IDENTIFICATION AND ANTIMICROBIAL PROPERTIES OF PSEUDOMONAS FROM SOIL

_____

A Thesis

Presented to

The Faculty of the Department of Biological Sciences

Sam Houston State University

_____

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

_____

by

Caitlyn Mary Gaffney

December 2020

IDENTIFICATION AND ANTIMICROBIAL PROPERTIES OF PSEUDOMONAS FROM SOIL

by

Caitlyn Mary Gaffney

_____

APPROVED:

Aaron Lynne, Ph.D.
Committee Advisor

Jeremy Bechelli, Ph.D.
Committee Member

Madhusudan Choudhary, Ph.D.
Committee Member

John Pascarella, Ph.D.
Dean, College of Science and Engineering Technology

ABSTRACT

Gaffney, Caitlyn Mary, *Identification and Antimicrobial Properties of Pseudomonas from Soil*. Master of Science (Biology), December 2020, Sam Houston State University, Huntsville, Texas.

As the emergence of multi-drug resistant bacteria are continuously rising, there is a dire need for novel antibiotics. As pharmaceutical companies have become less involved in the discovery of new antibiotics, alternative resources have been explored. The Small World Initiative (SWI) has teamed up with the Centers for Disease Control & Prevention (CDC) and the National Institutes of Health (NIH) to help alleviate the growing global antibiotic crisis of multi-drug resistant bacteria. The SWI has encouraged many around the country to search for novel antibiotics in the soil bacteria.

In this study we isolated antibiotic producing *Pseudomonas* isolates from soil in the Piney Woods. Over 300 isolates screened showed antibiotic properties against ESKAPE pathogens and of those, 30 isolates showed antibiotic properties against five clinical multi-drug resistant *Salmonella* strains. Additionally, we were able to extract antimicrobial compounds from the isolates and show inhibition towards five clinical multi-drug resistant *Salmonella* strains. Lastly, we did whole genome sequencing on ten of the antibiotic producing *Pseudomonas* isolates and compared the genomes of three to identify homologous genomic features for antibiotic production. Through genomic comparisons we were able to find unique features in the genomes that could potentially explain the varying inhibition abilities amongst the three isolates.

KEY WORDS: *Pseudomonas*, Multi-drug resistance, ESKAPE pathogens, Antibiotics, Drug discovery, Whole-genome sequencing, Comparative genomics

ACKNOWLEDGEMENTS

I would like to acknowledge everyone who has been a part of my academic journey to reach this achievement. To start with, my amazing parents, Denice and Don Gaffney, as well as, my grandparents Kathy and Gary Shiver for their constant love and support from the very beginning. Mom, I can never repay you for all the sacrifices you have made, or all you have done to get me here, I am forever grateful for you helping me achieve every dream I have. Also, to my brother, Chris Gaffney for always being someone I can look up to. Without you all, I could not have achieved all that I have, thank you!

Secondly, to my incredible committee members; Drs. Aaron Lynne, Jeremy Bechelli, and Madhusudan Choudhary, each of whom have provided advice and guidance throughout my entire graduate career. I would specifically like to express my gratitude to my advisor, Dr. Lynne, for putting up with my constant barging into his office and numerous hours spent with me to get here to the finish line. Without you, none of this would be possible, thank you for your incredible stories and the opportunity to do research with you. I am so grateful to have had you as a mentor who prepared for not only a career in STEM, but also the real world with your three “golden rules”. To my “co-advisor” Dr. Bechelli, that I forced to have shared custody of me, without you, I would have never started this project to begin with. Also, thank you for allowing me to join your lab as a “stepchild” and participate in different collaborations with other departments. I am forever grateful for everything I have learned from you. Dr. Lynne and Dr. Bechelli, thank you both for taking a chance on the girl who just showed up in your office. Dr. Choudhary, thank you for sharing your knowledge and ideas, and sitting with

me to explain all things genetics, especially phylogenetic trees. I am so grateful for these three men for their knowledge and skills they have passed down to me over the last two years.

Lastly, to all of my friends who have been there through all the meltdowns and tears. To Meghan, thank you for always allowing me to explain microbiology to you and for always being there for me. None of this would be possible without the amazing people I have met along the way, including Javier and Flo. From the late nights and early mornings to the walks across the street to Humps, I have enjoyed it all. Thank you for the help in the lab and in bioinformatics. You have all made this experience so much better!

TABLE OF CONTENTS

Page

ABSTRACT iii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS vi

LIST OF TABLES viii

LIST OF FIGURES x

CHAPTER I: INTRODUCTION 1

Salmonella Screening 8

Organic Extraction 11

Comparative Genomics 12

CHAPTER II: ISOLATION OF PSEUDOMONAS SPECIES WITH ANTIBIOTIC PROPERTIES 16

Abstract 17

Introduction 18

Materials and Methods 19

Results 23

Discussion 31

Conclusion 33

CHAPTER III: COMPARATIVE GENOMICS STUDY ON THREE UNIQUE *PSEUDOMONAS* SOIL ISOLATES 35

Abstract 36

Introduction 37

Materials and Methods 39

Results and Discussion 42

Conclusion 91

CHAPTER IV: CONCLUSION 93

REFERENCES 99

VITA 117

LIST OF TABLES

Table **Page**

1 MDR Salmonella and the corresponding antimicrobial agents 9

2 ESKAPE Pathogens and their safe relatives 24

3 Safe ESKAPE Pathogen Inhibition Screen 24

4 Multi-Drug Resistant Salmonella Strain Inhibition Screen 25

5 Organic Extraction using Ethyl Acetate 26

6 Multi-Drug Resistant Salmonella Strain Inhibition Screen 28

7 Multi-Drug Resistant Salmonella Strain Inhibition after Organic Extractions 28

8 Genomic Characteristics of ten antibiotic producing soil isolates 60

9 Genome Assembly 61

10 Genome Annotation 65

11 Genome Annotation: Protein Features 65

12 Genome Annotation: Specialty Genes 66

13 Genome Annotation: Antimicrobial Resistance Genes 67

14 Unique Pathways 75

15 Biosynthesis of secondary metabolites pathways in all 3 genomes associated with antimicrobial products 76

16 “Metabolite” Protein Families for Pseudomonas B217 and Pseudomonas B151 77

17 “Metabolite” Protein Families for Pseudomonas B228 77

18 “AMP” Protein Families for Pseudomonas B228, B151, and B217 78

19 “Antimicrobial” Protein Families for Pseudomonas B228, B151, and B217 78

20 “T6SS” Protein Families for Pseudomonas B217 79
21 “T6SS” Protein Families for Pseudomonas B228 80
22 “T6SS” Protein Families for Pseudomonas B151 81
23 “Polyketide” Protein Families for Pseudomonas B228 and B217..... 82
24 “Polyketide” Protein Families for Pseudomonas B151 82
25 “Phenazine” Protein Families for Pseudomonas B217..... 83
26 “Phenazine” Protein Families for Pseudomonas B228 and B151 83
27 “PCA” Protein Families for Pseudomonas B228, B151, and B217 83
28 “Pyoverdine” Protein Families for Pseudomonas B217and B228..... 84
29 “Pyoverdine” Protein Families for Pseudomonas B151 85
30 Summary of anti-SMASH results for Pseudomonas B228..... 87
31 Summary of antiSMASH 5.1.2 results for Pseudomonas B151 88
32 Summary of antiSMASH 5.1.2 results for Pseudomonas B217 89

LIST OF FIGURES

Page

Figure 1 16S tree on 26 soil isolates 29

Figure 2 16S tree on the 10 “unique” isolates 30

Figure 3 Pseudomonas B228 Closest Relatives Tree 47

Figure 4 Pseudomonas B151 Closest Relatives Tree 48

Figure 5 Pseudomonas B217 Closest Relatives Tree 49

Figure 6 Codon Tree of soil isolates and ten closest relatives 50

Figure 7 MLST Profiles 51

Figure 8 Pseudomonas B217 Allele Alignment MLST 52

Figure 9 Pseudomonas B151 Allele Alignment MLST 53

Figure 10 Pseudomonas B228 Allele Alignment MLST 54

Figure 11 GBDP tree (whole-genome sequence-based) 55

Figure 12 GBDP tree (16S rDNA gene sequence-based) 56

Figure 13 Circular representation of the Pseudomonas B228 genome structure 62

Figure 14 Circular representation of the Pseudomonas B151 genome structure 63

Figure 15 Circular representation of the Pseudomonas B217 genome structure 64

Figure 16 Biosynthetic Genes Similarity Clustering 90

CHAPTER I

Introduction

From deep in the soil, to inside our bodies, bacteria are abundant in almost every area in the world. Although, bacteria are known to be harmful and responsible for causing numerous infections, there are many that are very useful. In fact, the majority of antibiotics used in a clinical setting were discovered from different microorganisms commonly found in soil (1). Nearly two third of naturally occurring marketed antibiotics are derived from *Streptomyces* spp. (2), however, there are many soil bacteria with antimicrobial properties that remain unexplored (2).

Soil is complex and more than just “dirt”, as it is home to many microorganisms that can be utilized for various applications. Overall, there are numerous microorganisms that can produce a wide variety of antibiotics that have the ability to be used against many threatening infections and diseases in humans, animals, and agriculture (3). These microbes produce antibiotics that can kill or inhibit growth of other bacteria to protect themselves or to decrease the number of competitors for resources in their habitat (3). There are different components of soil that can be studied. In fact, researchers often study the rhizosphere when studying antibiotic producing bacteria. Rhizosphere is a soil component that is in close contact with plant roots (4). Plant growth-promoting rhizobacteria can be found in the rhizosphere and aid in plant growth, but interestingly can also produce antibiotics (4,5).

Researchers have found bacteria, specifically, *Pseudomonas* species in the rhizosphere soil that are able to secrete antibiotics and other secondary metabolites that limit growth of pathogens (5). Studies have been conducted to further investigate the

various species and their behaviors. Specifically, scientists have conducted genome analysis of antibiotic producing bacteria and comparative genomic analysis with various *Pseudomonas* species (5). Through their analysis, they were able to show that four similar bacteria from the same genus were able to produce different antibiotics (5). The data collected through these studies provide valuable information for further studies, as species comparison show different pathways and gene clusters among *Pseudomonas*.

For this reason, many researchers are testing the bacteria from various soil samples in hopes of finding new antibiotics. The ability for microbes to produce antibiotics creates a huge advantage in regard to fighting their competition and being able to thrive in a particular environment (6). Soil microbes live in very diverse environment with varying biotic and abiotic conditions that require them to adapt and develop strategies for survival and successful reproduction (6). A way to combat the competition for limited resources in the soil is by producing antibiotics, which has been a successful strategy used by different microorganisms.

Bacteria need nutrients in order to thrive and grow. These resources are not always easy to find, especially in certain environments, as there are limited resources in the soil for microorganisms to utilize. The ability to survive under the many stresses created in a soil environment requires specialty genes and adaptations. This limitation for nutrients, along with competition of other microorganisms, allows for evolution and mutations in the microorganisms. These changes can also be looked at in the lab. Many studies are conducted testing a microorganism's ability to adapt under certain stresses (7). Various experiments investigating inhibition abilities, as well as genomic changes over time are done to better understand a microorganism's evolution due to extensive stresses.

Bacteria handle competition and other environmental conditions in various ways. These methods often lead to the production of antimicrobial compounds. For example, many microorganisms are able to control the growth of other microbes using their own defense mechanisms, including production of secondary compounds from their cell walls or membranes (8). These specific defensive compounds are what can be found in the important chemotherapeutics used today (8). Additionally, during bacterial replication, bacteria are able to produce secondary compounds through varying pathways (8). These naturally made products can be found in many marketed antibiotics, as they are still the main source of drugs for fighting infections and cancer (9). Different methods can be used in order to find these chemical compounds and natural products.

Microbial genomics, including genome mining and metagenomics, as well as natural product biosynthesis using chemistry are often used for antibiotic discovery (9). From the late 1990's to 2014, bioactive secondary metabolites have been searched for and isolated from fungi and bacteria (8). As technology improved over time, more and more methods were examined. Whole-genome sequence mining became available and proved to be useful as, natural-product biosynthetic genes can be found in clusters in microbial genomes (9). In 2019, Lopes et al., were able to identify and compare genes and functions potentially associated with the soil niches of *Pseudomonas* species (4). Through whole genome sequencing, metagenomics and varying bioinformatic programs, they were able to show that different *Pseudomonas putida* populations differ at the phylogenetic, genomic, metabolic and gene levels (4).

Bacteria are resilient as their genomes and function can differ drastically. Even within the same species, bacteria can differ genomically and metabolically (5). As stated

earlier, certain closely related bacteria are able to excrete varying antimicrobial compounds (5). These varying extracellular products are not just secreted at just any given time. This process is highly regulated by not only the environment, but also through quorum sensing (10). Through quorum sensing, bacteria are able to regulate the genes they express because it allows them to sense their environment. Therefore, secretion of specific secondary metabolites allows for survival in varying environmental conditions. For example, the biosynthesis of multiple known *Burkholderia* antibiotics are controlled through quorum sensing (10). Over the last 10 years, researchers have looked at genes associated with biosynthesis and quorum sensing, in order to hopefully discover novel antibiotics (10).

Although soil microorganisms can produce a wide range of antibiotics, we still run into the problem of having pathogens that no antibiotics can treat (11). These pathogens often arise from mutations in bacteria that allow for them to become multi-drug resistant. Resistance can arise for numerous reasons, including naturally or induced. However, most resistance to drugs are often a result of the misuse from humans. For example, over prescribing, improper disposal, and prescription neglect, can allow bacteria to acquire preexisting resistance genes from another bacterium through horizontal gene transfer. This often occurs because, the given drugs are only killing the susceptible cells, and allowing resistant cells to survive and replicate.

This phenomenon has been occurring for years and is very well known. However, less known to the public; studies have found that in certain plants, their genomes are showing resistance genes for clinically used antibiotics (12). Studies further show that in transgenic plants, bacterial antibiotic resistance markers tend to be the most frequently

inserted genes (12). As plant DNA can be found in soil for extended periods of time, it is no surprise that we see these transgenes spreading horizontally to bacteria (12). What this means is that overtime these genes can be naturally transformed into bacteria. Again, highlighting the importance of proper disposable of antibiotics and other drugs.

Over the last few years antibiotic resistance has been rising to dangerously high levels all over the world (13). The increase in antibiotic resistance is turning into one of the biggest threats to global health, food security, and development around the world (13). In fact, the World Health Organization's (WHO) Global Antimicrobial Surveillance System reported in 2018 that there is a widespread occurrence of antibiotic resistance among 500,000 people with numerous bacterial infections in over 20 countries (13,14). Common infections like urinary tract infections (UTI) caused by *Escherichia coli,* are becoming harder to treat as resistance is constantly increasing. For example, Penicillin, a group of commonly used antibiotics for years has shown an increase of resistance up to 51% overall (13, 14). Although some bacteria are naturally resistant to antibiotics, the misuse of prescriptions make selecting for antibiotic-resistance bacteria occur more frequently. Additionally, bacteria have the unique ability to mutate and change overtime to become resistant. It has become quite evident that new antibiotics are needed in order to prevent and treat infectious disease for many years to come (15).

Each year, more than 700,000 people die due to various infections caused by infectious microorganisms that cannot be treated with existing drugs (16). If new antibiotics are not developed soon, we will be living in a world that existed before antibiotics were invented; where minor infections that normally could be treated with antibiotics will kill people. It is estimated that by 2050, superbugs causing various

infections will be killing more people than cancer and diabetes combined, resulting in over 300 million premature deaths (16).

Currently, there are 51 new antibiotics and biologicals in clinical development to treat the 12 priority pathogens causing infections (16). However, only 8 are classed by WHO as innovative treatments that add value to the current antibiotic treatment (16). Today, pharmaceutical companies and researchers are pushed towards creating more of an urgency towards development of new antibiotics. However, funding is limited and therefore it is becoming less common for pharmaceutical companies to put money into the development of new antibiotics. The production of natural products is what fuels the discovery of novel antibiotics, however, it is very costly to find a novel compound (9). For this reason, many people in the science community have joined forces in order to come to a greater solution.

For the last five years, The Small World Initiative (SWI) has teamed up with the Centers for Disease Control & Prevention (CDC) and the National Institutes of Health (NIH) to help alleviate this growing global antibiotic crisis. The SWI is an innovative program that not only encourages students to pursue careers in STEM, but also gets people on the search for new antibiotics, as well as becoming more aware of the scary future ahead. SWI goal is for students from all around the world to collect various soil samples and isolate bacteria in hopes to find novel antibiotics. Being that majority of antibiotics originated from soil bacteria and fungi, it is very hopeful that new antibiotics can be discovered using this method (16). This new and innovative project from SWI has led to many schools and universities stepping in to answer the question; are there new antibiotics to be discovered in the soil to solve the growing antibiotic resistance problem?

At Sam Houston State University, BIOL3440 General Microbiology students took part in the SWI project. Over the course of the semester, students collected soil sample in and around the Piney Woods region of Texas. Students were told to find “unique” areas to collect soil and to record GPS coordinates and location description. Students then began to isolate diverse bacterial through serial dilutions with growth on various growth medias. We wanted to select for bacterial isolates that were able to produce antimicrobial products. Therefore, serial dilutions were performed, and aliquots were placed on 10% TSA plates in order to ensure stress and competition among bacteria. Different growth condition allows for biodiversity among bacteria as well as differing secondary metabolites produced. Under certain environmental stresses, including lack of nutrients, bacteria will begin to compete with one another eventually secreting various compounds to eliminate their competition. Once students isolated individual colonies, they tested their isolates against numerous clinically relevant microorganisms as well as process isolates through biochemical testing.

