G. & GIRAUDEAU, P. 2020. Combined nuclear magnetic resonance spectroscopy and mass spectrometry approaches for metabolomics. *Analytical chemistry*, 93, 500-518.
- LI, X., XU, H., LI, Y., LIAO, S. & LIU, Y. 2023. Exploring diverse bioactive secondary metabolites from marine microorganisms using co-culture strategy. *Molecules*, 28, 6371.
- LV, J., HUANG, Z., LUO, L., ZHANG, S. & WANG, Y. 2022. Advances in molecular and microscale characterization of soil organic matter: current limitations and future prospects. *Environmental science & technology*, 56, 12793-12810.
- MARCARINO, M. O., ZANARDI, M. A. M., CICETTI, S. & SAROTTI, A. M. 2020. NMR calculations with quantum methods: development of new tools for structural elucidation and beyond. *Accounts of Chemical Research*, 53, 1922-1932.
- MAURYA, N. K. 2024. Microbial Architects of Soil Health: A Multifaceted Approach to Improving Plant Nutrition, Biocontrol, and Phytohormone Production. *Biotechnology*, 10, 29-42p.
- MCLAREN, M. R. & CALLAHAN, B. J. 2020. Pathogen resistance may be the principal evolutionary advantage provided by the microbiome. *Philosophical Transactions of the Royal Society B*, 375, 20190592.
- MHLONGO, M. I., PIATER, L. A., MADALA, N. E., LABUSCHAGNE, N. & DUBERY, I. A. 2018. The chemistry of plant-microbe interactions in the rhizosphere and the potential for metabolomics to reveal signaling related to defense priming and induced systemic resistance. *Frontiers in plant science*, 9, 112.
- MIGGIELS, P., WOUTERS, B., VAN WESTEN, G. J., DUBBELMAN, A.-C. & HANKEMEIER, T. 2019. Novel technologies for metabolomics: More for less. *TrAC Trends in Analytical Chemistry*, 120, 115323.
- MISHRA, S., SINGH, J. & SINGH, V. 2024. Types and applications of potential antibiotics produced by fungi. *Fungal Secondary Metabolites*. Elsevier.
- MOHAN, M. S., SALIM, S. A., PAKHIRA, P. & BUSI, S. 2024. Microbial Production of Polyketides and Non-ribosomal Peptides and Their Applications. *Microbial Products for Health and Nutrition*. Springer.

- MOHANTY, H., PACHPUTE, S. & YADAV, R. P. 2021. Mechanism of drug resistance in bacteria: efflux pump modulation for designing of new antibiotic enhancers. *Folia Microbiologica*, 66, 727-739.
- MOHIMANI, H., GUREVICH, A., SHLEMOV, A., MIKHEENKO, A., KOROBAYNIKOV, A., CAO, L., SHCHERBIN, E., NOTHIAS, L.-F., DORRESTEIN, P. C. & PEVZNER, P. A. 2018. Dereplication of microbial metabolites through database search of mass spectra. *Nature communications*, 9, 4035.
- MURUGAIYAN, J., KUMAR, P. A., RAO, G. S., ISKANDAR, K., HAWSER, S., HAYS, J. P., MOHSEN, Y., ADUKKADUKKAM, S., AWUAH, W. A. & JOSE, R. A. M. 2022. Progress in alternative strategies to combat antimicrobial resistance: Focus on antibiotics. *Antibiotics*, 11, 200.
- MUTEEB, G., REHMAN, M. T., SHAHWAN, M. & AATIF, M. 2023. Origin of antibiotics and antibiotic resistance, and their impacts on drug development: A narrative review. *Pharmaceuticals*, 16, 1615.
- NIU, X., ZHANG, J., XUE, X., WANG, D., WANG, L. & GAO, Q. 2021. Deacetoxycephalosporin C synthase (expandase): Research progress and application potential. *Synthetic and Systems Biotechnology*, 6, 396-401.
- OLADIPO, A., ENWEMIWE, V., EJEROMEDOGENE, O., ADEBAYO, A., OGUNYEMI, O. & FU, F. 2022. Production and functionalities of specialized metabolites from different organic sources. *Metabolites*, 12, 534.
- ONGPIATTANAKUL, C., DESORMEAUX, E. K., DICAPRIO, A., VAN DER DONK, W. A., MITCHELL, D. A. & NAIR, S. K. 2022. Mechanism of action of ribosomally synthesized and post-translationally modified peptides. *Chemical reviews*, 122, 14722-14814.
- OTHOUM, G., BOUGOUFFA, S., BOKHARI, A., LAFI, F. F., GOJOBORI, T., HIRT, H., MIJAKOVIC, I., BAJIC, V. B. & ESSACK, M. 2019. Mining biosynthetic gene clusters in *Virgibacillus* genomes. *BMC genomics*, 20, 1-10.
- OTHOUM, G., BOUGOUFFA, S., RAZALI, R., BOKHARI, A., ALAMOUDI, S., ANTUNES, A., GAO, X., HOEHNDORF, R., AROLD, S. T. & GOJOBORI, T. 2018. In silico exploration of Red Sea *Bacillus* genomes for natural product biosynthetic gene clusters. *BMC genomics*, 19, 1-11.
- PAL, P., SINGH, A. K., SARANGI, P. K., SAHOO, U. K., SINGH, H. B., SUBUDHI, S. & SINGH, T. A. 2024. Production of gamma-polyglutamic acid microgel by *Bacillus*

- species: Industrial applications and future perspectives. *Polymers for Advanced Technologies*, 35, e6565.
- PANTER, F., BADER, C. D. & MÜLLER, R. 2021. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chemical Science*, 12, 5994-6010.
- PÉREZ, J., CONTRERAS-MORENO, F. J., MARCOS-TORRES, F. J., MORALEDA-MUÑOZ, A. & MUÑOZ-DORADO, J. 2020. The antibiotic crisis: How bacterial predators can help. *Computational and Structural Biotechnology Journal*, 18, 2547-2555.
- PRASAD, S., MALAV, L. C., CHOUDHARY, J., KANNOJIYA, S., KUNDU, M., KUMAR, S. & YADAV, A. N. 2021. Soil microbiomes for healthy nutrient recycling. *Current trends in microbial biotechnology for sustainable agriculture*, 1-21.
- RAMABULANA, A.-T., PETRAS, D., MADALA, N. E. & TUGIZIMANA, F. 2021. Metabolomics and molecular networking to characterize the chemical space of four *Momordica* plant species. *Metabolites*, 11, 763.
- RAMABULANA, A.-T., PETRAS, D., MADALA, N. E. & TUGIZIMANA, F. 2023. Mass spectrometry DDA parameters and global coverage of the metabolome: Spectral molecular networks of *momordica cardiospermoides* plants. *Metabolomics*, 19, 18.
- RAMÍREZ-RENDON, D., PASSARI, A. K., RUIZ-VILLAFÁN, B., RODRÍGUEZ-SANOJA, R., SÁNCHEZ, S. & DEMAÍN, A. L. 2022. Impact of novel microbial secondary metabolites on the pharma industry. *Applied Microbiology and Biotechnology*, 106, 1855-1878.
- ROBERTSEN, H. L. & MUSIOL-KROLL, E. M. 2019. Actinomycete-derived polyketides as a source of antibiotics and lead structures for the development of new antimicrobial drugs. *Antibiotics*, 8, 157.
- RODRÍGUEZ-BELTRÁN, J., DELAFUENTE, J., LEON-SAMPEDRO, R., MACLEAN, R. C. & SAN MILLAN, A. 2021. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology*, 19, 347-359.
- ROGERS, S., ONG, C. W., WANDY, J., ERNST, M., RIDDER, L. & VAN DER HOOFT, J. J. 2019. Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. *Faraday Discussions*, 218, 284-302.

- ROMERO, C. A., KURTBÖKE, D. Í. & QUINN, R. J. 2024. Dereplication of the Termite Gut-associated Actinomycete Metabolome as a Source of Bioactive Secondary Metabolites. *Actinomycetes in Marine and Extreme Environments*. CRC Press.
- ROOPASHREE, K. & NAIK, D. 2019. Advanced method of secondary metabolite extraction and quality analysis. *Journal of Pharmacognosy and Phytochemistry*, 8, 1829-1842.
- ROSERO-CHASOY, G., RODRÍGUEZ-JASSO, R. M., AGUILAR, C. N., BUITRÓN, G., CHAIREZ, I. & RUIZ, H. A. 2021. Microbial co-culturing strategies for the production high value compounds, a reliable framework towards sustainable biorefinery implementation—an overview. *Bioresource technology*, 321, 124458.
- ROY, A., AHUJA, S. & GARG, S. 2021. Fungal secondary metabolites: Biological activity and potential applications. *Recent Trends in Mycological Research: Volume 1: Agricultural and Medical Perspective*, 159-188.
- SAHA, M. & SARKAR, A. 2021. Review on multiple facets of drug resistance: a rising challenge in the 21st century. *Journal of xenobiotics*, 11, 197-214.
- SALAM, L. B., OBAYORI, O. S. & MOHAMMED, M. O. 2021. Detection of biosynthetic genes of microbially-synthesized secondary metabolites in a contaminated tropical agricultural soil. *Biologia*, 76, 3449-3464.
- SARKAR, J., SINGH, R. & CHANDEL, S. 2024. Understanding LC/MS-Based Metabolomics: A Detailed Reference for Natural Product Analysis. *PROTEOMICS—Clinical Applications*, e202400048.
- SEGERS, K., DECLERCK, S., MANGELINGS, D., HEYDEN, Y. V. & EECKHAUT, A. V. 2019. Analytical techniques for metabolomic studies: A review. *Bioanalysis*, 11, 2297-2318.
- SELEGATO, D. M. & CASTRO-GAMBOA, I. 2023. Enhancing chemical and biological diversity by co-cultivation. *Frontiers in microbiology*, 14, 1117559.
- SHAO, S.-Y., TING, Y., WANG, J., SUN, J. & GUO, X.-F. 2020. Characterization and identification of the major flavonoids in *Phyllostachys edulis* leaf extract by UPLC–QTOF–MS/MS. *Acta Chromatographica*, 32, 228-237.
- SHARRAR, A. M., CRITS-CHRISTOPH, A., MÉHEUST, R., DIAMOND, S., STARR, E. P. & BANFIELD, J. F. 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *MBio*, 11, 10.1128/mbio.00416-20.
- SIGRIST, R., PAULO, B. S., ANGOLINI, C. F. & DE OLIVEIRA, L. G. 2020. Mass spectrometry-guided genome mining as a tool to uncover novel natural products. *JoVE (Journal of Visualized Experiments)*, e60825.

- SINGH, B., BHAT, A. & RAVI, K. 2024. Antibiotics Misuse and Antimicrobial Resistance Development in Agriculture: A Global Challenge. *Environment & Health*, 2, 618-622.
- SMITH, W. P., WUCHER, B. R., NADELL, C. D. & FOSTER, K. R. 2023. Bacterial defences: mechanisms, evolution and antimicrobial resistance. *Nature Reviews Microbiology*, 21, 519-534.
- SRIKANTH, P., SIVAKUMAR, D., SHARMA, A. & KAUSHIK, N. 2024. Recent developments in omics techniques for improving plant abiotic stress using microbes. *International Journal of Environmental Science and Technology*, 1-24.
- SUGIKI, T., FURUITA, K., FUJIWARA, T. & KOJIMA, C. 2018. Current NMR techniques for structure-based drug discovery. *Molecules*, 23, 148.
- SUŁKOWSKA-ZIAJA, K., TREPA, M., OLECHOWSKA-JARZĄB, A., NOWAK, P., ZIAJA, M., KAŁA, K. & MUSZYŃSKA, B. 2023. Natural compounds of fungal origin with antimicrobial activity—potential cosmetics applications. *Pharmaceuticals*, 16, 1200.
- TANG, K. W. K., MILLAR, B. C. & MOORE, J. E. 2023. Antimicrobial resistance (AMR). *British Journal of Biomedical Science*, 80, 11387.
- TIWARI, P. & BAE, H. 2022. Endophytic fungi: Key insights, emerging prospects, and challenges in natural product drug discovery. *Microorganisms*, 10, 360.
- VARELA, M. F., STEPHEN, J., LEKSHMI, M., OJHA, M., WENZEL, N., SANFORD, L. M., HERNANDEZ, A. J., PARVATHI, A. & KUMAR, S. H. 2021. Bacterial resistance to antimicrobial agents. *Antibiotics*, 10, 593.
- WANDY, J., ZHU, Y., VAN DER HOOFT, J. J., DALY, R., BARRETT, M. P. & ROGERS, S. 2018. Ms2lda. org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics*, 34, 317-318.
- WOLFENDER, J.-L., NUZILLARD, J.-M., VAN DER HOOFT, J. J., RENAULT, J.-H. & BERTRAND, S. 2018. Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography–high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Analytical Chemistry*, 91, 704-742.
- XU, F., WU, Y., ZHANG, C., DAVIS, K. M., MOON, K., BUSHIN, L. B. & SEYEDSAYAMDOST, M. R. 2019. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nature chemical biology*, 15, 161-168.
- YANG, J. Y., SANCHEZ, L. M., RATH, C. M., LIU, X., BOUDREAU, P. D., BRUNS, N., GLUKHOV, E., WODTKE, A., DE FELICIO, R. & FENNER, A. 2013. Molecular networking as a dereplication strategy. *Journal of natural products*, 76, 1686-1699.

- YIN, Q.-J., YING, T.-T., ZHOU, Z.-Y., HU, G.-A., YANG, C.-L., HUA, Y., WANG, H. & WEI, B. 2023. Species-specificity of the secondary biosynthetic potential in *Bacillus*. *Frontiers in Microbiology*, 14, 1271418.
- YU, G., GE, X., LI, W., JI, L. & YANG, S. 2024. Interspecific cross-talk: The catalyst driving microbial biosynthesis of secondary metabolites. *Biotechnology Advances*, 108420.
- ZHANG, X., LI, Z., PANG, S., JIANG, B., YANG, Y., DUAN, Q. & ZHU, G. 2021. The impact of cell structure, metabolism and group behavior for the survival of bacteria under stress conditions. *Archives of Microbiology*, 203, 431-441.

Chapter 3: Isolation and identification of antibiotic-producing bacteria from soil.

3.1 Summary

Background: Due to the emergence of drug-resistant microorganisms, the search for broad-spectrum antimicrobial compounds has become extremely crucial. Natural sources like plants and soils have been explored for diverse metabolites with antimicrobial properties.

Purpose: This study aimed to isolate microorganisms from agricultural soils exhibiting antimicrobial effects against known human pathogens, and to identify them through the 16S rRNA-based approach.

Methods: Herein, bacteria were extracted from soil samples and their antimicrobial potential was measured via the agar well diffusion method. Furthermore, 16S rRNA gene sequencing enabled the identification of active bacterial isolates.

Results: 16S rRNA gene sequencing aided the identification of bacterial isolates where isolate 1 and 2 were identified as *Bacillus pumilus*, whilst isolate 3 was found to be *Bacillus subtilis*. Interestingly, isolate 3 (*Bacillus subtilis*) displayed wide-ranging antimicrobial activity against the tested human pathogens. Additionally, the sprinkling technique accelerates the discovery of bioactive compounds and the isolation of strains which offer a promising avenue for discovering new clinical antimicrobials.

Conclusions: The results from this study showed that soil could serve as an alternative source of anti-microbial compounds that can assist in the fight against the emergence of multi-drug resistance bacterial pathogens.

Keywords: *Bacillus*, Soil microorganism, 16S rRNA, Bioprospection, Antibiotics.

3.2 Introduction

Globally, drug resistance and re-emerging pathogens have been a growing concern in the agriculture, food, and medical industry, necessitating efforts to discover new antibiotics and enhance treatment options (Shrivastava et al., 2018). The unprecedented rate of drug resistance has been attributed to indiscriminate use of antibiotics, inadequacies on the part of the manufactures and acquired resistance through microbial evolution (Podolsky, 2018). Though some bacteria are naturally resistant to antibiotics, other bacteria gain resistance through different mechanisms such as mutation in the genes, acquisition of new genetic elements encoding protective enzymes and protein (Bengtsson-Palme et al., 2018). As such, while the number of microorganisms gaining resistance grows, so should the need to discover new antibiotics to address the evolving challenges posed by infectious diseases. Conversely, problems with rediscovery of known strains and compounds led to the decline in the discovery efforts during the second half of the 20th century, hence making it less attractive for the pharmaceutical industry (Atanasov et al., 2021).

However, soil contains a high diversity of microorganisms (Nannipieri et al., 2003), some of which are essential for supporting life (Mhlongo et al., 2018). Nevertheless, our current understanding of soil-borne microbial populations remains limited, representing only a fraction of their actual diversity and composition, and their chemical space remains an untapped territory often described as 'dark matter' in metabolomics (Schultz et al., 2023). Consequently, a holistic understanding of the metabolites and various compounds produced by these microorganisms, including those classified as "known unknowns," "unknown knowns," and "unknown unknowns" remains a mystery (Hoskisson and Seipke, 2020). Thus, a notable research gap exists, which warrants concerted efforts to elucidate the composition of microbial populations in soil and their associated chemical characteristics. Unlocking the full potential of microorganisms necessitates innovative and efficient approaches for the systematic exploration of their metabolic capabilities. Such efforts not only enhance our understanding of soil microbiomes but also offer potential in developing novel antimicrobial agents, contributing to the fight against antimicrobial resistance.

Recently, researchers have used several combinations of interdisciplinary approaches for bioprospecting, such as bioassays (Cushnie et al., 2020), genome mining (Machado et al., 2015) and metagenomics (Thakur et al., 2020). Although these approaches provided a good

foundation for Microbial bioprospecting, their limitation is their inability to provide insights into how these genes are regulated and how they interact within the microbial community (Pierce and Dutton, 2022). However, the identification of bioactive compounds has been primary based on their extraction from the culture medium, often through traditional bioassays evaluating their inhibitory effects on other microorganisms (Sasidharan et al., 2011, Oberpaul et al., 2022). This approach is also accompanied or followed by analysis of their chemical composition and structure through various analytical techniques (Berthod et al., 2009, Granados-Chinchilla and Rodríguez, 2017, Feng et al., 2019). The approach enables the identification of potential biotechnological organisms, and the integration of genomics and metabolomics facilitates the discovery of previously unknown compounds (Mohana et al., 2018, Maghembe et al., 2020, Aborode et al., 2022).

Herein, we present a methodological systematic approach for an effective bioprospection from soil microorganisms. The present study is one such attempt to isolate soil-borne bacteria with the potential to produce anti-microbial compounds against known pathogenic bacteria. The findings of the current study highlight the use of sprinkling technique approach as a feasible approach for the identification of bacteria strains capable of producing anti-microbial compounds which may be for use in the management of drug-resistant pathogens in the future. Moreover, identification of the strains showing positive anti-microbial activities was achieved through 16S rRNA gene sequencing.

3.3 Experimental

3.3.1 Sample collection

Soil samples were collected from 40 different points in an agricultural setting at Tshakhuma village under Makhado Municipality (-23.083019, 30.301499) at a distance of varying proximity. The samples were collected from each site at depths of (0-20 cm) below the surface arbitrarily. Following collection, samples were placed in sterile plastic bags and stored, in the dark, at 4 °C and transferred to the Microbiology research laboratory, University of Venda where the entire research work was carried out. Processing was completed following the receipt of samples within 24 hours.

3.3.2 Isolation of antibiotic producing strains

In the present study, soil sprinkle technique was used to isolate antibiotic producing microorganism following the procedure described by Jamil et al., (2007). For this purpose, 1 mg of soil particles were sprinkled on the surface of non-selective nutrient agar plates seeded with the test organism, *Escherichia coli* (ATCC 35218). The plates were aerobically incubated at 37 °C for 24 hours. Antibiotic activity was checked by the zone of inhibition, surrounding a colony. Colonies displaying clear zones of inhibition around them were picked and streaked on separate nutrient agar plates to get pure cultures. Each isolated soil microorganism was assigned with specific codes, and these isolates were used as the source of antibiotic producing microbes. All strains were stored at -20 °C in 20 % glycerol Muller Hinton Broth (Merck, Kenilworth, NJ, USA) for further confirmation and characterization by PCR and were subcultured periodically.

3.3.3 Primary screening of antimicrobial compounds

Microorganisms isolated from different soil samples were screened for their antimicrobial activities. The test bacteria that were used for screening were *Staphylococcus aureus* ATCC (25923), *Pseudomonas aeruginosa* ATCC (9027), *Klebsiella pneumoniae* ATCC (BAA-1705) and *Escherichia coli* ATCC (35218). Antimicrobial activities were assessed using Muller Hinton agar. Each plate was spread with the test organism and a small amount of the bacterial isolate was spotted on the test organism. Isolates displaying clear zones of inhibition against the pathogens were isolated onto plates and those displaying little or no inhibitory properties were re-tested and discarded if scarce antibiotic biosynthesis persisted. Purified microorganisms obtained were further tested for growth-inhibitory properties by using the agar

well diffusion assay (Woappi et al., 2013), against all the test microorganisms. Isolates that showed inhibitory zones against one or more pathogens were further purified and maintained at -20 °C in 20 % glycerol Muller Hinton Broth (Merck, Kenilworth, NJ, USA).

3.3.4 Taxonomic identification of the bacterial isolates

Pure bacterial isolates were cultured in Mueller Hinton Broth for 24 hours. Genomic DNA was extracted from bacterial pellets using the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Polymerase chain reaction (PCR) amplification of the 16S rRNA gene was performed using the primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-CGGTTACCTTGTTACGACTT-3') as describe by (Chen et al., 2015). Each PCR reaction of 25 µl in total included 12.5 µl DreamTaq DNA polymerase mix (ThermoFisher Scientific, United States), 6.5 µl distilled water (dH₂O), the final concentration 0.4 µM of each primer and 4 µl DNA template. The cycling conditions were as follows: 5 min at 95 °C, 25 cycles at 94 °C for 40 s, at 55 °C for 30 s and 1 min at 72 °C, then followed by a final elongation step for 7 min at 72 °C. The PCR amplicons were sequenced by Inqaba-Biotech (Pretoria, South Africa), using the afore-mentioned PCR primers.

3.4 Results

3.4.1 Morphological characteristics

Based on the inhibition zone observed in the sprinkling technique, antibiotic-producing organisms were identified. Out of all the soil samples screened for antibiotic producers, only three isolates displayed colonies with a zone of inhibition. These cultures were then purified for further analysis. The colony morphology was dry, crumbly, and white to cream in color (Figure 10), with a gram-positive stain. Based on the colony morphology and gram-staining characteristics, the isolates were classified as members of the *Bacillus* genus and designated as isolates 1, 2, and 3.



Figure 10: Morphological characteristics of three soil-borne isolates on nutrient medium

3.4.2 Antimicrobial activity testing

Preliminary screening of soil isolates was carried out at 37 °C against the four 4 (four) ATCC strains. Out of 40 samples, only 3 isolates showed convincing activity against the test organisms. Maximum inhibition was achieved against *Staphylococcus aureus* ATCC (25923) by Isolate 3. Minimum inhibitory activity was noted against *Escherichia coli* ATCC (35218) by Isolate 1 as shown in Figure 11. Testing for growth-inhibitory properties was done using the agar well diffusion assay under standard culturing conditions at 37 °C. The diameter of the growth inhibition zones around each well was measured and was taken to be related to the susceptibility of the antimicrobial substances produced by isolate to the test organism, and the diffusion rate of the antimicrobial substance through the agar medium.

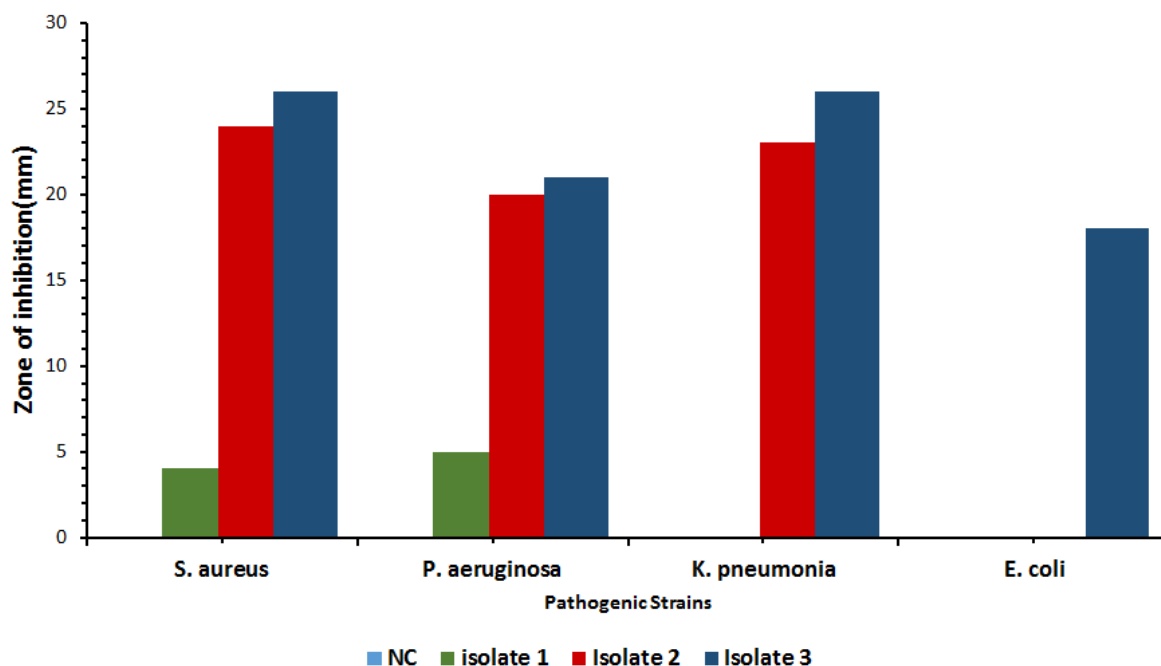


Figure 11: Infographic display of bacterial isolates extracts analyzed by agar diffusion technique against 4 pathogenic bacteria showing antimicrobial activity after 48 hours of incubation. Colors representing the isolates tested with NC being the negative control.

3.4.3 16S rDNA Aided Identification of Bacterial Isolates

For genome extraction, three bacterial isolates that demonstrated efficacy against pathogenic organisms were selected. Using nanodrop, the presence of DNA yield was evaluated. The DNA was then utilized as templates for PCR amplification that followed. In the present study, the 16s rRNA gene sequences of three active isolates were determined. The Basic Local Alignment Search Tool (BLAST) was employed, and the findings revealed that all three bacterial isolates belonged to the genus *Bacillus*. The identities of the 3 isolates were determined by comparing them to the available data in the Genbank (NCBI) and with high-scored rRNA gene sequences in the BLAST searches. BLAST search similarities scores range between 99%-100%. Phylogenetic analysis of the amplified 16S rRNA gene revealed that the two active isolates, here forth referred to isolate 1 and isolate 2 (3SIK, and 8SIK) share sequence homology *Bacillus pumilus* (Figure 12), on the other hand, isolate 3 had convincing sequence homology with *Bacillus subtilis* (Figure 13).

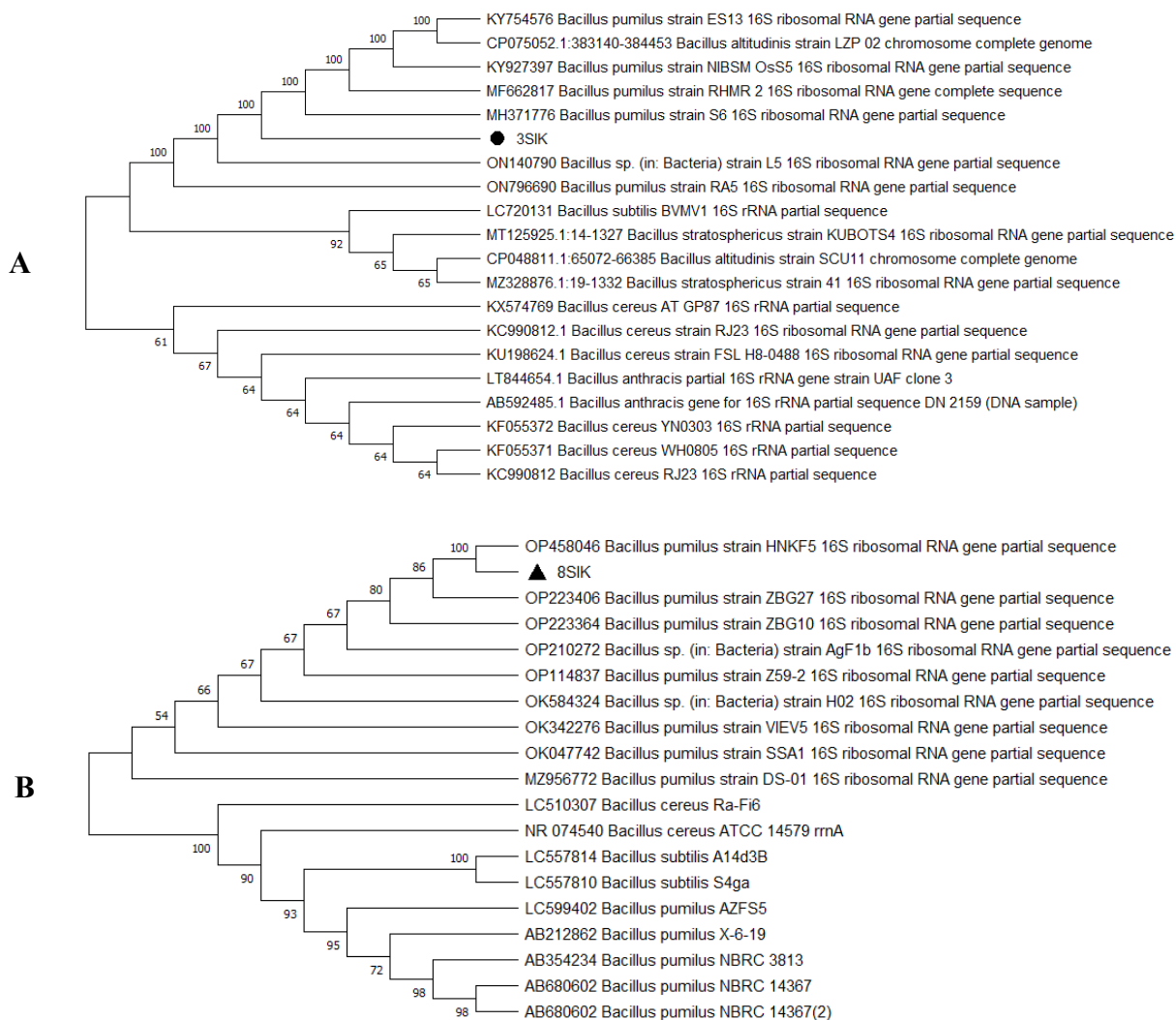


Figure 12: Phylogenetic tree based on 16s RNA gene fragment. The Neighbor-Joining technique was used to generate the tree. (A) In this study, the black triangles and dot represent soil isolates. Isolate 1 (3SIK) with the 19 DNA sequences randomly chosen from the GenBank. (B) Isolate 2 (8SIK) with 18 DNA sequences chosen at random from GenBank and assigned accession numbers. The analyses of evolution were conducted using MEGA X (10.0.5) d (Tamura et al., 2021)

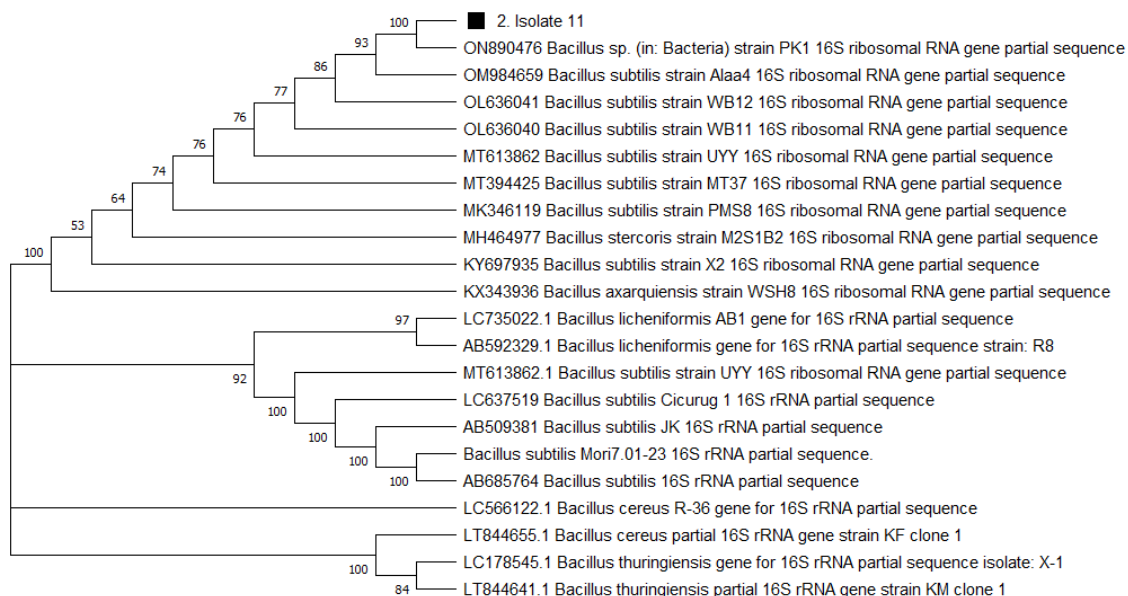


Figure 13: Phylogenetic tree based on the 16s RNA gene fragment. Evolutionary relationship inferred using the Neighbour-Joining method. In this study, the soil isolate is represented by a squared black dot. The tree was built using a total of 21 DNA sequences taken at random from GenBank and labelled with their accession number. The analyses of evolution were conducted using MEGA X (10.0.5) (Tamura et al., 2021).

3.5 Discussion

With the increasing number of antibiotic-resistant pathogens, it is becoming increasingly important to discover new antibacterial agents that are both highly active and have novel and diverse mechanisms of action. However, studies have been conducted from different sources in nature for instance marine, terrestrial organisms and medicinal plants to combat the tide of antibiotic resistance (Watts et al., 2017, Doss et al., 2017, Hayashi et al., 2013). Additionally, natural products from microbial origin are becoming increasingly important over the last few decades, as they are more renewable, and maybe reproducible source than plants or animals (Lam, 2007). Hence, the current study aimed at isolation and characterization of microorganisms with the potential to produce antimicrobial compounds from soil from various agricultural settings.

The soil bacteria isolates obtained and evaluated for antimicrobial production against known pathogenic organisms were shown to exhibit substantial activity against the pathogenic test organism in the current study. As shown in Fig. 11, these bacterial isolates have demonstrated a broader spectrum of activity against various pathogens, with isolate 3 having the greatest inhibitory impact against *Staphylococcus aureus* ATCC (25923), whereas isolate 1 had the least inhibitory effect against *Escherichia coli* ATCC (35218). This inhibition might be due to the presence of different secondary metabolites that have antimicrobial properties, the production of specific antimicrobial compounds, or differences in the mode of action. Nonetheless, our findings agree with the findings by Sethi et al., 2013, who also isolated *Bacillus subtilis* from soil which also showed to be active against some gram-positive including *Staphylococcus aureus* under Standard culturing conditions.

The 16S rRNA sequencing plays a crucial role in advancing our understanding of microbial diversity, evolution, and the discovery of novel bacteria mostly in cases of unusual phenotypic profiles, uncultivable or rare bacteria (Woo et al., 2008). In the present study, 16s RNA was performed to understand the evolutionary relationship and relatedness of different bacterial isolates. The identification of the three isolates was established by comparing them to the existing data in GenBank (NCBI) and cross-referencing with highly scored rRNA gene sequences in BLAST searches. The analysis of the amplified 16S rRNA gene through phylogenetics demonstrated that the two active isolates, henceforth referred to as isolate 1

(3SIK) and isolate 2 (8SIK), exhibit sequence homology with *Bacillus pumilus* (as illustrated in Figure 12). In contrast, isolate 3 displayed significant sequence homology with *Bacillus subtilis* (as depicted in Figure 13). These findings align with numerous prior studies indicating that *Bacillus* species are widespread and renowned for their antimicrobial production (Khalid et al 2016).

3.6 Conclusion

This study demonstrated that the use of sprinkling technique approach is a useful approach in the field of bio-guided bioprospection of microorganisms and has proved to be an effective and feasible towards unearthing potent antimicrobial compounds. Our results demonstrated that the approach accelerates the discovery of bioactive compounds and isolation of strains which offer a promising avenue for discovering new clinical antimicrobials. The antibacterial test results indicated that one of the three bacterial isolates demonstrated consistent inhibitory activity against both Gram-negative (*E. coli*) and Gram-positive (*S. aureus*) pathogenic bacteria. Amongst all the isolated bacteria, *Bacillus Subtilis* was found to be more potent. Therefore, in future bioreactors designed to cultivate these bacteria with stress inducers can be used to elevate the levels of these compounds. Nonetheless, the outstanding activity observed from these two *Bacillus* species, *Bacillus pumilus* and *Bacillus subtilis* motivate the search for new antimicrobial metabolites from these two species. Additionally, these could be a useful starting point for the discovery of clinically significant and new antimicrobial agents in the future.

3.7 References

- ABORODE, A. T., AWUAH, W. A., MIKHAILOVA, T., ABDUL-RAHMAN, T., PAVLOCK, S., KUNDU, M., YARLAGADDA, R., PUSTAKE, M., DA SILVA CORREIA, I. F. & MEHMOOD, Q. 2022. OMICs Technologies for Natural Compounds-based Drug Development. *Current topics in medicinal chemistry*, 22, 1751-1765.
- ATANASOV, A. G., ZOTCHEV, S. B., DIRSCH, V. M. & SUPURAN, C. T. 2021. Natural products in drug discovery: advances and opportunities. *Nature reviews Drug discovery*, 20, 200-216.
- BENGTSSON-PALME, J., KRISTIANSSON, E. & LARSSON, D. J. 2018. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS microbiology reviews*, 42, fux053.
- BERTHOD, A., RUIZ-ANGEL, M. & CARDA-BROCH, S. 2009. Countercurrent chromatography: People and applications. *Journal of Chromatography A*, 1216, 4206-4217.
- CUSHNIE, T. T., CUSHNIE, B., ECHEVERRÍA, J., FOWSANTEAR, W., THAMMAWAT, S., DODGSON, J. L., LAW, S. & CLOW, S. M. 2020. Bioprospecting for antibacterial drugs: a multidisciplinary perspective on natural product source material, bioassay selection and avoidable pitfalls. *Pharmaceutical research*, 37, 125.
- DOSS, J., CULBERTSON, K., HAHN, D., CAMACHO, J. & BAREKZI, N. 2017. A review of phage therapy against bacterial pathogens of aquatic and terrestrial organisms. *Viruses*, 9, 50.
- FENG, W., LI, M., HAO, Z. & ZHANG, J. 2019. Analytical methods of isolation and identification. *Phytochemicals in human health*. IntechOpen.
- GRANADOS-CHINCHILLA, F. & RODRÍGUEZ, C. 2017. Tetracyclines in food and feedingstuffs: from regulation to analytical methods, bacterial resistance, and environmental and health implications. *Journal of analytical methods in chemistry*, 2017, 1315497.
- HAYASHI, M. A., BIZERRA, F. C. & DA SILVA JUNIOR, P. I. 2013. Antimicrobial compounds from natural sources. *Frontiers in microbiology*, 4, 195.
- HOSKISSON, P. A. & SEIPKE, R. F. 2020. Cryptic or silent? The known unknowns, unknown knowns, and unknown unknowns of secondary metabolism. *MBio*, 11, 10.1128/mbio.02642-20.

- LAM, K. S. 2007. New aspects of natural products in drug discovery. *Trends in microbiology*, 15, 279-289.
- MACHADO, H., SONNENSCHN, E. C., MELCHIORSEN, J. & GRAM, L. 2015. Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. *BMC genomics*, 16, 1-12.
- MAGHEMBE, R., DAMIAN, D., MAKARANGA, A., NYANDORO, S. S., LYANTAGAYE, S. L., KUSARI, S. & HATTI-KAUL, R. 2020. Omics for bioprospecting and drug discovery from bacteria and microalgae. *Antibiotics*, 9, 229.
- MHLONGO, M. I., PIATER, L. A., MADALA, N. E., LABUSCHAGNE, N. & DUBERY, I. A. 2018. The chemistry of plant-microbe interactions in the rhizosphere and the potential for metabolomics to reveal signaling related to defense priming and induced systemic resistance. *Frontiers in plant science*, 9, 112.
- MOHANA, N. C., RAO, H. Y., RAKSHITH, D., MITHUN, P., NUTHAN, B. & SATISH, S. 2018. Omics based approach for biodiscovery of microbial natural products in antibiotic resistance era. *Journal of Genetic Engineering and Biotechnology*, 16, 1-8.
- NANNIPIERI, P., ASCHER, J., CECCHERINI, M. T., LANDI, L., PIETRAMELLARA, G. & RENELLA, G. 2003. The links between microbial diversity and soil functions. *Innovative Soil-Plant Systems for Sustainable Agricultural Practices*. Organisation for Economic Cooperation and Development (OECD).
- OBERPAUL, M., BRINKMANN, S., SPOHN, M. S., MIHAJLOVIC, S., MARNER, M., PATRAS, M. A., TOTI, L., KURZ, M., HAMMANN, P. E. & VILCINSKAS, A. 2022. Trichoderma-Derived Pentapeptides from the Infected Nest Mycobiome of the Subterranean Termite *Coptotermes testaceus*. *ChemBioChem*, 23, e202100698.
- PIERCE, E. C. & DUTTON, R. J. 2022. Putting microbial interactions back into community contexts. *Current Opinion in Microbiology*, 65, 56-63.
- PODOLSKY, S. H. 2018. The evolving response to antibiotic resistance (1945–2018). *Palgrave Communications*, 4.
- SASIDHARAN, S., CHEN, Y., SARAVANAN, D., SUNDRAM, K. & LATHA, L. Y. 2011. Extraction, isolation and characterization of bioactive compounds from plants' extracts. *African journal of traditional, complementary and alternative medicines*, 8.
- SCHULTZ, J., MODOLON, F., PEIXOTO, R. S. & ROSADO, A. S. 2023. Shedding light on the composition of extreme microbial dark matter: alternative approaches for culturing extremophiles. *Frontiers in Microbiology*, 14, 1167718.

- SHRIVASTAVA, S. R., SHRIVASTAVA, P. S. & RAMASAMY, J. 2018. Responding to the challenge of antibiotic resistance: World Health Organization. *Journal of Research in Medical Sciences*, 23, 21.
- TAMURA, K., STECHER, G. & KUMAR, S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38, 3022-3027.
- THAKUR, K., CHOWNK, M., KUMAR, V., PUROHIT, A., VASHISHT, A., KUMAR, V. & YADAV, S. K. 2020. Bioprospecting potential of microbial communities in solid waste landfills for novel enzymes through metagenomic approach. *World Journal of Microbiology and Biotechnology*, 36, 34.
- WATTS, J. E., SCHREIER, H. J., LANSKA, L. & HALE, M. S. 2017. The rising tide of antimicrobial resistance in aquaculture: sources, sinks and solutions. *Marine drugs*, 15, 158.
- WOO, P., LAU, S., TENG, J., TSE, H. & YUEN, K.-Y. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clinical Microbiology and Infection*, 14, 908-934.

Chapter 4: Assessing the metabolic landscape of soil-borne *Bacillus* spp. with an aid of UPLC-qTOF-MS and computational metabolomics

4.1 Summary

Despite the evidence that secondary metabolites from *Bacillus* spp. are generally used as bioproducts for agriculture and able to produce antimicrobial, anticancer compounds, pigments, and enzymes. We still have a relative few studies that focus on exploring the chemical space of this bacterial strains. Therefore, in this chapter, our aim was to assess the metabolic landscape of the potential antibiotic-producing strains and to highlight the chemical space of the responsible compounds through the computational metabolomics-based bioprospecting approach. Herein, methanolic extract from three bacterial isolates were subjected to UHPLC-qTOF-MS and, the resultants were analyzed using an integrative approach that relied on GNPS Molecular Networking and *in silico* dereplication tools. Molecular networking revealed the presence of Diketopiperazine compounds such as cyclo (D-Pro-D-Leu), cyclo (L-Tyr-L-Pro), cyclo (L-Pro-D-Phe), and cyclo (L-Pro-L-Val), alongside Surfactin C, Surfactin B, Pumilacidin E, and Isarrin D in the *Bacillus* strains as the main antimicrobial compounds. The application of the molecular networking approach represents an innovation in the field of bio-guided bioprospection of microorganisms and has proved to be an effective and feasible towards unearthing potent antimicrobial compounds. Additionally, the (computational metabolomics-based) approach accelerates the discovery of bioactive compounds and isolation of strains which offer a promising avenue for discovering new clinical antimicrobials.

KEYWORDS: *Bacillus*, Diketopiperazine, LC-QTOF-MS, Molecular networking, Surfactin.

4.2 Introduction

Bacillus species represent one of the most diverse genera within the class *Bacilli*. This genus encompasses aerobic and facultatively anaerobic, rod-shaped, gram-positive bacteria that are capable of forming spores, with guanine-cytosine (G+C) content ranging from 32% to 69% (Rabbee et al., 2019). *Bacillus* species are known as proficient producers of a variety of antibiotics (Ferreira et al., 2021), and tough and resistant endospores (Nicholson et al., 2000). This capability has established them as important biological control agents (Fira et al., 2018), effective in managing a broad spectrum of pathogens and microorganisms (including phytopathogens) (Shafi et al., 2017). These bacteria employ a range of competitive and antagonistic mechanisms to suppress co-existing species. However, despite the evidence that secondary metabolites are generally used as bioproducts in agriculture and medicine, mainly as antimicrobial (Mishra and Arora, 2018), anticancer compounds (Wadhwa et al., 2024), antimalarial (Khushboo et al., 2022), and antifungal (Rodrigo et al., 2021). We still have a relative few studies that focus on exploring the chemical space of this bacterial strains.

Studies on *Bacillus* species thus far have explored and reported their potential in producing compounds that can be used as antimicrobial (Khwathisi et al., 2024), antioxidant (Shahniani et al., 2024), and cytotoxicity (Abdel-Nasser et al., 2024) against various diseases of choice and interest to the authors. For other *Bacillus* spp. has focused on certain treatments such as cancer, antibacterial and most recently covid (Alam et al., 2022), to name but a few. However, most of these reports point to a rich and complex chemical space of the *Bacillus* spp., which is still poorly understood and explored (Tsotetsi et al., 2022). Microbial chemical space is indeed chemically complex and there may be synergistic effects of not only one compound across targets but also multiple compounds across multiple targets. This often leads to many compounds or extracts being identified as bioactive but with little explanation of mode of action.

For the longest time, the discovery of bioactive compounds and new natural product has been limited by the lack of high throughput analytical techniques (Zhu et al., 2010, Potterat and Hamburger, 2013). Recently, the emergence of advance techniques such as Liquid chromatography hyphenated to quadruple time of flight mass spectrometry (LC-qTOF-MS) has presented a promising avenue in the field of drug discovery particularly from natural products (Kumar and Madhusudanan, 2020, Nassar and Talaat, 2004). These advanced techniques enable comprehensive metabolite profiling, high-throughput screening, and the

integration of genomics data, facilitating the discovery of previously unknown antimicrobial compounds (Mohana et al., 2018, Maghembe et al., 2020, Aborode et al., 2022). Semantically, metabolomics is a challenging discipline to define due to the heterogeneity and dimensionality of the chemical space that it explores, but it can be described here as a multidisciplinary *omics* science that aims to define the entire complement of small molecular weight molecules, within a biological matrix of interest.

Moreover, the resulting data from the advanced instruments is also complex and impossible to analyze with traditional statistical methods. As such, multivariate statistical data models are used to decipher the inherited biological properties of these bioactive compounds by allowing an easy and computerized means of compound identification (Tugizimana et al., 2013). Additionally, an alternative approach to decipher the chemical data acquired through these analytical instruments, particularly mass spectrometry, is molecular networking (Aron et al., 2020) which enables the grouping of metabolites based on their similar fragmentation patterns owing to common structural moieties (Ramabulana et al., 2021). Molecular networking allows for easier visualization of complex data and identification of unknown compounds through their association with known chemical entities within a given sample (Clements et al., 2021). This technique has been positively applied to isolate anti-microbial compounds from plant growth-promoting bacterial strains (Nephali et al., 2022).

Herein, we present a computational metabolomics approach to comprehensively characterize the metabolic landscape of soil-borne *Bacillus* spp. Methanolic. In this chapter, methanolic extracts of bacteria strains isolated from agricultural soils were analyzed using a combination of UPLC-QTOF-MS and molecular networking approach to identify possible anti-microbial compounds thereof. The findings of the current study highlight the use of (computational metabolomics tool) molecular networking approach as a feasible approach for the identification of bacterial strains capable of producing anti-microbial compounds which may be for use in the management of drug-resistant pathogens in the future. Moreover, the molecular network approach allowed for a quick and easy-to-visualize approach for complex LC-MS/MS data.

4.3 Experimental

4.3.1 Metabolite extraction

Pure bacterial isolates cultured in Mueller Hinton Broth for 24 hours was used for bacterial cell harvesting. Erlenmeyer flasks were used to culture the bacterial isolates in triplicates and after 24 hours bacterial cells were harvested through centrifugation for 15 min at 5000 rpm at 4 °C, resulting in supernatants and pellets which were transferred into new conical tubes (50 mL). The damp weight of pellets were measured prior to extraction of the intracellular and extracellular metabolites by adding 80 % LC-Grade methanol (Romil SpS, Cambridge, UK). Consequently, the methanolic extract of the bacterial isolate were subjected to overnight agitation on a digital tube rotator at constant velocity maintained at 70 rpm to aid extraction. Centrifugation of the extracts was carried out for 15min at 5100 rpm at 4 °C. Therefore, the samples were filtered using a 0.22 µm nylon filters into LC-MS glass vials as described elsewhere (Ramabulana et al., 2021).

4.3.2 Liquid Chromatography Mass Spectrometry analysis

Bacterial methanolic extracts analysis was performed using a liquid chromatography-quadrupole time-of-flight tandem mass spectrometer (LC-MS-9030 q-TOF, Shimadzu Corporation, Kyoto, Japan) fitted with a Shim-pack Velox C18 column (100 mm x 2.1 mm with particle size of 2.7 µm). The column oven temperature was maintained at 50 °C. The injection volume was 5 µL, and the samples were analytically separated over a 30 min binary gradient. The flow rate was kept constant at 0.3 mL/min in the current study using a binary solvent mixture of water with 0.1% formic acid (Eluent A) and methanol with 0.1% formic acid (Eluent B). The gradient technique was gradually increased from 3 to 30 minutes to facilitate the separation of the compounds within the samples. Briefly, Eluent B was kept at 5% from 0 to 3 minutes, gradually increased from 5 to 40% between 3 and 8 minutes, and finally increased to 40- 95% between 8 and 23 minutes. Eluent B was then kept isocratic at 95% between 23 and 25 minutes. The gradient was returned to original conditions of 5% at 25 - 27 minutes, and re-equilibration at 5% occurred at 27 - 30 minutes. The liquid chromatographic eluents were then subjected to a Quadruple Time-of-Flight high-definition

mass spectrometer for analysis in positive ESI mode. The q-TOF-MS conditions were as follows: 400 °C heat block temperature, 250 °C Desolvation Line (DL) temperature, 42 °C flight tube temperature, and 3 L/min nebulization and dry gas flow. The data was acquired using the Data-dependent acquisition (DDA) mode, which simultaneously generated MS1 and MS2 data for all ions within a mass-to-charge ratio (m/z) range of 100–2000 (precursor m/z isolation window) and an intensity threshold above 5000. The MS2 Experiments were conducted out utilizing argon gas as the collision gas and a collision energy of 30 eV with a spread of 5 and sodium iodide (NaI) as a calibration solution to monitor high mass precision.

4.3.3 Metabolite annotation and Molecular Networking

Molecular networks were generated on the Global Natural Product Social Molecular Networking (GNPS) platform (<http://gnps.ucsd.edu>). Prior to uploading to the online workflow, LC-qTOF-MS/MS raw data was converted to an open-source format (.mzML). Once spectral data was uploaded, it was processed by filtering out all MS2 fragment ions that were within ± 17 Da of the precursor m/z . Only the top 4 fragment ions in the ± 50 Da window were selected in order to window filter MS/MS spectra. An MS-CLUSTER technique was used to cluster the data. The MS2 fragment ion and a mass tolerance for precursor ion were both set at 0.05 Da. Only when more than six matching peaks were present and 0.7 minimum cosine score was exceeded did edges (interconnections between metabolites/similarity) develop. In the network, only nodes that were each other's top 10 similar nodes were kept, which had a TopK value of 10. The maximum number of nodes that may be joined into a single molecular family was limited to 100. The spectral extracted from the mzML file was used to search through the GNPS in-house databases including Mass Spectral database (MassBank), Human metabolome Database (HMDB) and National Institute of Standards and technology (NIST). Cytoscape software was used to visualize the generated molecular networks from GNPS (Aron et al., 2020, Wang et al., 2016). Molecular networks were extended by computing in silico annotation through network annotation propagation (NAP), which utilizes molecular network topology, predicts spectra fragmentation in silico, gets, and rates candidate structures. Substructure annotation was conducted using MS2 latent Dirichlet allocation (MS2LDA) within the GNPS interface. The outputs from molecular networking (MN), network annotation propagation (NAP) and MS2 latent Dirichlet allocation (MS2LDA) was combined by MolNetEnhancer which performed chemical classification. All matched and some mismatched nodes were confirmed or annotated based on empirical formulae derived from accurate mass

and fragmentation patterns collected from MS2 studies. These annotated metabolites were also compared to available literature and searched against common dereplication databases for natural product such as KNApSAcK (http://www.knapsackfamily.com/knapsack_core/top.php)(accessed on 27th March 2022) and Dictionary of Natural Products (<http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>)(Accessed 8th February 2021). Metabolite annotation was completed at Metabolomics Standards Initiative (MSI) level 2 (Sumner et al., 2007).

4.4 Results

4.4.1 UPLC-QTOF-MS Analyses

A holistic untargeted metabolome analysis of antimicrobial producing bacteria was conducted. Bacterial cultures were grown on Mueller Hinton broth 24 hours. To determine whether these bioactive metabolites in the crude extracts feature some unique chemical diversity, the methanol extracts of the strains were analyzed through ultra-high liquid chromatography quadruple time of flight mass spectrometry (UPLC-QTOF-MS) operating in positive ESI-mode. The spectrum data was acquired using the Data-dependent acquisition (DDA) method, which creates MS2 fragmentation data for every precursor ion over a predetermined threshold. (Ramabulana et al 2023). **Supplementary Fig. 17** illustrates the base peak chromatogram of the samples indicating some similarities and differences in the metabolite profiles of the isolated *Bacillus* strains. Furthermore, these isolates showed to produce more extracellular than intracellular metabolites (**Supplementary Fig. 17**) and **Supplementary Fig. 17** shows that extracellular metabolites are more hydrophilic, whereas intracellular metabolites have a scarcity of hydrophilic compounds. To further investigate the chemical diversity of the bacterial isolates, the spectral data were analyzed through a publicly available database via the classical molecular networking workflow on the GNPS platform, enabling a broad overview of the chemical information deduced from the MS/MS data and the generated molecular networks were visualized in Cytoscape (**Supplementary Fig. 20**) (Doncheva et al., 2018). The computed classical molecular network resulted in 1325 nodes, with 831 grouped into 22 independent molecular families (with at least two nodes connected by an edge) based on GNPS spectral matching. Insights into the chemical identities of the bacterial species were revealed by the computed Molecular Network (MN), which revealed 7 of the nodes that were putatively annotated using an automatic library spectral matching against the public spectral libraries

within the GNPS platform. It was also evident that not every bioactive metabolites (Lipopeptides) were clustering in the same cluster (**Figure 14**), simply indicating that there are some significant differences in the MS/MS fragmentation of these metabolites. Additionally, these isolates produce a diverse range of unidentified compounds which could be vital for the antimicrobial activities observed against test organisms and beyond (**Figure 15**).

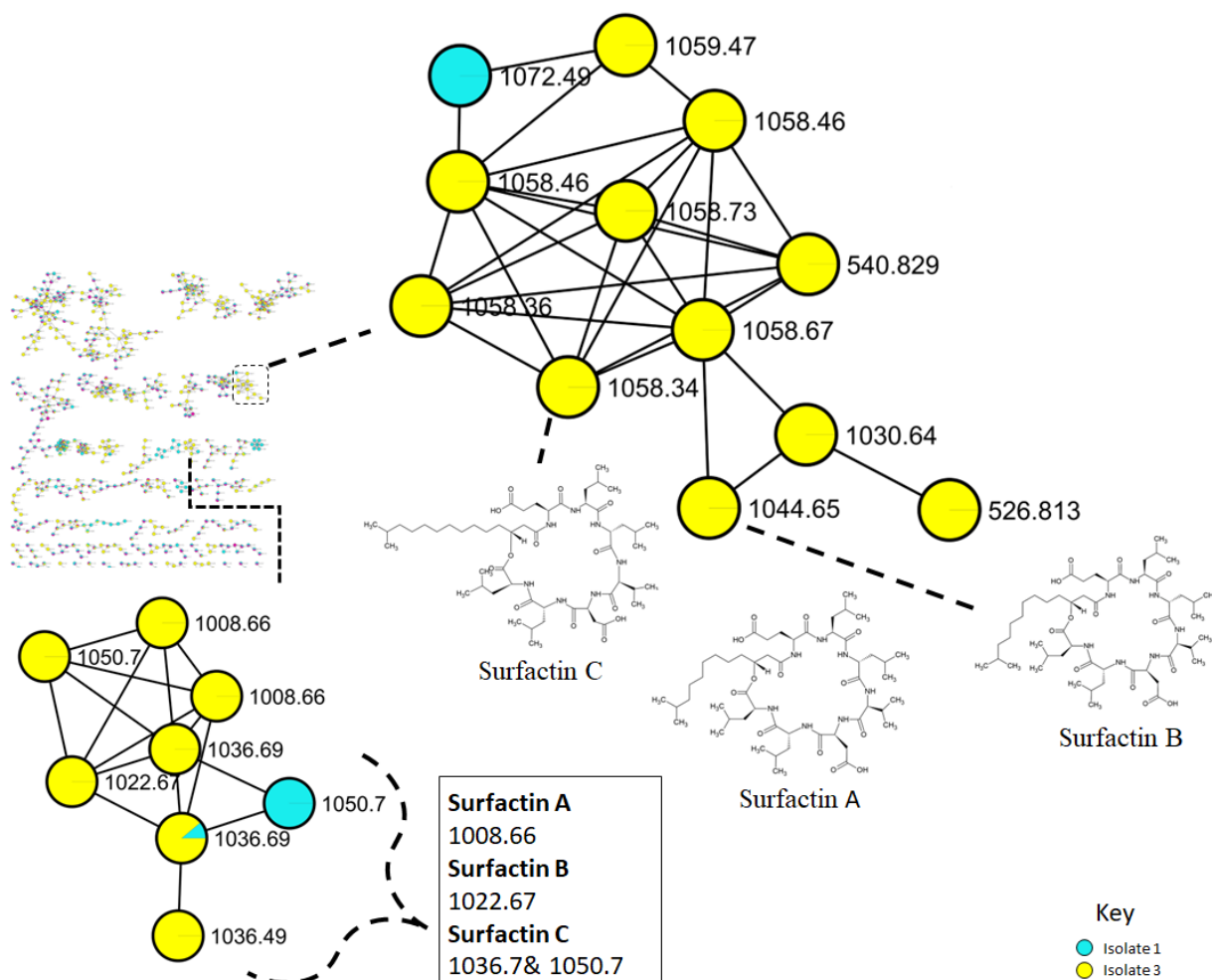


Figure 14: Molecular network of *Bacillus* species extracts analysed through liquid chromatography-tandem mass spectrometry using electrospray ionisation (ESI) in Positive mode. Surfactins (Lipopeptides) clustering in different molecular families. Node colours represent the studied *Bacillus* species and the nodes in the network represent distinct molecular ions detected in the samples, while the edge connecting the node indicate spectral similarity between the compounds.

The metabolome covered in this study included both metabolites that were unknown and classes that were known or putatively annotated. With the aid of various databases, the data

analysis resulted in the annotation of several compounds, these include Lipids, organic acid, alkaloids and derivatives, organoheterocyclic, benzenoids, and polyketides (Supplementary Fig 18). As described in the methodology, additional confirmatory scrutiny such as through NAP, MS2LDA and manually validated through MS-based accurate mass. **Figure 15** shows the metabolites in their different clusters based on their structural similarities. The surfactin molecular family distinguished itself for exhibiting ions found only in the *Bacillus subtilis* species. In this cluster surfactins were annotated and nodes representing the ions at m/z 526.813, 540.829 and 1072.49 (**Figure 15**) were not annotated, indicating similarity of spectra between the aforementioned spectral nodes and the surfactins family. Furthermore, Mass spectral datasets from all three bacterial strains, in silico annotation tools like dereplicator, MS2LDA and the network annotation propagation (NAP) were used to further investigate the chemical spaces of the various isolates.