The rise in multi-drug resistant (MDR) bacteria are increasing (16). Those of the upmost concern are ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterococcus* species), now recognized by the Infectious Disease Society of America as bacteria with the most significant risk to public health (16). According to the WHO, the most commonly reported antibiotic resistant bacteria are *Escherichia coli, Klebsiella pneumoniae, Staphylococcus aureus,* and *Streptococcus pneumoniae,* followed by *Salmonella* species, therefore testing against these clinically relevant bacteria was a good starting point to find novel antibiotics. Students tested their isolates against the ESKAPE

"safe relatives". The ESAKPE safe relatives or safe models used in lab are as follows, *Enterococcus raffinosus, Bacillus subtilis* (safe model), *Escherichia coli, Erwinia carotovora* (safe model), *Pseudomonas putida, and Enterobacter aerogenes* respectively. A safe model instead of the relative was used to do the stock not coming in on time. At SHSU, safe relatives along with safe models of the ESKAPE pathogens were used to be in compliance with BSL-2 and for the safety of our students.

Upon testing their bacterial isolates with known bacterial cultures, zones of inhibition were measured and noted. If the isolates were able to inhibit the growth of the ESKAPE relatives, then we could make the prediction that the isolate was able to produce antimicrobial products. Students were able to partake in the first step towards finding a novel antibiotic. Students' isolates that showed large, strong zones of inhibition were then made into glycerol stocks for further testing.

Successful discovery of soil isolates that are able to inhibit the growth of ESKAPE pathogens is possible (16). In this study, we aimed to take it further by seeing if soil isolates could inhibit known MDR clinical strains, as well as see if certain genomic features allowed this to occur.

Salmonella Screening

In this study, we propose to explore soil samples to find antibiotic producing bacteria, specifically *Pseudomonas* species capable of inhibiting multi-drug resistance (MDR) *Salmonella* pathogens. The CDC reports that in the United States, there are approximately 1.2 million cases of *Salmonella* infections. *Salmonella* infections can lead to many complications including bacteremia. Due to the increasing rates of reported hospitalizations and deaths, we decided to test several serotypes in this study.

Salmonella spp. has multiple strains that have been reported to be MDR (17-21). *Salmonella* strains; 853-Dublin, 49-Choleraesuis, 485-Typhimurium, 163-Heidelberg, and 64-Newport are all known to be MDR (17-21). Table 1 below highlights the antimicrobial agents that these strains of *Salmonella* have shown resistance to. These antimicrobial agents are used by The National Antimicrobial Resistance Monitoring System (NARMS) for Enteric Bacteria to track changes in antimicrobial susceptibility of certain enteric bacteria.

Table 1 MDR Salmonella and the corresponding antimicrobial agents

Antimicrobial Agent	**MDR Strains**				
	853-Dublin	49-Choleraesuis	64-Newport	163-Heidelberg	485-Typhimurium
Tetracycline	R		R	R	R
Streptomycin	R	R	R	R	R
Kanamycin	R				R
Ampicillin	R		R	R	R
Sulfisoxazole	R		R	R	R
Chloramphenicol	R		R	R	R
Amoxicillin/Clavulanic Acid	R		R	R	R
Cefoxitin			R		R
Trimethoprim/ Sulfamethoxazole			R	R	
Ceftiofur	R		R	R	R
Gentamicin	R	R		R	R
Ceftriaxone	R	R	R	R	
Nalidixic Acid			R		R
Amikacin			R		R
Ciprofloxacin	R		R		R

The *Salmonella* strains used were chosen first of all for their large number of resistant antibiotics, but also because of the growing heath concerns. Specifically, *Salmonella enterica* serovars Newport, Heidelberg, and Typhimurium are three of the most commonly reported serotypes that are responsible for invasive and even deadly *Salmonella* infections (18). Although not as common, *Salmonella enterica* serovars Dublin and Choleraesuis are also known to be one of the leading causes of infections and deaths associated with salmonellosis (17). Being that *Salmonella enterica* infections

continue to be a significant health problem, and less and less antibiotics seem to be able to treat these infections; it is crucial to discover novel antibiotics. There were approximately 300 isolates collected from SHSU students that went through the first round of processing. These isolates all showed inhibition of one or more ESKAPE safe relative pathogens. The isolated were subjected to 16s sequencing and were found to be primarily *Pseudomonas*. In a similar style, these isolates were screened against the five MDR *Salmonella* strains. If isolates are able to restrict the growth of known MDR *Salmonella* strains, it is highly probable that these isolates are capable of secreting bactericidal or bacteriostatic agents (22).

Hypothesis #1: Different strains of Pseudomonas isolated from the soil inhibit the growth of various multi-drug resistant Salmonella strains.

As previously mentioned, the genus *Pseudomonas* is known for secreting compounds that kill surrounding bacteria (23). These compounds secreted by *Pseudomonas* could act as bactericidal or bacteriostatic agents, thus these agents could help *Pseudomonas* by decreasing competition for resources in the soil community (22). By the use of these compounds, *Pseudomonas* is able to inhibit the growth of other microorganisms. These compounds could be used as novel antibiotics against MDR *Salmonella* strains. The objective of this experiment is to screen for *Pseudomonas* species from the General Microbiology student's isolates that inhibit MDR *Salmonella* strains.

As it is clear, we need new antibiotics, for years studies have tried to find novel antibiotics. Cain, *et al*., successfully found a naturally occurring, gram-negative,

nonobligate predator bacterial strain, that exhibits broad-spectrum antimicrobial activity (24). This study explored the extracellular compounds responsible for the antimicrobial activity. More recently, Yan, *et al*., provided a detailed exploration behind the evolution of antibiotic biosynthesis along with data that can utilized for antibiotic discovery and regulation (25). It was found that secondary metabolism along with competition can produce spontaneous mutants in the Gacs-Gaca regulatory system in *Pseudomonas protegens* (25). Additionally, studies have shown that *Pseudomonas* bacteria have numerous antimicrobial properties (26). For example, mupirocin, which is a polyketide antibiotic that is a mixture of four pseudomonic acids (26). Thus, supporting the idea; novel antibiotics can be isolated or constructed from various soil *Pseudomonas* species.

Organic Extraction

Various bacteria are capable of producing antibiotics to kill off surrounding bacteria. The *Pseudomonas* species is known for being able to kill off multiple types of bacteria with their antibiotics in order to survive and thrive (23). These compounds can be isolated through organic extractions using various methods. Compounds extracted can then be further analyzed and tested to evaluate antimicrobial activity. Additionally, identified compounds can be manipulated to create new synthetic antibiotics. In this study, we will extract antibiotics produced by *Pseudomonas* species to test growth inhibition against the *Salmonella* strains.

Hypothesis #2: Antimicrobial agents extracted from Pseudomonas strains show growth inhibition of Salmonella strains.

Many microorganisms produce antibiotics or antimicrobial peptides to reduce competition in their soil communities by inhibiting growth or killing (23). These antimicrobial agents usually target key structures or pathways of other microorganisms (23). Therefore, MDR *Salmonella* strains exposed to the organic extracted compounds from *Pseudomonas* should have growth inhibition or death (27). Thus, the objective is to extract antimicrobial agents.

The discovery of these antimicrobial agents has been extensively searched for in *Streptomyces* spp. Recently, Nandhini *et al*., aimed to find the antimicrobial compounds from terrestrial *Streptomyces* spp. (28). Through their isolation and extraction methods, they were able to conclude that the potent isolates from terrestrial soil could be a source for pharmaceutical industries to explore antibacterial and antifungal compounds (28). Recently, fermentation and medium optimization experiments are being conducted for the production of specific antimicrobials (29). It has been shown, that various extraction methods allow for differing antimicrobial compounds (29). For example, a study using *Streptomyces atrovirens,* showed that this strain utilized starch as the main carbon source and in turn increased the production of antibacterial compounds (30).

Comparative Genomics

Finally, in this study we aim to compare the genomes of the different *Pseudomonas* species to identify possible gene targets responsible for production of antibiotics. For this study, isolates will be selected based on how unique they were amongst each other based on inhibition zones, geographic location, and 16s sequencing results.

It is crucial to public health that we find new and innovative ways to find novel antibiotics. It is becoming increasingly difficult to find antibiotics that will work even through synthetic means, as those too are seeing an increase in resistance from bacteria (31). We are approaching a world with only a few antibiotics left that have any true antimicrobial activity that is effective. As stated before, pharmaceutical companies have avoided funding new antibiotic research, as there is more money put in, than profit received. In the early 1900s new antibiotics were being discovered constantly, and since then the addition of novel antibiotics has slowed down substantially. However, there was hope with the discovery of chemical modifications, but then again, there only so many chemical modifications that can be one without altering the bactericidal effects. For this reason, many have switched to a more molecular approach when searching for novel antibiotics.

Bacterial genomes can be used to help identify biosynthetic pathways that produce intrinsic antimicrobial compounds and peptides (31). Through the genetic data we now have access too, it has become easier to search for and discover novel compounds and products to use in the fight against MDR bacteria (9). Through comparative genomics technology and programs, it has also become easier to take different strains or even species of bacteria and further understand how one is able to inhibit growth of a pathogen while the other is not. Genome mining has opened a world to discovering many gene pathways and metabolites that lead to the productions of antimicrobial compounds. Specifically, finding genes associated with secondary metabolites have been a game changer.

Bacteria go through various stages of metabolism in order to survive. It has been shown that bacteria have the ability to produce numerous compounds with varying bacteriostatic or bactericidal activity (22). These compounds can then be extracted and used for controlling MDR bacteria. Bacteria also have varying specialty genes, including virulence genes that often aid them in surviving different conditions. Some of these genes specifically often aid in the regulation of the certain pathways that then control the production of antimicrobial compounds. In conclusion, through full genome sequencing, we can identify these genes and pathways in hopes to discovering novel antibiotics.

Hypothesis #3: Pseudomonas strains that inhibit the growth of closely related multidrug resistant Salmonella strains have homologous genomic features for antibiotic production.

Bacterial strains that inhibit or kill closely related microorganisms share orthologous genes (5). *Pseudomonas* strains that inhibit closely related MRD Salmonella strains should share antibiotic producing genes. Therefore, these *Pseudomonas* species should have a closer genetic distance. The objective is to identify the antibiotic producing genes of *Pseudomonas* strains that inhibit MDR *Salmonella* strains. Recently, whole genome sequencing studies showed that *Pseudomonas* sp. isolated from water showed antibiotic properties through genome mining and organic extractions (32). Additionally, Zhang *et al*., exhibited that closely related *Pseudomonas* species showed protein similarity and orthologous clusters through a comparative genomic study (5).

Full genome sequencing of bacterial genomes has the potential of changing the way to find novel antibiotics. Genome mining has allowed for the discovery of gene clusters associated with secondary metabolite biosynthesis, which helps identify proteins that produce antibiotics (32). Comparative studies have identified varying genetic features that produce different metabolic reactions in the same bacterial populations (4). This current study was a preliminary step into the discovery of novel antibiotics. In this study we have successfully sequenced ten isolates that are able to inhibit the growth of MDR *Salmonella* strains. Eight of the isolates are *Pseudomonas* species, while the remaining two are *Flavobacterium*. In the described study, we specifically aimed at comparing the genomes of three "unique" isolates that showed varying MDR pathogen inhibition abilities, metabolic pathways, specialty genes, secondary metabolite biosynthesis gene clusters, and other genomic features. As all three isolates were *Pseudomonas*, we wanted to find out what allowed these strains to inhibit different MDR *Salmonella* strains.

CHAPTER II

Isolation of Pseudomonas species with antibiotic properties

This thesis follows the style and format of *Wiley Microbiology*.

Abstract

Antibiotic resistance is a continuing threat we face worldwide. Each year, we see an increase in bacteria that are resistant to the antibiotics on the market. Combined efforts from different fields of science are needed to combat this ongoing battle with multi-drug resistant bacteria. Programs like Small World Initiative have been involved in the discovery of novel antibiotics. Through combined efforts of the Small World Initiative and the CDC, more research is being conducted on antibiotic producing bacteria isolated from the soil.

Here we report the isolation of *Pseudomonas* isolates with antibiotic properties against five clinical multi-drug resistant *Salmonella* strains. Over 300 antibiotic producing isolates were screened through and 30 were found to inhibit clinical multi-drug resistant *Salmonella* strains. Furthermore, we report that antibacterial compounds can be extracted through organic extractions using ethyl acetate. This study was a preliminary study for future studies involving comparative genomics.

Keywords: *Pseudomonas*, Organic extraction, Antibiotic production, Multi-drug resistance

Isolation of *Pseudomonas* species with Antibiotic Properties

Introduction

Pseudomonas bacteria are commonly found in numerous environments, especially in soil and in water habitats (33). While some species in *Pseudomonas* can cause serious infections, others remain useful in research. In fact, within *Pseudomonas,* there are many species that are able to produce secondary metabolites and antibiotics.

From the early days of Alexander Fleming, through the late 1900s, novel antibiotics were constantly being discovered from various soil microorganisms. However, with the over prescription and poor use of these new drugs, soon came multi-drug resistance. Antibiotic resistance is a growing concern for people of all ages all around the world. In fact, one of the biggest threats to public health is antibiotic resistance (13). As more and more bacteria are becoming resistant to multiple drugs on the market, it is very important that that we discover and/or develop novel antibiotics.

Combined efforts from programs like the Small World Initiative, have made a valiant effort in the discovery of antibiotic producing microorganism. As many drugs on the market now are derived from soil dwelling bacteria and fungi, it only made since to start the search from the ground up. Thousands of people across the country at various universities have all taken part in the initiative to find antibiotics from soil dwelling microorganisms.

Studies are being conducted to see if anything novel can be found among the microorganisms that inhabit soil. Numerous experiments are being ran that screen unknown isolates against pathogens in hopes of finding one that is able to inhibit its

growth (34). Additionally, further analyses are being conducted in hopes to isolate these compounds via organic extractions and liquid chromatography mass spectrometry (35).

Discovering a soil isolate that can inhibit growth of another pathogen is simple as bacteria often kill off its competitors in nature all the time. However, finding an isolate capable of inhibiting growth of known clinical MDR bacteria is more challenging. Even more challenging is successfully isolating, purifying, and repurposing that compound to be a useful antibiotic that can be used in the health industry.

In this study, we aimed to answer the question: Can bacteria isolated from soil samples inhibit the growth of known multi-drug resistant (MDR) *Salmonella* clinical isolates? And if yes, can the compounds extracted via organic extractions still have antibacterial properties?

Materials and Methods

Soil sample collection and bacterial isolation

At Sam Houston State University, approximately 100 soil samples were collected in February 2019 for the General Microbiology lab. Soil samples were collected into sterile conical tubes from various areas in Texas, including the Piney Woods region and Houston. The general location, along with GPS coordinates were recorded as well as a general description were noted for each soil sample. Soil samples were stored at 4°C until the bacterial isolation process began. Individual bacterial colonies were isolated as previously described in The Small World Initiative Research Protocols. Briefly, 1 gram of soil was mixed with 9mL of water and then serially diluted 5 times. Each dilution was plated on 10% Tryptic Soy Agar (TSA) and incubated at 30°C for 48 hours. From the

diluted samples, about 50 isolates from each dilution series were picked for further testing. These isolates were selected based on differing colony morphology.

ESKAPE pathogen screening

The individual colonies were then tested against safe relatives of the "ESKAPE" pathogens. ESKAPE pathogens are defined by the Infectious Diseases Society of America as *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* spp. In lab the students used their safe relatives or a safe alternative model, *Enterococcus raffinosus* (ATCC 49464), *Bacillus subtillis* (Handelsman Lab strain), *Escherichia coli* (ATCC 11775), *Erwinia carotovora* (Handelsman Lab strain), *Pseudomonas putida* (Handelsman Lab strain), *and Enterobacter aerogenes* (ATCC 51697) respectively. Students took these safe relatives and used the lawn streaking method to plate each pathogen on 10% TSA plates. Next, using the patch method described in The Small World Initiative Research Protocols, individual colonies were picked and patched onto the ESKAPE relative plates (36). Briefly, a sterile toothpick was used to scrape an individual colony, and then was transferred onto the plate with the lawn of the ESKAPE relative. These plates were then incubated at 30°C for 48 hours. After incubation, plates were examined for zones of inhibition in the lawn streaks. A clear halo indicated that the selected bacterial isolate was able to inhibit growth of the ESKAPE relative pathogen. Potential antibiotic producing isolates were selected and screened against the ESKAPE relative again to ensure inhibition properties. From these results, any bacterial isolates that inhibit growth of the ESKAPE relative pathogens were separated into two groups. Group A was used for the student's projects, while Group B was stored away for further analysis.

Multi-drug resistant Salmonella screening

Group B bacterial isolates that showed inhibition towards various ESKAPE pathogens were selected for screening against MDR *Salmonella* strains. The same process for screening was used again. All screens were performed in a certified BSL-2 lab. First, 10%TSA plates were used in order to ensure bacterial isolates will secrete antibiotics or secondary metabolite products. *Salmonella* strains: 853-Dublin, 21-Choleraesuis, 485-Typhimurium, 163-Heidelberg, and 64-Newport were streaked onto five 10%TSA plates. Next, using a toothpick, single colonies were picked and placed directly onto the plate. After 24 hours of incubating at 30°C, zones of inhibition were examined. The bacterial isolates that showed strong zones of inhibition with clear halos were retested with a smaller number of bacterial isolates per plate to insure inhibition. A total of 30 isolates were selected for further analysis.

Bacterial identification via 16s

The genus identification of 30 isolates was done by 16S sequencing. The 16S gene was amplified using universal primers 27f (5-AGAGTTTGATCCTGGCTCAG-3) and 1492r (5- TACGGYTACCTTGTTACGACTT-3). The PCR reaction was ran using both primers, as well as Apex 2X RED Taq Master Mix, 1.5mM MgCl2 [Final Conc.] with ammonium buffer (Genesee Scientific Corporation). Eight hundred seventy-five microliters of the master mix were added to 35 microliters of each primer, and 805 microliters of water. Next, each of the 30 bacterial isolates was picked and placed in a centrifuge tube and 50 microliters of the of the master mix solution was added. After solution and bacteria were mixed, PCR was ran on the thermocycler. PCR conditions were as follows: 94°C denaturing for 30s, 58°C annealing for 30s, elongation at 72°C for

60s for 30 cycles. Lastly, a 1% agarose gel was ran at 100V for 30 minutes, to ensure the 16s gene was amplified. The samples were then sent off for 16s rRNA Sanger sequencing to GenScript (Piscataway, NJ). Genus identification was determined using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) software.

Organic Extractions

Upon comparison between the 16s data and inhibition results, 14 unique isolates were selected for further analysis. Organic extractions were performed following the described methodology in The Small World Initiative Research Protocols (36). Briefly, using 50 milliliters of 10% TSA broth mixed with one of the isolates, the flasks were incubated on a shaker at room temperature. After 24 hours of incubation, 25 milliliters of ethyl acetate was added and shaking incubation at room temperature was continued overnight. The top organic layer was extracted from the flaks and placed into a vial to allow for evaporation. Once all vials were evaporated, they were re-suspended with 150 microliters of ethyl acetate. The antimicrobial properties of the organic extracts were tested using a disk diffusion method. Under aseptic conditions, the 5 MDR *Salmonella* strains were streaked on to 5 Nutrient Agar plates and 10 sterile paper disks were placed onto the agar, as well as a control disk. Ten microliters of extract was then placed on each disk and 10 microliters of ethyl acetate was placed as a control. After a 24-hour incubation, the zones of inhibition were measured for all extracts against all 5 MDR *Salmonella* strains.

Results

ESKAPE pathogen screening

For the preliminary studies done by the students, safe relatives were used instead of the designated of ESKAPE pathogens. ESKAPE pathogens, *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter* spp have safe relatives, *Enterococcus, Staphylococcus aureus, Escherichia coli, Acinetobacter baumannii, Pseudomonas putida, and Enterobacter aerogenes* respectively. However, safe models of two of the ESKAPE safe relatives were used due to stock not coming in on time. For *Staphylococcus aureus* and *Acinetobacter baumannii,* the safe model organisms *Bacillus subtilis* and *Erwinia carotovora were* used respectively as seen in Table 2. Soil isolates tested against the ESKAPE relatives showed inhibition abilities to both Gram-negative and Gram-positive isolates. Total (300) isolates inhibited 77% of Gram negative and 95% of Gram-positive ESKAPE relative pathogens. Majority of the soil isolates tested were able to inhibit the growth of *Bacillus subtilis* as seen in Table 3.

Table 2 ESKAPE Pathogens and their safe relatives

ESKAPE Pathogens	**Safe ESKAPE Pathogen Relatives**
Enterococcus faecium	*Enterococcus raffinosus*
Staphylococcus aureus	*Bacillus subtilis**
Klebsiella pneumoniae	*Escherichia coli*
Acinetobacter baumannii	*Erwinia carotovora**
Pseudomonas aeruginosa	*Pseudomonas putida*
Enterobacter species	*Enterobacter aerogenes*

**species are considered a safe model for ESKAPE pathogens, not the traditional relative*

Table 3 Safe ESKAPE Pathogen Inhibition Screen

Safe ESKAPE Pathogen Relatives	**Gram Reaction**	**Inhibition %**
Enterococcus raffinosus	Positive	18
Bacillus subtilis	Positive	62
Escherichia coli	Negative	28
Erwinia carotovora	Negative	28
Pseudomonas putida	Negative	6
Enterobacter aerogenes	Negative	15

Multi-drug resistant Salmonella screening

In determining the antimicrobial properties of the isolates, clear halos in the lawn of bacteria were examined. Isolates were considered to have good antimicrobial properties if a distinct clear halo was present around the isolate in the lawn of ESKAPE relative. Of the 300 soil isolates tested against MDR *Salmonella* strains, only 10% showed antimicrobial properties capable of inhibiting the MDR *Salmonella* strains. However, those 30 soil isolates were able to inhibit the growth of 80% of MDR *Salmonella* strains that are resistant to 10 or more antimicrobial agents seen in Table 4. More than 60% of soil isolates showed inhibition of *Salmonella* strain, Dublin. As previously mentioned in Table 1, the 5 *Salmonella* strains are known to be resistant against a variety of antibiotics. While some strains are more resistant to known antibiotics than others, it was clear the each of the isolates from the soil were able to produce a broad spectrum of antibacterial compounds to inhibit the *Salmonella* strains growth.

Table 4 Multi-Drug Resistant Salmonella Strain Inhibition Screen

MDR *Salmonella* Strain	**Number of antibiotics spp. is resistant to**	**Inhibition %**
Dublin	11	62
Choleraesuis	3	9
Typhimurium	13	34
Heidelberg	10	46
Newport	13	40

Organic Extractions

Using the patch method, we were able to see clear zones of inhibition when soil isolates were grown with lawn streaks of MDR *Salmonella* strains. To further test these soil isolates ability to produce antimicrobial compounds, organic extractions using ethyl acetate were performed. As seen in Table 5, we successfully isolated antimicrobial compounds that still showed inhibition of MDR *Salmonella* strains. Successful inhibition via organic extracts was considered through comparison to the control. Overall, inhibition abilities increased after extraction.

Table 5 Organic Extraction using Ethyl Acetate

MDR *Salmonella* Strain	**Number of antibiotic spp. is resistant to**	**Inhibition %**
Dublin	11	73
Choleraesuis	3	73
Typhimurium	13	91
Heidelberg	10	73
Newport	13	91

Bacterial identification via 16s

In order to identify these unknown soil isolates, we wanted to perform 16S rRNA sequencing. Raw reads were read in BLAST, and 92% of soil sample collected was found to be *Pseudomonas*, while the remaining 8% was *Flavobacterium.* Figure 1 shows the remaining 26 isolates that made it through the sequencing process and were able to be sorted into a tree. We created a Neighbor-Joining consensus tree on Geneious Prime using default parameters. It was clear that the tree branched two very distinct groups, and within the *Pseudomonas*, there was different branching.

As these soil isolates were all taken in similar regions of the Piney Woods, it was likely that a few could be clones of each other. Therefore, of the 30 soil isolates, 10 "unique" isolates were chosen for further analysis for future studies. These ten isolates were picked out based on a few characteristics. The first "unique" feature examined was how many MDR *Salmonella* strains they were able to inhibit, along with location the isolate was found. Additionally, results from the organic extractions, and 16s sequencing data was taken into consideration. Table 6 depicts the 10 isolates, along with specific MDR *Salmonella* strains they inhibited growth of. Isolates B28 and B151, both had the ability to inhibit growth of all 5 MDR *Salmonella* strains. Table 7 shows the unique ten isolates and their inhibition abilities after organic extractions. We tried to pick 10 isolates that would potentially be different *Pseudomonas* species or strains. We again created a Neighbor-Joining consensus tree on Geneious Prime using default parameters to create Figure 2 with the unique isolates.

Table 6 Multi-Drug Resistant Salmonella Strain Inhibition Screen

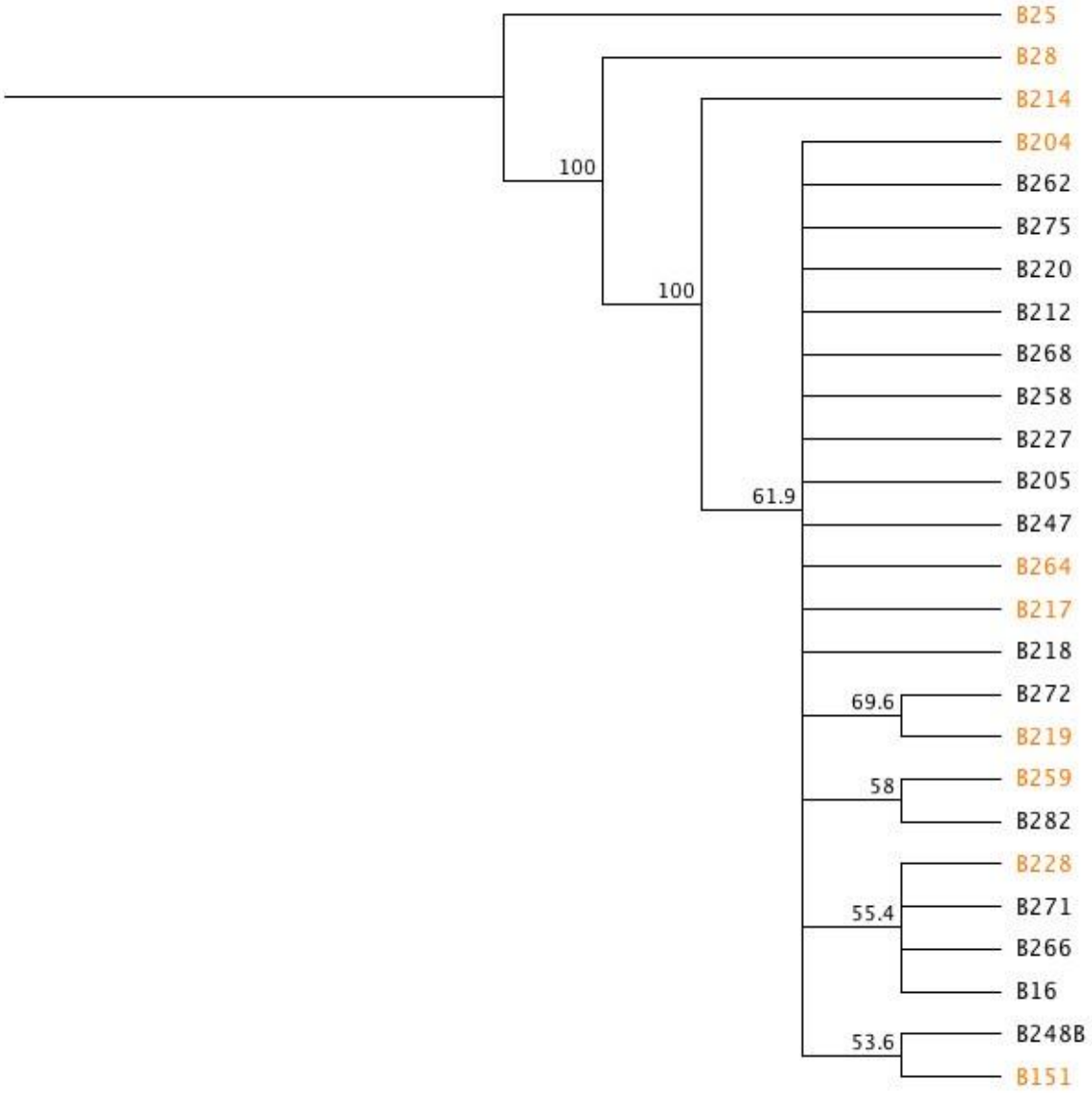
Sample ID	MDR Strains				
	853-Dublin	49-Choleraesuis	64-Newport	163-Heidelberg	485-Typhimurium
B214	INHIBITED				
B217	INHIBITED		INHIBITED	INHIBITED	
B25	INHIBITED		INHIBITED	INHIBITED	INHIBITED
B28	INHIBITED	INHIBITED	INHIBITED	INHIBITED	INHIBITED
B204	INHIBITED				
B219	INHIBITED		INHIBITED	INHIBITED	
B151	INHIBITED	INHIBITED	INHIBITED	INHIBITED	INHIBITED
B228	INHIBITED		INHIBITED	INHIBITED	
B259	INHIBITED				INHIBITED
B264	INHIBITED				

Table 7 Multi-Drug Resistant Salmonella Strain Inhibition after Organic Extractions

Sample ID	MDR Strains				
	853-Dublin	49-Choleraesuis	64-Newport	163-Heidelberg	485-Typhimurium
B214	-	-	-	-	-
B217	INHIBITED	INHIBITED	INHIBITED	INHIBITED	INHIBITED
B25			INHIBITED		
B28			INHIBITED	INHIBITED	INHIBITED
B204	-	-	-	-	-
B219		INHIBITED	INHIBITED	INHIBITED	INHIBITED
B151	INHIBITED		INHIBITED	INHIBITED	
B228	INHIBITED	INHIBITED	INHIBITED	INHIBITED	INHIBITED
B259	-	-	-	-	-
B264	INHIBITED	INHIBITED	INHIBITED		INHIBITED

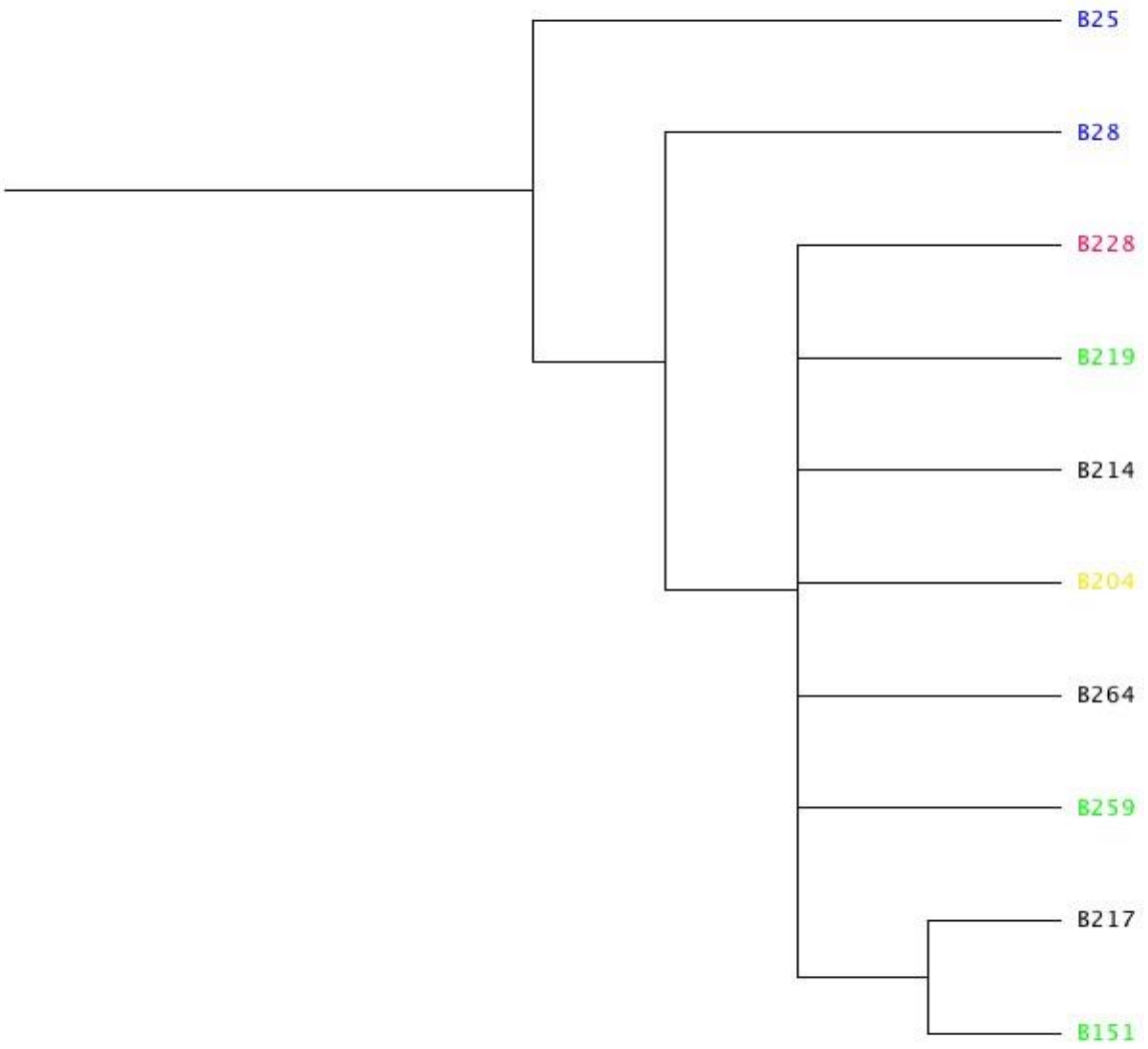
**Isolate B214, B259 and B204 got contaminated during organic extraction process.*

Figure 1 16S tree on 26 soil isolates



16s sequencing identified two major groups. Approximately 92% of soil sample collected was found to be *Pseudomonas*, while the remaining 8% was *Flavobacterium*, isolates B25 and B28. The two *Flavobacterium* isolates show different inhibition patterns, as well as lineages. Of the remaining 24 isolates, the tree exhibits different linages and the isolates show different inhibition results, suggesting they are different strains in the *Pseudomonas* family. The orange numbers are the 10 unique isolates chosen for whole genome sequencing.