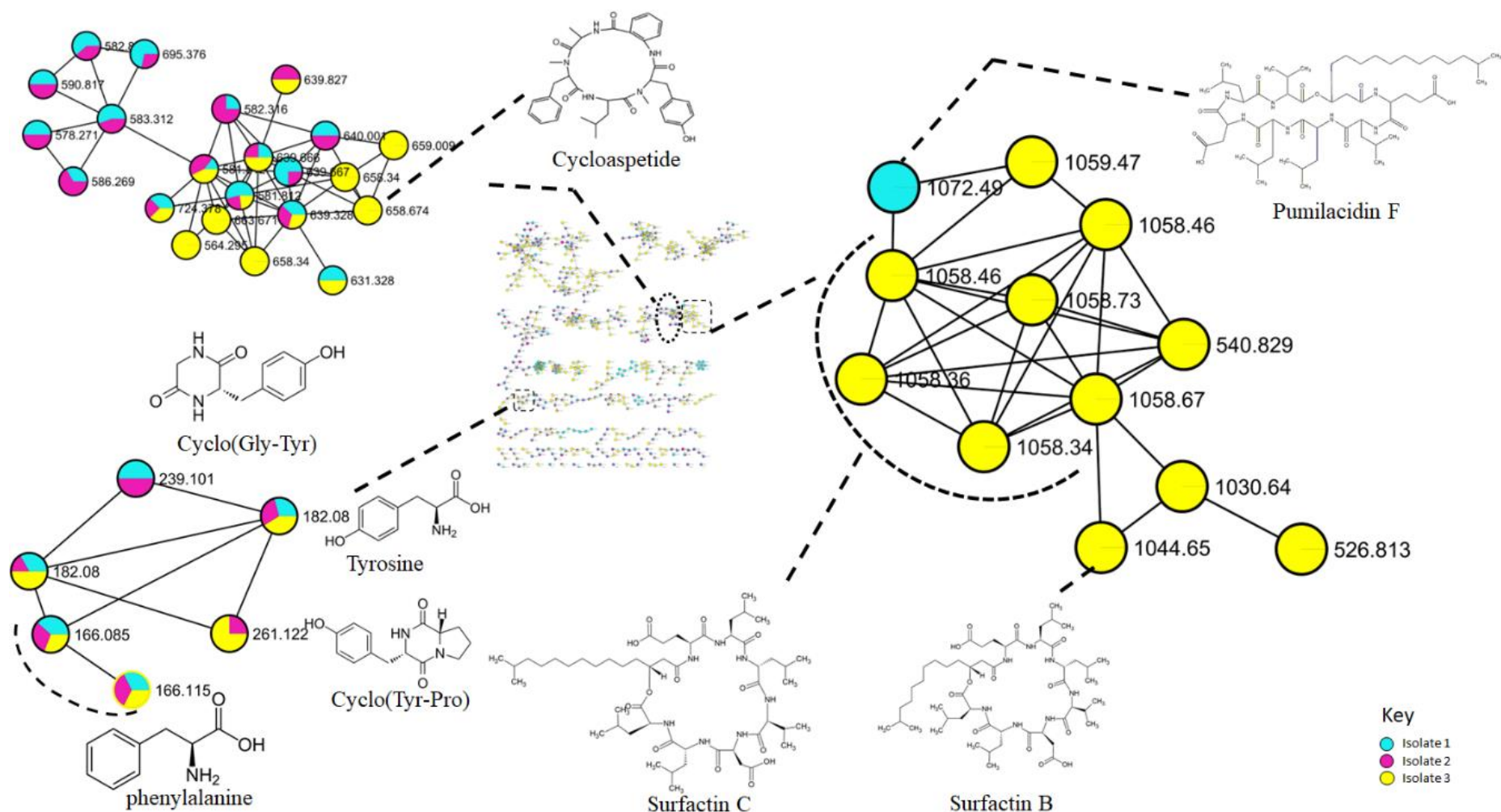
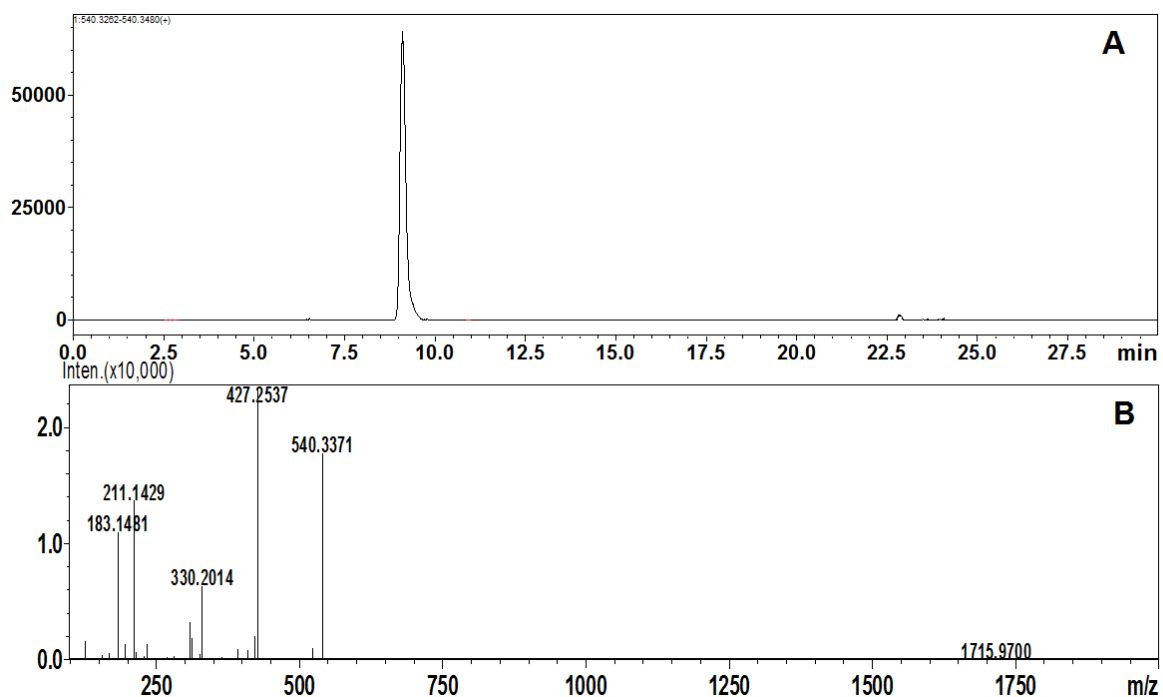


Figure 15: Molecular network of *Bacillus* species extracts analysed through liquid chromatography-tandem mass spectrometry using electrospray ionisation in Positive mode. Lipopeptides (right) and the Cyclodipeptides (CDP) molecular family (left). Node colours represent the studied *Bacillus* species and the nodes in the network represent distinct molecular ions detected in the samples, while the edge connecting the node indicate spectral similarity between the compounds.



Supplementary Figure 16: Representative UHPLC-qTOF-MS/MS chromatogram showing single ion chromatogram of putatively identified Isarrin D (A) at m/z 540, with Y-axis showing peak intensity and X- axis showing retention time. Fragmentation spectra of the metabolite putatively identified, precursor at m/z 540 (B), with product ions at m/z 427, 330, 211 and at m/z 183.

4.5 Discussion

Molecular and UPLC-QTOF-MS were used to analyze the diversity and classification of the metabolites produced by these bacterial isolates. Notably, these isolates produced a higher quantity of extracellular metabolites compared to intracellular metabolites (see Supplementary Fig. 17). Moreover, to further investigate the chemical diversity of the bacterial isolates, the spectral data were analyzed through a publicly available database *via* the classical molecular networking workflow on the GNPS platform, enabling a broad overview of the chemical information deduced from the MS/MS data and the generated molecular networks were visualized in Cytoscape (Supplementary Fig. 20) (Doncheva et al., 2018). As a result, this information was employed to delve deeper into the identities of these nodes generated.

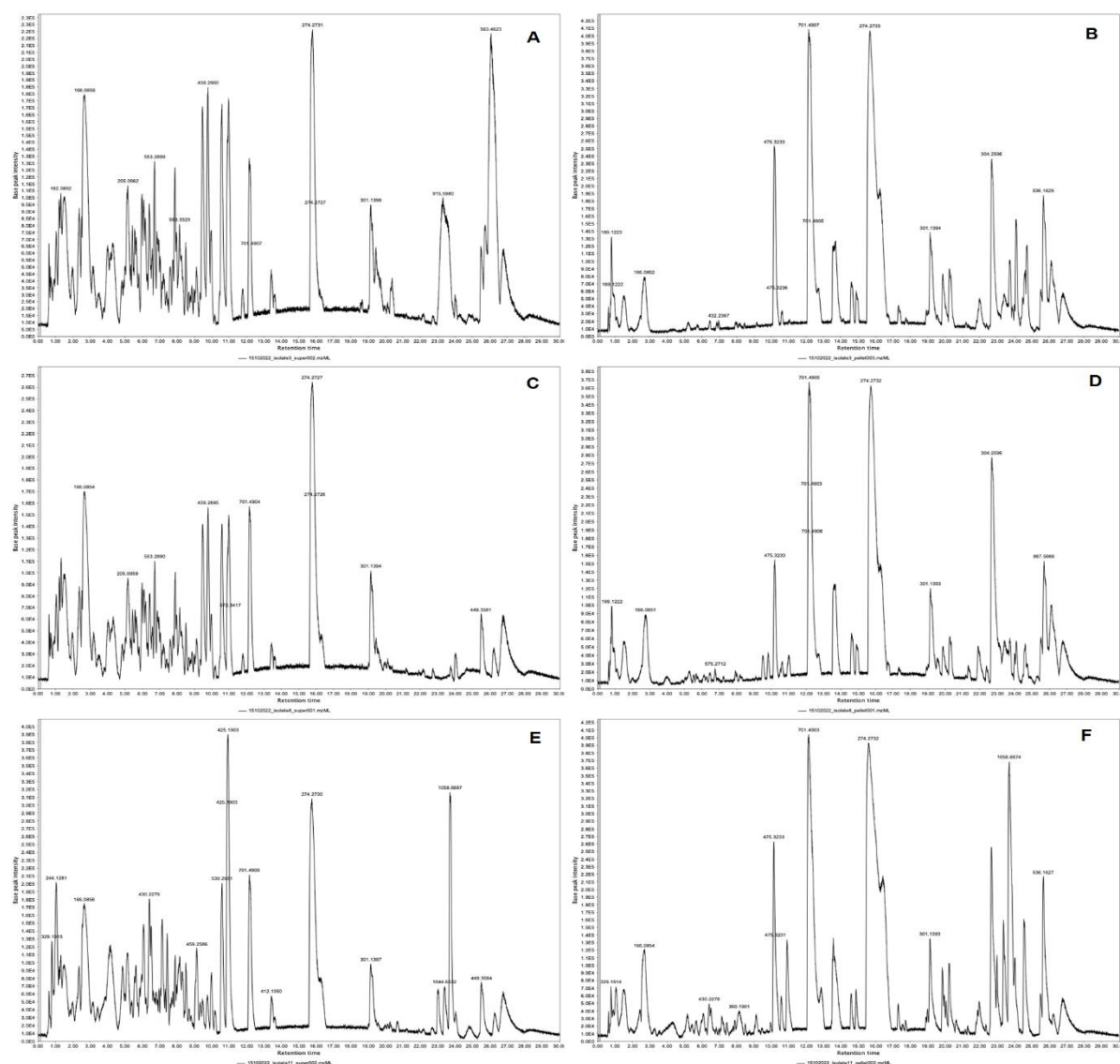
Molecular networking not only helps in annotating known compounds but also facilitates the discovery of novel metabolites. As such, the annotation of surfactins in the present study facilitated with the identification or annotation of other nodes which were previously unidentified features at m/z 1072.49 which was in the same cluster with Surfactin, and it was putatively identified as Pumilacidin F (Naruse et al., 1990). This lipopeptide consists of a cyclic ring connected to a fatty acid chain and has been discovered to possess strong antimicrobial characteristics (Jemil et al., 2017). Thus far, several studies have demonstrated the antibacterial (Saggese et al., 2018), antifungal (de Oliveira et al., 2020), antiviral (Singh and Cameotra, 2004), antimotility activities (Xiu et al., 2017) of Pumilacidin. Saggese et al. (2018) also showed that Pumilacidin has antimicrobial activities against *Staphylococcus aureus* and *Listeria monocytogenes*. Similarly, in the present study, isolate 1 showed to have convincing activity against *S. aureus*. Notably, in the cluster Pumilacidin F was shown to be unique to *Bacillus pumilus*, which therefore gives confidence in the annotation of the compound. Moreover, the findings from the present study suggest that the unannotated nodes within these clusters are related to Surfactins structurally, which can aid in the assignment of names to additional unannotated nodes. In addition, a combination of molecular network and manual confirmatory scrutiny, led us to identify another cyclodipeptide known as Isarrin D at m/z 540 $[M + H]^+$, a cyclodipeptides from the isaridine family. It showed a fragment ion at m/z 427, m/z 330, m/z 211 and m/z 183. Isariins derivatives have been shown to have insecticidal (Langenfeld et al., 2011), antiplasmodial (Lemmens-Gruber et al., 2009) and anti-inflammatory activities (Wang et al., 2018).

The outcomes of high-throughput molecular networking did not match the extensive GNPS repository dataset in our study, giving an indication of the occurrence of new class of compounds among active isolates. It was also evident that not every bioactive metabolites (Lipopeptides) were clustering in the same cluster, suggesting that there are some significant differences in the MS/MS fragmentation which therefore suggest the presence of different type of metabolites in the extracts. These isolates produce a diverse range of unidentified compounds which could be vital for the antimicrobial activities observed against test organisms and beyond. This does, however, draw attention to one of the most important drawbacks of untargeted metabolomics, as many metabolites are still unknown or poorly characterized, particularly in microbial metabolomics (Vinaixa et al., 2016). These unknown metabolites are known as dark matter in metabolomics and efforts to generate algorithms such as those aimed at establishing the identity of these metabolites and the identity of the microbes that makes them are being made (Wang et al., 2020).

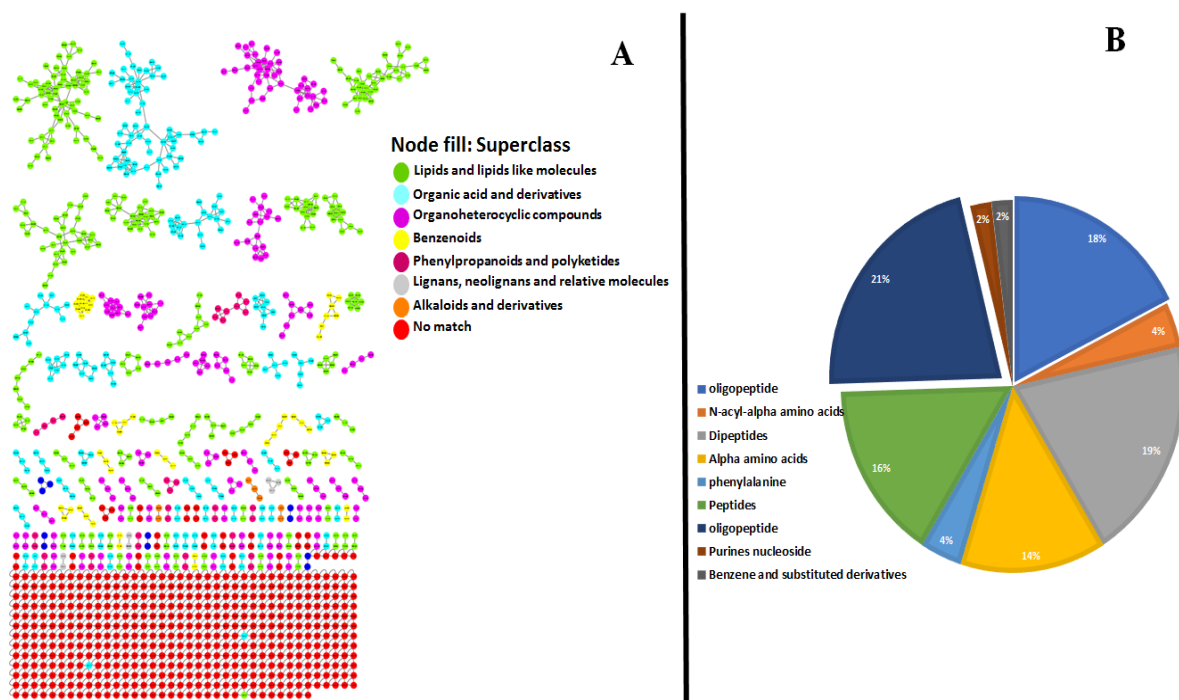
The recent development in the GNPS ecosystems has transformed the discovery of metabolites in high-throughput technology. The contribution of this *in silico* tools has enabled us to identify several cyclodipeptides such as Cyclo (Tyr-pro), Cyclo (Leu -pro), Cyclo (Pro-Val) and Cyclo (Pro-Phe). These cyclodipeptides compounds have relevance in quorum sensing (González et al., 2017), cell-cell signaling or as drug delivery systems (Bojarska and Wolf, 2020). However, the putative presence of these known bioactive compounds suggests that some of the antimicrobial activity might be attributed to these compounds. Overall, we were able to putatively identify several metabolites (Supplementary Table 1) from *Bacillus* isolates using Molecular Networking approach. However, isolate 3 had the greatest level of activity out of the three isolates, thus it can therefore be suggested that based on the results observed from the molecular network, the combination of the cyclopeptides and lipopeptides (Surfactins) together can confer an enhanced antimicrobial activity, and given that isolates 3 and 8 produce these cyclodipeptide molecules instead of the lipopeptides. The other bacteria can withstand the effects of cyclodipeptides, as compared against the most potent surfactins. Therefore, these are resourceful and easy-to-find bacteria that has the potential to serve as a valuable source of antimicrobial compounds. As a consequence, future bioreactors can be designed to cultivate these bacteria to produce high titers of the antimicrobial compounds. Moreover, stress inducer can be used to elevate the levels of this compounds.

4.6 Conclusion

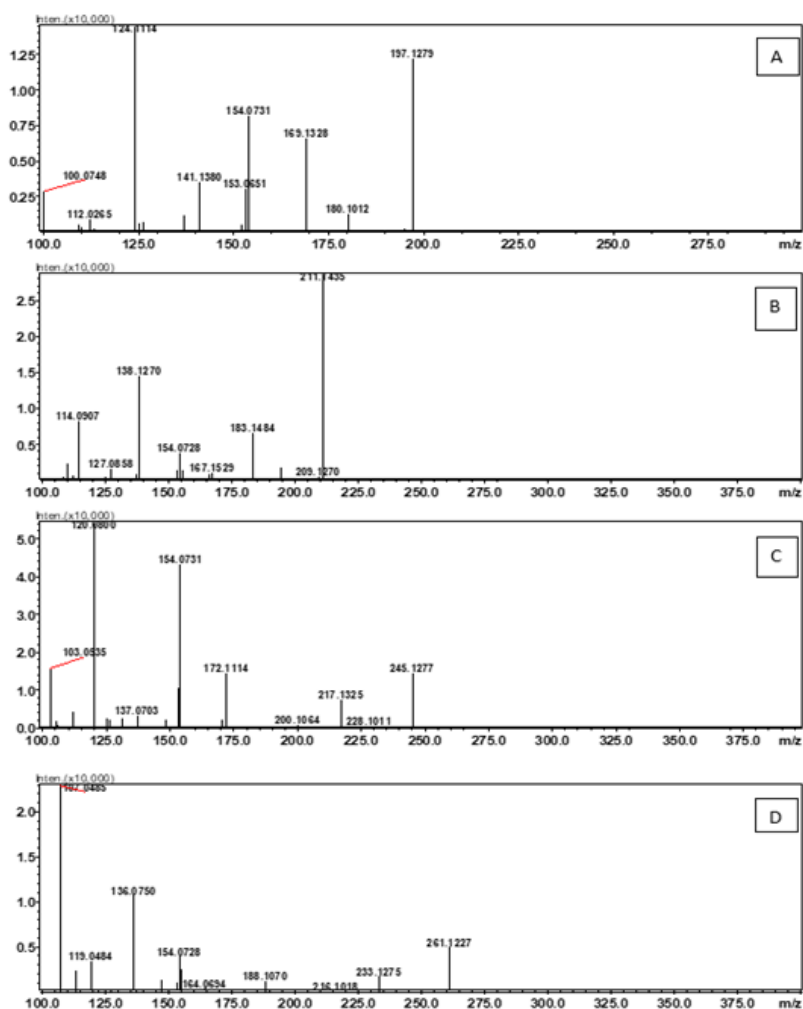
The present study shows that the use of molecular networking approach (computational metabolomics tools) in the field of bio-guided bioprospection of microorganisms and has proved to be an effective and feasible towards unearthing potent antimicrobial compounds. Our results demonstrated that the approach accelerates the discovery of bioactive compounds and isolation of strains which offer a promising avenue for discovering new clinical antimicrobials. However, molecular networks generated from the methanol extracts enabled annotation of various metabolites, the existence of abundance of novel compounds which couldn't be identified (dark matter) has been noted and is major drawbacks of this approach, which could be remedied through a multi-analytical platform approach. The chemical overview on *Bacillus spp.* elucidated in this study will provide valuable insights for further exploration and research on these and other microbial sources. Nonetheless, the chemical diversity from these two *Bacillus* species, *Bacillus pumilus* and *Bacillus subtilis* motivate the search for new antimicrobial metabolites from these two species. Additionally, these could be a useful starting point for the discovery of clinically significant and new antimicrobial agents in the future.



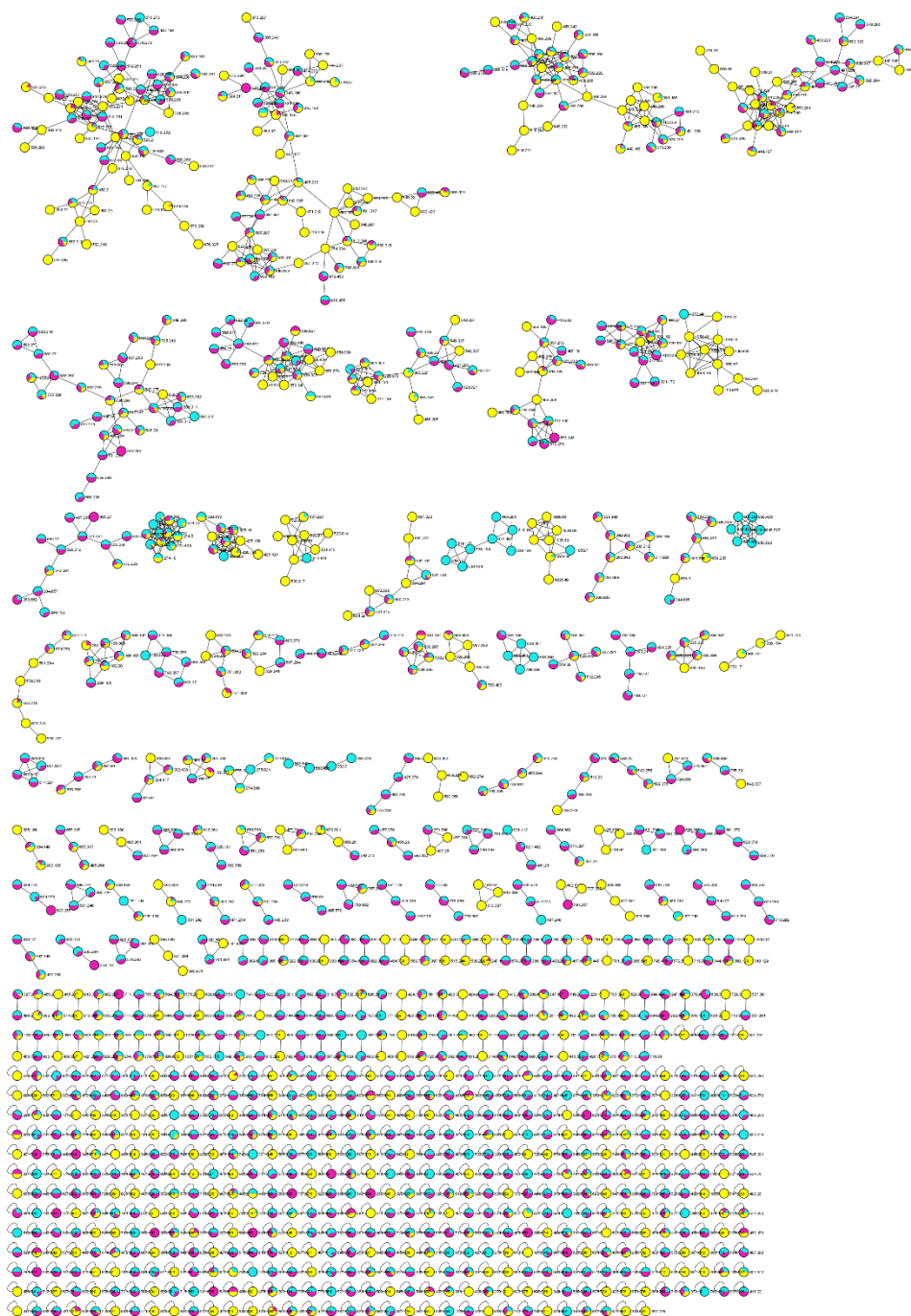
Supplementary Figure 17: Representative base peak intensity chromatograms (BPC) indicating the separation of metabolites in methanol extracts of isolate 3 supernatant (A) and pellet (B), isolate 11 supernatant (C) and pellet (D), and isolate 8 supernatant (E) and pellet (F), indicating some differences in the metabolite profiles of these stains.



Supplementary Figure 18: Chemical classification by MolNetEnhancer of Positive ionization datasets acquired through UPLC-QTOF-MS (A), as indicated in the figure legend revealed 7 different chemical superclasses. Pie chart (B) representing the Class which were identified by MolNetEnhancer. It can be observed that bacterial strains can produce a diverse compound with wide spectra of subclasses from the chemical classification. We can also observe that the larger molecular families are mostly annotated with Lipids and lipids (Green) and Organic acids and derivatives (Blue).



Supplementary Figure 19: Typical mass spectra of the fragmentation patterns of cyclo (L-Pro-L-Val) (A), cyclo (D-Pro-D-Leu) (B), cyclo(L-Tyr-L-Pro) (D) and cyclo (L-Pro-D-Phe) (C) with their typical structures.



Supplementary Figure 20: Molecular network Analysis of antimicrobial compounds from three bacterial isolates. Blue, purple, and yellow represent the secondary metabolites from isolate 1, isolate 2 and isolate 3, respectively. Nodes in the network represent individual compounds detected in the LC-QTOF-MS

Table 2: Metabolite annotation table. Profiling of metabolites reported to level 2 or 3 of the Metabolomics standards initiative.

<i>m/z</i>	Rt	Elemental composition	Fragment ions	Putative annotation	<i>m/z</i>
116.0700	0.84	C5H9NO2	116	Proline	116.0700
118.0854	0.92	C5H11NO2	118	Glycine betaine	118.0854
182.0799	1.34	C9H11NO3	165,136,119	L-Tyrosine	182.0799
132.1010	1.60	C6H13NO2	120120	Leucine	132.1010
166.0853	2.76	C9H11NO2	120	Phenylalanine	166.0853
229.1535	3.48	C11H20N2O	229, 183, 118	Leucylproline	229.1535
342.2371	8.69	C17H31N3O4	229, 183	Isoleucyl-proline	342.2371
205.0959	5.15	C11H12N2O2	188, 170 159, 146, 132 132, 118	L-Tryptophan	205.0959
602.29	0.82	C23H39N9O10	531, 389, 197	Grgesp	602.29
260.12	0.83	C10H17N3O5	213, 156, 122, 110	L-alanine, N[N(N-acetyl-L-alanine)]	260.12
284.14	1.02	C12H19N3O5	152, 135, 110	L-proline, N[N(N-acetyl-L-alanine)]	284.14
301.15	1.05	C12H17N3O5	244, 155, 127	L-proline, 1- [N(5-oxo-L-proly) glycyl]	301.15
244.13	1.05	C10H17N3O4	187, 155, 127	Glycylprolylalanine	244.13
247.13	1.09	C10H18N2O5	229, 183, 118	Valyl-glutamate	247.13
269.16	1.21	C12H20N4O3	223, 110	Histidylleucine	269.16
235.12	1.21	C11H14N4O2	207, 162, 110	Cyclo(his-pro)	235.12
229.12	1.32	C10H18N2O5	229, 183 118	pyroglutamylvaline	229.12
247.13	1.32	C11H20N2O5	229, 183, 155, 138, 118, 102	Glutamylvaline	247.13
261.14	1.50	C11H20N2O5	197, 148, 130, 102	Leucyl-Glutamate	261.14
432.24	1.51	C19H33N3O8	319, 301, 206, 188, 170, 153	ACMC-20lebal	432.24
261.14	1.73	C11H20N2O5	197, 148, 130, 102	Leucyl-Glutamate	261.14
471.22	1.86	C19H30N6O8	357,244, 187	Atrial peptide	471.22
284.10	1.98	C10H13N5O5	152, 135, 110	Guanosin	284.10
364.16	2.03	C16H21N5O5	235, 209, 181, 156,152, 110	TRH, free acid	364.16
377.15	2.06	C22H20N2O4	341, 241, 189	Opreal-082153	377.15
251.17	2.14	C12H18N4O2	223, 178, 110	Cyclo(his-leu)	251.17
281.11	2.48	C13H16N2O5	234, 120, 103	Phenylalanylasparic acid	281.11
291.10	2.51	C10H10N8O3	268, 255, 232,203,187,159	CB Micro-033252	291.10
1030.66	23.3	C51H89N7O13	917, 786,707,643,481,391,279,186	Surfactin A (C13)	1030.66
1044.66	23.6	C52H91N7O13	931, 800,707, 594, 481, 391,316,227	Surfactin B (C14)	1044.66
1058.67	23.6	C53H93N7O13	945, 814,707,594,463,320	Surfactin C (C15)	1058.67
1022.67	23.3	C52H91N7O13	909,794,685,582,441,338,227	Surfactin C14 homolog	1022.67
1036.69	23.6	C53H93N7O13	923,810,685,596,483,441,352,227	Surfactin C14	1036.69

239.103	1.9	C11H14N2O4	165, 147, 136, 119	Gly-Tyr	239.103
1088.076	24.3	C55H99N7O12	1086, 707, 608, 477	Pumilacidin E	1088.076
1078.73		C56H99N7O13	965, 852, 699, 56, 8, 455, 227	Pumilacidin C	1078.73
282.279	0.78	C18H35NO	264, 195, 177, 128	9-Octadecenamide, (Z)	282.279
366.21	2.63	C17H27N5O4	251, 234, 166, 15, 6, 110	Angiotensin1/2	366.21
369.11	2.69	C16H20N2O6 S	222, 204, 186, 178, 160, 138	Penicilloyl V	369.11
260.16	3.01	C11H21N3O4	242, 197, 132, 129	L-isoleucine, L-glutaminy-	260.16
304.13	3.06	C12H21N3O4 S	247, 155, 127	ACMC-20mcas	304.13
328.14	3.23	C15H21NO7	310, 292, 264, 24, 6, 178, 166, 143, 132, 120, 103	fructose-phenylalanine	328.14
295.13	3.31	C14H18N2O5	248, 120	phenylalanylgutamate	295.13
316.19	3.76	C14H25N3O5	215, 199, 171, 12, 7	threoylvalylproline	316.19
227.10	3.90	C10H14N2O4	209, 181, 139, 125	L-proline, 5-oxo-L-prolyl	227.10
325.12	4.31	C19H12N6	296, 273, 264	4-(benzotriazol-1-yl)-2-pyridin-3-ylquinazoline	325.12
302.17	4.73	C13H23N3O5	143, 132	Gly-Leu-Hydroxyproline	302.17
223.11	4.75	C11H14N2O3	166, 131, 120, 103	Maybridge1_006768	223.11
279.13	4.88	C14H18N2O4	232, 214, 158, 13, 2	D-Proline, D-tyrosyl-	279.13

4.7 References

- ABDEL-NASSER, A., BADR, A. N., FATHY, H. M., GHAREEB, M. A., BARAKAT, O. S. & HATHOUT, A. S. 2024. Antifungal, antiaflatoxic, and cytotoxic properties of bioactive secondary metabolites derived from *Bacillus* species. *Scientific Reports*, 14, 16590.
- ALAM, S., SADIQI, S., SABIR, M., NISA, S., AHMAD, S. & ABBASI, S. W. 2022. *Bacillus* species; a potential source of anti-SARS-CoV-2 main protease inhibitors. *Journal of Biomolecular Structure and Dynamics*, 40, 5748-5758.
- DE OLIVEIRA, J. A., WILLIAMS, D. E., ANDERSEN, R. J., SARRAGIOTTO, M. H. & BALDOQUI, D. C. 2020. Pumilacidins AE from sediment-derived bacterium *Bacillus* sp. 4040 and their antimicrobial activity evaluation. *Journal of the Brazilian Chemical Society*, 31, 357-363.
- DONCHEVA, N. T., MORRIS, J. H., GORODKIN, J. & JENSEN, L. J. 2018. Cytoscape StringApp: network analysis and visualization of proteomics data. *Journal of proteome research*, 18, 623-632.
- FERREIRA, W. T., HONG, H. A., HESS, M., ADAMS, J. R., WOOD, H., BAKUN, K., TAN, S., BACCIGALUPI, L., FERRARI, E. & BRISSON, A. 2021. Micellar antibiotics of *Bacillus*. *Pharmaceutics*, 13, 1296.
- FIRA, D., DIMKIĆ, I., BERIĆ, T., LOZO, J. & STANKOVIĆ, S. 2018. Biological control of plant pathogens by *Bacillus* species. *Journal of biotechnology*, 285, 44-55.
- JEMIL, N., MANRESA, A., RABANAL, F., AYED, H. B., HMIDET, N. & NASRI, M. 2017. Structural characterization and identification of cyclic lipopeptides produced by *Bacillus methylotrophicus* DCS1 strain. *Journal of Chromatography B*, 1060, 374-386.
- KHUSHBOO, KUMAR, P., DUBEY, K. K., USMANI, Z., SHARMA, M. & GUPTA, V. K. 2022. Biotechnological and industrial applications of *Streptomyces* metabolites. *Biofuels, Bioproducts and Biorefining*, 16, 244-264.
- KHWATHISI, A., MADALA, N. E., TRAORE, A. N. & SAMIE, A. 2024. Bioprospecting of soil-borne microorganisms and chemical dereplication of their anti-microbial constituents with the aid of UPLC-QTOF-MS and molecular networking approach. *PeerJ*, 12, e17364.