Figure 2 16S tree on the 10 "unique" isolates



Ten unique antibiotic producing isolates will be used to whole genome sequencing. Above shows their relationship when compared to one another. The colored numbers are categorized by what the 16s data suggested the isolates species were. The blue represents the *Flavobacterium* isolates, the green were identified as *P. syringae*, yellow was *P. moraviensis*, and pink was *P. koreensis*, while the black was unknown.

Discussion

Every year, the number of infections caused by antimicrobial resistant pathogens is increasing. This trend creates the urgency for novel drugs in hopes to control these MDR pathogens. Although most naturally occurring antibiotics are derived from *Streptomyces*, there is potential to derive antimicrobial compounds from other bacterial species as well. As many antibiotic producing *Streptomyces* species can be found in soil environment, other microorganisms with antibiotic producing capabilities may also be found there. This search for useful secondary metabolites produced from bacteria was inspired by the hypothesis that other soil bacteria can inhibit MDR *Salmonella* strains.

The piney woods of Texas is full of diverse tree, shrubs, wildflowers, and grasses; making it a good spot to isolate potential antibiotic producing bacteria. In fact, many plants are able to produce their own secondary metabolites that interact or even harm different soil microorganisms, thus allowing these microorganisms to produce their own unique secondary metabolites (37). In the present study, 1000s of bacteria were found during serial dilutions. Only isolates that were observed after 48 hours in 37°C were selected for this study. The isolates were observed after 48 hours to allow for the slower growing bacteria to grow, without causing an overgrowth on the plates.

During this study, majority of samples collected throughout the piney woods showed antimicrobial properties against ESKAPE relatives. However, only 30 of the 300 isolates stored showed antimicrobial properties against MDR *Salmonella* strains. These results might be due to the fact that these MDR *Salmonella* strains are clinical isolates that are already resistant to multiple antimicrobial agents, as well as *Salmonella* being a Gram-negative bacterium. The clinical isolates used have been exposed to varying

transposons and plasmids that further allow their virulence as well as an increase in their resistant genes. Additionally, Gram-negative bacteria are composed of an outer membrane that is impermeable to lipophilic solutes as well as a thin peptidoglycan layer (37). In contrast, Gram-positive bacteria do not have this extra outer membrane, but instead a thick peptidoglycan layer that is not an effective permeability barrier for antimicrobial agents (37). Which would also explain why *Bacillus subtilis* showed the most inhibition from the various soil isolates, as it is a Gram-positive bacterium.

We have successfully shown that soil bacteria can be isolated and can actively inhibit growth of MDR *Salmonella* strains. We have also shown that using ethyl acetate, we can extract antimicrobial compounds that are still able inhibit growth. Upon extraction, more of the isolates showed antibacterial properties against the MDR *Salmonella* strains. The soil isolate organic extracts were able to inhibit 21-Choleraesuis more effectively than the soil isolates alone. Inhibition percentage went from 9% to 73%, supporting that antibacterial compounds can be extracted using ethyl acetate. This is not surprising as ethyl acetate is not very polar and allows for separation of a target secondary metabolites (antibiotics) from the primary metabolites (sugars and amino acids) (36). Additionally, this trend we see of increasing inhibition percentages could be explained through organic extractions allow for the extraction of compounds that are not normally actively secreted. When comparing Table 6 and 7 you will also notice that the strains the soil isolates could inhibit before extraction, are not the exact same after extraction. For example, isolates B217 and B228 went from inhibiting 3 of the MDR *Salmonella* strains to all five. Whereas, isolates B24, B28, B151 and B219, lost the ability to inhibit one or more of the MDR *Salmonella* strains after extraction. These

results suggest that the antimicrobial compounds that allow inhibition in those isolates were lost during organic separation. Therefore, using another solvent or separation technique could allow for isolation of different antibacterial compounds.

Ten “unique” isolates were chosen for future studies; B214, B217, B25, B28, B204, B219, B151, B228, B259, and B264. As seen in the trees created, the soil isolates are most likely different species within *Pseudomonas* (Figure 1). However, it is clear that these isolates are very closely related as it is shown in Figure 2. As these isolates are from the genus *Pseudomonas*, future studies including full genome sequencing would be needed to help understand why certain strains are capable of inhibiting 5 MDR *Salmonella* strains, while others can just inhibit one. Whole genome sequencing will be beneficial as *Pseudomonas* bacteria are known to have high levels of genomic diversity (7). Furthermore, *Pseudomonas* can be found in very diverse environments, allowing for unique metabolic pathways to be adapted for producing various secondary metabolites or antibiotic compounds (7). Thus, comparing genomes will help in identifying these pathways and secondary metabolites for commercial use.

Conclusion

In this present study we showed that antibiotic producing bacteria can be isolated from the soil. Furthermore, we have shown that soil isolates have the ability to inhibit the growth of MDR clinical strains. Specifically, *Pseudomonas* species have antibacterial properties against 5 MDR *Salmonella* strains. This study was limited as certain conditions were used to isolate the soil species. We are in the middle of current studies using different initial conditions in hopes to isolate a more diverse range of antibiotic producers.

Additionally, in this study we have shown that extracts from antimicrobial producing bacteria can also be used as an antimicrobial agent. However, with every study comes its limitations as we did not identify the specific antimicrobial compounds responsible for inhibition. However, future studies would focus on metabolite identification. Other researchers have done this starting with the preliminary step we did; extraction of crude compounds by ethyl acetate (28). Through extraction and purification via thin layer chromatography of these compounds, Nandhini, S, *et al*. were able to identify these antimicrobial compounds via UV-Visible spectrometer, FT-IR, Gas chromatography and mass spectrometry techniques (28). Other studies to find specific antibiotic compounds is through transposon mutagenesis experiments and genome mining (7). In order to find a true novel antibiotic for MDR bacteria that can be used in the heath field, all listed experiments would need to be conducted.

CHAPTER III

Comparative Genomics study on three unique *Pseudomonas* soil isolates

---

This thesis follows the style and format of *Wiley Microbiology*.

Abstract

Each year, thousands of people die due to multi-drug resistant pathogens. Controlling the growing multi-drug resistant pathogen outbreaks can be nearly impossible. However, through new developments in technologies, we have a better chance in the discovery and development of new antibiotics. Whole genome technologies have drastically changed how we analyze microorganisms. We now have easy access to identify pathways, drug targets, virulence genes, and other genomic features. Fast and cheaper whole genome sequencing capabilities has allowed for more research involved in the fight against multi-drug resistant bacteria.

Here we report three soil isolates belonging to the intrageneric group *Pseudomonas fluorescens* that possess different antibacterial properties. These three isolates are capable of inhibiting five multi-drug resistant *Salmonella* stains. Through whole genome sequencing and comparative genomic studies, we have identified unique genomic features, along with homologous genomic features that allow for antibacterial production. Various pathways, genes, secondary metabolites, and protein families were found in each genome that potentially contribute to the *Pseudomonas* isolates successful inhibition of the five multi-drug resistant *Salmonella* stains.

Keywords: *Pseudomonas fluorescens*, Multi-drug resistance, Antibiotics, Drug discovery, Whole-genome sequencing, Comparative genomics

Introduction

The “CDC’s Antibiotic Resistance Threats in the United States, 2019 Report” details that more than 2.8 million antibiotic-resistant infections occur in the United States each year, and more than 35,000 people die as a result (15). Controlling the growing multi-drug resistant (MDR) pathogen outbreaks can be very challenging. However, through advances in molecular technologies, we are able to get one step closer to reducing the catastrophic events that follow an increase in MDR pathogens.

Whole-genome sequencing (WGS) has been a huge advancement in not only the control of MDR pathogens, but also in discovery of novel antibiotics to combat MDR pathogens (34). WGS allows for the identification of antimicrobial resistance (AMR) genes, virulence genes, as well as metabolic pathways and the genes involved. These genomic features allow us to control MDR pathogens by identifying potential antibiotics or antibacterial compounds and also allow us to treat these MDR pathogens by identifying specific targets antibiotics can attack.

As it has become easier to analyze, and sequencing technologies are becoming more cost efficient, bacterial genomes from a wide variety of isolates are being sequenced and becoming available to the public. This, along with the innovative techniques that allow global transcriptional and proteomic profiling of bacterial, makes discovering novel antibiotics more accessible (31). Within the last few years, many different methods utilizing bacterial genomes have been uncovered in hopes of finding novel antibiotics. Target based antibiotic discovery, comparative genomics, essential gene targeting, transcriptomics, proteomics, and identification of ribosomally encoded peptide antibiotics have all been explored for the development of new antibiotics (31).

We have come a long way since the first bacterial genome was sequenced in the 1990's. Now, sequencing can be done within hours, and through the development of various bioinformatic applications, assembly, annotation, and analysis can be achieved in days. In fact, from 2009 to 2014, there was a huge increase in complete bacterial genomes from 1,000 to 14,000 (31). In the beginning, the hopes of comparative genomics were to be able to quickly identify potential targets for new antibiotics (31). The idea was, through the 1000s of available genomes, one could compare genomes of pathogenic strains and find a highly conserved bacterial component that an antibiotic could target. However, after years of trying, many studies fell short, as only a few clinically approved antibiotics were able to be produced through JUST comparative studies (31).

Researchers soon realized that comparative genomics was just a piece to the puzzle in antibiotic discovery. WGS is still a valuable tool as the available of thousands of genomes led to determination of a core genome for a given species. For example, a study using *Pseudomonas* found 5,233 genes in the core genome, and of those 1840 genes encode metabolic functions (38). This leads us to using additional approaches and methods for antibiotic discovery.

Comparison between numerous bacterial genomes has uncovered a large number of biosynthetic pathways that are not always highly expressed under "normal" conditions (31). These "cryptic" pathways are found in gene clusters of antibiotic producing bacteria. This discovery has led to the hypothesis that these pathways code for enzymes involved in the synthesis of molecules with antimicrobial activity (31). Thus, leading to additional methodologies for discovering novel antibiotics.

Genome mining can lead to the discovery of peptides with novel mechanisms of action (31). Specifically, bacteriocins, can be targeted through allowing competition to occur between bacteria. Bacteriocins are genetically encoded bacterial antimicrobial peptides that can have a narrow or broad spectrum of activity (31). Other non-ribosomal antibiotics and secondary metabolites can also be found through genome mining. Tools, such as anti-SMASH, allow for the identification of these genes in various genomes (39). However, these tools are often considered as just a "prediction step", as connecting bacteriocin genes to their peptide product requires rigorous experiments including isolation of the peptide and mass spectrometry (31).

Nonetheless, various bioinformatic tools are a great way to start the process of antibiotic discovery. In this study we did comparative genomics on three *Pseudomonas* soil isolates that showed antimicrobial properties. Through different bioinformatic tools we were able to predict the species of these soil isolates, along with compare various genomic features, pathways, protein features, and gene clusters encoding biosynthesis of secondary metabolites. These comparisons help us better understand the diverse genomic feature within the *Pseudomonas* species and further emphasizes the need for future studies regarding *Pseudomonas* soil isolates an antibiotic discovery.

Materials and Methods

DNA Extraction

DNA was extracted from 10 antibiotic producing *Pseudomonas* and *Flavobacterium* strains using Qiagen's DNeasy Blood and Tissue Kit, following the protocol: "Pretreatment for Gram-Negative Bacteria". Briefly, all isolates were grown in

10% TSA broth for 24 hours at 30°C in a shaking incubator. Following the incubation, 3mL was centrifuged at 7500 rpm for ten minutes and then the pellet was resuspended in 180μL of ATL buffer. Next, 20μL of proteinase K was added, and the mixture was vortexed and incubated at 56°C in a water bath overnight. Afterwards, the tubes were vortexed and 200μL of AL buffer and 200μL of 100% ETOH were added respectively with vortexing in between. The mixture was then placed into a spin column and collection tube and centrifuged at 8000 rpm for one minute. After discarding the flow through 500μL of AW1 buffer was added and again the mixture was centrifuged at 8000 rpm for one minute. Again, the flow through was discarded and 500μL of AW2 buffer was added and the mixture was centrifuged at 14000 rpm for three minutes. Lastly, 200μL of AE buffer was added and incubated for one minute at room temperature, then centrifuged for one minute at 8000 rpm. Before sending samples to Psomagen, purity was checked using the Nanodrop.

Whole Genome Sequencing

Samples were sent to Psomagen (Rockville, MD) for whole genome sequencing on an Illumina platform. Paired-end reads with a read length of 301 were achieved using TruSeq DNA PCR Free (350) library kit. The library kit used was TruSeq DNA PCR-Free kit/TruSeq DNA PCR-Free, Sample Preparation Guide, Part # 15036187 Rev. A. Briefly, after quality control is performed on samples, the library is constructed. The sequencing library was prepared by random fragmentation of the DNA sample, followed by 5’ and 3’ adapter ligation. These adapter-ligated fragments were then PCR amplified and gel purified. This library was then loaded into a flow cell where fragments were captured on a lawn of surface-bound oligos complementary to the library adapters. Each

fragment was then amplified through bridge amplification. Lastly, sequencing data was then converted into raw data. The Illumina sequencer generates raw images using RTA (Real Time Analysis). The BCL (base calls) binary is converted into FASTQ utilizing Illumina package bcl2fastq.

Comparative Genomics

PATRIC was used for some of the analysis of the isolates. PATRIC is the Bacterial Bioinformatics Resource Center, designed to support the biomedical research community's work on bacterial infectious diseases (40). Three isolates, *Pseudomonas* B217 (hereafter, B217), *Pseudomonas* B151 (hereafter, B151), and *Pseudomonas* B228 (hereafter, B228), were selected for comparative studies. Raw data was uploaded into PATRIC and the Comprehensive Genome Analysis service was used for the comparative genomic studies, default settings were used. The Comprehensive Genome Analysis services performs a comprehensive analysis including; assembly, annotation, identification of nearest neighbors, a basic comparative analysis that includes a subsystem summary, phylogenetic tree, and the features that distinguish the genome from its nearest neighbors (40). The service "Assembly" was also used separately to use other programs needed for comparative studies. PATRIC was used to compare genomic features using the full genome report provided (41-59). Within PATRIC, specific protein features, pathways, and specialty genes were investigated.

Species Identification

The genus *Pseudomonas* is known to have a high level of genetic diversity amongst species, thus making classification down to the species difficult. For this study, a variety of programs were investigated to help identify the *Pseudomonas* isolates. The

assembled FASTA contig file for all three isolates were uploaded onto the Type Strain Genome Server (TYGS) (60-67). Additionally, Multi-Locus Sequence Typing (MLST) 2.0 (68), Species Finder-2.0 Server (69), and Reads2Type (70) were all used.

Identification of antimicrobial compounds

The assembled genomes of the three isolates, B217, B151, and B228, was analyzed using Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) 5.0 webserver (39) and we also used Antibiotic Resistant Target Seeker- ARTS 2.0 in combination with antiSMASH (71,72).

Results and Discussion

Species Prediction

Whole genome sequencing of bacterial strains has allowed for more discriminatory power to differentiate between species like *Pseudomonas* (73). *Pseudomonas* has been described as one of the most diverse and ubiquitous bacterial genera worldwide (74). From 2009 to 2018, more than 70 novel species have been added to the genus *Pseudomonas*, with an average of 10 new species identified every year (74). As of 2020, there are more than 220 different species that have been identified and characterized (75). As these new tools allow for better taxonomy identification, it also creates confusion and reclassification of various species. As a result, researchers often have trouble classifying these closely related species in *Pseudomonas* (75).

In this study, we isolated three unique *Pseudomonas* species from soil in the Piney Woods of Texas, specifically, Huntsville and The Woodlands. These isolates showed antimicrobial properties against ESKAPE relatives and MDR *Salmonella* strains.

After whole genome sequencing, we aimed to classify these three isolates at the species level. Through various bioinformatic programs, we have predicted that these three isolates belong to *Pseudomonas fluorescens.*

Within the genera *Pseudomonas*, there are numerous groups and subgroups, one being *Pseudomonas fluorescens* (77). The diversity grows even more within *Pseudomonas fluorescens* as there are more than 50 named species that differ from multilocus sequence analysis and phylogenomic analysis (77). For this reason, strains belonging to the *Pseudomonas fluorescens* complex are often difficult to taxonomically classify. However, through the use of PATRIC, Reads2Type-2.0, Species Finder 2.0 Server, MLST 2.0, and TYGS, we believe that the isolates belong to one of the 50 species described in the *Pseudomonas fluorescens* complex.

We first used Reads2Type as it is a web-based tool for taxonomy identification based on whole bacterial genome sequence data (70). This tool takes the uploaded FASTQ files of the whole genome sequence and maps it against marker probes that are derived from currently available bacteria complete genomes, giving a species prediction with 99.5 % accuracy. Read 1 and read 2 of the *Pseudomonas* isolates were uploaded separately and both reads results in *Pseudomonas fluorescens* for all *Pseudomonas* isolates.