- KUMAR, B. & MADHUSUDANAN, K. 2020. Role of mass spectrometry in modern herbal drug research. *Biophysical and Computational Tools in Drug Discovery*. Springer.
- LANGENFELD, A., BLOND, A., GUEYE, S., HERSON, P., NAY, B., DUPONT, J. & PRADO, S. 2011. Insecticidal cyclodepsipeptides from *Beauveria felina*. *Journal of natural products*, 74, 825-830.
- LEMMENS-GRUBER, R., KAMYAR, M. & DORNETSHUBER, R. 2009. Cyclodepsipeptides-potential drugs and lead compounds in the drug development process. *Current medicinal chemistry*, 16, 1122-1137.
- MISHRA, J. & ARORA, N. K. 2018. Secondary metabolites of fluorescent pseudomonads in biocontrol of phytopathogens for sustainable agriculture. *Applied Soil Ecology*, 125, 35-45.
- NARUSE, N., TENMYO, O., KOBARU, S., KAMEI, H., MIYAKI, T., KONISHI, M. & OKI, T. 1990. Pumilacidin, a complex of new antiviral antibiotics production, isolation, chemical properties, structure and biological activity. *The Journal of antibiotics*, 43, 267-280.
- NASSAR, A.-E. F. & TALAAT, R. E. 2004. Strategies for dealing with metabolite elucidation in drug discovery and development. *Drug discovery today*, 9, 317-327.
- NICHOLSON, W. L., MUNAKATA, N., HORNECK, G., MELOSH, H. J. & SETLOW, P. 2000. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and molecular biology reviews*, 64, 548-572.
- POTTERAT, O. & HAMBURGER, M. 2013. Concepts and technologies for tracking bioactive compounds in natural product extracts: generation of libraries, and hyphenation of analytical processes with bioassays. *Natural product reports*, 30, 546-564.
- RABBEE, M. F., ALI, M. S., CHOI, J., HWANG, B. S., JEONG, S. C. & BAEK, K.-H. 2019. *Bacillus velezensis*: a valuable member of bioactive molecules within plant microbiomes. *Molecules*, 24, 1046.
- RODRIGO, S., GARCÍA-LATORRE, C. & SANTAMARIA, O. 2021. Metabolites produced by fungi against fungal phytopathogens: Review, implementation and perspectives. *Plants*, 11, 81.
- SAGGESE, A., CULURCIELLO, R., CASILLO, A., CORSARO, M. M., RICCA, E. & BACCIGALUPI, L. 2018. A marine isolate of *Bacillus pumilus* secretes a pumilacidin active against *Staphylococcus aureus*. *Marine drugs*, 16, 180.

- SHAFI, J., TIAN, H. & JI, M. 2017. Bacillus species as versatile weapons for plant pathogens: a review. *Biotechnology & Biotechnological Equipment*, 31, 446-459.
- SHAHNIANI, A., BAMZADEH, Z., MAHMOUDNIA, F. & ROUHI, L. 2024. Evaluation of antibacterial and anticancer properties of secondary metabolites isolated from soil Bacillus spp focusing on two strains of Bacillus licheniformis and Bacillus siamensis. *BMC Molecular and Cell Biology*, 25, 21.
- SINGH, P. & CAMEOTRA, S. S. 2004. Potential applications of microbial surfactants in biomedical sciences. *TRENDS in Biotechnology*, 22, 142-146.
- TSOTETSI, T., NEPHALI, L., MALEBE, M. & TUGIZIMANA, F. 2022. Bacillus for plant growth promotion and stress resilience: what have we learned? *Plants*, 11, 2482.
- VINAIXA, M., SCHYMANSKI, E. L., NEUMANN, S., NAVARRO, M., SALEK, R. M. & YANES, O. 2016. Mass spectral databases for LC/MS-and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends in Analytical Chemistry*, 78, 23-35.
- WADHWA, K., KAPOOR, N., KAUR, H., ABU-SEER, E. A., TARIQ, M., SIDDIQUI, S., YADAV, V. K., NIAZI, P., KUMAR, P. & ALGHAMDI, S. 2024. A Comprehensive Review of the Diversity of Fungal Secondary Metabolites and Their Emerging Applications in Healthcare and Environment. *Mycobiology*, 1-53.
- WANG, M., JARMUSCH, A. K., VARGAS, F., AKSENOV, A. A., GAUGLITZ, J. M., WELDON, K., PETRAS, D., DA SILVA, R., QUINN, R. & MELNIK, A. V. 2020. Mass spectrometry searches using MASST. *Nature biotechnology*, 38, 23-26.
- WANG, X., GONG, X., LI, P., LAI, D. & ZHOU, L. 2018. Structural diversity and biological activities of cyclic depsipeptides from fungi. *Molecules*, 23, 169.
- XIU, P., LIU, R., ZHANG, D. & SUN, C. 2017. Pumilacidin-like lipopeptides derived from marine bacterium Bacillus sp. strain 176 suppress the motility of Vibrio alginolyticus. *Applied and Environmental Microbiology*, 83, e00450-17.
- ZHU, Y., ZHANG, Z., ZHANG, M., E MAIS, D. & WANG, M.-W. 2010. High throughput screening for bioactive components from traditional Chinese medicine. *Combinatorial Chemistry & High Throughput Screening*, 13, 837-848.

Chapter 5: Genomic Insights into three soil-borne *Bacillus* sp.: Unveiling their Potential as a plant growth promoting and Antibiotic-producing Bacterium

5.1 Summary

Soilborne bacteria exhibit a range of unique traits that facilitate their survival and adaptation to diverse environments. One key characteristic is their production of a variety of secondary metabolites, which serve as mechanisms for communication and defense against coexisting microorganisms. In this study, the phylogenetic relationship of three *Bacillus* strains (KH1, KH2, and KH3) were analysed, based on digital DNA-DNA hybridization (dDDH) and average nucleotide identity (ANI) methods, indicated that these strains are closely related to *Bacillus altitudinis* 41KF2b, *Bacillus pumilus* NCTC 10337, and *Bacillus subtilis* PY79. Whole-genome sequencing of these *Bacillus* strains revealed the presence of genes associated with quorum sensing, stress response, and motility. Genome annotation using RAST and eggNOG-mapper identified genes linked to secondary metabolite biosynthesis, auxin production, and stress response pathways, highlighting the potential antimicrobial and plant growth-promoting properties of these strains. Secondary metabolite analysis via the antiSMASH tool uncovered gene clusters related to various compound classes, including cyclic peptides, saccharides, polyketides, and bacilysin. Notably, several of these clusters showed low similarity to known metabolites such as lichensin and surfactin, suggesting the presence of novel, uncharacterized natural products with potential biotechnological applications. Additionally, the genome analyses revealed resistance to multiple antibiotics and disinfectants, with genes such as *qacJ* and *qacG* contributing to environmental adaptability. This study underscores the potential of *Bacillus* strains for agricultural biotechnology, including biocontrol, biofertilization, and bioremediation. Further investigation into the functional roles of these genes could provide valuable insights into biotechnological applications.

Keywords: *Bacillus* sp., whole genome sequencing, antimicrobial compound, Biosynthetic gene clusters.

5.2 Introduction

In the 21st century, the rise of antimicrobial resistance pathogens poses a pivotal global challenge in both the agriculture and medical sectors (Adimpong et al., 2012, Naghavi et al., 2024). This phenomenon is driven by several key factors, including selective pressure from extensive antimicrobial use, genetic mutations within microbial populations, and the inappropriate application of existing antimicrobial compounds (Michael et al., 2014, Li et al., 2017, Dadgostar, 2019). Consequently, there is an urgent need for novel antimicrobial agents to counteract antimicrobial resistance effectively. Recently, to address this challenge, researchers are turning to microbial natural product as they offer promising solutions due to their vast structural and chemical diversity (Nainu et al., 2021, Uddin et al., 2021).

Microorganisms produce antimicrobial compounds as their natural defense system against the co-existing microorganisms (Baishya and Wakeman, 2019). The majority of commercially available antibiotics are derived from a relatively small group of microorganisms, primarily within the genera *Streptomyces* (Procópio et al., 2012), *Cephalosporium* (Khan et al., 2019), *Micromonospora* (Hifnawy et al., 2020), *Penicillium* (Elander, 2003), and *Bacillus* (Elshaghabee et al., 2017). Of these, species of *Bacillus* are particularly notable for their broad range of biotechnological applications, such as food processing (Payne et al., 2024), plant growth promotion (Sansinenea, 2019), and crop protection (Adimpong et al., 2012). However, this versatility is largely due to their ability to stimulate the immune system and their capacity to synthesize a wide range of bacterial secondary metabolites with potential applications as either antiviral (Wang et al., 2023), anticancer (Aimaier et al., 2023), antifungal (Devi et al., 2019), and antibacterial agents (Khwathisi et al., 2024).

Previous research has highlighted several *Bacillus* species, including *B. subtilis* (Hashem et al., 2019), *B. licheniformis* (Sukkasem et al., 2018), *B. amyloliquefaciens* (Abdallah et al., 2018), *B. velezensis* (Myo et al., 2019), *B. siamensis* (Xie et al., 2021), and *Bacillus altitudinis* (Hudson et al., 2021) for their antimicrobial production and biocontrol properties. Genomic analyses of various *Bacillus* strains have shown that a significant portion of their genome is dedicated to the biosynthesis of these compounds (Chen et al., 2007, Zhu et al., 2020). Some estimates suggest that between 4-5% of the genome in certain *Bacillus* strains may indeed be involved in the production of antimicrobial compounds (Caulier et al., 2019, Nannan et al., 2021). The majority of these compounds include antibiotics, lipopeptides, and other metabolites. Moreover, the classification of these compounds is cumbersome, due to their

structural diversity, and they can be further classified based several criteria such as biosynthetic machinery, sources, biological function as well as their biological properties.

The emergence of -omics science has opened new avenues for uncovering the molecular mechanisms of action mediated by antimicrobial-producing bacteria. For example, the use of liquid chromatographic-mass spectrometric-based metabolomics has led to the identification of specific secondary metabolites produced by these bacteria that offers a great biotechnological application such as gibberellins, iturins and surfactin (Nephali et al., 2022). Additionally, advancements in next-generation sequencing (NGS) technology have facilitated whole-genome sequencing of various antimicrobial producing bacterial strains, including *Bacillus* species (Maumela et al., 2024, Hafeez et al., 2023). Furthermore, genome mining through *in silico* bioinformatics approaches has been employed to explore the biosynthetic gene clusters (BGC) present in a number of microorganisms (Alam et al., 2022, Maumela et al., 2024).

Among the reported BGCs across different species within the *Bacillus* genus (Secondary biosynthetic potential), there is a limited amount of complete genomic information available for *Bacillus* species in relation to biosynthetic potential (Yin et al 2023). Therefore, conducting a comprehensive genomic analysis of these bacterial isolates is essential to enhance our understanding of their biosynthetic potential. While sequencing can help identify the genes responsible secondary metabolites production, there remains a gap in our understanding of the regulatory mechanisms governing these genes in these isolates. Consequently, this study aims to explore the whole genome sequence of a three isolated *Bacillus* strain, along with its chemical space to assess their antimicrobial biosynthetic potential. The genome has been mined for genes related to secondary metabolite production based on the results of the whole genome sequence analyses. The genetic potential of these isolates was validated through the identification of biosynthetic gene clusters (BGCs), which were compared with those of commercially applied strains of *Bacillus* species. Thus, the objectives of this study are: (1) to determine the precise taxonomic position of bacterial strain through phylogenetic analysis and genomic data among reported strains of *Bacillus* spp., utilizing an approach based on digital DNA-DNA hybridization (dDDH) and Average Nucleotide Identity (ANI); and (2) to predict gene clusters encoding potential secondary metabolites in these bacterial isolates.

5.3 Experimental

5.3.1 Isolation and culture maintenance

The bacterial strain was isolated from an agricultural setting through sprinkling techniques, as detailed in a previous study that highlighted its potential antimicrobial properties against various pathogenic strains (Khwathisi et al., 2024). A 30% glycerol stock of the bacterial culture was plated on nutrient agar (NA) and incubated at 30 degrees Celsius for 24 hours to ensure routine culture maintenance.

5.3.2 DNA extraction and Genome sequencing

Genomic DNA was extracted from solid colonies using the NucleoSpin microbial DNA extraction kit, following the manufacturer's instructions (Macherey-Nagel, Germany). The extracted DNA was sequenced on an Illumina platform at a commercial service provider, Biotechnology Platform, Agricultural Research Council, located in Onderstepoort, South Africa. Paired-end libraries (2×150 bp) were prepared using the NextEra DNA sample preparation kit (Illumina, United States), and sequencing was conducted on the HiSeq 2,500 platform.

5.3.3 Genome Assembly and bioinformatics analysis

Pre-annotation analyses were conducted on the Galaxy Web server (<https://usegalaxy.org/>) using default settings. Quality control of the raw sequence reads was performed with FastQC (version 0.72.0), followed by trimming using Trimmomatic (version 0.38.0). The sequence reads were then *de novo* assembled with Unicycler (version 0.4.8.0), and quality assessment was carried out using Quast (Galaxy Version 5.0.2). The draft genome was annotated using the National Center for Biotechnology Information's Prokaryotic Genome Annotation Pipeline and Rapid Annotations using Subsystems Technology. Additionally, functional annotation of genes included scanning for cluster of orthologous groups (COGs) annotations with eggNOG-mapper v2 (<http://eggno-mapper.embl.de>).

5.3.4 Genome Comparison

Genome comparison was conducted based on the average nucleotide identity of KH1 with closely related strains using the OrthoANI algorithm. The closely related type strains of *Bacillus* sp. were selected for comparative genomic analysis. This process involved obtaining complete genome sequences from various databases, followed by quality control and

alignment. Pairwise distances between the genomes were calculated using the GBDP algorithm with the $d\%$ distance formula, incorporating a 'trimming' approach to enhance data quality. To ensure robustness, multiple distance replicates were utilized. Digital DNA-DNA hybridization (DDH) values and their confidence intervals were computed using the GGDC 4.0 tool with default parameters for *Bacillus* sp. in TYGS.

5.4 Results

5.4.1 Genomic Characteristics and Assembly Metrics

5.4.1.1 Basic genomic characteristics of the strain

Using the HiSeq 2500 platform (Illumina, United States), we successfully sequenced the complete genome of *Bacillus* isolates, isolated from an agricultural setting. The sequencing data was processed and assembled using the SPAdes algorithm, yielding high-quality genomic assemblies, as shown in Table 3. Quality and completeness were further assessed using the Quast Genome assembly tool, which confirmed an impressive 99.59% genome completeness and only 0.21% contamination. The *de novo* assembly produced total genome sizes of 3,733,820 bp for KH1, 3,688,736 bp for KH2, and 4,176,670 bp for KH3. The N50 values were 440,094 bp, 378,244 bp, and 1,056,378 bp, respectively, with each genome comprising 26, 22, and 24 contigs (all ≥ 500 bp). The G+C content of the genomes was 41.2% for KH1 and 2, and 43.7% for KH3. The assembled draft genomes have been deposited in NCBI-GenBank under the accession numbers JBPXED000000000.

Table 3: Genome characteristic for *Bacillus* sp. strain KH1

GENOME CHARACTERISTICS	ISOLATE 1	ISOLATE 2	ISOLATE 3
TOTAL SIZE (BP)	3,733,820 bp	3,688,736 bp	4,176,670 bp
CONTIGS	26	22	24
% G+C CONTENT	41.2%	41.2	43.7
GENES (TOTAL)	4025	3896	4382
CDSS (TOTAL)	3954	3825	4318
GENES (RNA)	71	71	64

5.4.1.2 Whole-Genome-Based Taxonomic Analysis

In the present study the whole genome analysis identified the strain KH1 which was originally classified as a *Bacillus* strain through comparing the overall genome-based relatedness indices such as ANIb, ANIm, dDDh and GC content differences. During the analysis *Escherichia coli* 11368 was used as an outlier. The Type (Strain) Genome Server (TYGS) was utilized to elucidate the phylogenomic relationships and accurately identify *Bacillus* sp. strain KH1. A pairwise genome comparison with related species was conducted to perform a comprehensive whole-genome phylogenetic analysis. The results revealed that *Bacillus* sp. strain KH1 is closely related to *Bacillus altitudinis* 41KF2b, with a digital DNA-DNA hybridization (dDDH) value of 86.1% (Table 4). Strain KH2 is closely related to *Bacillus subtilis* PY79, with a dDDH value of 92.8% (Table 5), while KH3 is closely related to *Bacillus pumilus* NCTC 10337, with a dDDH value of 89.5% (Table 6). This substantial dDDH value suggests a significant genetic similarity between the two strains, reinforcing their phylogenetic relationship. The dDDH, calculated using the Genome–Genome Distance Calculator (GGDC), significantly exceeds the species cut-off of 70%, indicating a strong phylogenetic affinity between these strains. Such a high dDDH value not only signifies genetic closeness but also has implications for species delineation. Strains exhibiting dDDH values above 70% are typically classified within the same species, while those below this threshold are considered distinct species. Therefore, the observed dDDH values of 86.1%, 92.8%, and 89.5% support the classification of *Bacillus* sp. strain KH1 as closely related to *Bacillus altitudinis* 41KF2b, strain KH2 as closely related to *Bacillus subtilis* PY79, and strain KH3 as closely related to *Bacillus pumilus* NCTC 10337, confirming the genetic relationships between these strains.

Table 4: Pairwise comparisons of user genomes (KH1) vs. type strain genomes

Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
<i>Bacillus altitudinis</i> 41KF2b	94.0	(91.4 - 95.8)	86.1	(83.5 - 88.4)	95.0	(93.0 - 96.4)	0.08
<i>Bacillus safensis subsp. safensis</i> FO-36b	86.9	(83.3 - 89.8)	36.5	(34.0 - 39.0)	74.8	(71.4 - 78.0)	0.55
<i>Bacillus pumilus</i> NCTC 10337	84.8	(81.1 - 87.9)	36.1	(33.6 - 38.6)	72.9	(69.5 - 76.1)	0.53
<i>Bacillus australimaris</i> NH7I 1	89.2	(85.9 - 91.8)	35.8	(33.4 - 38.3)	76.3	(72.9 - 79.5)	0.16
<i>Bacillus cereus</i> ATCC 14579	12.8	(10.1 - 16.1)	27.5	(25.1 - 29.9)	13.2	(10.8 - 15.9)	5.9
<i>Bacillus anthracis</i> M21	12.7	(10.0 - 15.9)	22.4	(20.1 - 24.8)	13.0	(10.7 - 15.8)	6.01
<i>Bacillus subtilis</i> PY79	15.6	(12.7 - 19.0)	18.0	(15.8 - 20.3)	15.5	(13.0 - 18.4)	2.65
<i>Escherichia coli</i> 11368	12.6	(9.9 - 15.8)	17.5	(15.4 - 19.9)	13.0	(10.6 - 15.7)	9.46
<i>Bacillus stercoris</i> X2(2017)	15.4	(12.4 - 18.8)	17.5	(15.3 - 19.8)	15.3	(12.8 - 18.2)	2.51

Table 5: Pairwise comparisons of user genomes (KH2) vs. type strain genomes

Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
<i>Bacillus pumilus</i> NCTC 10337	89,7	(86.5 - 92.3)	63,5	(60.6 - 66.3)	87,7	(84.7 - 90.1)	0,55
<i>Bacillus safensis subsp. safensis</i> FO-36b	86,9	(83.3 - 89.8)	42,2	(39.7 - 44.7)	78	(74.5 - 81.1)	0,57
<i>Bacillus australimaris</i> NH7I 1	88,3	(84.8 - 91.0)	40,9	(38.4 - 43.4)	78,5	(75.1 - 81.6)	0,18
<i>Bacillus altitudinis</i> 41KF2b	89,2	(85.9 - 91.8)	34,8	(32.4 - 37.4)	75,7	(72.2 - 78.9)	0,1
<i>Bacillus cereus</i> ATCC 14579	12,8	(10.1 - 16.1)	28	(25.6 - 30.5)	13,2	(10.9 - 16.0)	5,88
<i>Bacillus anthracis</i> M21	12,7	(10.0 - 16.0)	21,8	(19.5 - 24.2)	13,1	(10.8 - 15.8)	5,99
<i>Bacillus subtilis</i> PY79	15,2	(12.3 - 18.6)	18,4	(16.3 - 20.8)	15,2	(12.7 - 18.1)	2,67
<i>Bacillus stercoris</i> X2(2017)	15	(12.1 - 18.4)	17,8	(15.6 - 20.1)	15	(12.5 - 17.9)	2,53
<i>Escherichia coli</i> 11368	12,6	(9.9 - 15.8)	17,5	(15.4 - 19.9)	13	(10.6 - 15.7)	9,48

Table 6: Pairwise comparisons of user genomes (KH3) vs. type strain genomes

Subject strain	dDDH (d0, in %)	C.I. (d0, in %)	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
<i>Bacillus subtilis</i> PY79	92.8	(90.0 - 94.9)	87.2	(84.7 - 89.4)	94.3	(92.2 - 95.8)	0.05
<i>Bacillus stercoris</i> X2	88.9	(85.6 - 91.6)	62.4	(59.5 - 65.2)	86.7	(83.7 - 89.3)	0.09
<i>Bacillus cereus</i> ATCC 14579	87.3	(83.7 - 90.1)	45.1	(42.5 - 47.7)	79.6	(76.2 - 82.7)	0.02
<i>Bacillus anthracis</i> M21	86.8	(83.2 - 89.7)	36.1	(33.7 - 38.6)	74.6	(71.1 - 77.7)	0.45
<i>Bacillus safensis</i> subsp. <i>safensis</i> FO-36b	15.2	(12.3 - 18.7)	19.4	(17.2 - 21.8)	15.3	(12.8 - 18.2)	2.12
<i>Bacillus pumilus</i> NCTC 10337	14.9	(12.0 - 18.3)	17.8	(15.7 - 20.1)	14.9	(12.5 - 17.8)	2.35
<i>Bacillus australimaris</i> NH71 I	15.3	(12.4 - 18.7)	17.5	(15.3 - 19.8)	15.2	(12.7 - 18.1)	2.44
<i>Escherichia coli</i> 11368	12.6	(9.9 - 15.8)	17.4	(15.3 - 19.7)	12.9	(10.6 - 15.7)	6.9
<i>Bacillus altitudinis</i> 41KF2b	15.0	(12.1 - 18.4)	18.3	(16.2 - 20.7)	15.0	(12.6 - 17.9)	2.03

To further analyze the taxonomy of the genome, Average Nucleotide Identity (ANI) analysis was conducted, incorporating an outlier (*E. coli*). The results indicated that strain KH1 is most closely aligned with *Bacillus altitudinis* 41KF2b, yielding an ANI value of 98.45% (Figure 21), strain KH2 is closely related to *Bacillus subtilis* PY79 with an ANI value of 95.51% (Figure 22), and strain KH3 is closely related to *Bacillus pumilus* NCTC 10337 with an ANI value of 95.31% (Figure 23). This value surpasses the speciation boundary threshold of ANI > 95-96%, which is generally recommended for a strain to be considered as belonging to the same species. Although clusters indicating a close relationship with *Bacillus pumilus* NCTC 10337, *Bacillus australimaris* NH71, and *Bacillus australimaris* NH71 are observed in KH1, KH2, and KH3, respectively, the ANI values for these clusters are below the threshold, thereby reinforcing the confidence in the taxonomic identification derived from the analysis.

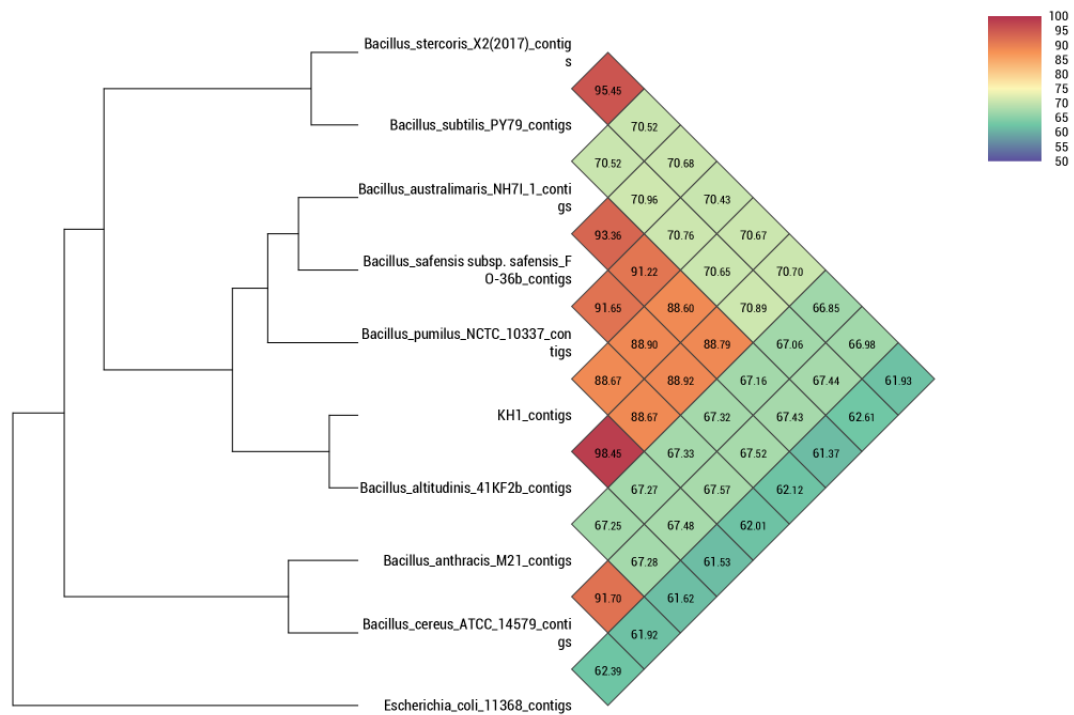


Figure 21: Heat-map, based on ANI values, for *Bacillus* sp. Strain KH1 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness (indicated by dark red colour).

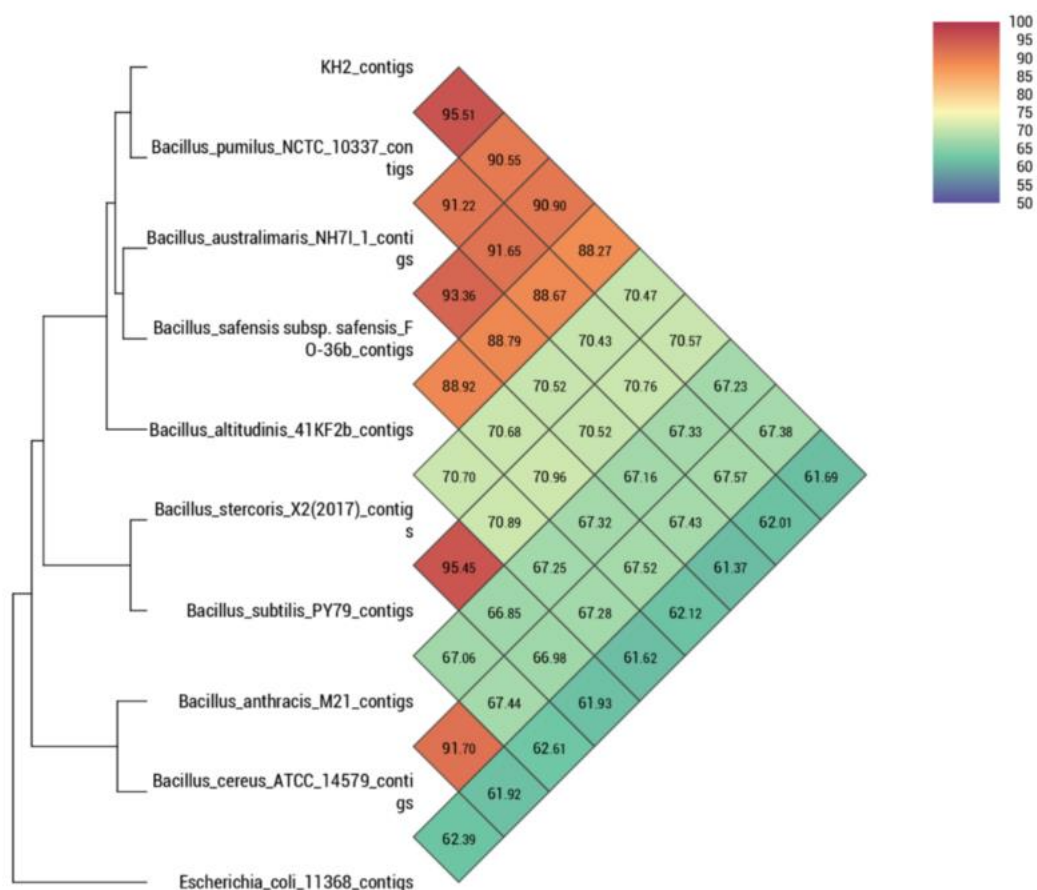


Figure 22: Heat-map, based on ANI values, for *Bacillus* sp. Strain KH2 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness (indicated by dark red colour).

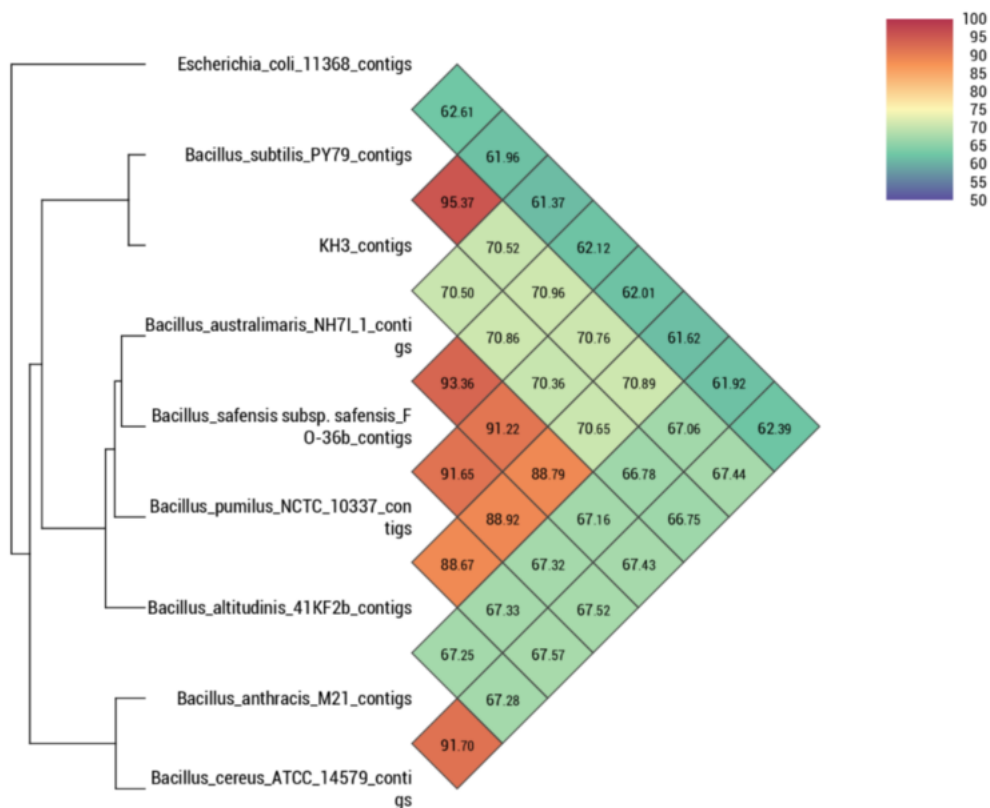


Figure 23: Heat-map, based on ANI values, for *Bacillus* sp. Strain KH2 and related species. The heat map visually represents the genomic similarity between the strains, with higher ANI values indicating greater genetic relatedness (indicated by dark red colour).

5.4.2 Genome annotation of the *Bacillus* spp strains

Based on the RAST annotation, this analysis predicted a variety of putative genes that are crucial for bacterial survival in its natural harsh environment and interaction with other organisms through various relationships such as Motility and chemotaxis. The genome annotation revealed several key traits that contribute to their antimicrobial production such as amino acid and derivatives, a vital defence mechanism in their natural environment. With *B. subtilis* showing highest (292) number of these genes while *B. altitudinis* showing the lowest (178). Antimicrobial production serves as a primary strategy for survival, enabling the bacteria to inhibit the growth of competing microorganisms and establish dominance within their ecological niche. The analysis identified a set of putative genes organized into functional

categories linked to antimicrobial production, as well as plant growth promotion. These functional groups included Metabolic Pathways, Stress Response, Cellular Processes, Nitrogen Fixation, and Secondary Metabolite Production (as shown in Figure 24, 25, and 26). *Bacillus subtilis* consistently exhibited the greatest number of genes associated with nitrogen metabolism. In contrast, *Bacillus altitudinis* had the fewest genes related to nitrogen metabolism. Plant growth promotion is significantly influenced by the complex interactions between plants and various microorganisms, particularly PGPB. These beneficial bacteria enhance plant growth through mechanisms such as phytohormone regulation and stress mitigation, which are often mediated by specific genetic pathways within the bacteria. Interestingly, the genome annotation analysis revealed four genes associated with auxin biosynthesis, including IAR, which encodes indole-acetaldehyde reductase. This enzyme plays a vital role in the production of indole-related compounds by catalysing the reduction of indole-3-acetaldehyde to indole-3-acetic acid, an essential plant growth hormone (Jirakkakul et al., 2023). Additionally, the analysis identified the ATFS gene, which encodes flavonol 3-sulfotransferase (EC 2.8.2.25). This enzyme is crucial for modifying flavonols to enhance their stability, thereby playing an important role in plant defence mechanisms, signalling, and interactions. Furthermore, the presence of 3-dehydroquinate dehydratase II was noted. This enzyme is essential for synthesizing aromatic amino acids as it catalyses the conversion of 3-dehydroquinate to 3-dehydroshikimate, a critical step in the biosynthesis of aromatic compounds and various secondary metabolites (Hu et al., 2025).

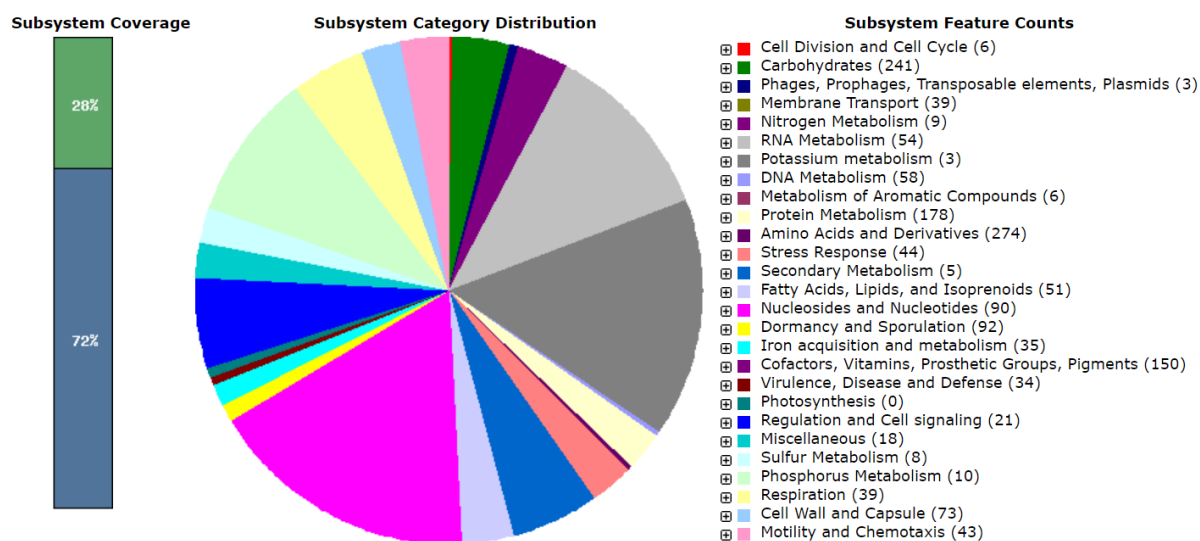


Figure 24: RAST annotation results for the genome of *Bacillus* sp. strain KH1. The annotation categorizes the identified genes into various functional subsystems.

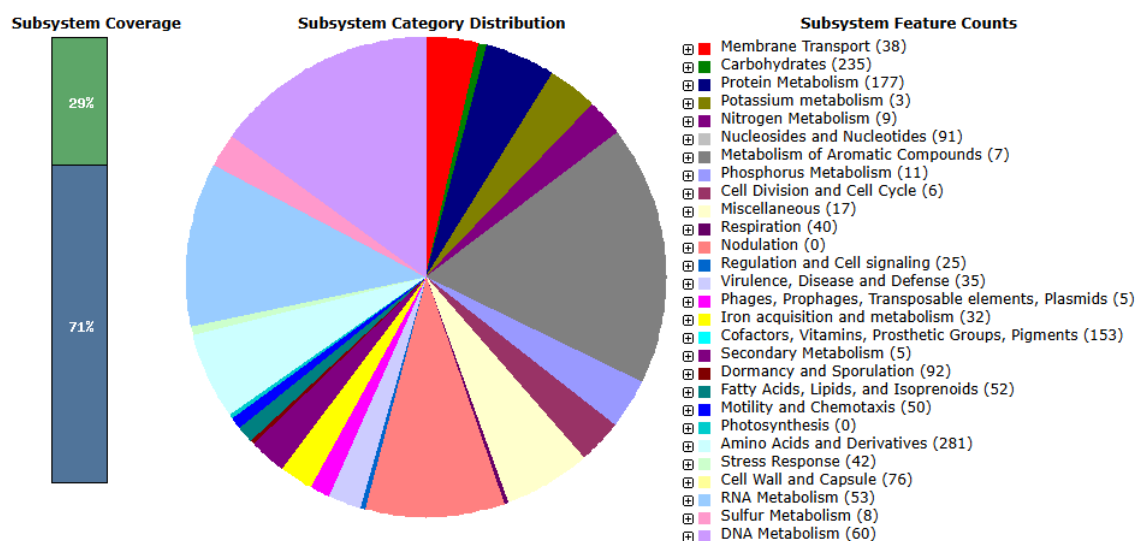


Figure 25: Rast annotation results for the genome *Bacillus* sp. Strain KH2. The annotation categorizes the identified genes into various functional subsystems.

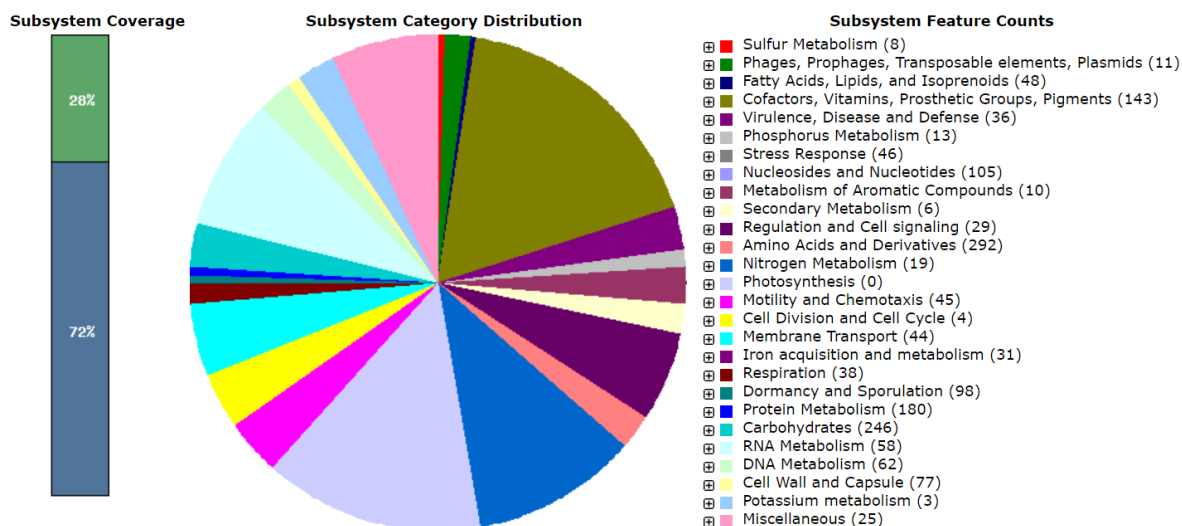


Figure 26: RAST annotation results for the genome of *Bacillus* sp. strain KH1. The annotation categorizes the identified genes into various functional subsystems.

The functional annotation of the Cluster of Orthologous Groups (COG) was conducted using eggnoG-Mapper v2.1.10, which involved comparing the target genome sequence with known functional sequences. The predicted COG functional annotations were classified into four main groups: Metabolism, Cellular Processes and Signalling, Information Storage and Processing, and Poorly Characterized Genes. In total, we annotated 3,660, 3,966, and 3,937 genes likely essential for the functionality of *B. Altitudinis* (Figure 27), *B. Pumilus* (Figure 28), and *B. Subtilis* (Figure 29), respectively. Notably, the annotation of *B. Altitudinis* resulted in 3,660 genes, with significant contributions from the following categories: cellular processes (560 genes), information storage (632 genes), metabolism (1,271 genes), and poorly characterized genes (1,199 genes). The annotation of *B. Pumilus* resulted in 3,966 genes, with notable contributions from: cellular processes (1,360 genes), information storage (786 genes), metabolism (1,487 genes), and poorly characterized genes (333 genes). Finally, the annotation of *B. Subtilis* resulted in 3,937 genes, with significant contributions from: Cellular processes (1,454 genes), information storage (818 genes), metabolism (1,428 genes), and poorly characterized genes (237 genes). These findings highlight the potential importance of these genes in the biosynthesis of secondary metabolites, including antimicrobial compounds, which enhance the ecological role of *Bacillus* species strains. The Gene Ontology (GO) annotations revealed that a significant majority of the genes were primarily associated with biological processes, closely followed by molecular functions and cellular components. Notably, within

the metabolism category, amino acid transport and metabolism and carbohydrate transport and metabolism were the most prominent. Among the three main functional categories, genes related to cellular processes were the least abundant, while those associated with metabolism were the most numerous, followed by information storage. Within the cellular processes, the two most abundant functions identified were cell wall/membrane/envelope biogenesis and signal transduction mechanisms.

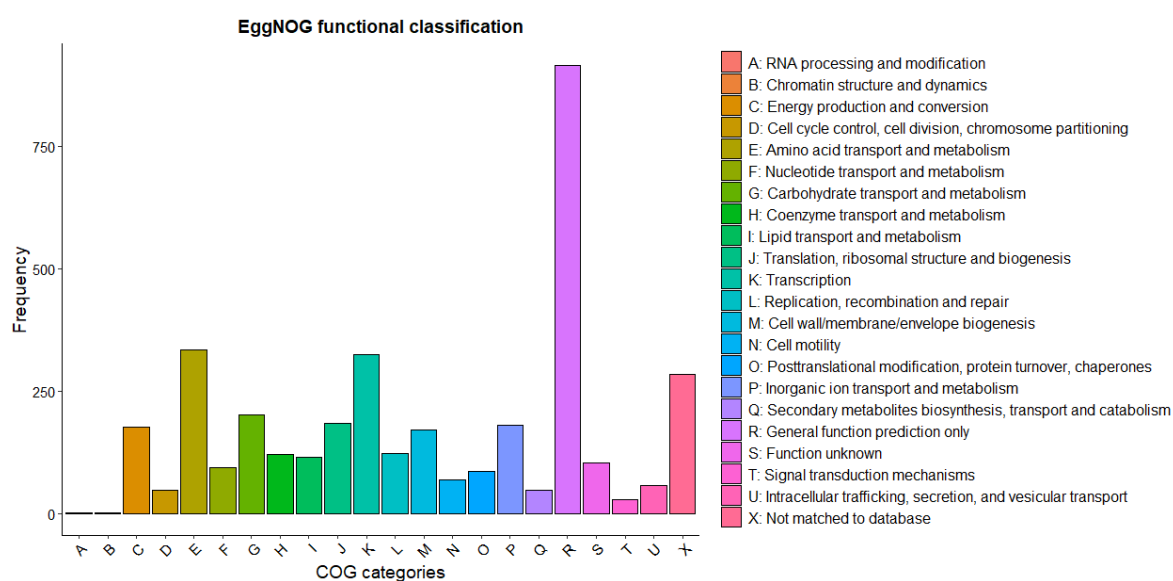


Figure 27: EggNOG based functional classification of the genome of *Bacillus sp.* strain KH1. The classification is organized into various Cluster of Orthologous Groups categories, highlighting the distribution of genes across key functional groups.

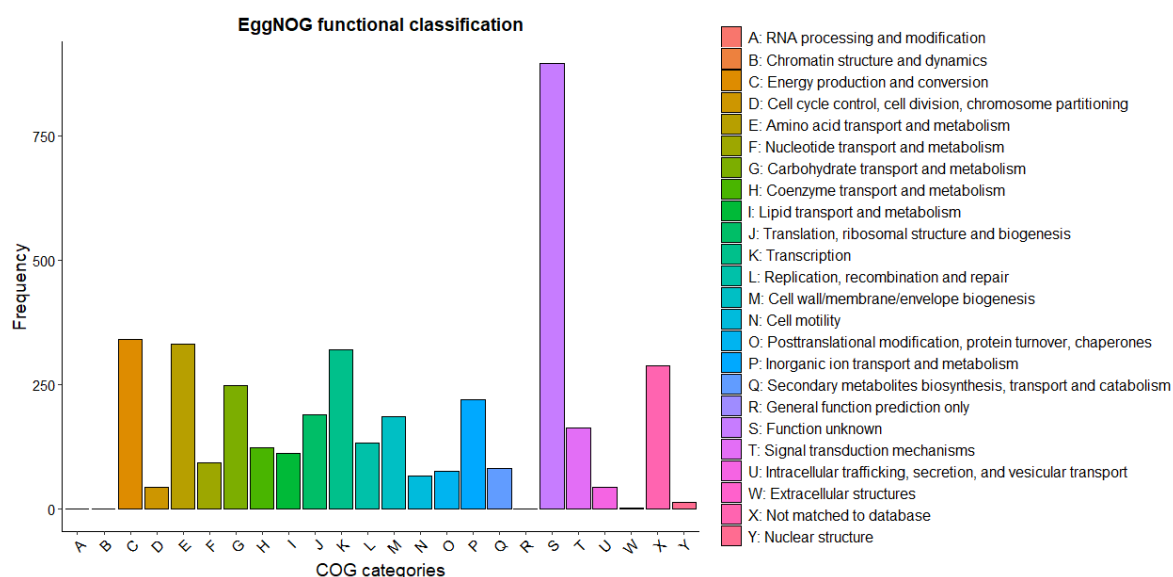


Figure 28: EggNOG based functional classification of the genome of *Bacillus sp.* strain KH2 based on the EggNOG. The classification is organized into various Cluster of Orthologous Groups categories, highlighting the distribution of genes across key functional groups

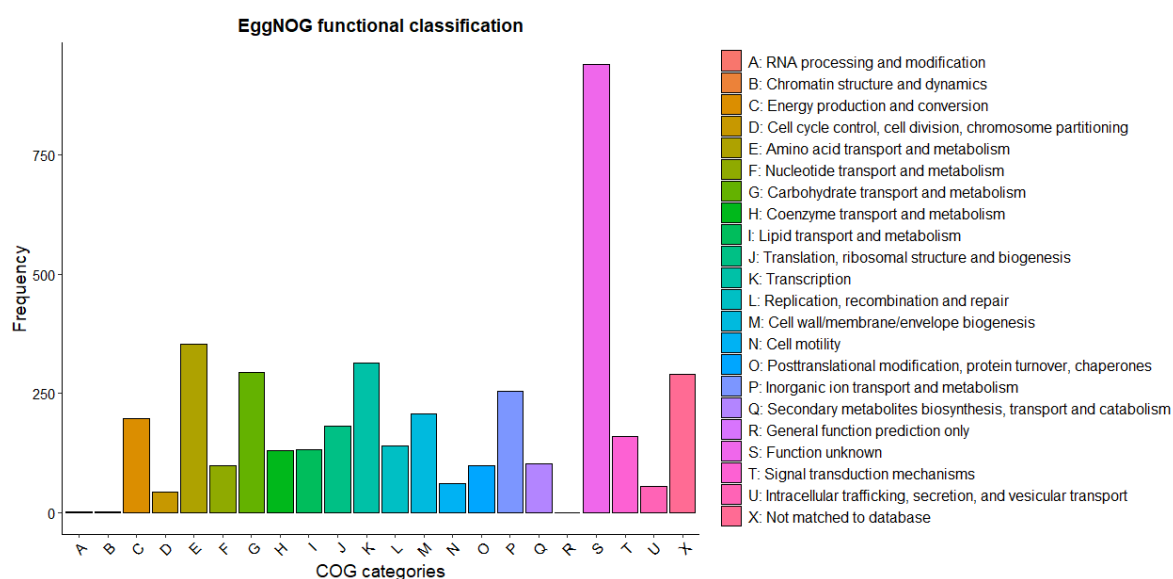


Figure 29: EggNOG based functional classification of the genome of *Bacillus spp.* strain KH3. The classification is organized into various Cluster of Orthologous Groups categories, highlighting the distribution of genes across key functional groups

5.4.3 The predictive analysis of secondary metabolites

The three bacterial genomes of *Bacillus* sp. strains were predicted for their secondary metabolite synthesis gene clusters using the anti-SMASH analysis. This analysis identified 21, 29, and 38 gene clusters in *B. altitudinis*, *B. pumilus*, and *B. subtilis*, respectively, associated with the synthesis of secondary metabolites such as Bacillibactin, Subtilosin A, Bacilysin, Fengycin, Plipastatin, and Bacillibactin (Figure 30). These metabolites belong to various classes of biosynthetic gene clusters, specifically NRP + Polyketide, NRP, NRP, Polyketide, and NRP BGCs, respectively (see Table 7). However, the majority results of this study revealed that the majority of findings exhibited low similarity (less than 70%) to known natural products, indicating that these soil-borne *Bacillus* strains may have the genetic capacity to produce novel, unidentified natural products with potential applications in various industries. The ability of these bacteria to thrive in harsh conditions, such as those found in soil and competitive environments, may facilitate the production of new bioactive compounds that are crucial for their adaptation and survival.

Table 7: Predicted secondary metabolites gene cluster in the *Bacillus* sp. Strain

Cluster	Similar cluster	TYPE	<i>B. altitudinis</i>	<i>B. pumilus</i>	<i>B. subtilis</i>
1	3-(Z-2'-isocyanoethenyl)-indole	NRPS	-	√	-
2	Ambiguine H isonitrile	PKS	-	√	-
4	Aurantinin B	PKS	-	-	√
5	Bacillaene	PKS	-	-	√
6	Bacillibactin	PKS	√	-	√
7	Bacillomycin D	PKS	-	-	√
8	Bacilysin	NRPS (Siderophore)	√	√	
10	Benarthin	NRPS	-	-	√
11	Bengamide	NRPS	-	√	-
12	Butyrolactol A	NRPS	-	√	√
13	Calyculin A	NRPS	-	-	√
14	Carotenoid	NRPS	√	√	-
15	Chondrochloren A	PKS	-	√	√
16	Corynecin III	PKS	-	√	√
18	Elansolid A1	Other (Terpenoid, involves isoprenoid biosynthesis)	-	-	√
19	Fengycin	PKS	√	√	√
20	Fischerindole	NRPS	-	√	-
21	Fischerindole L	PKS	-	√	-

22	Fuscachelin A	PKS	√	-	√
23	Griseobactin	NRPS	-	-	√
24	Jamaicamide A	NRPS	-	-	√
25	Kedarcidin	NRPS	-	√	√
26	Lagriene	NRPS	-	-	√
27	Lichenysin	NRPS	√	√	√
28	Macrobrevin	NRPS	-	-	√
29	Mycalamide A	PKS	-	-	√
30	Mycosubtilin	PKS	√	√	√
32	Nosperin	NRPS	-	-	√
33	Paenibactin	NRPS	-	-	√
34	Paenilamicin	NRPS	-	√	-
35	Pellasoren	NRPS	-	√	-
36	Plipastatin	NRPS	-	√	√
37	Pseudomonic acid A	NRPS	-	-	√
38	Schizokinen	NRPS (Siderophore)	-	√	-
39	S-layer glycan	NRPS	√	√	-
40	Subtilosin A	NRPS	-	-	√
41	Surfactin	NRPS	√	√	√
42	Synechobactin C9	NRPS	-	√	-
43	Xenocoumacin 1	NRPS (Siderophore)	-	-	√
44	Zwittermicin A	Other (Glycan/Polysaccharide)	-	√	√
45	Bacillibactin	NRPS	√	-	-
46	Bacilysin	NRPS	-	-	-
47	Basiliskamide A	NRPS (Siderophore)	√	-	√
48	Bottromycin A2	NRPS	√	-	-
50	Chejuenolide A	NRPS	√	-	-
51	Dipeptide aldehydes	NRPS (Siderophore)	√	-	-
54	Molybdenum cofactor	NRPS	√	-	-
55	Mycosubtilin	NRPS	-	-	-
56	Myxochelin A	NRPS	√	-	-
57	Paenibactin	PKS	√	-	-
58	Plipastatin	NRPS	-	-	√
59	S-layer glycan	Other (Inorganic compound, non- biosynthetic)	-	-	-
60	Sporulation killing factor	NRPS	√	-	-
61	Teichuronic acid	NRPS (Siderophore)	√	-	-
62	Zwittermicin A	NRPS (Siderophore)	√	-	-

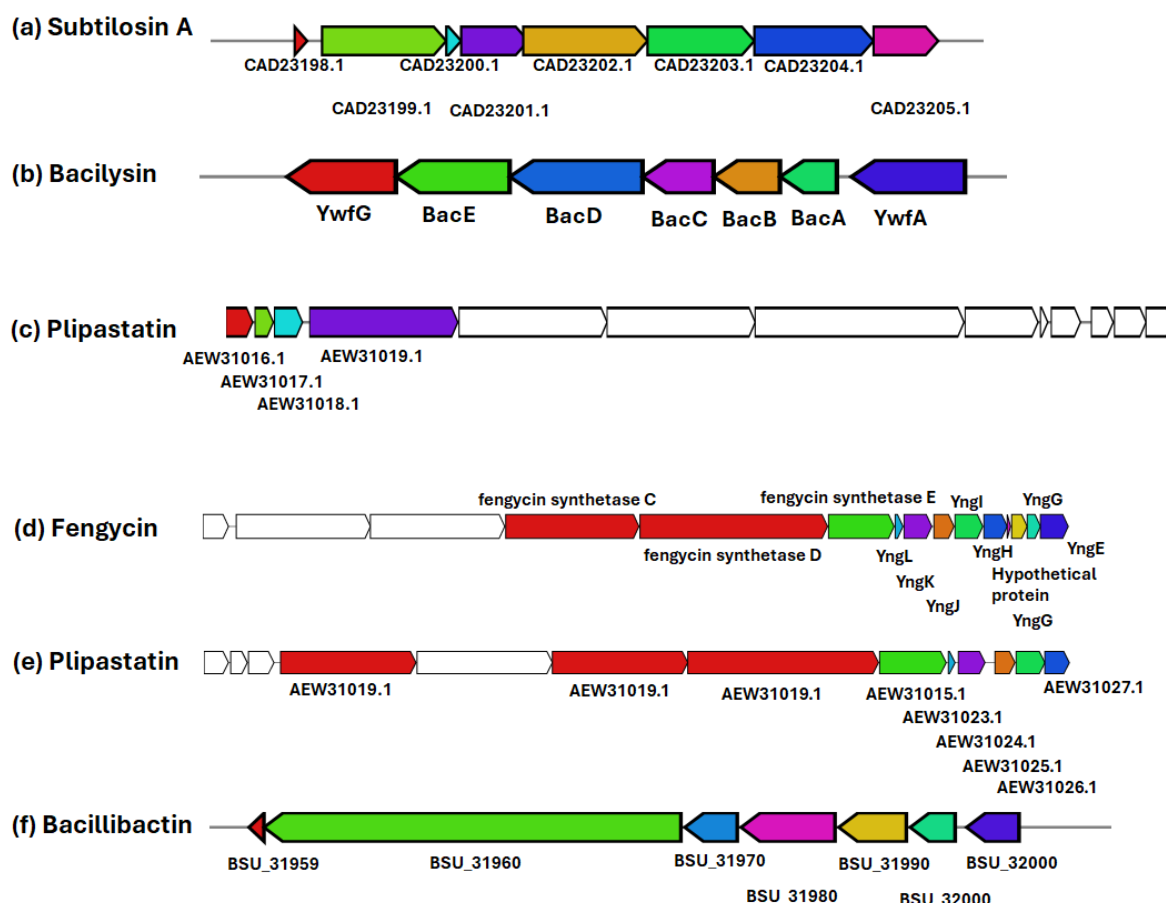


Figure 30: Identification of Gene Clusters for Antimicrobial Secondary Metabolites in *Bacillus subtilis* Bbv57 Using antiSMASH 6.0: (a) Bacillibactin, (b) Subtilosin A, (c) Bacilysin, (d) Fengycin, (e) Plipastatin, and (f) Bacillibactin.

5.4.5 Antibiotic-resistant gene prediction

For the analysis of antimicrobial resistance genes, the Comprehensive Antibiotic Resistance Database (CARD) was utilized to identify genes associated with resistance mechanisms. The genome was found to encode eight genes responsible for various classes of drugs, including phenicol antibiotics, glycopeptide antibiotics, disinfectants, and antiseptics. The CARD analysis identified the genes *qacJ* and *qacG* (Table 8), which encode proteins that function as multidrug efflux pumps, providing resistance to quaternary ammonium compounds (QACs). Additionally, the genome encodes the *cat86* gene, which produces a catechol 1,2-dioxygenase enzyme involved in the degradation of aromatic compounds, including chlorinated hydrocarbons and chlorinated phenols, as well as certain herbicides. Notably, the genome also included genes encoding for chloramphenicol and vancomycin, both of which are utilized in veterinary medicine and human healthcare, respectively. Among these classes, the identified

resistance mechanisms included antibiotic efflux, antibiotic inactivation, and antibiotic target alteration. The genome contained a greater number of genes associated with antibiotic target alteration, followed by those related to antibiotic efflux and antibiotic inactivation.

Table 8: Potential antimicrobial resistance genes predicted from the genome of *Bacillus* spp., using CARD

No	ARO Term	AMR Gene Family	Drug Class	Resistance Mechanism	% Identity of Matching Region	% Length of Reference Sequence
1	qacJ	small multidrug resistance (SMR) antibiotic efflux pump	disinfecting agents and antiseptics	antibiotic efflux	43.88	114.02
2	qacG	small multidrug resistance (SMR) antibiotic efflux pump	disinfecting agents and antiseptics	antibiotic efflux	45.0	99.07
3	<i>Bacillus pumilus</i> cat86	chloramphenicol acetyltransferase (CAT)	phenicol antibiotic	antibiotic inactivation	88.18	99.55
4	vanY gene in vanG cluster	vanY, glycopeptide resistance gene cluster	glycopeptide antibiotic	antibiotic target alteration	37.11	98.92
5	vanT gene in vanG cluster	glycopeptide resistance gene cluster, vanT	glycopeptide antibiotic	antibiotic target alteration	32.53	55.20
6	vanY gene in vanM cluster	vanY, glycopeptide resistance gene cluster	glycopeptide antibiotic	antibiotic target alteration	34.93	125.32
7	vanT gene in vanG cluster	glycopeptide resistance gene cluster, vanT	glycopeptide antibiotic	antibiotic target alteration	33.88	55.20
8	vanG	glycopeptide resistance gene cluster, Van ligase	glycopeptide antibiotic	antibiotic target alteration	39.32	103.15

5.5 Discussion

In microbial taxonomy, 16S rRNA has traditionally been the primary method for identifying microorganisms (Petti et al., 2005). However, a major limitation of 16S rRNA-based taxonomy is the lack of sufficient taxonomic resolution, particularly for closely related species (Poretsky et al., 2014, Ellegaard and Engel, 2016). In contrast, whole-genome sequencing offers a more comprehensive approach by providing the complete genetic information of an organism, thus enabling a more accurate and reliable taxonomic classification (Fuentes-Pardo and Ruzzante, 2017). In the present study, *in silico* analysis was used to analyse and determine the antimicrobial biosynthesis potential of three strains of *Bacillus* sp. isolated from soil. The genomic analysis of *Bacillus* sp., performed using the HiSeq 2500 platform, provided valuable insights into its genetic composition and phylogenetic relationships. The *de novo* assembly resulted in genome sizes of 3,733,820 bp for KH1, 3,688,736 bp for KH2, and 4,176,670 bp for KH3. The corresponding N50 values were 440,094 bp, 378,244 bp, and 1,056,378 bp, respectively, with each genome comprising 26, 22, and 24 contigs (all ≥ 500 bp). The G+C content of the genomes was 41.2% for KH1 and 2, and 43.7% for KH3. These results, particularly the favorable N50 values, indicate that the genomes are well-structured and suitable for further analyses.

In the present study, *Bacillus* sp. strains were classified using the TYGS method for prokaryotic species identification. Although, *Bacillus* sp. Strain KH1 was previously identified as *B. pumilus*, the results revealed that *Bacillus* sp. strains are closely related to *Bacillus altitudinis* 41KF2b, with a digital DNA-DNA hybridization (dDDH) value of 86.1%, indicating a close genetic relationship and supporting their classification within the same species. Additionally, strain KH2 showed a close relationship to *Bacillus subtilis* PY79, with a high dDDH value of 92.8%, while KH3 was closely related to *Bacillus pumilus* NCTC 10337, with a dDDH value of 89.5%. According to the literature, a dDDH value above 70% typically indicates that organisms belong to the same species, and the value observed in the present study comfortably exceeds this threshold, thus providing strong evidence for their species-level classification (Richter and Rosselló-Móra, 2009, Hu et al., 2022). These findings are further corroborated by Orthologous Average Nucleotide Identity (OrthoANI) analysis. The results show that strain KH1 is most closely aligned with *Bacillus altitudinis* 41KF2b, yielding an ANI value of 98.45%, strain KH2 is closely related to *Bacillus subtilis* PY79 with an ANI value of 95.51%, and strain KH3 is closely related to *Bacillus pumilus* NCTC 10337 with an ANI value of

95.31%. Similarly, this value surpasses the commonly accepted threshold of 95% for species delineation (Jain et al., 2018) suggesting that these two organisms are likely the same species. In addition, the use of an outlier strain, *Escherichia coli* 11368, in these analyses provides a robust understanding of the taxonomy of *Bacillus* sp. Furthermore, all the above observations indicate that strain exhibits a high degree of similarity to *Bacillus altitudinis*, *Bacillus pumilus*, and *Bacillus subtilis*.

Generally, genome functional annotation aims to assign biological meaning to proteins or genes, and there is a variety of pipelines to choose from. However, there are no defined standards for evaluating functional annotation tools. In the present study of two pipelines were used, i.e, RAST and (COG) eggnoG-Mapper. The genome annotation revealed the bacterial strains as potential antimicrobial-producing and plant growth-promoting bacteria, with abilities that could produce a wide array of antimicrobial compounds and also enhance bacteria's resilience in competitive environments. This analysis aided the identification of putative genes organized into functional groups related to antimicrobial production and plant growth promotion, such as metabolic pathways, stress response, nitrogen fixation, and secondary metabolite production. Notably, the presence of genes associated with auxin biosynthesis, including IAR, which encodes indole-acetaldehyde reductase, suggests that *Bacillus* strains can synthesize indole-3-acetic acid (IAA). This critical plant growth hormone regulates various physiological processes in plants (Duca and Glick, 2020). A previous study by Duca et al. (2020), has demonstrated that IAA produced by PGPB, not only that they enhance bacterial stress tolerance, but they also stimulate plant growth through various mechanisms. However, Jahn et al. (2021) also demonstrated that IAA serves as a key signal in the communication between host plants and endophytes, with endophytes synthesizing this hormone to influence IAA homeostasis in plants. Moreover, this ability to produce IAA may enhance root development and overall plant vigor, thus promoting growth under suboptimal conditions (Zhao et al., 2020).