Species Finder-2.0 web-server was also used to help classify these isolates (69). After uploading the contig FASTA files, all three isolates had a 98% identity match with *Pseudomonas* sp. MS586, which is also referred to as *Pseudomonas glycinae*. In July 2020 *Pseudomonas* sp. MS586 was identified as a novel species belonging to the *Pseudomonas koreensis* subgroup (SG) (77). The *Pseudomonas* genus can be divided into

two main intrageneric groups (IG), IG *Pseudomonas aeruginosa* and IG *Pseudomonas fluorescens* (78,79). *P. fluorescens* group is very complex as it is comprised of nine subgroups and six main groups. The main groups are represented by the species *P. fluorescens, P. syringae, P. lutea, P. putida, P. anguilliseptica* and *P. straminea* groups (78,79). The study that identified *Pseudomonas glycinae*. found that the concatenated 16S rRNA, *rpoB, rpoD, and gyrB* gene sequences shares the highest similarity with the *Pseudomonas kribbensis* strain, which belongs to the *Pseudomonas koreensis* SG (77).

The Comprehensive Genome Analysis Service in PATRIC created a "Closest relative tree" for all three isolates (Figures 3-5) and determined that the closest relative to each isolate was *Pseudomonas fluorescens* Pf0-1. The trees generated in PATRIC used Mash/MinHash for the closest reference and representative genomes. To determine the phylogenetic placement of these genomes, PATRIC global protein families were selected from the genomes. MUSCLE was used for the alignment of the protein sequences from the protein families used for phylogenetic placement. The nucleotides for each of those sequences were then mapped to the protein alignment and then RaxML, with fast bootstrapping was used to analyze the joint set of amino acid and nucleotide alignments that were concatenated into a data matrix; generated the support values in the tree. When comparing the three trees, they are pretty similar except B228, as the isolate is not grouped with the other two. We then decided to use the Phylogenic Tree builder service on PATRIC using the 10 closest relatives generated from the Comprehensive Genome Analysis plus all three of the soil isolate genomes. We used the same method that was used to create the tree in the Comprehensive Genome Analysis, except we used 1000, instead of 100 amino acid and nucleotide sequences for the alignment and the tree,

because *Pseudomonas* species are closely related. We found that the B217 and B151 are grouped together, while B228 is branched separately. However, again *Pseudomonas fluorescens* Pf0-1was the closest relative with a branch support of 100 (Figure 6).

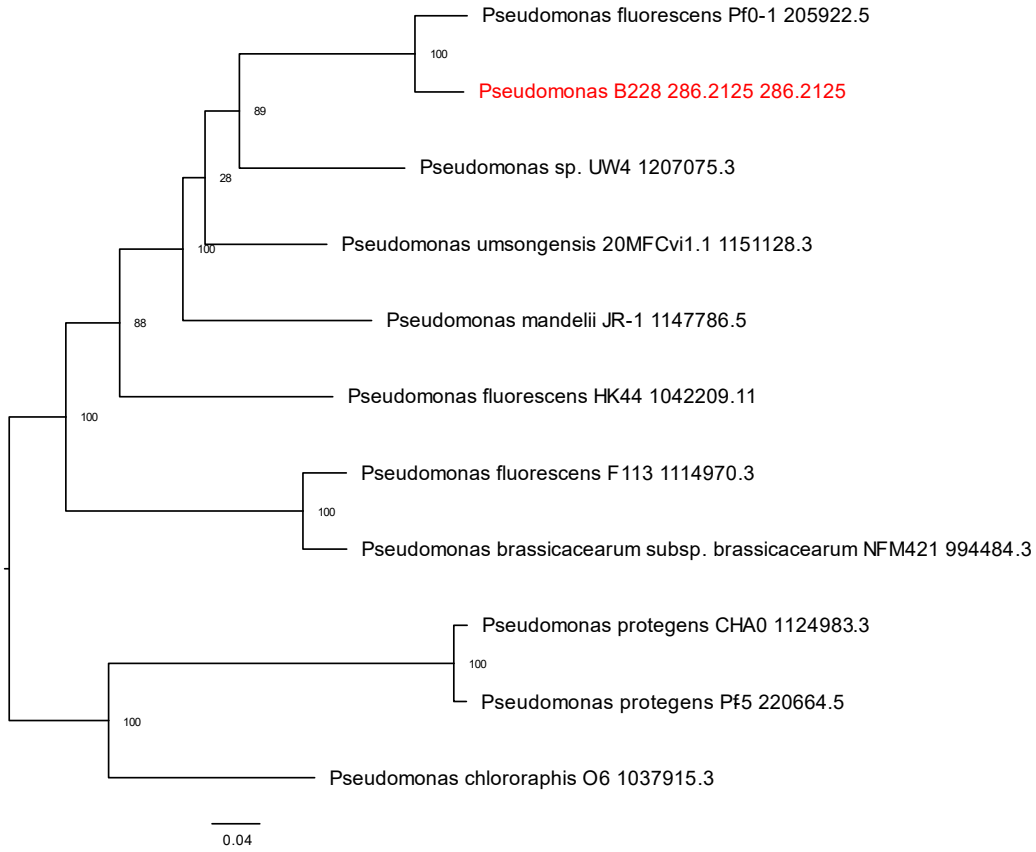
A multi-locus sequence analysis was also completed on the three genomes, the results are outlined in Figures 7-10 (68). All three of the isolates were mapped against *P. fluorescens* locus *glnS, gyrB, ileS, nuoD, recA, rpoB, and rpoD*. *Pseudomonas* B217 had 99.0253% identity 99.4186% coverage on locus *nuoD*, making it an imperfect match. Additionally, red in the figures indicates mismatching amongst the nucleotides. However, from the data we can assume that the three *Pseudomonas* isolates belong to the IG *P. fluorescens.*

Lastly in looking for species identification support, we used TYGS (60-67). After uploading the three assembled sequences, TYGS generated two trees with 17 of the closest relatives to the three soil isolates (Figure 11 & 12). Results from the TYGS analysis predicted that all three isolates may be a novel species. The first tree is using the whole genome sequences (Figure 11), while the second tree (Figure 12) is 16S rDNA gene sequenced based. The two trees differ two trees by the phylogenetic location of the soil isolates. Both trees were inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The tree was rooted at the midpoint and has an average branch support of 92.2% and 67.4% respectively. Additionally, the tree branches have GBDP pseudo-bootstrap support values > 60 % from 100 replications. Figure 11 shows a 100-branch support of all three isolates to *Pseudomonas kribbensis* KCTC 32541T. As discussed earlier, *P. kribbensis* belongs to the SG *Pseudomonas koreensis*, which is a part of IG *Pseudomonas fluorescens.* When looking at Figure 12, the three

isolates are not all in the same branch. B228 has a branch support of 82 with *Pseudomonas kribbensis* KCTC 32541T, while both B151 and B217 share a branch support of 82 with *Pseudomonas atagosis* PS14. In March 2020, *Pseudomonas atagosis* PS14 was reported as a novel soil bacterium (80). It was concluded through a phylogenic study that *Pseudomonas atagosis* PS14 belongs to the *P. fluorescens* lineage but is distinct from other species in that lineage (80).

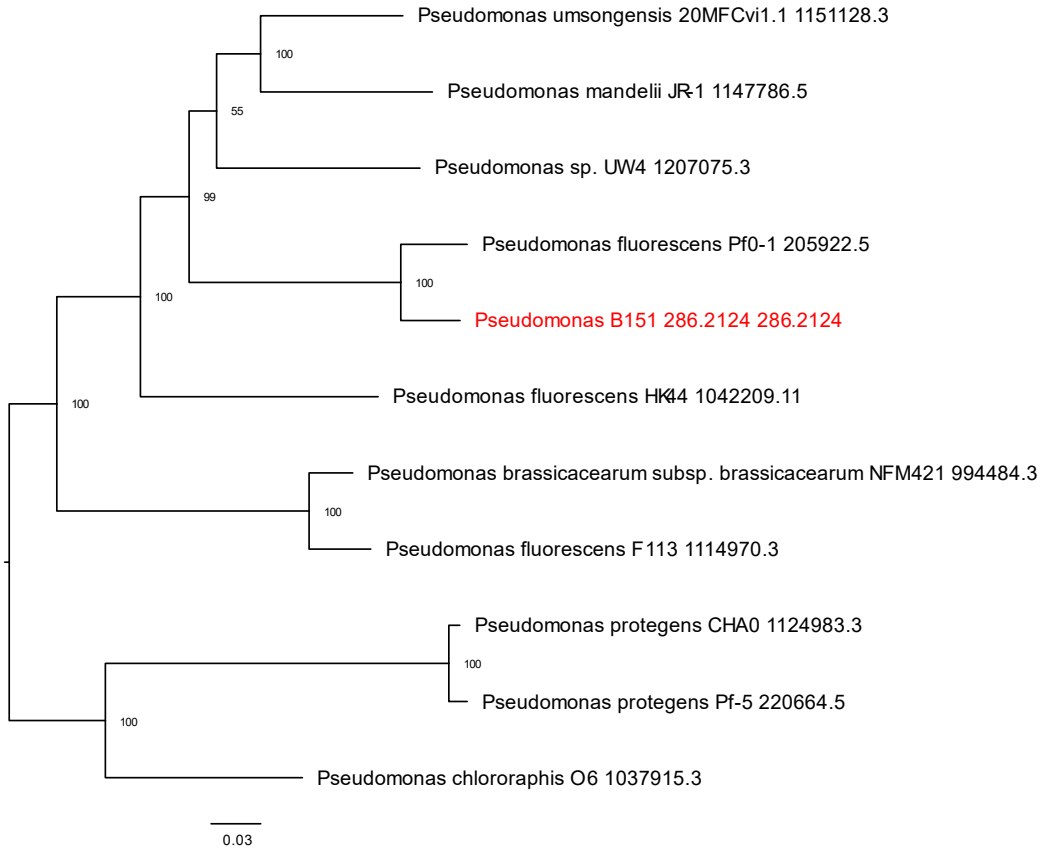
In conclusion, we recognize that the three *Pseudomonas* soil isolates belong to the *P. fluorescens* lineage. However, it is clear that the three isolates are most likely different strains belonging to the various groups and subgroups that compromise IG *P. fluorescens.* Numerous studies point out the difficulty of confidently identifying *Pseudomonas* species and the constant discovery of novel species along with reclassification (74). Further phylogenetic testing, such as amplification and partial sequencing of housekeeping genes, as well as DNA fingerprinting (78) would need to be conducted in order to get an exact match on what SG each isolate belong to.

Figure 3 Pseudomonas B228 Closest Relatives Tree



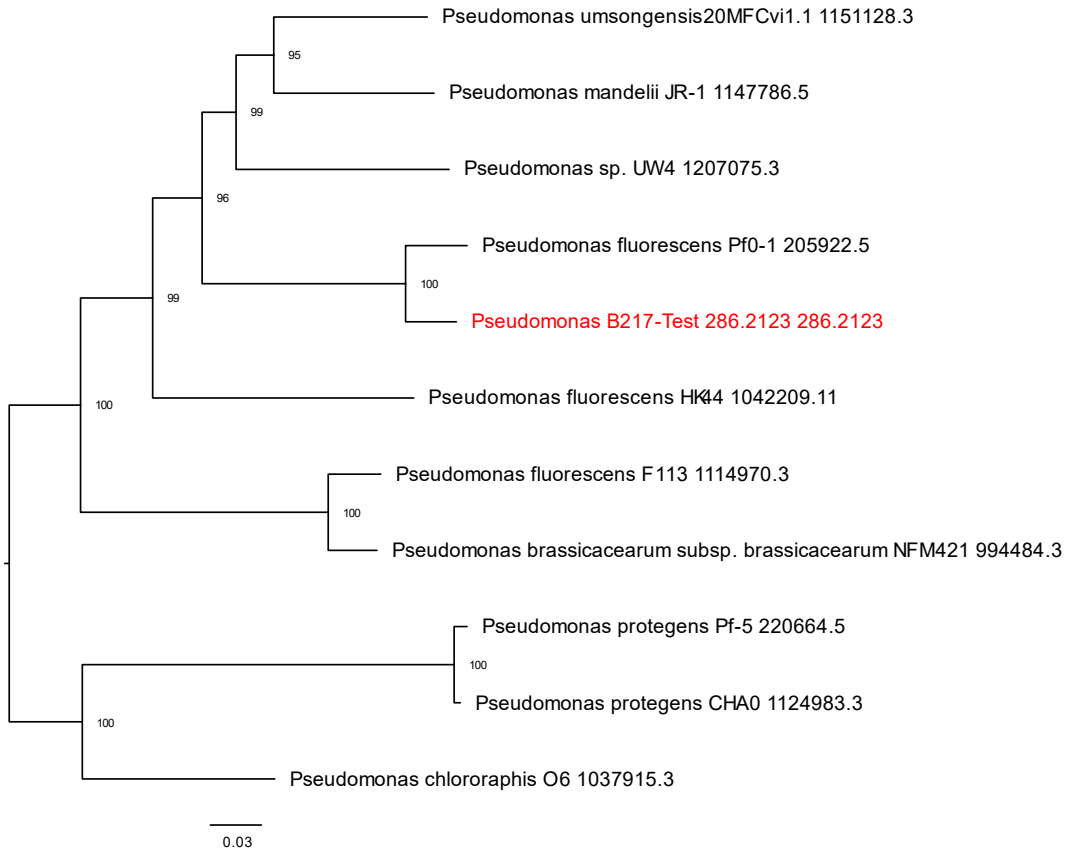
The closest reference and representative genomes to B228 (in red) were identified by Mash/MinHash. PATRIC global protein families (PGFams)were selected from these genomes to determine the phylogenetic placement of this genome

Figure 4 Pseudomonas B151 Closest Relatives Tree



The closest reference and representative genomes to B151 (in red) were identified by Mash/MinHash. PATRIC global protein families (PGFams)were selected from these genomes to determine the phylogenetic placement of this genome

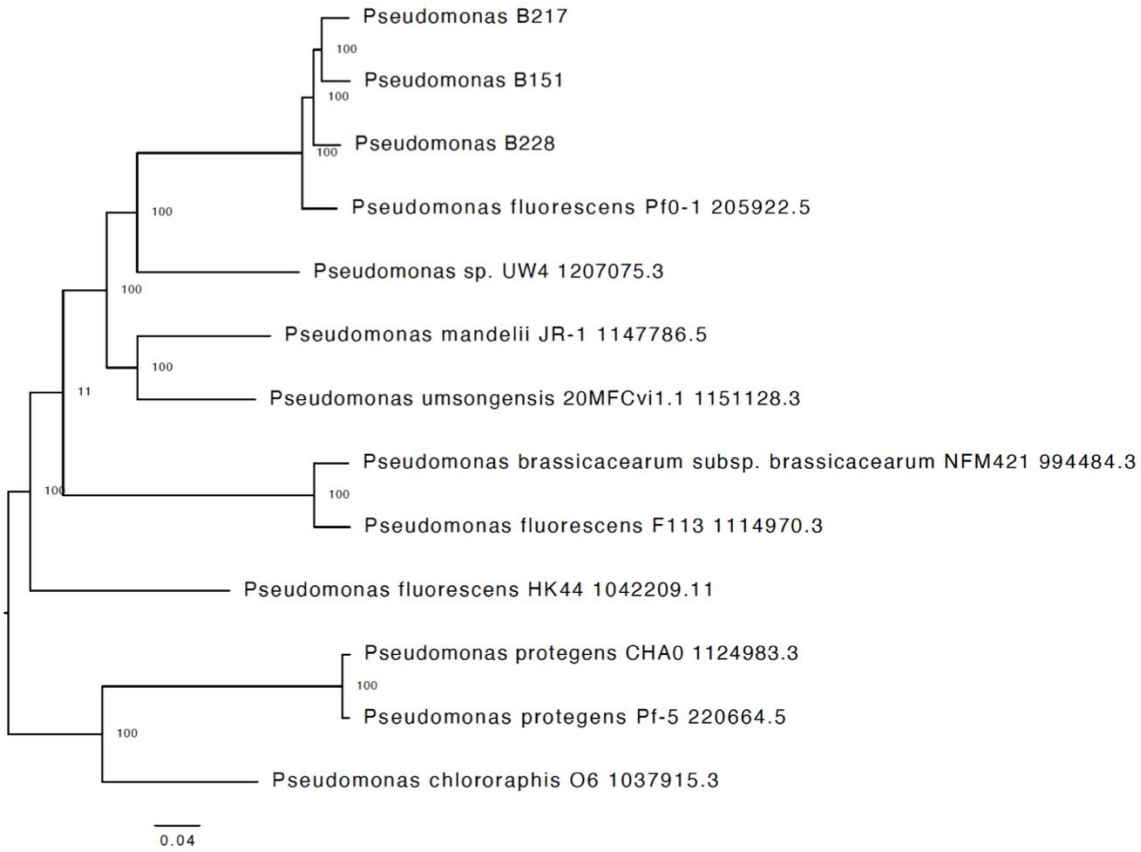
Figure 5 Pseudomonas B217 Closest Relatives Tree



The closest reference and representative genomes to B217 (in red) were identified by Mash/MinHash. PATRIC global protein families (PGFams)were selected from these genomes to determine the phylogenetic placement of this genome

.

Figure 6 Codon Tree of soil isolates and ten closest relatives



A codon tree generated in PATRIC of the three *Pseudomonas* isolates and ten of the closest relatives generated. The three soil isolates are all closely related at the top of the tree.

Figure 7 MLST Profiles

A

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
glnS	97.6048	100	501	501	0	glnS_2*
gyrB	97.1545	100	492	492	0	gyrB_11*
ileS	98.0072	100	552	552	0	ileS_2*
nuoD	99.0253	99.4186	513	516	0	nuoD_94?*
recA	98.6207	100	435	435	0	recA_2*
rpoB	99.1614	100	477	477	0	rpoB_2*
rpoD	98.3333	100	480	480	0	rpoD_2*

B

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
glnS	97.8044	100	501	501	0	glnS_2*
gyrB	96.5447	100	492	492	0	gyrB_31*
ileS	97.8261	100	552	552	0	ileS_2*
nuoD	99.6124	100	516	516	0	nuoD_94*
recA	98.1609	100	435	435	0	recA_2*
rpoB	99.7904	100	477	477	0	rpoB_2*
rpoD	98.3333	100	480	480	0	rpoD_2*

C

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
glnS	98.004	100	501	501	0	glnS_92*
gyrB	97.1545	100	492	492	0	gyrB_11*
ileS	99.0942	100	552	552	0	ileS_2*
nuoD	99.2248	100	516	516	0	nuoD_94*
recA	98.1609	100	435	435	0	recA_2*
rpoB	99.5807	100	477	477	0	rpoB_2*
rpoD	98.3333	100	480	480	0	rpoD_2*

Locus: MLST locus against which the input sequence has been aligned. Identity: Percentage of nucleotides that are identical between the best matching MLST allele in the database and the corresponding sequence in the genome. Coverage:Percent coverage. Alignment Length: Length of the alignment between the best matching MLST allele in the database and the corresponding sequence in the genome. Allele Length: Length of the best matching MLST allele in the database. Gaps: Number of gaps in the HSP. Allele: Name of the best matching MLST allele A) MLST profile of *Pseudomonas* B217, B) MLST profile of *Pseudomonas* B151, and C) MLST profile of *Pseudomonas* B228.

Figure 8 Pseudomonas B217 Allele Alignment MLST

Green indicates matching nucleotides, red indicates mismatches, and the grey is the part of the MLST allele sequence which is not part of the HPS.

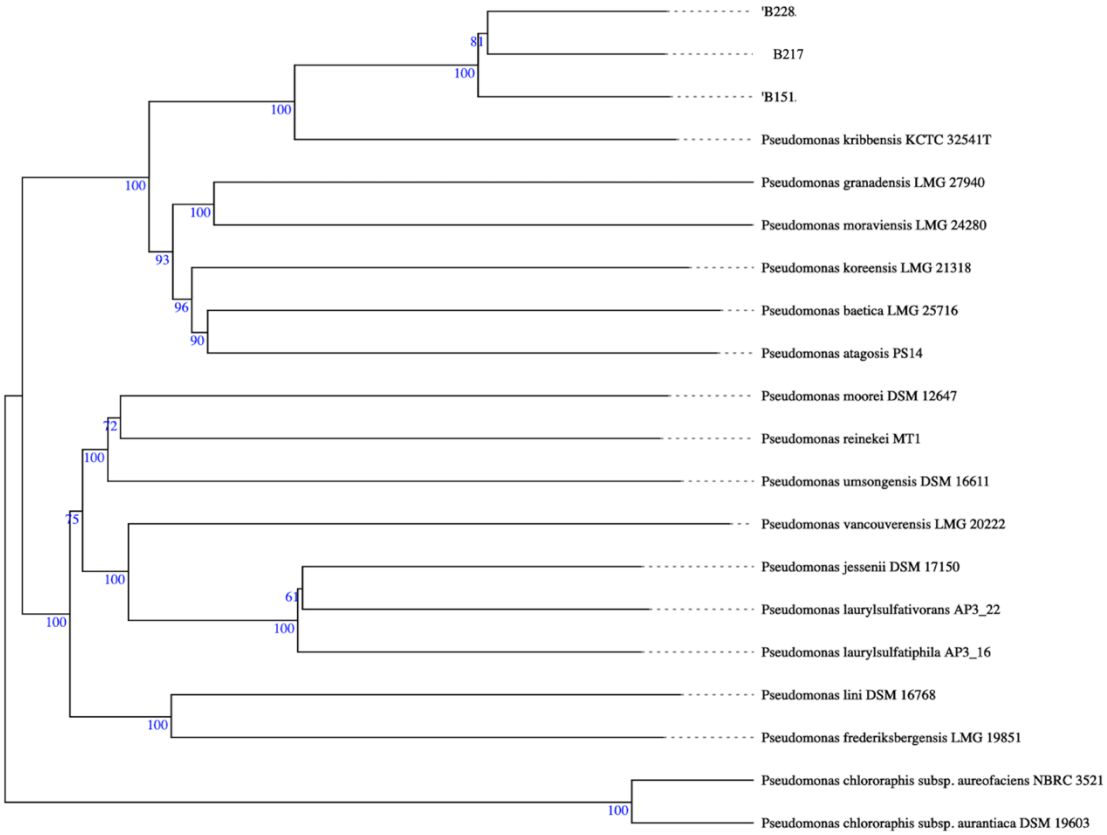
Figure 9 Pseudomonas B151 Allele Alignment MLST

Green indicates matching nucleotides, red indicates mismatches, and the grey is the part of the MLST allele sequence which is not part of the HPS.

Figure 10 Pseudomonas B228 Allele Alignment MLST

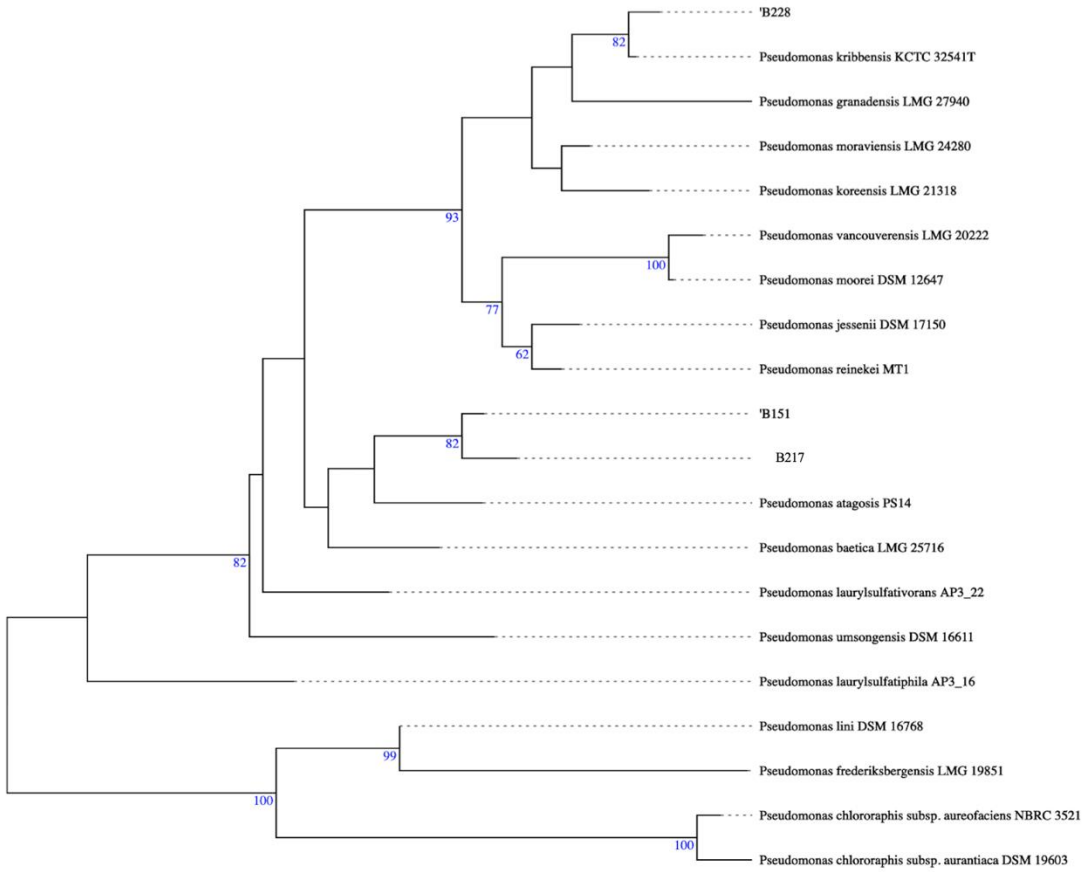
Green indicates matching nucleotides, red indicates mismatches, and the grey is the part of the MLST allele sequence which is not part of the HPS.

Figure 11 GBDP tree (whole-genome sequence-based)



Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 92.2 %. The tree was rooted at the midpoint.

Figure 12 GBDP tree (16S rDNA gene sequence-based)



Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 67.4 %. The tree was rooted at the midpoint.

General genome features

Ten soil isolates were subjected to Illumina paired-end sequencing. The total number of bases, reads, GC (%), Q20 (%), and Q30 (%) are calculated for all 10 samples in Table 8. Two of the samples are *Flavobacterium* and show significantly different data when compared to the similar results between the remaining *Pseudomonas* species. The three isolates B217, B151, and B228 had a range of 933,210,768bp, 945,273,644bp, and 809,322,178bp in total read bases, and their total reads ranged from 3,100,368, 3,140,444, and 2,688,778 respectively. The GC% and AT% for B217, B151, and B228 were 60.56 and 39.44, 60.64 and 39.36, and 60.42 and 39.58 respectively. Lastly, the Q20% and Q30% for B217, B151, and B228 were 91.09 and 82.81, 90.84 and 82.35, and 91.18 and 83.58 respectively.

The Comprehensive Genome Analysis Service was used in PATRIC for the three isolates. The isolates were assembled using SPAdes (41) and annotated using RAST tool kit RASTtk (42). There was on average 27 contigs, an estimated genome length of 6,367,580 bp and a G+C content of 60.69% between the isolates. The N50 length, which is defined as the shortest sequence length at 50% of the genome, averaged to 515,485 bp, while the L50 count, the smallest number of contigs whose length sum produces N50, averaged to be 4 (Table 9). Genomes B228, B151, and B217 were annotated and assigned unique genome identifier of 286.2125, 286.2124, 286.2123 respectively.

The Comprehensive Genome Analysis Service in PATRIC creates Figures 13 through 15, which show a circular representation of the genomes, and the varying genes within each subsystem. Worth noting are both the metabolism and the stress response subsystem as both can potentially contain genes that help the isolates inhibit growth of

the MDR *Salmonella* strains. *Pseudomonas* isolates B228, B151, and B217 contain 1105, 1133, and 1106 genes involved in metabolism respectively and 195, 188, and 187 genes involved in stress response respectively. The slight differences among each genome could account for the ability to inhibit one strain over another.

Assembly, annotation, protein features, specialty genes, and antimicrobial resistant genes data were analyzed. Tables 10 through 13 show a side by side comparison of all three isolates, highlighting their similar genomic features. Although, the isolates show similar genomic features, there are a few notable differences that support the three isolates may be very closely related but probably in different subgroups of IG *P. fluorescens,* or at least different strains within the SG.

The genomes averaged 5,789 protein coding sequences (CDS), 64 transfer RNA (tRNA) genes, and 4 ribosomal RNA (rRNA) genes (Table 10). Upon annotation, certain protein features were analyzed. The genomes averaged 1,146 hypothetical proteins and 4,644 proteins with functional assignments. The proteins with functional assignments included on average 1,276 proteins with Enzyme Commission (EC) numbers, 1,092 with Gene Ontology (GO) assignments, and 957 proteins that were mapped to KEGG pathways. PATRIC annotation also found these genomes averaged 5,646 proteins that belong to the genus-specific protein families (PLFams), and 5,674 proteins that belong to the cross-genus protein families (PGFams) (Table 11).

In Table 9, the genome lengths vary by at least 100,000bp between B228 with 6,374,780bp, B151 with 6,466,605bp, and B217 with 6,261,355bp. Another varying factor between the genomes is the Contig N50; which is the shortest contig length needed to cover 50% of the genome. The N50 values for B228, B151, and B217 are 422,896,

445,171, and 678,417 respectively. These differences are expected as bacteria can undergo numerous rounds of horizontal gene acquisition, creating varying genome contents between closely related strains within a bacterial species (81).

In Table 10, the isolates genes and their annotated features are described. Majority of the annotated genome features are very similar except for the coding region of the genes. The CDS values for B228, B151, and B217 range from 5,381, 5,875, and 5,680 respectively. Often times to compare the extent of diversity within and between species researchers will use computational approaches based on comparisons between: (i) genomic alignments, (ii) coding sequences (CDS) and genomic alignments, and (iii) protein alignments (82). The CDS values of the tree genomes tell us there are differences between that should further be investigated to further identify the specific SG and strain of all three isolates.

Tables 11 and 12 highlight specific protein features and specialty genes in the genomes respectively. Both of which can be further investigated for potential antibiotic properties. Specifically, Table 11 addresses the proteins with pathway assignments for B228 (948), B151 (961), and B217 (961). These pathway assignments were later investigated for potential antibiotic properties, as many antibiotics are produced from secondary metabolites (132). Additionally, Table 12 highlights the total number of specialty genes that were compared individually that will be later discussed in the next section. Lastly, Table 13 shows the antimicrobial resistance genes for the three isolates.

Table 8 Genomic Characteristics of ten antibiotic producing soil isolates

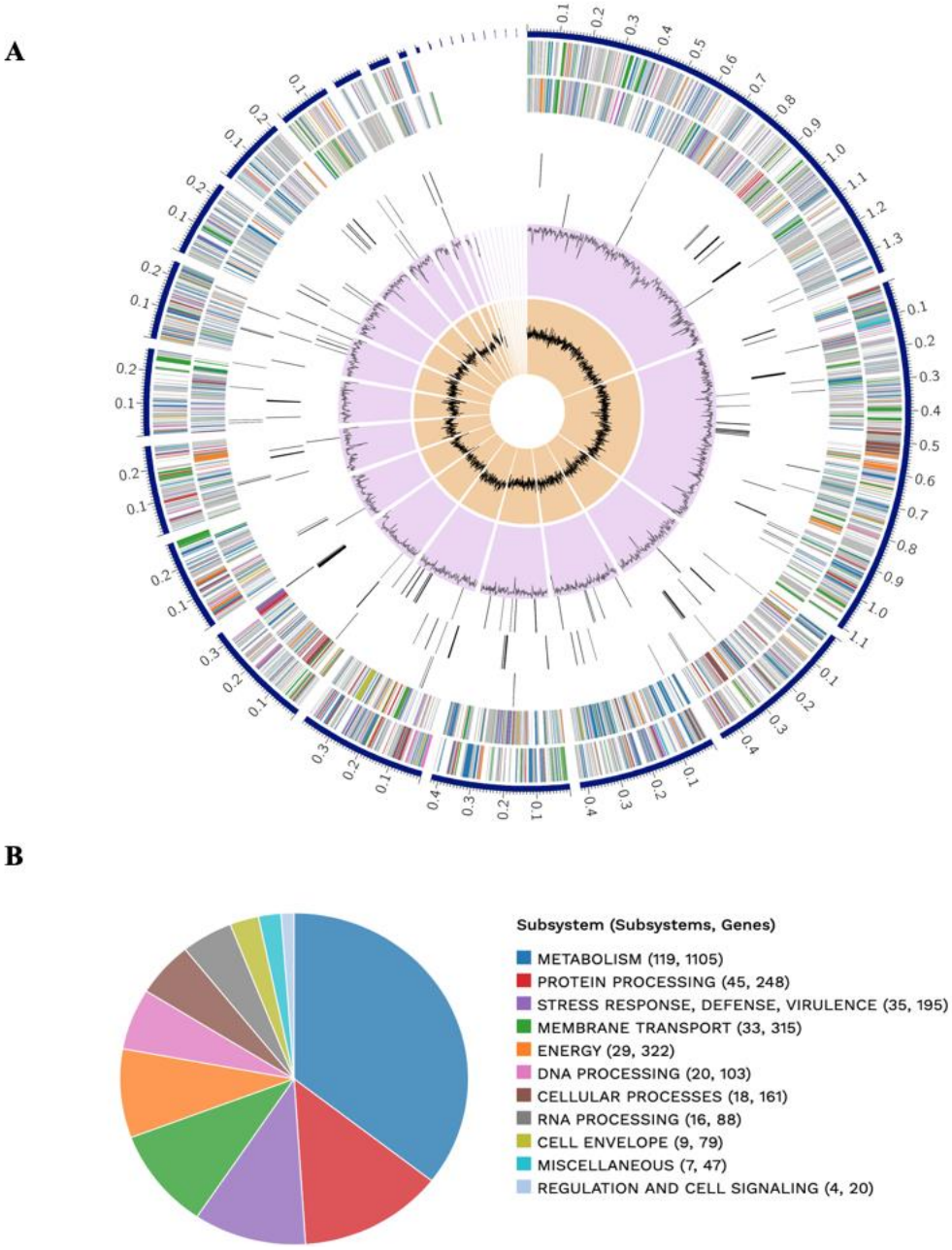
Sample ID	Total read bases (bp)	Total reads	GC (%)	AT (%)	Q20 (%)	Q30 (%)
B214	1,886,981,642	6,269,042	60.56	39.44	91.09	82.81
B217*	933,210,768	3,100,368	60.3	39.7	91.52	83.72
B25	1,015,908,712	3,375,112	35.92	64.08	95.39	89.07
B28	731,493,210	2,430,210	35.94	64.06	95.98	89.98
B204	1,517,341,000	5,041,000	60.52	39.48	91.29	83.05
B219	1,341,041,688	4,455,288	60.71	39.29	91.71	83.82
B151*	945,273,644	3,140,444	60.64	39.36	90.84	82.35
B228*	809,322,178	2,688,778	60.42	39.58	91.51	83.58
B259	1,613,789,828	5,361,428	60.63	39.37	91.18	82.95
B264	1,160,469,380	3,855,380	60.71	39.29	89.59	80.58