Furthermore, the genome annotation revealed additional traits that contribute to plant interactions and defence mechanisms. For instance, the identification of several genes such as the ATFS gene, which encodes flavonol 3-sulfotransferase, indicates a role in modifying flavonols to enhance their stability and efficacy in plant defence signalling. Additionally, the genome analysis revealed various cofactors, vitamins, and pigments, including quinone cofactors, riboflavin, and pyridoxine, which are essential for numerous metabolic processes and contribute to the overall adaptation of the bacteria even in harsh environments. Moreover,

the presence of enzymes involved in the biosynthesis of aromatic amino acids and secondary metabolites further highlights the strain's potential to produce antimicrobial compounds through diverse biochemical pathways. The identification of genes responsible for siderophore synthesis and iron acquisition in *Bacillus* sp. strains is particularly noteworthy, as iron is a critical nutrient for both plants and microorganisms. Siderophores facilitate iron uptake in nutrient-limited environments, thereby enhancing plant growth and resilience (Lal et al., 2024). In addition, these findings align with existing literature that emphasizes the significance of microbial interactions in promoting plant growth and resilience against biotic and abiotic stresses (Kumar and Verma, 2018, Malviya et al., 2020, Enebe and Babalola, 2018, Jahn et al., 2021). For example, studies have shown that beneficial bacteria can enhance nutrient availability and modulate phytohormone levels, leading to improved plant performance (Poveda and González-Andrés, 2021).

For a more comprehensive analysis, functional annotation of the Cluster of Orthologous Groups (COGs) was conducted using eggNOG-mapper. This approach provided a deeper understanding of the genomic capabilities of the strains, particularly regarding their potential for antimicrobial production. The annotation identified a total of 3,660 (KH1), 3966 (KH2) and 3937 (KH3) genes across four main functional categories: metabolism, cellular processes and signalling, information storage and processing, and poorly characterized genes. Notably, the largest group was associated with metabolism for all three strains, indicating that metabolic processes play a significant role in the strain's functionality. Similar studies on *Bacillus* species have demonstrated that metabolic pathways related to nutrient cycling, such as nitrogen fixation and phosphate solubilization, are crucial for promoting plant health and productivity (Yadav et al., 2021, Mashabela et al., 2022).

The Gene Ontology (GO) annotations further emphasize the functional capabilities of *Bacillus* sp. strains, with majority of genes linked to biological processes. However, within the metabolism category, our analysis showed that the amino acid transport and metabolism functions emerged as prominent functions, which are essential for synthesizing key compounds that could serve as antimicrobial agents. Moreover, the identification of genes related to cellular processes, particularly those involved in cell wall/membrane/ envelope biogenesis and signal transduction mechanisms, suggests that this strain may also enhance plant-microbe interactions through effective communication and structural support. The presence of auxin biosynthesis-related genes, such as IAR for indole-acetaldehyde reductase and ATFS for flavonol 3-

sulfotransferase, highlights the strain's potential to regulate phytohormones that promote root development and overall plant vigor.

The anti-SMASH (Blin et al., 2017) analysis of the *Bacillus* sp. strains genome revealed a significant potential for secondary metabolite production, identifying gene clusters associated with various biosynthetic pathways, these findings were compared with existing secondary metabolite gene clusters in the NCBI database. Notably, the presence of clusters related to well-known metabolites such as Bacillibactin, Subtilosin A, Bacilysin, Fengycin, Plipastatin, and Bacillibactin highlights the strain's capability to synthesize bioactive compounds that may have applications in agriculture and biotechnology. The classification of these metabolites into different biosynthetic gene clusters, including non-ribosomal peptide (NRP) and polyketide synthases (PKS), highlights the genetic diversity and metabolic potential of this soil-borne bacterium. However, the low similarity (less than 70%) of most predicted metabolites to known natural products suggests that *Bacillus* sp. strains may possess the genetic machinery to produce novel compounds that have yet to be characterized. Through comprehensive genome sequencing, Hudson et al. (2021) demonstrated that *Bacillus altitudinis* contains a cluster of six genes implicated in the synthesis of secondary metabolites with antimicrobial properties.

The analysis of antimicrobial resistance genes (ARGs) in the genome of *Bacillus* sp. strains, conducted using the Comprehensive Antibiotic Resistance Database (CARD), reveals a significant capacity for resistance against various classes of antibiotics and disinfectants. The identification of genes associated with resistance mechanisms, including those responsible for phenicol antibiotics and glycopeptides, highlights the strain's potential adaptability to its environment. Notably, the presence of multidrug efflux pump genes, such as *qacJ* and *qacG*, indicates that *Bacillus* sp. strains can effectively expel quaternary ammonium compounds (QACs), which are commonly used disinfectants even in agricultural settings (Jennings et al., 2015). This mechanism is crucial for survival in environments exposed to antimicrobial agents, suggesting that the strains have developed robust strategies to mitigate the effects of chemical stressors. Furthermore, the genome encodes the *cat86* gene, which produces catechol 1,2-dioxygenase, an enzyme involved in degrading aromatic compounds, including chlorinated hydrocarbons (Kumari et al., 2023.) and certain herbicides (Xu et al., 2021). However, this capability not only contributes to the strain's resistance profile but also highlights its potential role in bioremediation efforts. The analysis revealed that antibiotic target alteration mechanisms were more prevalent than those related to antibiotic efflux or inactivation, indicating a complex resistance strategy that may involve modifying the target sites of

antibiotics. This finding aligns with trends observed in other bacterial species where target alteration is a significant mechanism of resistance (Lambert, 2005, Darby et al., 2023).

5.6 Conclusion

This study highlights the use of whole genome sequencing as a useful technique in microbial taxonomy and has proved to be effective for clarifying species boundaries among closely related bacterial strains. Our results have shown that 16S rRNA gene sequencing was sufficient to identify KH1 as *B. pumilus*, however, whole genome analysis reclassified the taxonomic identity as *B. altitudinis*. Moreover, *Bacillus* sp. genome possesses genes encoding for bioproducts with potential application in biocontrol, biofertilization, and bioremediation. The findings demonstrate the potential to produce diverse chemical compounds through biochemical interactions, thereby improving their productivity and resilience against environmental stresses. Additionally, the presence of antimicrobial resistance genes (ARGs) such as *qacJ* and *qacG* suggests that this strain possesses a versatile arsenal for surviving in competitive and chemically challenging environments, which could have important implications for its use in agricultural and biotechnological applications. The insights gained from this study open avenues for discovering new bioactive substances that could be harnessed for various industrial applications, including pharmaceuticals and agricultural biocontrol agents. Therefore, further research is recommended to explore the functional roles of these resistance genes and their potential impacts on microbial ecology and plant health, as well as to investigate the practical applications of *Bacillus* sp. strains in sustainable agricultural practices.

5.7 References

- ADIMPONG, D. B., SØRENSEN, K. I., THORSEN, L., STUER-LAURIDSEN, B., ABDELGADIR, W. S., NIELSEN, D. S., DERKX, P. M. & JESPERSEN, L. 2012. Antimicrobial susceptibility of *Bacillus* strains isolated from primary starters for African traditional bread production and characterization of the bacitracin operon and bacitracin biosynthesis. *Applied and environmental microbiology*, 78, 7903-7914.
- AIMAIER, R., LI, H., CAO, W., CAO, X., ZHANG, H., YOU, J., ZHAO, J., ZHANG, Q., YIN, L. & MEI, Q. 2023. The secondary metabolites of *Bacillus subtilis* strain Z15 induce apoptosis in hepatocellular carcinoma cells. *Probiotics and Antimicrobial Proteins*, 1-11.
- ALAM, K., ISLAM, M. M., GONG, K., ABBASI, M. N., LI, R., ZHANG, Y. & LI, A. 2022. In silico genome mining of potential novel biosynthetic gene clusters for drug discovery from *Burkholderia* bacteria. *Computers in Biology and Medicine*, 140, 105046.
- BAISHYA, J. & WAKEMAN, C. A. 2019. Selective pressures during chronic infection drive microbial competition and cooperation. *npj Biofilms and Microbiomes*, 5, 16.
- BLIN, K., WOLF, T., CHEVRETTE, M. G., LU, X., SCHWALEN, C. J., KAUTSAR, S. A., SUAREZ DURAN, H. G., DE LOS SANTOS, E. L., KIM, H. U. & NAVE, M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic acids research*, 45, W36-W41.
- CHEN, X. H., KOUMOUTSI, A., SCHOLZ, R., EISENREICH, A., SCHNEIDER, K., HEINEMEYER, I., MORGENSTERN, B., VOSS, B., HESS, W. R. & REVA, O. 2007. Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nature biotechnology*, 25, 1007-1014.
- DADGOSTAR, P. 2019. Antimicrobial resistance: implications and costs. *Infection and drug resistance*, 3903-3910.
- DARBY, E. M., TRAMPARI, E., SIASAT, P., GAYA, M. S., ALAV, I., WEBBER, M. A. & BLAIR, J. M. 2023. Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology*, 21, 280-295.
- DEVI, S., KIESEWALTER, H. T., KOVÁCS, R., FRISVAD, J. C., WEBER, T., LARSEN, T. O., KOVÁCS, Á. T. & DING, L. 2019. Depiction of secondary metabolites and antifungal activity of *Bacillus velezensis* DTU001. *Synthetic and Systems Biotechnology*, 4, 142-149.

- DUCA, D. R. & GLICK, B. R. 2020. Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. *Applied microbiology and biotechnology*, 104, 8607-8619.
- ELANDER, R. 2003. Industrial production of β -lactam antibiotics. *Applied microbiology and biotechnology*, 61, 385-392.
- ELLEGAARD, K. M. & ENGEL, P. 2016. Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. *Frontiers in microbiology*, 7, 1475.
- ELSHAGHABEE, F. M., ROKANA, N., GULHANE, R. D., SHARMA, C. & PANWAR, H. 2017. Bacillus as potential probiotics: status, concerns, and future perspectives. *Frontiers in microbiology*, 8, 1490.
- ENEBE, M. C. & BABALOLA, O. O. 2018. The influence of plant growth-promoting rhizobacteria in plant tolerance to abiotic stress: a survival strategy. *Applied microbiology and biotechnology*, 102, 7821-7835.
- FUENTES-PARDO, A. P. & RUZZANTE, D. E. 2017. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular ecology*, 26, 5369-5406.
- HAFEEZ, A. B., PEŁKA, K., BUZUN, K., WOROBO, R. & SZWEDA, P. 2023. Whole-genome sequencing and antimicrobial potential of bacteria isolated from Polish honey. *Applied Microbiology and Biotechnology*, 107, 6389-6406.
- HIFNAWY, M. S., FOU DA, M. M., SAYED, A. M., MOHAMMED, R., HASSAN, H. M., ABOUZID, S. F., RATEB, M. E., KELLER, A., ADAMEK, M. & ZIEMERT, N. 2020. The genus *Micromonospora* as a model microorganism for bioactive natural product discovery. *RSC advances*, 10, 20939-20959.
- HU, S., LI, K., ZHANG, Y., WANG, Y., FU, L., XIAO, Y., TANG, X. & GAO, J. 2022. New insights into the threshold values of multi-locus sequence analysis, average nucleotide identity and digital DNA–DNA hybridization in delineating *Streptomyces* species. *Frontiers in Microbiology*, 13, 910277.
- JAHN, L., HOFMANN, U. & LUDWIG-MÜLLER, J. 2021. Indole-3-acetic acid is synthesized by the endophyte *Cyanodermella asteris* via a tryptophan-dependent and-independent way and mediates the interaction with a non-host plant. *International Journal of Molecular Sciences*, 22, 2651.
- JAIN, C., RODRIGUEZ-R, L. M., PHILLIPPY, A. M., KONSTANTINIDIS, K. T. & ALURU, S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications*, 9, 5114.

- JENNINGS, M.C., MINBIOLE, K.P. AND WUEST, W.M., 2015. Quaternary ammonium compounds: an antimicrobial mainstay and platform for innovation to address bacterial resistance. *ACS infectious diseases*, 1(7), pp.288-303.
- JIRAKKAKUL, J., KHOIRI, A.N., DUANGFOO, T., DULSAWAT, S., SUTHEEWORAPONG, S., PETSONG, K., WATTANACHAISAREEKUL, S., PAENKAEW, P., TACHALEAT, A., CHEEVADHANARAK, S. AND PROMMEENATE, P., 2023. Insights into the genome of *Methylobacterium* sp. NMS14P, a novel bacterium for growth promotion of maize, chili, and sugarcane. *Plos one*, 18(2), p.e0281505.
- KHAN, D. A., BANERJI, A., BERNSTEIN, J. A., BILGICER, B., BLUMENTHAL, K., CASTELLS, M., EIN, D., LANG, D. M. & PHILLIPS, E. 2019. Cephalosporin allergy: current understanding and future challenges. *The Journal of Allergy and Clinical Immunology: In Practice*, 7, 2105-2114.
- KHWATHISI, A., MADALA, N. E., TRAORE, A. N. & SAMIE, A. 2024. Bioprospecting of soil-borne microorganisms and chemical dereplication of their anti-microbial constituents with the aid of UPLC-QTOF-MS and molecular networking approach. *PeerJ*, 12, e17364.
- KUMAR, A. & VERMA, J. P. 2018. Does plant—microbe interaction confer stress tolerance in plants: a review? *Microbiological research*, 207, 41-52.
- KUMARI, S. AND DAS, S., 2023. Bacterial enzymatic degradation of recalcitrant organic pollutants: catabolic pathways and genetic regulations. *Environmental Science and Pollution Research*, 30(33), pp.79676-79705.
- LAMBERT, P. A. 2005. Bacterial resistance to antibiotics: modified target sites. *Advanced drug delivery reviews*, 57, 1471-1485.
- LI, J., XIE, S., AHMED, S., WANG, F., GU, Y., ZHANG, C., CHAI, X., WU, Y., CAI, J. & CHENG, G. 2017. Antimicrobial activity and resistance: influencing factors. *Frontiers in pharmacology*, 8, 364.
- LAL, M.K., TIWARI, R.K., ADAVI, S.B., KUMAR, A., BAMATOV, I., IVANOVA, E., BEHERA, L., JENA, R. AND KUMAR, R., 2024. Deciphering phenomics approaches for understanding plant–microbe interactions in nutrient absorption and utilization. *Plant Physiology Reports*, pp.1-17.
- MALVIYA, D., SINGH, U. B., SINGH, S., SAHU, P. K., PANDIYAN, K., KASHYAP, A. S., MANZAR, N., SHARMA, P. K., SINGH, H. & RAI, J. P. 2020. Microbial

- interactions in the rhizosphere contributing crop resilience to biotic and abiotic stresses. *Rhizosphere Microbes: Soil and Plant Functions*, 1-33.
- MASHABELA, M. D., TUGIZIMANA, F., STEENKAMP, P. A., PIATER, L. A., DUBERY, I. A. & MHLONGO, M. I. 2022. Untargeted metabolite profiling to elucidate rhizosphere and leaf metabolome changes of wheat cultivars (*Triticum aestivum* L.) treated with the plant growth-promoting rhizobacteria *Paenibacillus alvei* (T22) and *Bacillus subtilis*. *Frontiers in Microbiology*, 13, 971836.
- MAUMELA, P., KHWATHISI, A., MADALA, N. E. & SEREPA-DLAMINI, M. H. 2024. In silico biotechnological potential of *Bacillus* sp. strain MHSD_37 bacterial endophyte. *BMC genomics*, 25, 399.
- MICHAEL, C. A., DOMINEY-HOWES, D. & LABBATE, M. 2014. The antimicrobial resistance crisis: causes, consequences, and management. *Frontiers in public health*, 2, 145.
- NAGHAVI, M., VOLLSET, S. E., IKUTA, K. S., SWETSCHINSKI, L. R., GRAY, A. P., WOOL, E. E., AGUILAR, G. R., MESTROVIC, T., SMITH, G. & HAN, C. 2024. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet*, 404, 1199-1226.
- NAINU, F., PERMANA, A. D., DJIDE, N. J. N., ANJANI, Q. K., UTAMI, R. N., RUMATA, N. R., ZHANG, J., EMRAN, T. B. & SIMAL-GANDARA, J. 2021. Pharmaceutical approaches on antimicrobial resistance: prospects and challenges. *Antibiotics*, 10, 981.
- NEPHALI, L., STEENKAMP, P., BURGESS, K., HUYSER, J., BRAND, M., VAN DER HOOFT, J. J. & TUGIZIMANA, F. 2022. Mass spectral molecular networking to profile the metabolome of biostimulant bacillus strains. *Frontiers in plant science*, 13, 920963.
- PAYNE, J., BELLMER, D., JADEJA, R. & MURIANA, P. 2024. The Potential of *Bacillus* species as probiotics in the food industry: A Review. *Foods*, 13, 2444.
- PETTI, C., POLAGE, C. & SCHRECKENBERGER, P. 2005. The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. *Journal of clinical microbiology*, 43, 6123-6125.
- PORETSKY, R., RODRIGUEZ-R, L. M., LUO, C., TSEMENTZI, D. & KONSTANTINIDIS, K. T. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PloS one*, 9, e93827.
- POVEDA, J. & GONZÁLEZ-ANDRÉS, F. 2021. *Bacillus* as a source of phytohormones for use in agriculture. *Applied Microbiology and Biotechnology*, 1-17.

- PROCÓPIO, R. E. D. L., SILVA, I. R. D., MARTINS, M. K., AZEVEDO, J. L. D. & ARAÚJO, J. M. D. 2012. Antibiotics produced by *Streptomyces*. *Brazilian Journal of Infectious Diseases*, 16, 466-471.
- RICHTER, M. & ROSSELLÓ-MÓRA, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, 106, 19126-19131.
- SANSINENEA, E. 2019. *Bacillus* spp.: As plant growth-promoting bacteria. *Secondary metabolites of plant growth promoting rhizomicroorganisms: Discovery and applications*, 225-237.
- UDDIN, T. M., CHAKRABORTY, A. J., KHUSRO, A., ZIDAN, B. R. M., MITRA, S., EMRAN, T. B., DHAMA, K., RIPON, M. K. H., GAJDÁCS, M. & SAHIBZADA, M. U. K. 2021. Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *Journal of infection and public health*, 14, 1750-1766.
- WANG, X., HAO, G., ZHOU, M., CHEN, M., LING, H. & SHANG, Y. 2023. Secondary metabolites of *Bacillus subtilis* L2 show antiviral activity against pseudorabies virus. *Frontiers in Microbiology*, 14, 1277782.
- XU, A., ZHANG, X., WU, S., XU, N., HUANG, Y., YAN, X., ZHOU, J., CUI, Z. AND DONG, W., 2021. Pollutant degrading enzyme: catalytic mechanisms and their expanded applications. *Molecules*, 26(16), p.4751.
- YADAV, R., ROR, P., RATHORE, P., KUMAR, S. & RAMAKRISHNA, W. 2021. *Bacillus subtilis* CP4, isolated from native soil in combination with arbuscular mycorrhizal fungi promotes biofortification, yield and metabolite production in wheat under field conditions. *Journal of Applied Microbiology*, 131, 339-359.
- ZHAO, T., DENG, X., XIAO, Q., HAN, Y., ZHU, S. & CHEN, J. 2020. IAA priming improves the germination and seedling growth in cotton (*Gossypium hirsutum* L.) via regulating the endogenous phytohormones and enhancing the sucrose metabolism. *Industrial Crops and Products*, 155, 112788.
- ZHU, Z., PENG, Q., MAN, Y., LI, Z., ZHOU, X., BAI, L. & PENG, D. 2020. Analysis of the antifungal properties of *Bacillus velezensis* B-4 through a bioassay and complete-genome sequencing. *Frontiers in Genetics*, 11, 703.

Chapter 6: Multi-parametric optimization of Data Dependent Acquisition towards more holistic bacterial metabolite coverage through Molecular networking

6.1 Summary

Despite its undeniable importance and biotechnological potential application, *Bacillus subtilis* strain has historically received limited attention, particularly in relation to its unique arsenal of specialized metabolites. The structural diversity of these compounds offers great analytical challenges, since there is no universal acquisition setting that can be applied to achieve their comprehensive coverage. Therefore, the use of single experimental setup inevitably hinders the comprehensive metabolite coverage which would affect the outputs. To address this, we propose employing a design of experiment (DoE) approach through central composite design (CCD) to enhance the metabolites detection and broaden the coverage of data-dependent acquisition (DDA) mode of the UHPLC-qTOF-MS technique. Our study reveals that altering collision energy significantly enhances metabolite coverage compared to adjusting DDA threshold of detection. Furthermore, the ability of GNPS-based molecular network models to annotate metabolites is greatly influenced by data acquisition settings, particularly affecting MS² data. Interestingly, molecular networks constructed from averaged spectral data obtained through randomly selected DDA settings outperform those generated using customized settings through DoE modelling. This study demonstrates that in untargeted LC-MS metabolomics, both collision energy and intensity threshold independently enhance metabolite coverage in untargeted metabolomics. However, their combined use results in even greater coverage. Consequently, we recommend adopting group-based optimization over single-point optimization for more comprehensive metabolite coverage and in-depth exploration. However, cautioned should be taken in order to balance between robust data and redundancy.

Keywords: *Bacillus subtilis*, Metabolites, LC-MS, Molecular Networking, Metabolomics, Data-Dependent Acquisition (DDA)

6.2 Introduction

Microorganisms, particularly bacteria, play crucial roles in the natural environment through various interactions with plants. These interactions can be mutualistic, such as nitrogen fixation (Maseko et al., 2020), or parasitic, affecting plant health (Bui and Desaegeer, 2021). The interactions are facilitated by environmental sensing and the exchange of signals into molecular and genetic information (Gruet et al., 2023, Kreve and Dos Reis, 2021). However, to ensure their survival and adaptation to their environment, microorganism must produce a diverse range of secondary metabolites, which significantly impact their interaction with both biotic and abiotic factors (Mhlongo et al., 2020). Key mechanisms include the exchange of secondary metabolites (Tong et al., 2023) production of antimicrobial compounds (Meyer et al., 2024), signalling molecules (Khoshru et al., 2023), and genetic elements (Anguita-Maeso et al., 2023), influencing microbial colonization and establishment (Pang et al., 2021). These bacterial metabolites serve as antibiotics, pigments, growth hormones, and antitumor agents, highlighting their significance in maintaining human health and well-being (Minhas et al., 2024).

Given that these compounds are not essential for the microorganism's growth and development and their production is regulated by cryptic gene clusters, which are inactive under normal conditions, their comprehensive analysis is challenging (Hoskisson and Seipke, 2020). Previous studies have shown that the number of metabolites in bacteria can be affected by various factors, including the type of bacteria, the growth conditions, and the analytical methods used for detection (Gavriilidou et al., 2022). However, the use of advance analytical techniques such as liquid chromatography hyphenated to mass spectrometry (LC-MS) in biological systems has emerged as an effective technique for analysing metabolites due to its capability to detect and characterize compounds with diverse structural differences (Masike et al., 2017, Nengovhela et al., 2022, Nephali et al., 2022). Additionally, these instruments generate large datasets, and handling and analysing such large datasets can be challenging and may require specialized data analysis strategies and computational resources (Aron et al., 2020).

Recently, efforts have been centered on tackling the challenge of mining mass spectral data, which has emerged as a key focus in initiatives aimed at improving metabolite profiling (Jarmusch et al., 2021). These recent approaches usually define the substructure and their relationship to molecular fingerprints, annotating chemical compounds classes and create mass

spectral networks from metabolomics data (Beniddir et al., 2021). Molecular networking, housed in the Global Natural Product Social Science (GNPS) is an effective tool for analysing and visualizing mass spectral data (Yang et al., 2013, Aron et al., 2020). This approach involves grouping compounds based on their spectral similarity, allowing for the visualization of chemical structural relationships between detected metabolites solely through their MS² spectral data (Aron et al., 2020, Wang et al., 2016). Previously, studies have used molecular networking to identify and explore the structural relationship between various compounds, aiding the discovery of novel metabolites and their potential application in medical fields (Nephali et al., 2022, Moyo et al., 2023). This technique provides insight into the diverse chemical profiles of bacterial population and facilitating the identification of unique bioactive compounds (Tlhapi et al., 2024). However, molecular networking approach heavily relies on the MS/MS spectra generated to deduce spectral similarities, which implies that the quality of MS² data is fundamentally essential for reducing redundancy in generating informative molecular networks (Ramabulana et al., 2023).

Over the last decade advanced MS data acquisition approaches such as data dependant acquisition (DDA) (Defossez et al., 2023) and data-independent acquisition (DIA) approaches, have been increasingly used to provide both high resolution and high throughput for a full metabolome composition of the organisms (Guo et al., 2020). DDA involves selecting specific ions for fragmentation based on their intensity or abundance, while DIA involves fragmenting all ions within a predetermined m/z range, regardless of their intensity or abundance. However, DDA is a more widely used technique and preferred for its high spectral quality and time efficiency, especially for metabolite profiling and annotation in untargeted metabolomics. Despite its limitations in capturing the complete spectrum of metabolites (metabolome), (Defossez et al., 2023) introduced eight guidelines for selecting technical parameters to enhance the effectiveness of DDA analyses. Given all these improvements, the chemical heterogeneity of metabolites poses a significant challenge in achieving holistic metabolite coverage even with an optimal LC-MS conditions, which therefore requires the continuous development of new technologies and standards to expand the coverage of metabolites and improve the quality and robustness of the data. However, effort to optimize the DDA parameters have been done through UHPLC-qTOF-MS analyses of plant extracts (Ramabulana et al., 2023) and orbitrap-based mass spectrometers (Stincone et al., 2023) for plant compounds, but none has been done for bacterial metabolites. In this study, we propose using a design of experiment (DoE) approach with central composite design (CCD) to improve

metabolite detection and broaden the coverage of DDA mode within the UHPLC-qTOF-MS technique. Specifically, we aim to assess the metabolic diversity of a *Bacillus subtilis* strain isolated from soil, using a multi-parametric approach. Despite its extensive study, particularly regarding its genome, *B. subtilis*'s specialized metabolites have been less thoroughly explored. While its genome spans approximately 4.2 million base pairs, with an estimated 4-5% devoted to antimicrobial compound production, research on its unique metabolite arsenal remains limited. Hence the present chapter aim to employ the multivariate analysis to enhance the detection of metabolites from *B. subtilis*.

6.3 Experimental

6.3.1 Bacterial cultivation and Metabolite extraction

The bacterial strain used in this study was *Bacillus subtilis* strain KH3 isolated from an agricultural environment as part of a previous study that reported its potential antimicrobial activity against several pathogenic strains (Khwathisi et al., 2024). Pure bacterial isolates were cultured in Nutrient Broth and incubated at 30 °C for 24 hours to facilitate bacterial cell harvesting. Each Erlenmeyer flask containing 200 mL of Nutrient broth received an inoculation of 1 mL from the bacterial stock solutions. Following incubation at 30 °C for 24 hours, bacterial cells were harvested by centrifugation at 5000 rpm for 15 minutes at 4 °C. The yielded supernatants were subsequently transferred into new 50 mL conical tubes. The supernatants were then filtered through 0.22 µm nylon filters into LC-MS glass vials, following established protocols (Ramabulana et al., 2021).

6.3.2 Liquid Chromatography Mass Spectrometry analysis

The analysis of bacterial methanolic extracts was conducted using a Shimadzu LC- q-TOF-MS, 9030 model equipped with a Shim-pack Velox C18 column (100 mm x 2.1 mm, particle size 2.7 µm), as previously described by (Maumela et al., 2024). The column temperature was maintained at 50 °C throughout the analysis. A 5 µL injection volume was used, and samples were separated over a 30-minute binary gradient at a constant flow rate of 0.3 mL/min. The gradient employed a solvent mixture of water with 0.1% formic acid (Eluent A) and methanol with 0.1% formic acid (Eluent B). Specifically, the gradient elution was as follows: Eluent B was maintained at 5% from 0 to 3 minutes, increased linearly from 5% to 40% between 3 and 8 minutes, further increased to 95% between 8 and 23 minutes, held isocratically at 95% from 23 to 25 minutes, returned to 5% from 25 to 27 minutes, and re-equilibrated at 5% from 27 to

30 minutes. The LC eluents were then introduced into a Quadrupole Time-of-Flight high-definition mass spectrometer operating in positive Electrospray ionization (ESI) mode. The q-TOF-MS conditions included a 400 °C heat block temperature, 250 °C desolvation line temperature, 42 °C flight tube temperature, and a nebulization and dry gas flow rate of 3 L/min. Data acquisition was performed in Data-dependent acquisition (DDA) mode, simultaneously capturing both MS¹ and MS² spectra for ions within a mass-to-charge ratio (m/z) range of 100-2000 and an intensity threshold exceeding 1000. During MS² experiments, collision-induced dissociation employed argon gas with a collision energy of 30 eV and a spread of 5, utilizing sodium iodide (NaI) as a calibration solution to ensure high mass precision.

Table 9: Parameters that were varied for MS² scan in a DDA method

Intensity threshold optimization experiment	
Intensity thresholds (counts)	Collision energy (eV) with a spread of 5 ev
1000	30
2000	30
3000	30
4000	30
Collision energy optimisation experiment	
1000	10
1000	20
1000	30

6.3.3 Metabolite annotation and Molecular Networking

Molecular networks were generated using the Global Natural Product Social Molecular Networking (GNPS) platform (<http://gnps.ucsd.edu>). Prior to uploading to the online workflow, UHPLC-qTOF-MS/MS raw data was converted to. mzML format. Upon uploading spectral data, MS² fragment ions within +/- 17 Da of the precursor m/z were filtered out. Only the top 4 fragment ions within a +/- 50 Da window were selected for MS/MS spectra window filtering. An MS-CLUSTER technique was used to cluster the data. The MS² fragment ion and a mass tolerance for precursor ion were both set at 0.02 Da. Edges (representing interconnections/similarities between metabolites) were formed only when more than six matching peaks were present, and a minimum cosine score of 0.7 was exceeded. In the network, only nodes that ranked within each other's top 10 similar nodes (TopK value of 10) were retained, with a maximum of 100 nodes per molecular family. All matched and some

mismatched nodes were annotated based on empirical formulas derived from accurate mass and fragmentation patterns obtained from MS² studies. These annotated metabolites were further validated against literature and searched against common dereplication databases for natural products (e.g., KNApSAcK). Metabolite annotation adhered to the Metabolomics Standards Initiative (MSI) level 2 guidelines (Sumner et al., 2007). Classical molecular networking and Spec2vec methodologies were employed to assess the number of nodes, library IDs, and self-loops within the networks. The generated networks were visualized using Cytoscape 3.10.1.

6.3.4 Statistical analysis

The central composite design response surface methodology (CCD RSM) was employed for experimental design and analysis to establish the relationship between factors, aiming to optimize metabolite coverage using Design-Expert. A two-level full factorial CCD was implemented, comprising a total of 10 experimental runs (including two repetitions). Model parameters and significance were assessed at $p < 0.005$, with the model's adequacy evaluated based on the coefficient of regression (R^2) obtained from the analysis of variance (ANOVA). The model fit produced a response surface, depicting the behaviour of the response variable. These plots facilitated the identification of optimized factor ranges that maximize the desired response. Upset plots and Venn diagrams were created using the online platform (<https://www.bioinformatics.com.cn>) for data analysis and visualization.

6.4 Results

Bacterial extracts were analysed using liquid chromatography-mass spectrometry (LC-MS) operating in Data-Dependent Acquisition (DDA) mode. The generated data were processed using molecular networking and visualized with Cytoscape. In the current study, the effect of collision energy was assessed to examine its effect on metabolite coverage when the threshold of intensity was kept constant. Using GNPS-based molecular networking, the MS/MS data were organised into nodes and clusters, with the nodes indicating number of metabolite and clusters representing molecular families of metabolites/compounds that showed spectral similarities (Aron et al., 2020). Herein, the number of nodes showed to increase with the increase of the collision energy, with 10eV exhibiting the fewest metabolites and 30eV displaying the highest when threshold of intensity was kept at 1000, as depicted in Figure 31.

To further explore the impact of threshold of intensity on metabolite coverage, molecular networks were constructed using the MS/MS data when the collision energy was kept constant at 10eV. The influence of the threshold of intensity was distinctly evident in the generated molecular networks (Figure 32). Lowering the threshold allowed more ions to be detected, resulting in a broader coverage of metabolites within the network. Specifically, Figure 32D, where the threshold of intensity was set at 4000, detected the fewest number of metabolites. In contrast, Figure 32B, with a threshold of intensity was set at 2000, detected the highest number of metabolites. This demonstrates that a lower intensity threshold enhances the detection sensitivity, leading to a more comprehensive identification of metabolites in the analysis.