Each category is respectively described as: Total read bases: Total number of bases sequenced, Total reads: Total number of reads for Illumina paired-end sequencing, (this value refers to the sum of read1 and read2), GC (%): GC content, AT (%): AT content, Q20(%): Ratio of bases that have phred quality score greater than or equal to 20, Q30(%): Ratio of bases that have phred quality score greater than or equal to 30.
** indicates the three isolates used for comparative genomic studies*

Table 9 Genome Assembly

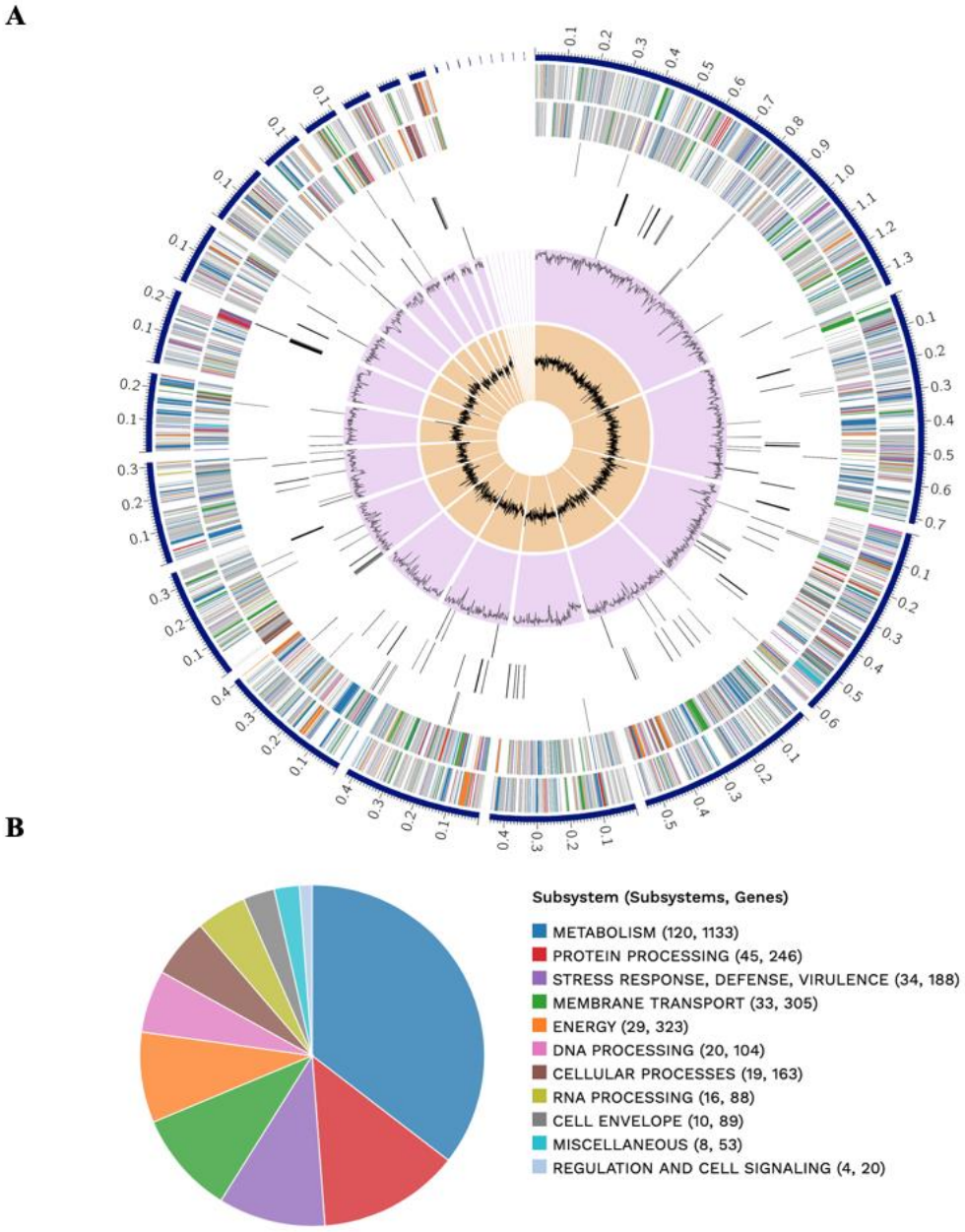
Genome Assembly Details	B228	B151	B217
Contigs	27	27	28
GC content	60.62	60.68	60.77
Plasmids	0	0	0
Contig L50	4	5	4
Genome length	6,374,780 bp	6,466,605 bp	6,261,355 bp
Contig N50	422,896	445,171	678,417
Chromosomes	0	0	0

Figure 13 Circular representation of the Pseudomonas B228 genome structure



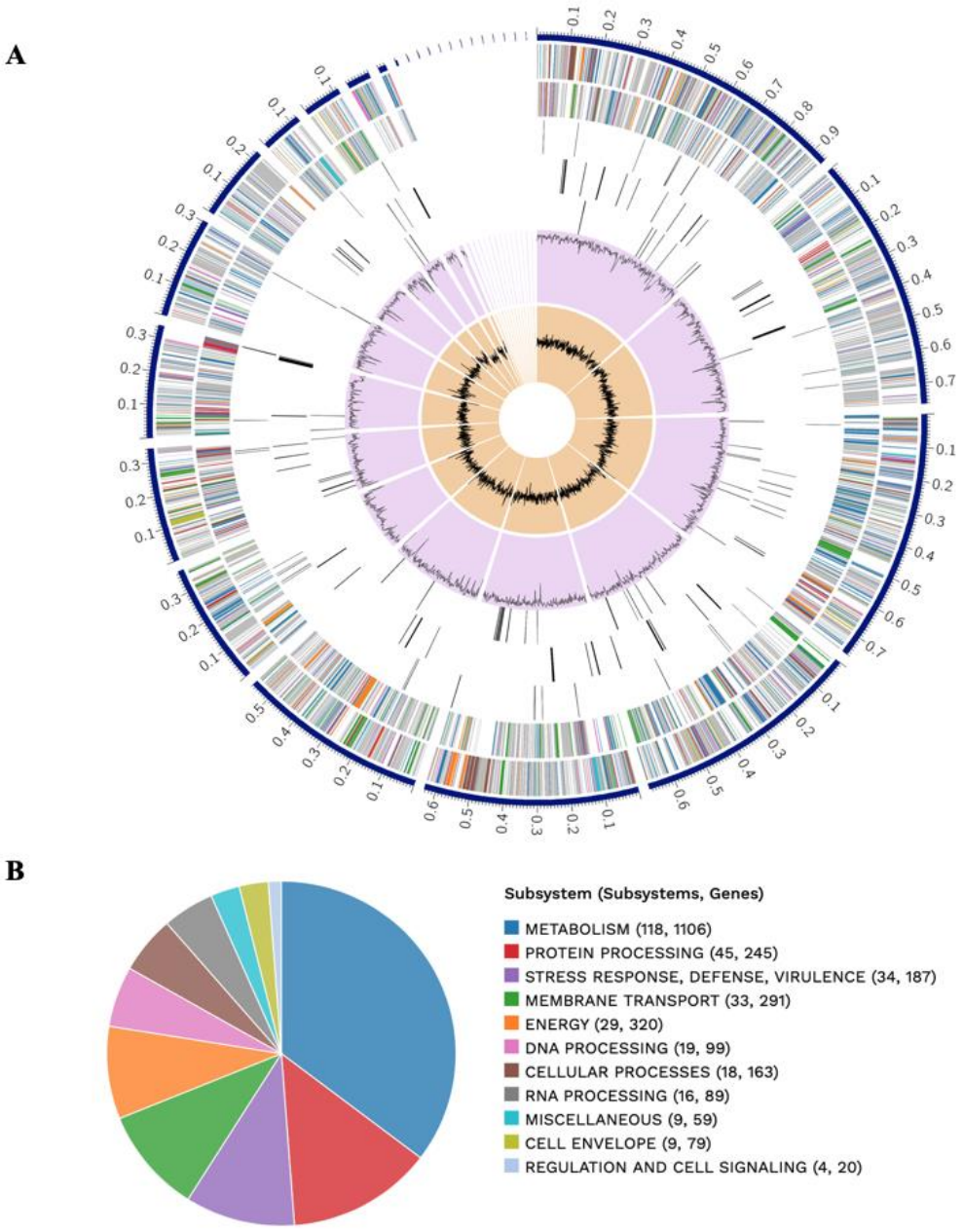
A. A circular graphical display of the distribution of the genome annotations. From outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to know virulence factors, GC content and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem that these genes belong to.
B. An overview of the subsystems for the genome. A subsystem is a set of proteins that together implement a specific biological process or structural complex.

Figure 14 Circular representation of the Pseudomonas B151 genome structure



A. A circular graphical display of the distribution of the genome annotations. From outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to know virulence factors, GC content and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem that these genes belong to.
B. An overview of the subsystems for the genome. A subsystem is a set of proteins that together implement a specific biological process or structural complex.

Figure 15 Circular representation of the Pseudomonas B217 genome structure



A. A circular graphical display of the distribution of the genome annotations. From outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to know virulence factors, GC content and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem that these genes belong to.
B. An overview of the subsystems for the genome. A subsystem is a set of proteins that together implement a specific biological process or structural complex.

Table 10 Genome Annotation

Annotated Genome Features	B228	B151	B217
CDS	5,813	5,875	5,680
tRNA	66	62	63
rRNA	4	4	5
Partial CDS	0	0	0
Miscellaneous RNA	0	0	0
Repeat regions	0	0	0

Table 11 Genome Annotation: Protein Features

Protein Features	B228	B151	B217
Hypothetical proteins	1,187	1,158	1,092
Proteins with functional assignments	4,626	4,717	4,588
Proteins with EC number assignments	1,262	1,290	1,275
Proteins with GO assignments	1,083	1,103	1,091
Proteins with pathway assignments	948	961	961
Proteins with PATRIC genus-specific family assignments	5,665	5,734	5,540
Proteins with PATRIC cross-genus family assignments	5,693	5,761	5,569

Table 12 Genome Annotation: Specialty Genes

Specialty Genes	B228	B151	B217
Antibiotic Resistance Genes (CARD)	4	3	4
Antibiotic Resistance Genes (PATRIC)	73	74	72
Drug Target Genes (Drug Bank)	27	27	26
Drug Target Genes (TTD)	6	6	7
Transporter Genes (TCDB)	80	80	80
Virulence Factor Genes (VFDB)	26	25	25
Virulence Factor Genes (Victors)	27	27	26

The specific source database where homology was found is listed in parentheses.

Table 13 Genome Annotation: Antimicrobial Resistance Genes

Antimicrobial Resistance Genes: B228, B151, B217

Antibiotic activation enzyme	KatG
Antibiotic target in susceptible species	Alr, Ddl, dxr, EF-G, EF-Tu, folA, Dfr, folP, gyrA, gyrB, inhA, fabI, Iso-tRNA, kasA, MurA, rho, rpoB, rpoC, S10p, S12p
Antibiotic target replacement protein	FabG, HtdX
Efflux pump conferring antibiotic resistance	EmrAB-OMF, EmrAB-TolC, MacA, MacB, MdtABC-OMF, MdtABC-TolC, MexAB-OprM, MexEF-OprN, MexEF-OprN system, MexHI-OpmD, MexHI-OpmD system, MexJK-OprM/OpmH, MexVW-OprM, TolC/OpmH, TriABC-OpmH
Genes conferring resistance via absence	gidB
Protein altering cell wall charge conferring antibiotic resistance	GdpD, PgsA
Protein modulating permeability to antibiotic	OccD1/OprD, OccD2/OpdC, OccD3/OpdP, OccD4/OpdT, OccD6/OprQ, OccD7/OpdB, OccK10/OpdN, OccK8/OprE, OprB, OprD family, OprF
Regulator modulating expression of antibiotic resistance genes	OxyR

The Genome Annotation Service in PATRIC uses k-mer-based AMR genes detection method.

Potential antibiotic properties within genome (PATRIC)

The Comparative Pathway Tool on PATRIC was used to compare the three genomes, B228, B151, and B217 pathways, with the ten closely related species generated in the comprehensive genome analysis tool, as well as species predictions from Species Finder-2.0 Server, *Pseudomona*s sp. MS586, and TYGS, *P. kribbensis* and *P. koreensis* were used.

PATRIC generates a list of all of the pathways that have any genes assigned to them from the genomes in that taxon level. As previously mentioned, Table 11 highlighted the protein features of all the genomes. As there were differences in the proteins associated with pathways, we decided to look at the specific pathways in each genome. As seen in Table 14, the three genomes differed. With both B217 and B151 containing their own unique pathways that the other genomes did not. Isolate B217 contains the biphenyl degradation pathway which belongs to the xenobiotics biodegradation and metabolism pathway class. Isolate 151 contained two unique pathways, brassinosteroid biosynthesis and isoflavonoid biosynthesis which both belongs to the biosynthesis of secondary metabolites pathway class. Although the biphenyl degradation pathway and brassinosteroid biosynthesis are not typically associated with antibiotic production, it is worth noticing that these two pathways are not found in each of the genomes. The isoflavonoid biosynthesis pathway found in B151 is intriguing as recent studies concluded it would be extremely useful to synthesize antibacterial isoflavones in the future as flavonoids and related compounds have potent antimicrobial activities (83).

Additionally, between the three genomes, there are 24 pathways that belong to the class “Biosynthesis of Secondary Metabolites”. Among these pathways, 18 are specifically related to antibiotics and/or antibiotic production (Table 15). Many of these pathways can also be found in *Streptomyces* species, which are known to secrete antibiotic products that can be found in medicines we use today (2). Pathways, puromycin biosynthesis, tetracycline biosynthesis, novobiocin biosynthesis, streptomycin biosynthesis, terpenoid backbone biosynthesis, diterpenoid biosynthesis, and isoquinoline alkaloid biosynthesis all contain genes associated with the *Pseudomonas* soil isolates, but also varying *Streptomyces* isolates. Studies using different *Streptomyces* isolates have shown that these pathways result in compounds that have antimicrobial properties (84-89). The remaining pathways have been shown to be involved in antimicrobial production as well (90-96). The identity of these pathways in the *Pseudomonas* soil isolates supports the idea that these pathways could be associated with different antibiotic producing genes that allow for the inhibition of MDR *Salmonella* strains. Additionally, these pathways highlight that the three isolates do in fact have homologous genomic features associated with antibacterial production.

The Purine metabolism pathway and the Pyrimidine metabolism pathway are both associated with nucleotide metabolism; however, it has also been linked to production of antibiotics (i.e coformycin and bacimethrin) (97, 98). Pyrimidine has been extensively studied as researchers have found that that its various derivatives can be used as antimicrobial agents (99). Furthermore, the purine metabolism has been linked to the production of antibiotics in *Streptomyces* (100). Both pathways can be found in all three

genomes, supporting the idea, that closely related strains of *Pseudomonas* have homologous features that allow for the inhibition of MDR *Salmonella* strains.

When investigating the different subsystem classes (Figures 13-15), we noticed many genes are associated with efflux systems. Specifically, isolates B228, B151, and B217 had 77, 86, and 77 genes associated with efflux systems. This association is interesting as multi-drug efflux transporters are known to extrude a variety of antimicrobials from the bacterial cell to facilitate their survival in stressful environments (101).

Protein families were also investigated between all three genomes. Pseudomonas B228, Pseudomonas B151, and Pseudomonas B217 contain 5,559, 5,634, and 5,438 different protein families respectively. Keywords were used to filter through all of the families, including “metabolite”, “AMP”, “T6SS”, “antimicrobial”, “polyketide”, “phenazine”, “PCA”, and “pyoverdine”. Differences between the genomes annotated protein families are outlined in Tables 16-29

Table 16 and 17, show the different proteins family involved in Permease of the drug/metabolite transporter (DMT) superfamily, with B228 having one less family matching this description. The DMT superfamily is a large group of membrane transporters and includes exporters for a wide range of substrates, including toxic compounds and metabolites (102). In combination, these subclasses and protein families further support that these soil isolates have the means to secrete compounds capable of inhibiting the growth of MDR *Salmonella* strains via proteins that are a part of DMT family, including transporters and exporters.