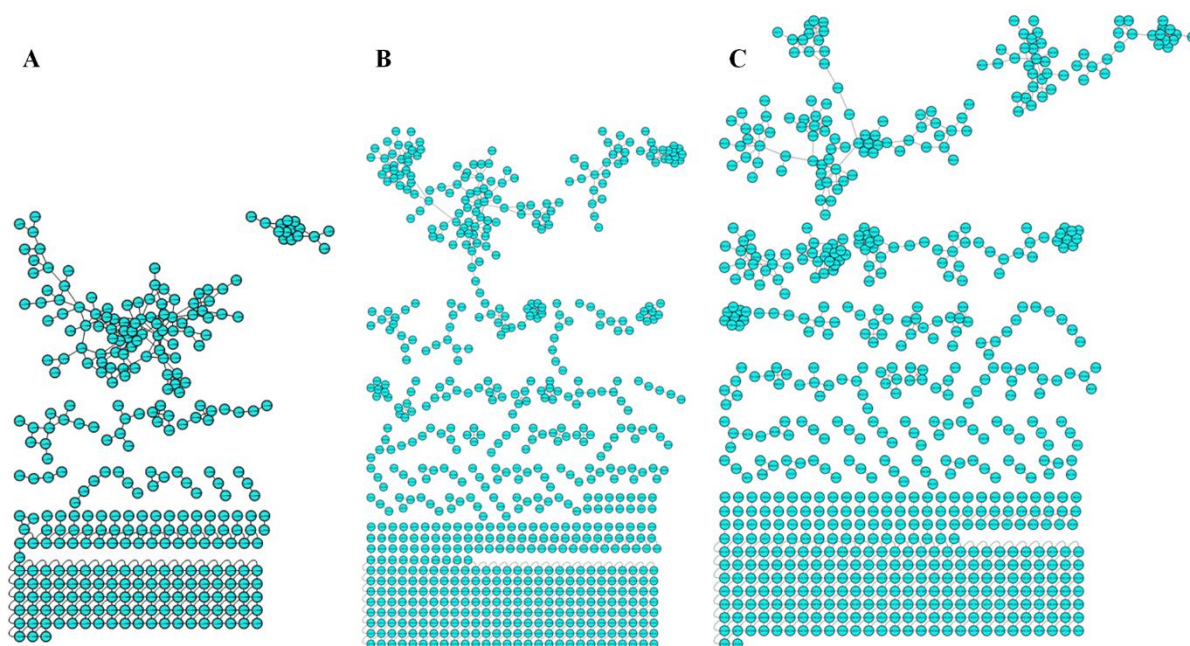


Figure 31: Molecular networks derived from bacterial extracts illustrating the effect of collision energy on MS data acquisition. (A) collision energy set at 10eV with a 1000 threshold. (B) collision energy set at 20eV with a 1000 threshold. (C) collision energy set at 30eV with a 1000 threshold.

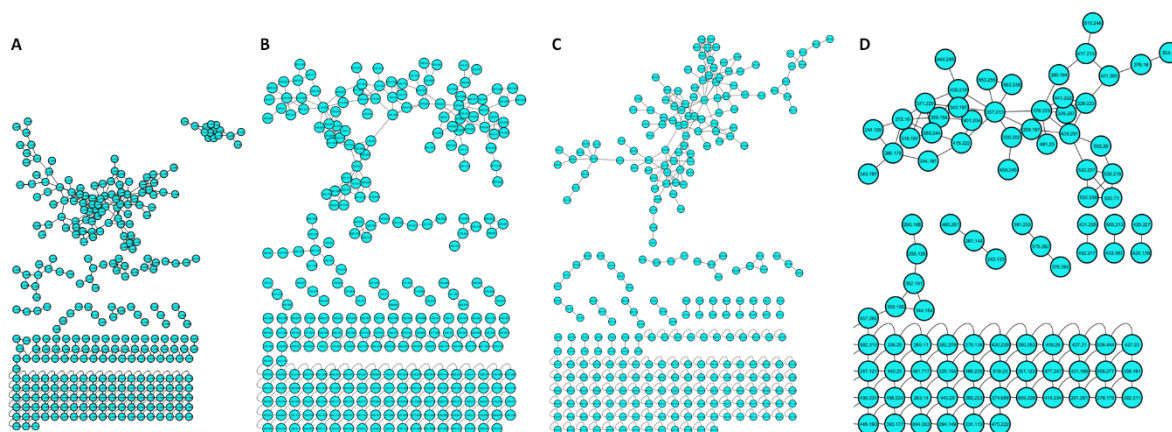


Figure 32: Molecular networks derived from bacterial extracts illustrating the effect of threshold intensity on MS data acquisition. (A) 1000 threshold with collision energy set at 10eV. (B) 2000 threshold with collision energy set at 10eV. (C) 3000 threshold with collision energy set at 10eV. (D) 4000 threshold with collision energy set at 10eV.

Additionally, to enable the identification of optimal conditions that would improve the metabolite coverage, we opted to investigate the two factors (collision energy and threshold of intensity) simultaneously using a Design of Experiments (DoE) approach. Subsequently, the models generated were used to acquire MS data, which was then used to construct molecular networks for analysis and visualization. Figure 33, shows the different molecular networks generated from the DoE analysis of threshold and collision energy, visualized in Cytoscape: (A) threshold of 2000 with 20eV, (B) threshold of 3000 with 10eV, (C) threshold of 1000 with 30eV, (D) threshold of 3000 with 30eV, and (E) threshold of 2000 with 20eV. Evidently, among the various metabolites detected in each condition, there exists a subset of shared metabolites, suggesting a common pool of metabolites detected across all conditions. These shared metabolites may possess structural features that enable their detectability regardless of experimental conditions. In addition, the effect of collision energy on the data is evident, as the collision energy increased, there was a corresponding rise in the number of detected metabolites, attributed to the increase in fragment ions. Furthermore, the impact of threshold

of intensity was also observed, such that decreasing the threshold of intensity resulted in a greater number of metabolites being detected (Figure 33F).

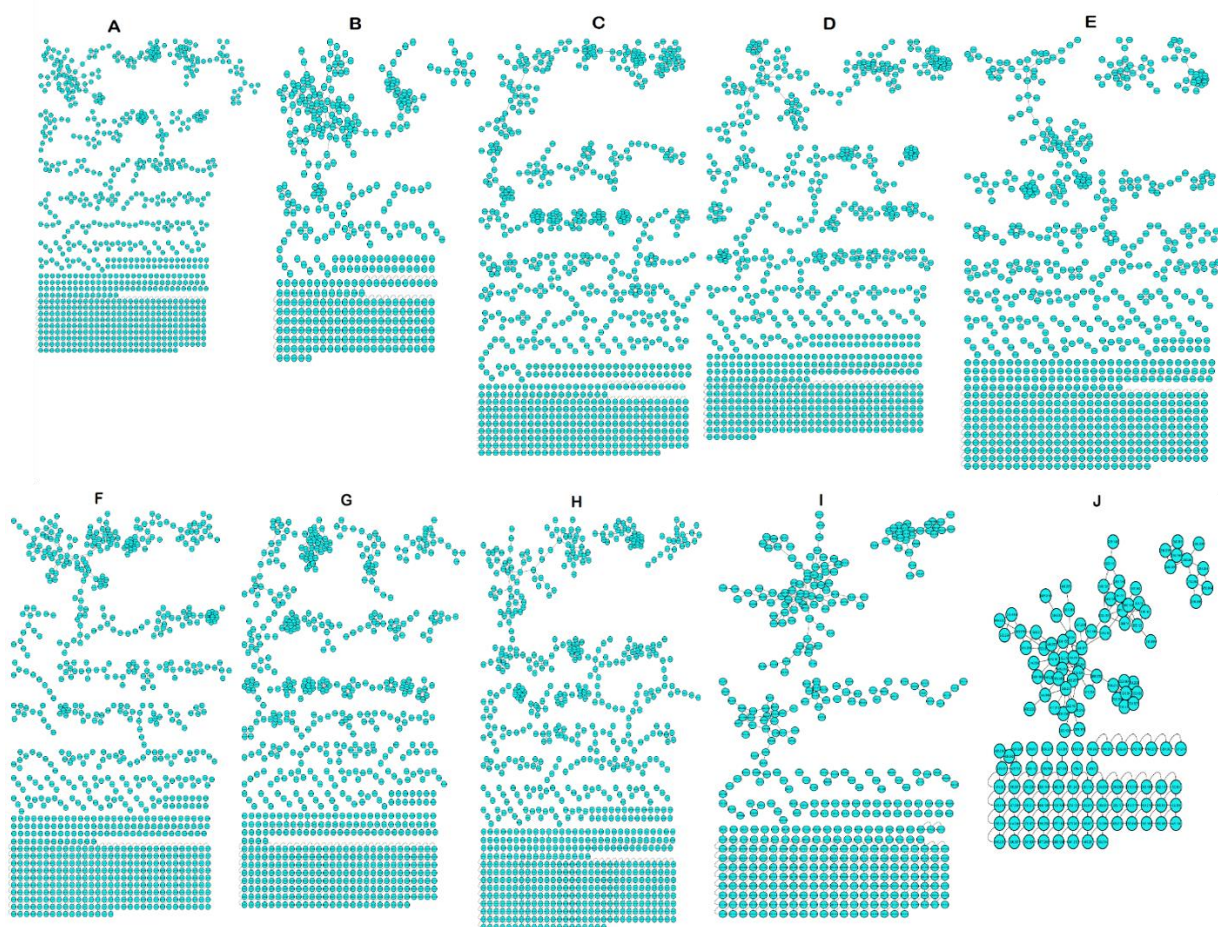


Figure 33: Showing the different molecular networks generated from the Design of Experiment (DOE) analysis of threshold and collision energy, visualized in Cytoscape. (A) threshold of 2000 with 20ev. (B) threshold of 3000 with 10ev. (C) threshold of 1000 with 30ev. (D) threshold of 3000 with 30ev. (E) threshold of 2000 with 20ev. (F) threshold of 585 with 20ev. (G) threshold of 2000 with 34ev. (H) Threshold of 3414 with 20ev. (I) threshold of 1000 with 10ev. (J) threshold of 2000 with 5ev.

To have an in-depth understanding of the synergistic effect of collision energy and threshold of intensity on metabolite coverage, the data derived from the molecular network was utilized to visually represent the distribution of metabolites captured using various conditions through an upset plot, (Figure 34A). The results showed that the shared metabolites across all conditions were 71, which approximately accounted 13.23%, while unique metabolites specific to individual conditions ranged from 0.18% to 1.67%. In addition, combining the spectral data in

the GNPS platform results in an increased coverage of the detected metabolites. This observation aligns with the findings depicted in the Venn diagram in Figure 34B, highlighting the interactions between the highest, lowest, and combined events. Similarly, when all conditions were combined, there was a significant increase in metabolite coverage. This suggests that while collision energy and threshold of intensity independently enhance metabolite coverage, the coverage is even greater when both parameters are combined. The results highlight the importance of optimizing both collision energy and threshold of intensity for comprehensive metabolite coverage.

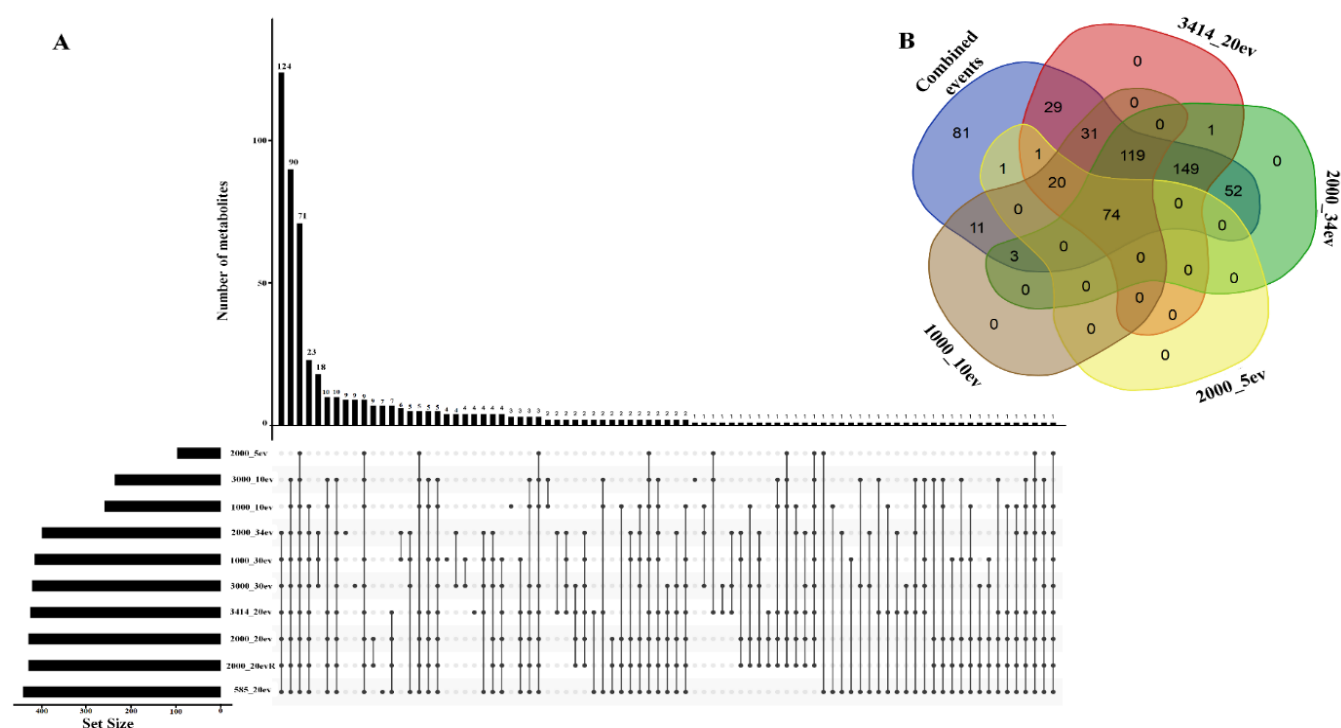


Figure 34: (A) Upset plot illustrating the overlap of different metabolites among the ten experimental conditions. Each circle represents a set of metabolites associated with a specific condition, with the size of the circle indicating the number of metabolites in that set. The bars along the top of the plot show the number of metabolites unique to each condition, while the intersections between circles represent the overlap of metabolites between conditions. (B) Venn diagram showed the overlapped metabolites between four different conditions with the combined events.

As mentioned, the present study aimed to maximize the coverage of metabolites from *Bacillus subtilis* by quantitatively and systematically assessing the effects of collision energy and threshold of intensity. The analysis revealed that *Bacillus subtilis* contained a diverse array of metabolite classes, including dipeptides, lipopeptides, and organic compounds. The annotation of the detected metabolites was facilitated by matching the data against the GNPS library. Notably, the annotation of the metabolites showed a marked increase when the data from all the experimental conditions were combined. While there were significant improvements in metabolite annotation observed with individual condition optimizations, such as increasing the collision energy and decreasing the threshold of intensity, the greatest improvement in annotation was achieved when the heterogeneous data from all the conditions were combined. This suggests that leveraging the complementary information captured across the varied experimental settings led to a more comprehensive metabolite coverage and identification. Furthermore, a cluster of biologically significant compounds (surfactins) were used as model to analyse the effect of threshold and collision energy simultaneously. As observed in figure 36, the collision energy and threshold of intensity were found to have a significant effect on the detection of metabolites. Precisely, lower collision energies and higher threshold of intensities led to fewer compounds being detected, whereas higher collision energies and lower threshold of intensities resulted in a greater number of compounds being detected. Additionally, some compounds were observed to form sodium (Na) adduct. Notably, these compounds with different adducts were clustered in distinct groups within the molecular network, with a greater number of metabolites in the M+Na group compared to the M+H group.

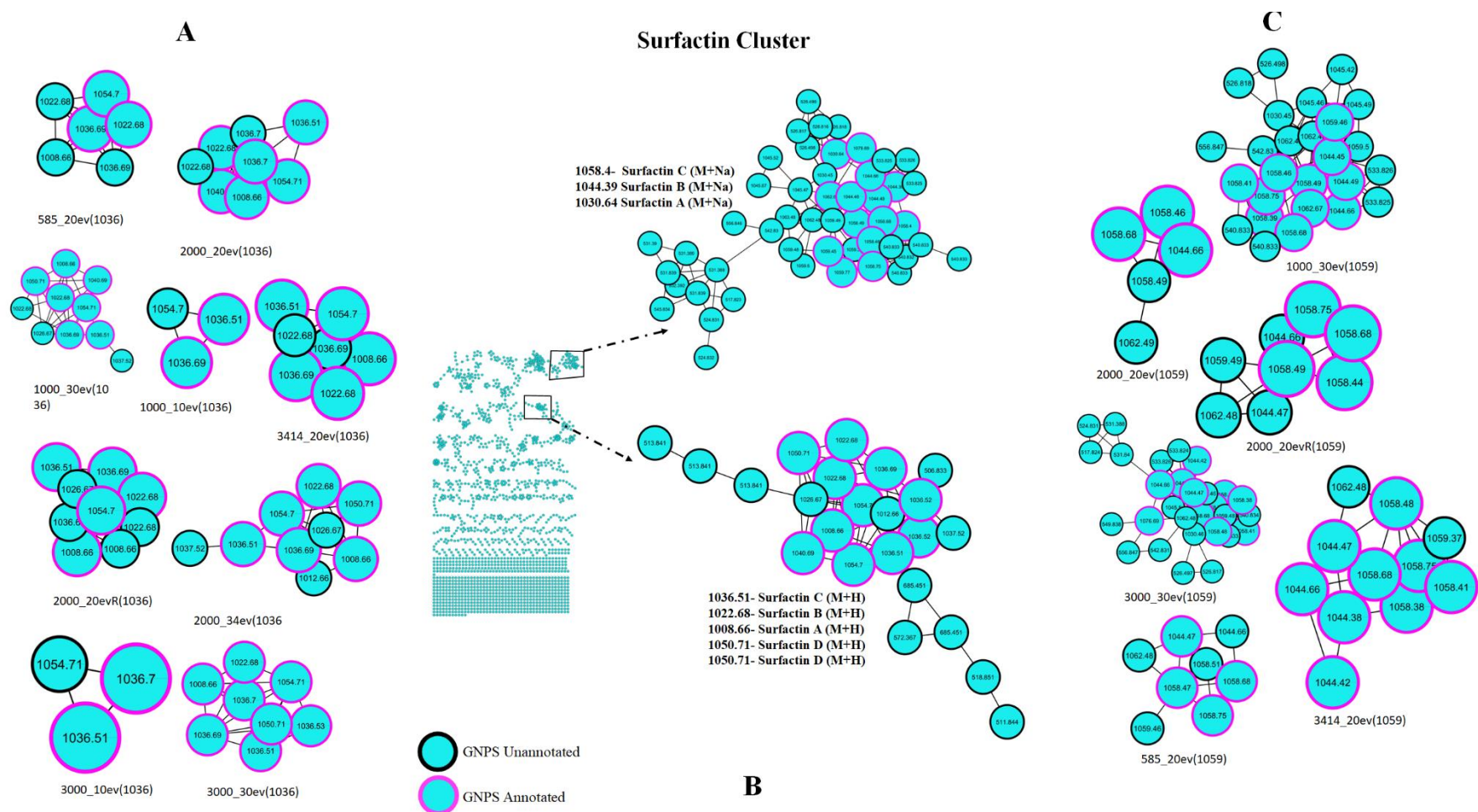


Figure 35: Molecular networking analysis of the methanolic extracts of *B. subtilis* generated from various conditions. (A) The molecular network shows distinct clusters of Surfactin at varying collision energies and thresholds of detected at M+H. (B) Improved clustering of surfactin is depicted when combining different conditions. (C) The clustering of surfactin detected in M+Na is influenced by varying collision energy and ion detection threshold across different conditions.

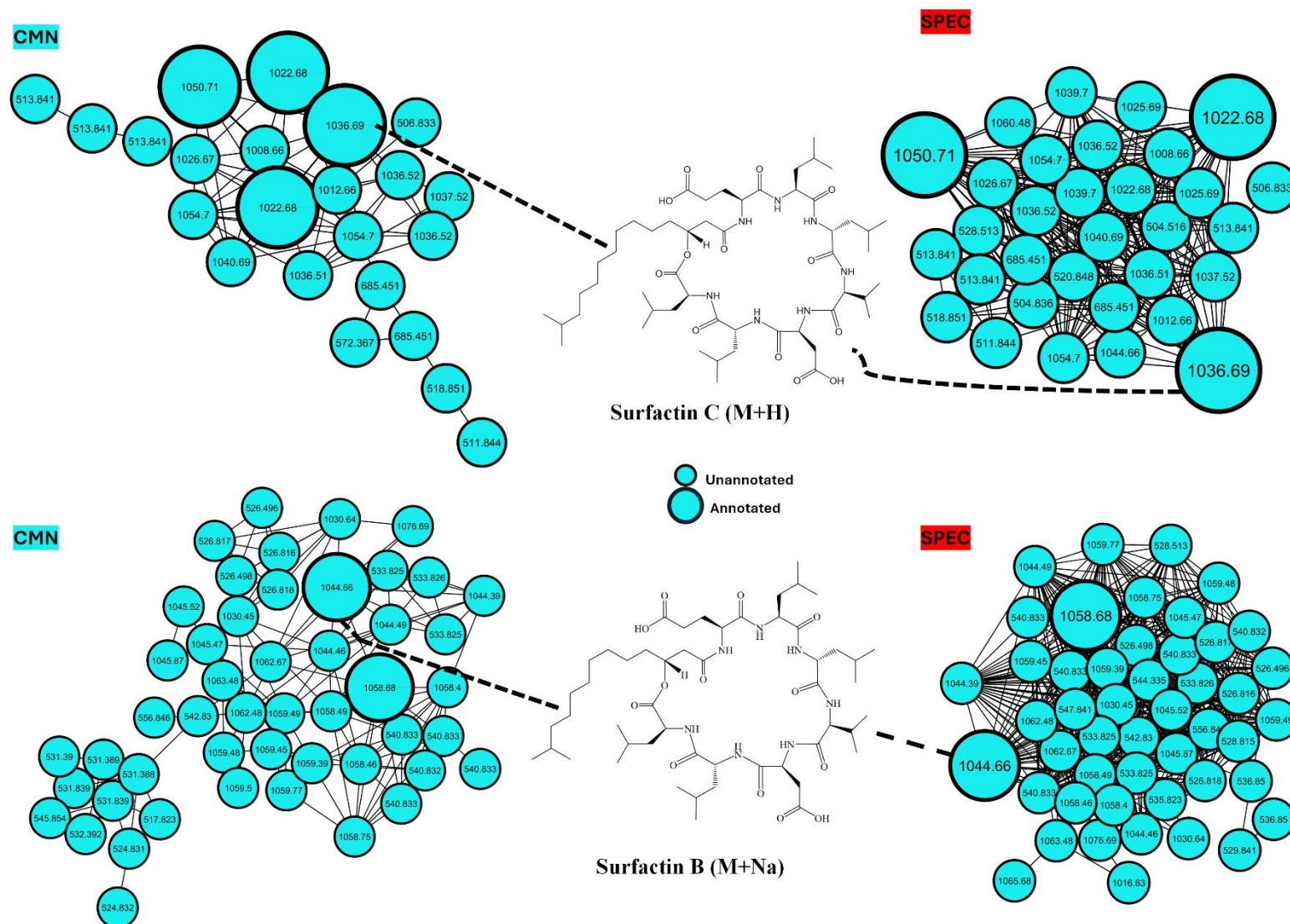


Figure 36 : Displays the classical molecular network and Spec2vec networks generated from ESI-positive MS/MS spectra, utilizing consolidated data settings to highlight variations in the clustering of related ion species across different methods. On the left side, the surfactin-extracted cluster is shown from classical molecular networking, while the right side depicts the surfactin-extracted cluster from Spec2vec. This comparison underscores the differences in how these methodologies organize and present molecular relationships within the *Bacillus subtilis* extracts.

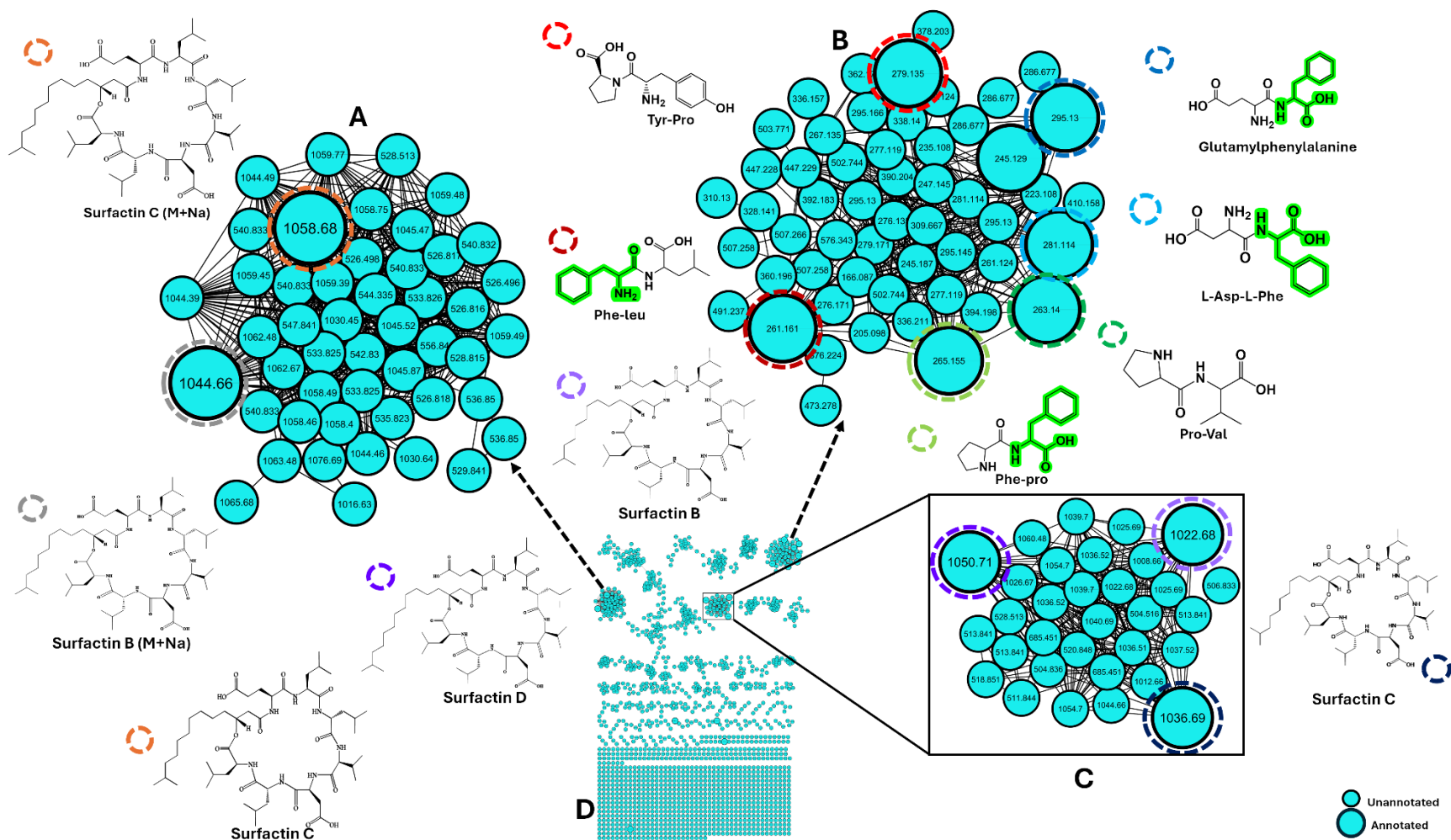


Figure 37: The molecular network of *Bacillus subtilis* extracts was analysed using liquid chromatography-tandem mass spectrometry with electrospray ionization in Positive mode. In the network, (A) Surfactins with M+Na adduct, (B) Dipeptides, (C) Surfactin with M+H adduct and (D) full molecular network. Each node represents a compound detected in the samples, with edges connecting nodes indicating spectral similarity between the compounds.

The findings from classical molecular networking showed that most clustered nodes were distributed across all DDA settings. However, a significant proportion of these nodes appeared to originate exclusively from settings characterized by high collision energy and low intensity of thresholds. To complement these observations, the Spec2vec workflow was employed in this study. Spec2vec enhances mass spectral similarity scoring by learning structural relationships, thereby providing additional insights into the molecular networking results (Huber et al., 2021). Inspecting the number of metabolites in both classical molecular networking and the Spec2vec network, there was minimal variation in clustering across individual settings (Figure 36). However, a notable finding was the increased inter- and intra-node connectivity observed in the Spec2vec network, indicating enhanced connections among nodes within each cluster. Supplementary Figure 40 highlights that acquisition settings used for (Supplementary Figure 40 K and J) detected the highest number of metabolites. Furthermore, significant improvements in the number of detected metabolites were evident when data from all settings were consolidated, as depicted in Supplementary Figure 40A. The Spec2vec network facilitated putative annotations, offering comprehensive and detailed insights into the chemistry and metabolome of *Bacillus subtilis*. Analysis of the spectral network, particularly focusing on dipeptides subclusters (Figure 37B), revealed distinct spectral signatures associated with phenylalanine (i.e Phe-pro, Phe-Leu, Glu-Phe, Asp-Phe). Moreover, as anticipated with *Bacillus subtilis*, surfactin A, B, C and D were identified as one of the prominent metabolite classes in the bacterial extract (Figure 37A&C). These surfactins were observed in two different clusters, highlighting their structural diversity. Within these clusters, certain nodes were not annotated via GNPS spectral library matching but are hypothesized to be structurally similar to surfactins.

As a results of the Response Surface Methodology (RSM), the study established the relationship between two variables, collision energy, and threshold of intensity detection, and the response value to determine the most favourable conditions for maximizing metabolite coverage. In the statistical analysis of RSM, 10 assays were conducted under varied experimental conditions (refer to Table 10). Design-Expert V8.0.6.1 software was utilized for visualization, generating 3D surface plot diagrams illustrating the interaction between the two factors based on the regression model of metabolite coverage yield from *Bacillus subtilis*. The 3D surface plots response in Figure 38 indicate that an increase in the threshold of intensity led to a decrease in metabolite coverage. Additionally, the increase of collision energy resulted in the increase of number of metabolites detected. The observation herein, aligns with the

observation from the *Pareto* chart (Figure 38C), indicating a significant ($p < 0.05$) linear and quadratic effect of collision energy on number of metabolites (Table 2). In addition, the predicted optimal conditions for comprehensive metabolite coverage from *Bacillus subtilis* (Figure 38B) were determined to be 26.93 eV and Threshold of intensity at 1137.5, with the desirability score of 1 (as seen in figure 38B).

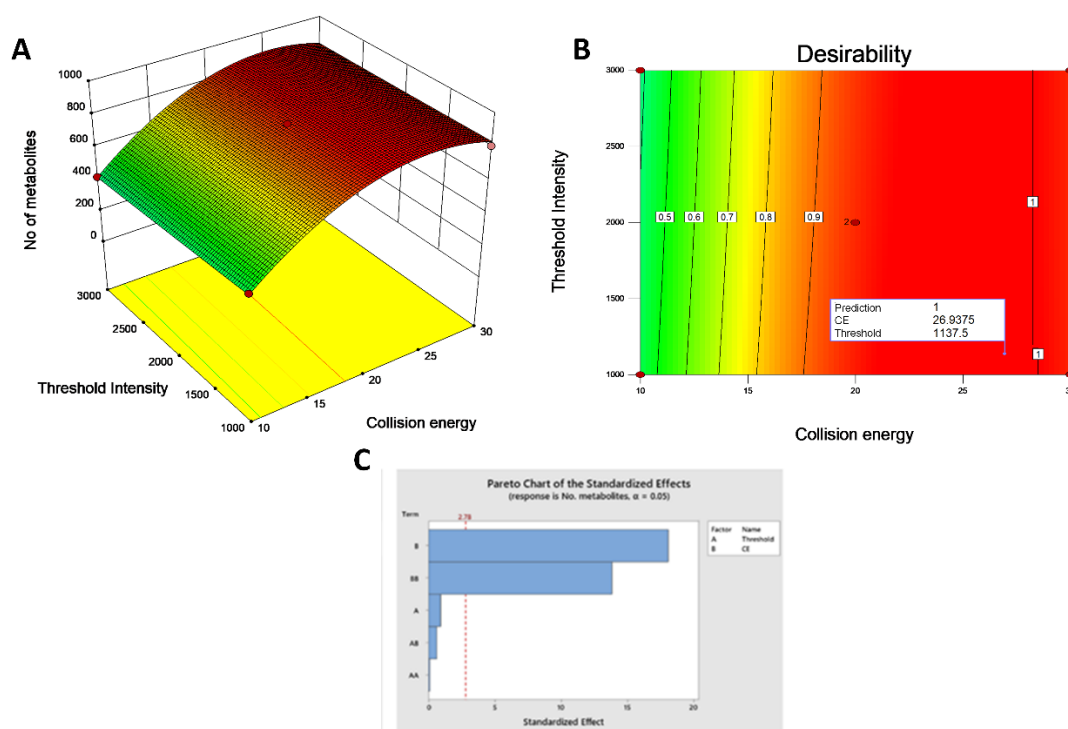


Figure 38: Response surfaces evaluating the Synergistic effect of collision energy and Threshold of intensity on the metabolite coverage. (A) 3D surface plot showing the response of Number of metabolites. (B) contour plot showing the predicted optimal conditions for a wider coverage. (C) Pareto chart indicating the significant factors affecting the response.

Table 10: Showing the "Model Summary Statistics": Focus on the model maximizing the "Adjusted R-Squared" and the "Predicted R-Squared".

Summary (detailed tables shown below)					
	Sequential	Lack of Fit	Adjusted	Predicted	
Source	p-value	p-value	R-Squared	R-Squared	
Linear	0.0493	0.0275	0.4559	0.1038	
2FI	0.9261	0.0249	0.3662	0.0167	
Quadratic	0.0003	0.1446	0.9842	0.9502	Suggested
Cubic	0.3768	0.1187	0.9881	0.8357	Aliased

6.5 Discussion

The metabolite profiling of biological sample remains a global challenge, this is due to the structural diversity that these metabolites possess (Wolfender et al., 2018). This diversity poses difficulties for developing innovative approaches to effectively profile metabolites within complex biological samples (Perez de Souza et al., 2021). In general, the selection of acquisition settings in untargeted metabolomics studies may introduce biases into the data, potentially leading to redundancy and compromising data robustness (Ivanisevic and Want, 2019). Considering that in metabolite profiling, the metabolites of interest are often unknown, which therefore means there is no acquisition settings that could be designed for them (Wolfender et al., 2015). As such, there is no universal acquisition settings that can be applied to achieve comprehensive coverage of metabolites effectively. Hence, the present study proposes the use of multi-parametric settings of DDA approach to improve the bacterial metabolite coverage through molecular networking.

In the field of metabolomics, the use of computational tools has become a solution for analysing the large datasets generated by advanced analytical instruments. Recently, molecular networking has emerged as a cornerstone in metabolomics data analysis, this is due to its capacity for data processing, dereplication, and visualization of chemical space (Nothias et al., 2020). In the current study, bacterial extracts were analysed using UHPLC-qTOF-MS in positive ionization mode, and the resulting MS data were processed and interpreted using molecular networking visualized on Cytoscape. As depicted in Figure 31 and 32, the molecular networks generated show that each condition captures a distinct set of metabolites. While the number of metabolites varies significantly among the conditions, this simply indicates that each condition independently captures its own set of metabolites. As such, the results from the current study corroborate with a study by Jia et al. (Jia et al., 2023) who reported that selecting specific tandem MS settings might lead to the loss of a limited number of metabolites. This underscores the importance of careful consideration when conducting comprehensive metabolite profiling studies.

Network-based methodologies in metabolomics have garnered significant attention and have emerged as a focal point in the space of natural products (Jarmusch et al., 2021). However, these network-based strategies heavily rely on MS² data for their analyses. In Data-Dependent Acquisition (DDA), one of the key parameters that significantly influences the appearance of tandem mass spectra is the collision energy. In our study, we investigated the impact of collision

energy on the generation of tandem mass spectra to achieve comprehensive coverage of the bacterial metabolome. Our results show that lower collision energy promotes fragmentation of compounds that require lower energy, resulting in some fragments having less spectral intensity. Comparatively, higher collision energy promotes the fragmentation of a larger number of metabolites, resulting in informative MS² spectra, including high molecular weight molecules. Our results align with previous work, such as the study by Ramabulana et al., (Ramabulana et al., 2023), which reported that increasing the collision energy generates a higher number of fragments in the MS² spectra for a given precursor ion. However, this could be due to the structural characteristics of many natural products such as diverse bond strength; meaning that the optimal collision energy for generating informative MS² spectra can vary significantly across different classes of natural products.

In comprehensive chemical analysis involving both known and unknown compounds with structural diversity, it is essential to fine-tune the threshold of intensity. In the present study, we investigated the effect of threshold of intensity and found that a lower threshold increases the likelihood of detecting noise, while a higher threshold can filter out noise but may miss lower-abundance ions as seen in Figure 32. This, however, influences the sensitivity and robustness of the analysis. The result from our study agrees with the result reported by Barupal et al., (Barupal et al., 2021), who reported that higher thresholds can lead to the rejection of a significant proportion of low-abundance peaks. Additionally, Defossez et al., (Defossez et al., 2023) highlighted that the threshold of intensity strongly depends on the signal intensity and background noise level of the mass spectrometer, and that setting the threshold too high or too low can lead to limitations in metabolite coverage.

Central composite design (CCD) is a widely used experimental design approach in optimization studies as it allows researchers to explore the effects of multiple independent variables (factors) on a dependent variable (response) efficiently. This approach has also been extensively used in metabolomics studies, from investigating the effect of ESI source parameters on metabolome coverage in untargeted LC/MS-based metabolomics (Tugizimana et al., 2018) to optimizing protocols for processing LC-MS-based metabolomic data (Zheng et al., 2013). In the present study, CCD was used to systematically investigate which are the model or conditions that can capture the full or comprehensive metabolite coverage. Building on previous experiments that used a one-variable-at-a-time (OVAT) approach, we further investigated the effect of these two variables, since the (OVAT) overlooked the interaction

between the two factors. Moreover, the two variables are key factors affecting the detection of compounds in mass spectrometry. The independent variables (collision energy and threshold intensity) were defined with the range from 10-30 eV and 1000-3000, respectively. In the present study, a total of ten (10) independent experimental setups consisting of varying settings of collision energy and threshold of intensity were generated as seen in figure 34. As evident in the vertex position of the 3D surface plot, the optimal conditions for metabolite coverage were 26.937eV and threshold of intensity of 1100 as seen in figure 36B. Additionally, this further confirms the observations by Ramabulana et al., (Ramabulana et al., 2023), which also indicated that increasing the collision energy enhances the robustness of the molecular network and expands metabolite coverage. Moreover, lower coverage of MS² fragmentation were observed when high threshold of intensities was used, this resulted in fewer nodes in the cluster. This could be attributed to the fact that lower collision energies often lead to less intense MS² spectra, thereby restricting the ability of the molecular network to classify them as sufficiently fragmented for inclusion.

The GNPS-based molecular networking approach provides the opportunity to average as many MS/MS spectrum files as possible within a single spectrum group, provided that the data are of the same chemical background, thus containing similar metabolites. Interestingly, this resulted in a wider coverage of metabolites, with over 50% of the metabolites being detected. This not only improved the overall molecular network and its clustering, but it also enhanced the annotation of the metabolites against the GNPS library database. This improvement can be attributed to the increased number of spectra available for the algorithm to search, as having more spectral data to reference provided the algorithm with greater confidence in the metabolite identifications. One way of experimentally standardising this approach could be through increasing events time through data acquisition, since this is also known to increase the data points needed to generate robust data output (Defosse et al., 2023). This indicates that an increased coverage of MS² acquisition, thus overcoming the limitation encountered through the semi-stochastic selection of abundant ion for MS² fragmentation during DDA.

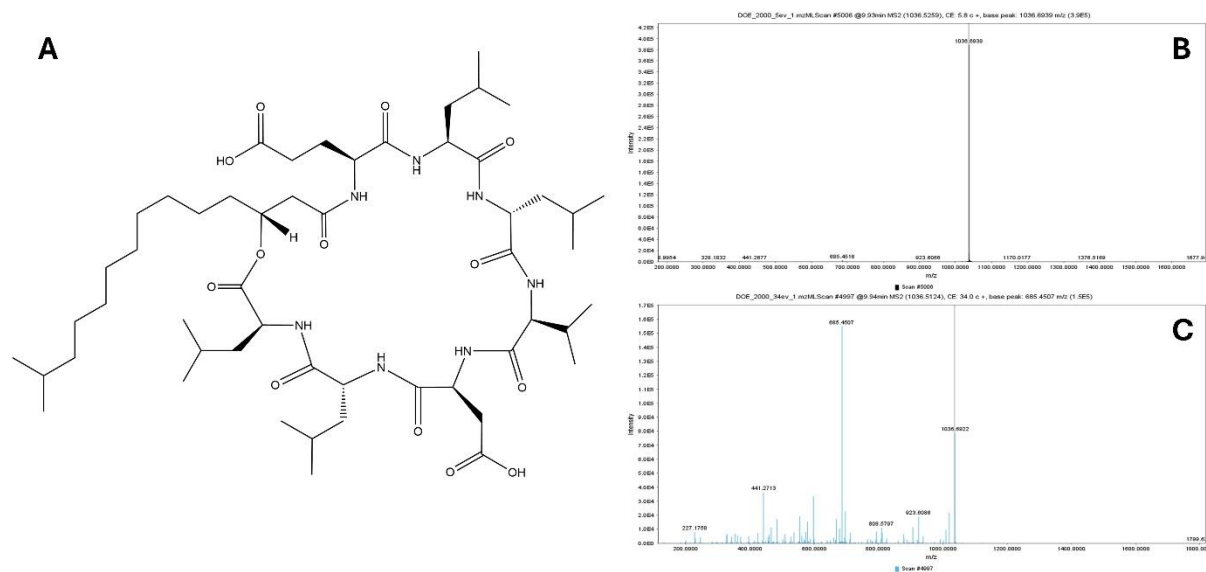
In our analysis, we focused on a cluster of known lipopeptides (Surfactins) to examine how the improved data quality impacted the clustering within the molecular network. As shown in Figure 35, the different experimental conditions from the CCD each captured their own unique set of metabolites. Classical molecular networking revealed a broad distribution of clustered nodes across all DDA settings, with a notable concentration originating from settings

characterized by high collision energy and low intensity thresholds. To complement this, Spec2vec was utilized to improve mass spectral similarity scoring through structural relationship learning, offering further insights into the molecular networking outcomes. While minimal variation in clustering across individual settings was noted between classical molecular networking and Spec2vec networks (Figure 35), Spec2vec demonstrated enhanced inter- and intra-node connectivity within clusters.

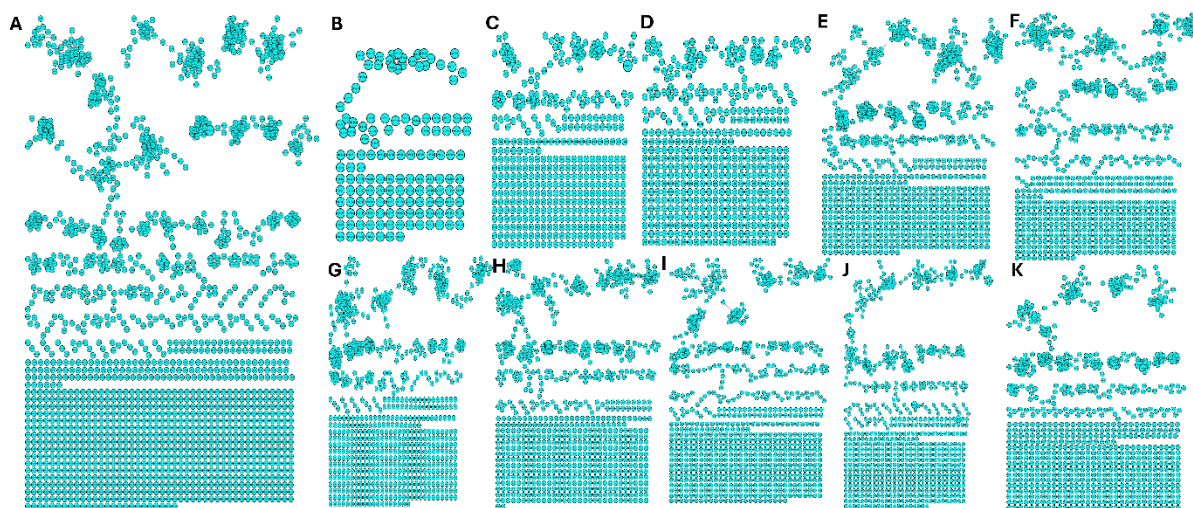
The Spec2vec network facilitated comprehensive and detailed putative annotations, enhancing understanding of the chemistry and metabolome of *Bacillus subtilis*. Analysis focusing on dipeptides subclusters revealed distinct spectral signatures associated with phenylalanine derivatives (Figure 37B). Surfactins (A, B, C, and D) were identified as prominent metabolites in *Bacillus subtilis* extracts (Iqbal et al., 2023), observed in different clusters highlighting their structural diversity (Figure 37A&C). Furthermore, certain nodes within these clusters were speculated to share structural similarities with surfactins, despite not being annotated through GNPS spectral library matching. This suggests the presence of potentially unique metabolites, underscoring the ongoing challenge in metabolomics due to the limited spectra available in the databases for annotation. Furthermore, the annotation of the nodes was enhanced when the data from all the conditions were combined. This suggests that the molecular networking approach relies heavily on the quality of the MS² data for effective clustering and visualization of related metabolites. This finding introduces a new dimension to the optimization of DDA experiments. Moreover, it raises the question of whether it is more beneficial to thoroughly optimize a single experimental condition or to systematically vary the conditions and then combine the data to achieve a more comprehensive coverage of the metabolome. By leveraging the advantages of molecular networking and the insights gained from the combined CCD data, we were able to improve the clustering and annotation of the Surfactin-related metabolites. This highlights the importance of considering the overall quality and complementarity of the MS² data when employing molecular networking approaches for metabolite profiling and identification.

6.6 Conclusion

In conclusion, molecular networking is the method of choice for analysing mass spectral data in untargeted metabolomics studies due to its ability to simultaneously analyse numerous compounds. Furthermore, soil bacteria represent a promising reservoir of pharmacologically significant metabolites. However, due to the complexity and diverse chemistry of a metabolome, no single acquisition setting can effectively cover the entire metabolome. As such, the present study has demonstrated both collision energy and threshold of intensities not only significantly influence the number of metabolites detected but also overall data robustness and quality of acquired mass spectral data. Through CCD, the present study has demonstrated that both collision energy and threshold of intensity independently enhance metabolite coverage. However, it is noteworthy that when both parameters are combined, the comprehensiveness in maximizing metabolome coverage is even greater. Moreover, molecular networking has proven to be an effective tool for analysing large datasets, enabling the visualization and comprehensive profiling of the metabolome. Therefore, to enhance exploration and maximize the potential of untargeted metabolomics, this study recommends transitioning from single-point optimization to group-based optimization. This approach is more viable for achieving comprehensive metabolite coverage and facilitating in-depth exploration.



Supplementary Figure 39: Showing the effect of collision energy on the fragment ions. (A) Structure of Surfactin C, (B) MS/MS fragmentation patterns of Surfactin C at an intensity threshold of 2000 with a collision energy of 5 eV, (C) MS/MS fragmentation patterns of Surfactin C at an intensity threshold of 2000 with a collision energy of 34 eV.



Supplementary Figure 40: Spec2vec analysis generated from ESI-positive MS/MS data, highlighting the variations in detected metabolites under different acquisition settings. (A) Combined data, (B) Intensity threshold of 2000 with 5 eV collision energy, (C) Intensity threshold of 3000 with 10 eV collision energy, (D) Intensity threshold of 1000 with 10 eV collision energy, (E) Intensity threshold of 3000 with 30 eV collision energy, (F) Intensity threshold of 2000 with 34 eV collision energy, (G) Intensity threshold of 1000 with 30 eV collision energy, (H) Intensity threshold of 2000 with 20 eV collision energy, (I) Intensity threshold of 585 with 20 eV collision energy, (J) Intensity threshold of 2000 with 20 eV collision energy and (K) 2000 with 20 eV collision energy.

6.7 Reference

- ANGUITA-MAESO, M., NAVAS-CORTÉS, J. A. & LANDA, B. B. 2023. Insights into the methodological, biotic and abiotic factors influencing the characterization of xylem-inhabiting microbial communities of olive trees. *Plants*, 12, 912.
- ARON, A. T., GENTRY, E. C., MCPHAIL, K. L., NOTHIAS, L.-F., NOTHIAS-ESPOSITO, M., BOUSLIMANI, A., PETRAS, D., GAUGLITZ, J. M., SIKORA, N. & VARGAS, F. 2020. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature protocols*, 15, 1954-1991.
- BARUPAL, D. K., BAYGI, S. F., WRIGHT, R. O. & ARORA, M. 2021. Data processing thresholds for abundance and sparsity and missed biological insights in an untargeted chemical analysis of blood specimens for exposomics. *Frontiers in public health*, 9, 653599.
- BENIDDIR, M. A., KANG, K. B., GENTA-JOUVE, G., HUBER, F., ROGERS, S. & VANDER HOOFT, J. J. 2021. Advances in decomposing complex metabolite mixtures using substructure-and network-based computational metabolomics approaches. *Natural product reports*, 38, 1967-1993.
- BUI, H. X. & DESAEGER, J. A. 2021. Volatile compounds as potential bio-fumigants against plant-parasitic nematodes—a mini review. *Journal of nematology*, 53, 1-12.
- DEFOSSEZ, E., BOURQUIN, J., VON REUSS, S., RASMANN, S. & GLAUSER, G. 2023. Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 42, 131-143.
- GAVRIILIDOU, A., KAUTSAR, S. A., ZABURANNYI, N., KRUG, D., MÜLLER, R., MEDEMA, M. H. & ZIEMERT, N. 2022. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nature microbiology*, 7, 726-735.
- GRUET, C., ALAOUI, M., GERIN, F., PRIGENT-COMBARET, C., BÖRNER, A., MULLER, D. & MOËNNE-LOCCOZ, Y. 2023. Genomic content of wheat has a higher influence than plant domestication status on the ability to interact with *Pseudomonas* plant growth-promoting rhizobacteria. *Plant, Cell & Environment*, 46, 3933-3948.
- GUO, Z., HUANG, S., WANG, J. & FENG, Y.-L. 2020. Recent advances in non-targeted screening analysis using liquid chromatography-high resolution mass spectrometry to explore new biomarkers for human exposure. *Talanta*, 219, 121339.

- HOSKISSON, P. A. & SEIPKE, R. F. 2020. Cryptic or silent? The known unknowns, unknown knowns, and unknown unknowns of secondary metabolism. *MBio*, 11, 10.1128/mbio. 02642-20.
- HUBER, F., RIDDER, L., VERHOEVEN, S., SPAAKS, J. H., DIBLEN, F., ROGERS, S. & VAN DER HOOFT, J. J. 2021. Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS computational biology*, 17, e1008724.
- IQBAL, S., BEGUM, F., RABAAN, A. A., ALJELDAH, M., AL SHAMMARI, B. R., ALAWFI, A., ALSHENGETI, A., SULAIMAN, T. & KHAN, A. 2023. Classification and multifaceted potential of secondary metabolites produced by *Bacillus subtilis* group: a comprehensive review. *Molecules*, 28, 927.
- IVANISEVIC, J. & WANT, E. J. 2019. From samples to insights into metabolism: Uncovering biologically relevant information in LC-HRMS metabolomics data. *Metabolites*, 9, 308.
- JARMUSCH, S. A., VAN DER HOOFT, J. J., DORRESTEIN, P. C. & JARMUSCH, A. K. 2021. Advancements in capturing and mining mass spectrometry data are transforming natural products research. *Natural product reports*, 38, 2066-2082.
- JIA, Q., YANG, Z., WANG, Q., YANG, H., TANG, X., ZHANG, H., CAO, L. & ZHANG, G. 2023. A liquid chromatography-tandem mass spectrometry method for comprehensive determination of metabolites in the purine pathway of rat plasma and its application in anti-gout effects of *Lycium ruthenicum* Murr. *Journal of Separation Science*, 46, 2300090.
- KHOSHURU, B., MITRA, D., JOSHI, K., ADHIKARI, P., RION, M. S. I., FADIJI, A. E., ALIZADEH, M., PRIYADARSHINI, A., SENAPATI, A. & SARIKHANI, M. R. 2023. Decrypting the multi-functional biological activators and inducers of defense responses against biotic stresses in plants. *Heliyon*, 9.
- KHWATHISI, A., MADALA, N. E., TRAORE, A. N. & SAMIE, A. 2024. Bioprospecting of soil-borne microorganisms and chemical dereplication of their anti-microbial constituents with the aid of UPLC-QTOF-MS and molecular networking approach. *PeerJ*, 12, e17364.
- KREVE, S. & DOS REIS, A. C. 2021. Bacterial adhesion to biomaterials: What regulates this attachment? A review. *Japanese dental science review*, 57, 85-96.

- MASEKO, S. T., MAREDI, M. P., MATHEWS, C. & DAKORA, F. D. 2020. Harnessing ecosystem services from biological nitrogen fixation. *The role of ecosystem services in sustainable food systems*. Elsevier.
- MASIKE, K., MHLONGO, M. I., MUDAU, S. P., NOBELA, O., NCUBE, E. N., TUGIZIMANA, F., GEORGE, M. J. & MADALA, N. E. 2017. Highlighting mass spectrometric fragmentation differences and similarities between hydroxycinnamoyl-quinic acids and hydroxycinnamoyl-isocitric acids. *Chemistry Central Journal*, 11, 1-7.
- MAUMELA, P., KHWATHISI, A., MADALA, N. E. & SEREPA-DLAMINI, M. H. 2024. In silico biotechnological potential of *Bacillus* sp. strain MHSD_37 bacterial endophyte. *BMC genomics*, 25, 399.
- MEYER, C., JEANBILLE, M., BREUIL, M.-C., BRU, D., HÖFER, K., SCREPANTI, C. & PHILIPPOT, L. 2024. Soil microbial community fragmentation reveals indirect effects of fungicide exposure mediated by biotic interactions between microorganisms. *Journal of Hazardous Materials*, 470, 134231.
- MHLONGO, M. I., PIATER, L. A., STEENKAMP, P. A., LABUSCHAGNE, N. & DUBERY, I. A. 2020. Metabolic profiling of PGPR-treated tomato plants reveal priming-related adaptations of secondary metabolites and aromatic amino acids. *Metabolites*, 10, 210.
- MINHAS, B., KATOCH, P., SINGHA, A., MINHAS, N. & KAUSHIK, N. 2024. Microbial Metabolites and Their Role to Improve Soil Health. *Advancements in Microbial Biotechnology for Soil Health*. Springer.
- MOYO, B., TAVENGWA, N. T., DUBERY, I. & MADALA, N. E. 2023. Exploration of chemical interactions between *Viscum combreticola* Engl. and its hosts through a metabolic profiling approach and molecular networking. *Journal of Plant Interactions*, 18, 2266511.
- NENGOVHELA, N., MBEDZI, D. T., NDHLALA, A. R., MATHOMU, L. M., MHLONGO, M. I. & MADALA, N. E. 2022. LC-MS based metabolite profiling reveals hydroxycinnamoyl conjugation as a discriminatory chemical factor between two closely related *Coccinia* species. *South African Journal of Botany*, 145, 199-206.
- NEPHALI, L., STEENKAMP, P., BURGESS, K., HUYSER, J., BRAND, M., VAN DER HOOFT, J. J. & TUGIZIMANA, F. 2022. Mass spectral molecular networking to profile the metabolome of biostimulant *Bacillus* strains. *Frontiers in plant science*, 13, 920963.

- NOTHIAS, L.-F., PETRAS, D., SCHMID, R., DÜHRKOP, K., RAINER, J., SARVEPALLI, A., PROTSYUK, I., ERNST, M., TSUGAWA, H. & FLEISCHAUER, M. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nature methods*, 17, 905-908.
- PANG, Z., CHEN, J., WANG, T., GAO, C., LI, Z., GUO, L., XU, J. & CHENG, Y. 2021. Linking plant secondary metabolites and plant microbiomes: a review. *Frontiers in plant science*, 12, 621276.
- PEREZ DE SOUZA, L., ALSEEKH, S., SCOSSA, F. & FERNIE, A. R. 2021. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nature Methods*, 18, 733-746.
- RAMABULANA, A.-T., PETRAS, D., MADALA, N. E. & TUGIZIMANA, F. 2021. Metabolomics and molecular networking to characterize the chemical space of four *Momordica* plant species. *Metabolites*, 11, 763.
- RAMABULANA, A.-T., PETRAS, D., MADALA, N. E. & TUGIZIMANA, F. 2023. Mass spectrometry DDA parameters and global coverage of the metabolome: Spectral molecular networks of *momordica cardiospermoides* plants. *Metabolomics*, 19, 18.
- STINCONE, P., PAKKIR SHAH, A. K., SCHMID, R., GRAVES, L. G., LAMBIDIS, S. P., TORRES, R. R., XIA, S.-N., MINDA, V., ARON, A. T. & WANG, M. 2023. Evaluation of data-dependent MS/MS acquisition parameters for non-targeted metabolomics and molecular networking of environmental samples: Focus on the Q exactive platform. *Analytical Chemistry*, 95, 12673-12682.
- SUMNER, L. W., AMBERG, A., BARRETT, D., BEALE, M. H., BEGER, R., DAYKIN, C. A., FAN, T. W.-M., FIEHN, O., GOODACRE, R. & GRIFFIN, J. L. 2007. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics*, 3, 211-221.
- TLHAPI, D., RAMAITE, I., ANOKWURU, C., VAN REE, T., MADALA, N. & HOPPE, H. 2024. Effects of seasonal variation on phytochemicals contributing to the antimalarial and antitrypanosomal activities of *Breonadia salicina* using a metabolomic approach. *Heliyon*, 10.
- TONG, C., HONDA, K. & DEREK, C. 2023. A review on microalgal-bacterial co-culture: The multifaceted role of beneficial bacteria towards enhancement of microalgal metabolite production. *Environmental research*, 115872.
- TUGIZIMANA, F., STEENKAMP, P. A., PIATER, L. A. & DUBERY, I. A. 2018. Mass spectrometry in untargeted liquid chromatography/mass spectrometry metabolomics:

- Electrospray ionisation parameters and global coverage of the metabolome. *Rapid Communications in Mass Spectrometry*, 32, 121-132.
- WANG, M., CARVER, J. J., PHELAN, V. V., SANCHEZ, L. M., GARG, N., PENG, Y., NGUYEN, D. D., WATROUS, J., KAPONO, C. A. & LUZZATTO-KNAAN, T. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature biotechnology*, 34, 828-837.
- WOLFENDER, J.-L., MARTI, G., THOMAS, A. & BERTRAND, S. 2015. Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A*, 1382, 136-164.
- WOLFENDER, J.-L., NUZILLARD, J.-M., VAN DER HOOFT, J. J., RENAULT, J.-H. & BERTRAND, S. 2018. Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography–high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Analytical Chemistry*, 91, 704-742.
- YANG, J. Y., SANCHEZ, L. M., RATH, C. M., LIU, X., BOUDREAU, P. D., BRUNS, N., GLUKHOV, E., WODTKE, A., DE FELICIO, R. & FENNER, A. 2013. Molecular networking as a dereplication strategy. *Journal of natural products*, 76, 1686-1699.
- ZHENG, H., CLAUSEN, M. R., DALSGAARD, T. K., MORTENSEN, G. & BERTRAM, H. C. 2013. Time-saving design of experiment protocol for optimization of LC-MS data processing in metabolomic approaches. *Analytical chemistry*, 85, 7109-7116.

Chapter 7

7.1 General conclusion

Microorganisms have always been the primary source for the production of antibiotics and continue to maintain their significance. The present study highlighted the effectiveness of a multifaceted approach for the bio-guided bioprospection of soilborne microorganisms, highlighting its potential to uncover novel antimicrobial compounds with significant applications in both clinical and biotechnological fields. The sprinkling technique, employed in this study, proved to be an effective and feasible method for isolating potent bacterial strains, demonstrating its value in accelerating the discovery of bioactive compounds. Among the bacterial isolates, *Bacillus pumilus* and *Bacillus subtilis* exhibited strong antimicrobial properties, with one isolate showing consistent inhibitory activity against both Gram-negative (*E. coli*) and Gram-positive (*S. aureus*) pathogenic bacteria. These findings suggest that stress-induced bioreactor conditions could further elevate antimicrobial compound production from these strains, providing a promising avenue for the development of new clinically relevant antimicrobials. Additionally, the present study has shown that, the use of UHPLC-q-TOF-MS and molecular networking approach, integrating computational metabolomics tools, is an efficient method for annotating metabolites from bacterial extracts. This approach accelerated the discovery of bioactive compounds, although the presence of novel, unidentified metabolites ("dark matter") highlights the limitations of current methodologies, which could be addressed by combining multiple analytical platforms.

The chemical diversity identified within *Bacillus pumilus* and *Bacillus subtilis* emphasizes the potential for discovering new antimicrobial agents and motivates further exploration of these species as sources of novel bioactive compounds. Furthermore, the present study explored the Whole genome analysis and further contributed valuable insights into the accurate taxonomy of *Bacillus sp.*, revealing that KH1 was miss identified as *B. pumilus* and whole genome analysis reclassified the strains as *B. altitudinis*. From the genome annotation results this study revealed genes associated with bioproducts relevant to biocontrol, biofertilization, and bioremediation from the bacterial isolates. These findings suggest the potential of the strains for application in biotechnology. Moreover, the presence of antimicrobial resistance genes (ARGs), such as *qacJ* and *qacG*, within the genomes indicates the strains resilience in

chemically challenging environments, making it a versatile candidate for agricultural and biotechnological applications.

Lastly, this study also demonstrates the critical role of optimizing acquisition parameters in untargeted metabolomics, particularly collision energy and intensity thresholds, to improve metabolite detection and data quality. The use of molecular networking has proven invaluable in analysing complex mass spectral data, offering a comprehensive means of visualizing and analysing the metabolome. To maximize metabolome coverage and enhance data robustness, the study advocates for transitioning from single-point optimization to group-based optimization, which provides a more holistic approach for untargeted metabolomics studies. In addition, this study not only highlights the potential of *Bacillus* species in discovering new antimicrobial agents and biotechnological products but also underscores the need for advanced methodologies and optimization strategies in microbial bioprospection. The findings open new avenues for further research into the functional roles of resistance genes and their broader implications for microbial ecology, plant health, and sustainable agricultural practices.

7.2 Challenges and future studies

Despite these successes in the use of computational metabolomics for analysis and visualization of the metabolome, understanding chemical intercommunication between the microorganisms is still inscrutable. The limitation of this study was small sample size and the unavailability of comprehensive databases for the microbial mass spectral annotation of secondary metabolites, commonly referred to as 'dark matter,' which were identified. Furthermore, the absence of downstream functional validation through bioassays represents a further limitation, as the biological activity of the identified compounds could not be confirmed. These limitations have affected the depth of our analysis and the ability to fully characterize the range of antimicrobial compounds present. Suggesting the need for more advanced analytical approaches, such as NMR and multi-analytical platforms, to fully explore the microbial metabolome.

Future research should focus on isolating and characterizing the unidentified metabolites to expand the repertoire of bioactive compounds. Optimizing bioreactor conditions and utilizing multi-omics approaches will be essential to elucidate the biosynthesis pathways of these compounds and enhance their production. Additionally, expanding spectral databases and incorporating diverse analytical platforms will be critical for overcoming the limitations in

metabolite identification. Further investigation into the functional roles of antimicrobial resistance genes will be necessary to ensure the safety and efficacy of these strains in agricultural and clinical settings.