Due to no new antibiotics being created recently, Antimicrobial Peptides (AMP) have been investigated to be a solution to the issue at hand. AMPs are getting more attention and being explored for their potential antimicrobial applications (103). Therefore, AMPs were searched for among the three isolates. Table 18, shows that the three isolates contain different proteins involved in the regulation of AMP and the metabolism involved. One of the protein families found in all three isolates was AmpG permease which has been linked to the production of β-lactamase (104). This enzyme is known to provide multi- resistance to β-lactam antibiotics. Therefore, we hypothesize that this helps protect the isolates from the antibiotics it is secreting. There were 12 families found that are related to AMPs. The AMPs found in the isolates should be further investigated as their properties make them potential candidates for therapeutic application (103).

Table 19 shows the ABC type transport systems and the multi-antimicrobial extrusion (MATE) protein families found in the three genomes. ABC transporters are involved in the secretion of the antibiotic and contribute to self-resistance to the produced antibiotic (105). MATE protein families aid in again protecting bacteria from the antibiotics they secrete, as they function as drug/sodium antiporters (106). Polyketides were also mined for in the genomes as they have been reported as a key subject of antibiotic screening from soil samples (107). Tables 23 and 24 show that the genomes have multiple protein families involved in polyketide synthase modules and related proteins.

Type VI secretion systems (T6SS) were found through the genomes of the three isolates. Tables 20-22, show all of the involved protein families associated with T6SS.

Isolate B151 and B228 showed over 12 more families involved in T6SS than B217. Data has shown that these systems are involved in the regulation of bacterial interactions and competition (108). Specifically, T6SS are able to kill neighboring, non-immune bacterial cells by secreting anti-bacterial proteins directly into the periplasm of the target cells (108). The difference in T6SS protein families between the isolates is expected as B151 was able to inhibit the growth of all 5 MDR *Salmonella* strains, whereas B217 was only to inhibit 3 MDR *Salmonella* strains.

We also wanted to search the genomes for phenazines as naturally derived phenazines exhibit broad-spectrum antibiotic activity against bacteria, fungi, and parasites (109). Additionally, we searched for "PCA" as simple phenazines such a phenazine-1-carboxylic acid (PCA) are known for their broad-spectrum antibiotic activity as well (109). Tables 25-27 show that the genomes contain families of phenazine biosynthesis protein PhzF like and Pca regulon regulatory protein PcaR. These findings should be further investigated as their properties make them potential candidates for protentional antibiotics (109).

Lastly, pyoverdines were mined for in the genomes because it is a fluorescent nonribosomal peptide siderophore made by fluorescent pseudomonads (110). These nonribosomal peptides are known to be medically and industrially relevant as they can be used for antibiotics, immunosuppressants, and anticancer agents (110). Table 28 and 29, show all of the protein families related to pyoverdines. These findings highlight that these isolate's compound have the potential to become an antibiotic.

From the Comprehensive Genome Analysis Service, various protein features and specialty genes for each genome was annotated (Tables 11 and 12). When looking

through the specialty genes analyzed by PATRIC, there were a few unique ones amongst the three isolates. Specialty genes in PATRIC can be defined as antibiotic resistant genes, virulence genes, drug target genes, and human homologs. *Pseudomonas* B228, *Pseudomonas* B151, and *Pseudomonas* B217 contain 243, 240, and 240 specialty genes respectively.

Pseudomonas B217 was missing one of the *OprD* family genes that the other two isolates contained. *OprD* gene is an antibiotic resistance gene belonging to the *OprD* gene family with a product of an outer membrane low permeability porin. This gene has been explored for the new therapeutic targets as *OprD* is responsible for binding and passage of carbapenems, histidine, lysine, and arginine (111). Carbapenems are a class of antibiotics and the *OprD* genes has been found to be responsible for *Pseudomonas* spp. resistance against carbapenems (111). Therefore, excess of these gene families could explain one reason why B151 is able to inhibit more MDR Salmonella strains than B217.

In comparison with the other genomes, *Pseudomonas* B151 genome contained an extra *dxr* gene. The *dxr* gene is involved in Terpenoid backbone biosynthesis (44). Terpenes and their derivatives have been described to have potent antimicrobial activity (112). This extra gene could be part of the reason why B151 is able to inhibit more MDR Salmonella strains than the other two genomes.

Additionally, *Pseudomonas* B228 and *Pseudomonas* B151 both contained genes only found in their genomes when compared to each other. In *Pseudomonas* B151, the *modA* gene was unique to the genome, while in *Pseudomonas* B228 genes *PA1157* and *algA* were unique to the genome. *modA* is a periplasmic binding protein of an ABC transporter (111). ABC transporters are involved in the secretion antibiotics through the

cell membrane and also contribute to self-resistance to the produced antibiotic (114). Therefore, the *modA* gene could be a significant benefit in secreting antibiotics against MDR *Salmonella* strains. *PA1157* is a probable two-component response regulator that allows the regulation of response to changes in different environmental conditions and *algA* produces a precursor for alginate polymerization (115). *PA1157* is a part of the *OmpR* family which is a transcriptional controller of porin expression (116). Researchers have found that altered or absent porins reduce access of polar antibiotics across the outer membrane (116). Therefore, the addition of this gene can protect B228 from the antibiotics it produces. The finding of *algA* was interesting as alginate has been found to be useful in wound healing, drug delivery, and tissue engineering applications to date (117).

The three genomes also had many similar specialty genes (i.e. transporter, antibiotic resistant, and virulence) as well, including *MacA, PA4115, algR, Pvds*, and *waaF*. These specialty genes found in all three isolates are known to be associated with antimicrobial functions. *MacA* is a is a membrane fusion protein that forms an antibiotic efflux complex with MacB and TolC (118, 119). *PA4115* genes are known to possess pyrimidine/purine nucleotide 5′-monophosphate nucleosidase activity (120). *algR* genes are transcriptional regulators that controls alginate production, type IV pilus function, and virulence (121). *Pvds* is a sigma factor that controls pyoverdine biosynthesis (122). Lastly, *waaF* encodes proteins involved in the formation of LPS (123). All of these genes contribute to these *Pseudomonas* isolates ability to produce antibiotics and/or protect itself from the antibiotics secreted.

Table 14 Unique Pathways

Soil Isolate	Unique Pathway ID	Pathway Name	Pathway Class
B217	00621	Biphenyl degradation	Xenobiotics Biodegradation & Metabolism
B151	00905	Brassinosteroid biosynthesis	Biosynthesis of Secondary Metabolites
B151	00943	Isoflavonoid biosynthesis	Biosynthesis of Secondary Metabolites

Table 15 Biosynthesis of secondary metabolites pathways in all 3 genomes associated with antimicrobial products

Pathway ID	Pathway Name
00231	Puromycin biosynthesis
00232	Caffeine metabolism
00253	Tetracycline biosynthesis
00311	Penicillin and cephalosporin biosynthesis
00312	beta-Lactam resistance
00401	Novobiocin biosynthesis
00521	Streptomycin biosynthesis
00900	Terpenoid backbone biosynthesis
00901	Indole alkaloid biosynthesis
00904	Diterpenoid biosynthesis
00940	Phenylpropanoid biosynthesis
00941	Flavonoid biosynthesis
00942	Anthocyanin biosynthesis
00943	Isoflavonoid biosynthesis
00944	Flavone and flavanol biosynthesis
00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis
00950	Isoquinoline alkaloid biosynthesis
00965	Betalain biosynthesis

Table 16 "Metabolite" Protein Families for Pseudomonas B217 and Pseudomonas B151

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001177	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	300	300	300
PLF_286_00001533	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297
PLF_286_00006314	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	316	316	316
PLF_286_00006706	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	302	302	302
PLF_286_00007447	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297
PLF_286_00009205	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	287	287	287
PLF_286_00009977	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	154	154	154
PLF_286_00010518	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	315	315	315
PLF_286_00010821	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	304	304	304
PLF_286_00024749	**1**	**1**	**Permease of the drug/metabolite transporter (DMT) superfamily**	**298**	**298**	**298**
PLF_286_00107418	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297

When the keyword "Metabolite" was used to filter through the protein families, 11 families were identified. In bold represents the one family not found in Pseudomonas B228

Table 17 "Metabolite" Protein Families for Pseudomonas B228

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001177	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	300	300	300
PLF_286_00001533	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297
PLF_286_00006314	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	291	291	291
PLF_286_00006706	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	302	302	302
PLF_286_00007447	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297
PLF_286_00009205	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	287	287	287
PLF_286_00009977	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	154	154	154
PLF_286_00010518	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	315	315	315
PLF_286_00010821	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	304	304	304
PLF_286_00107418	1	1	Permease of the drug/metabolite transporter (DMT) superfamily	297	297	297

When the keyword "Metabolite" was used to filter through the protein families, 10 families were identified.

Table 18 "AMP" Protein Families for Pseudomonas B228, B151, and B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00000465	1	1	AMP nucleosidase (EC 3.2.2.4)	487	487	487
PLF_286_00000837	1	1	Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)	133	133	133
PLF_286_00001092	1	1	Threonylcarbamoyl-AMP synthase (EC 2.7.7.87)	185	185	185
PLF_286_00005806	1	1	HAMP domain protein	502	502	502
PLF_286_00005866	1	1	AmpG permease	519	519	519
PLF_286_00006168	1	1	Uncharacterized protein EC-HemY in Proteobacteria (unrelated to HemY-type PPO in GramPositives)	412	412	412
PLF_286_00006283	1	1	FIGfam138462: Acyl-CoA synthetase, AMP-(fatty) acid ligase / (3R)-hydroxymyristoyl-[ACP] dehydratase (EC 4.2.1.-)	559	559	559
PLF_286_00008094	1	1	Glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46) periplasmic (secreted in GramPositives)	375	375	375
PLF_286_00010929	1	1	Transcriptional regulator AmpR, LysR family	291	291	291
PLF_286_00097704	1	1	Cyclic AMP receptor protein @ Vfr transcriptional regulator	214	214	214
PLF_286_00099532	1	1	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	260	260	260
PLF_286_00157717	1	1	Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases	554	554	554

When the keyword "AMP" was used to filter through the protein families, 12 families were identified.

Table 19 "Antimicrobial" Protein Families for Pseudomonas B228, B151, and B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00000684	1	1	ABC-type antimicrobial peptide transport system, permease component	421	421	421
PLF_286_00052293	1	1	Multi antimicrobial extrusion protein (Na(+)/drug antiporter), MATE family of MDR efflux pumps	469	469	469
PLF_286_00070244	1	1	ABC-type antimicrobial peptide transport system, ATPase component	227	227	227
PLF_286_00112228	1	1	ABC-type antimicrobial peptide transport system, ATPase component	236	236	236

When the keyword "Antimicrobial" was used to filter through the protein families, 4 families were identified.

Table 20 "T6SS" Protein Families for Pseudomonas B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001402	1	1	T6SS component TssM (IcmF/VasK)	1179	1179	1179
PLF_286_00001403	1	1	T6SS outer membrane component TssL (ImpK/VasF)	291	291	291
PLF_286_00001433	1	1	T6SS component TssF (ImpG/VasA)	595	595	595
PLF_286_00001434	1	1	T6SS component TssC (ImpC/VipB)	491	491	491
PLF_286_00001439	1	1	T6SS component TssB (ImpB/VipA)	167	167	167
PLF_286_00001440	1	1	T6SS component TssK (ImpJ/VasE)	443	443	443
PLF_286_00001464	1	1	T6SS component TssG (ImpH/VasB)	335	335	335
PLF_286_00001557	1	1	T6SS lysozyme-like component TssE	135	135	135
PLF_286_00001571	1	1	T6SS secretion lipoprotein TssJ (VasD)	166	166	166
PLF_286_00012155	1	1	T6SS Serine/threonine protein kinase (EC 2.7.11.1) PpkA	333	333	333
PLF_286_00015869	1	1	T6SS component Hcp	513	513	513
PLF_286_00037841	1	1	T6SS protein serine/threonine phosphatase PppA	242	242	242
PLF_286_00108270	1	1	T6SS sigma-54-dependent regulator VasH	508	508	508
PLF_286_00155730	**4**	**1**	**T6SS component Hcp**	**22**	**171**	**103**
PLF_286_00279238	1	1	T6SS component Hcp	158	158	158
PLF_286_00279796	1	1	T6SS AAA+ chaperone ClpV (TssH)	885	885	885
PLF_286_00364932	1	1	T6SS forkhead associated domain protein ImpI/VasC	397	397	397
PLF_286_00367574	1	1	T6SS PAAR-repeat protein	168	168	168
PLF_286_00373843	1	1	T6SS PAAR-repeat protein / RhaS protein	1564	1564	1564
PLF_286_00374554	**3**	**1**	**T6SS PAAR-repeat protein**	**164**	**165**	**164**
PLF_286_00436076	1	1	T6SS component TssA (ImpA)	520	520	520
PLF_286_00443528	**2**	**1**	**T6SS PAAR-repeat protein / RhaS protein**	**540**	**687**	**613**

When the keyword "T6SS" was used to filter through the protein families, 22 families were identified. In bold represents the multiple proteins not found in the other two genomes.

Table 21 "T6SS" Protein Families for Pseudomonas B228

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001402	1	1	T6SS component TssM (IcmF/VasK)	1179	1179	1179
PLF_286_00001403	1	1	T6SS outer membrane component TssL (ImpK/VasF)	291	291	291
PLF_286_00001433	1	1	T6SS component TssF (ImpG/VasA)	595	595	595
PLF_286_00001434	1	1	T6SS component TssC (ImpC/VipB)	491	491	491
PLF_286_00001439	1	1	T6SS component TssB (ImpB/VipA)	167	167	167
PLF_286_00001440	1	1	T6SS component TssK (ImpJ/VasE)	443	443	443
PLF_286_00001464	1	1	T6SS component TssG (ImpH/VasB)	335	335	335
PLF_286_00001557	1	1	T6SS lysozyme-like component TssE	135	135	135
PLF_286_00001571	1	1	T6SS secretion lipoprotein TssJ (VasD)	166	166	166
PLF_286_00001834	**1**	**1**	**T6SS component TssG (ImpH/VasB)**	**341**	**341**	**341**
PLF_286_00001835	**1**	**1**	**T6SS component TssK (ImpJ/VasE)**	**443**	**443**	**443**
PLF_286_00001847	**1**	**1**	**T6SS component TssC (ImpC/VipB)**	**494**	**494**	**494**
PLF_286_00001853	**1**	**1**	**T6SS component TssF (ImpG/VasA)**	**597**	**597**	**597**
PLF_286_00001861	**1**	**1**	**T6SS component Hcp**	**165**	**165**	**165**
PLF_286_00001862	**1**	**1**	**T6SS component TssB (ImpB/VipA)**	**180**	**180**	**180**
PLF_286_00012155	1	1	T6SS Serine/threonine protein kinase (EC 2.7.11.1) PpkA	333	333	333
PLF_286_00015869	1	1	T6SS component Hcp	513	513	513
PLF_286_00016026	**1**	**1**	**T6SS lysozyme-like component TssE**	**147**	**147**	**147**
PLF_286_00037841	1	1	T6SS protein serine/threonine phosphatase PppA	242	242	242
PLF_286_00075352	**1**	**1**	**T6SS component TssM (IcmF/VasK)**	**1273**	**1273**	**1273**
PLF_286_00095010	**1**	**1**	**T6SS AAA+ chaperone ClpV (TssH)**	**846**	**846**	**846**
PLF_286_00108270	1	1	T6SS sigma-54-dependent regulator VasH	508	508	508
PLF_286_00155730	1	1	T6SS component Hcp	171	171	171
PLF_286_00231985	**2**	**1**	**T6SS PAAR-repeat protein**	**179**	**181**	**180**
PLF_286_00279238	1	1	T6SS component Hcp	161	161	161
PLF_286_00279796	1	1	T6SS AAA+ chaperone ClpV (TssH)	885	885	885
PLF_286_00297984	**1**	**1**	**T6SS component TssA (ImpA)**	**362**	**362**	**362**
PLF_286_00364932	1	1	T6SS forkhead associated domain protein ImpI/VasC	397	397	397
PLF_286_00367574	1	1	T6SS PAAR-repeat protein	174	174	174
PLF_286_00373843	1	1	T6SS PAAR-repeat protein / RhaS protein	1619	1619	1619
PLF_286_00374554	2	1	T6SS PAAR-repeat protein	159	198	178.5
PLF_286_00382154	**1**	**1**	**T6SS outer membrane component TssL (ImpK/VasF)**	**254**	**254**	**254**
PLF_286_00436076	1	1	T6SS component TssA (ImpA)	520	520	520
PLF_286_00443528	1	1	T6SS PAAR-repeat protein / RhaS protein	1588	1588	1588

When the keyword "Metabolite" was used to filter through the protein families, 34 families were identified. In bold represents the families not found in Pseudomonas B217.

Table 22 "T6SS" Protein Families for Pseudomonas B151

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00017681	**2**	**1**	**T6SS PAAR-repeat protein**	**87**	**87**	**87**
PLF_286_00155730	**2**	**1**	**T6SS component Hcp**	**171**	**171**	**171**
PLF_286_00001402	1	1	T6SS component TssM (IcmF/VasK)	1179	1179	1179
PLF_286_00001403	1	1	T6SS outer membrane component TssL (ImpK/VasF)	291	291	291
PLF_286_00001433	1	1	T6SS component TssF (ImpG/VasA)	595	595	595
PLF_286_00001434	1	1	T6SS component TssC (ImpC/VipB)	491	491	491
PLF_286_00001439	1	1	T6SS component TssB (ImpB/VipA)	167	167	167
PLF_286_00001440	1	1	T6SS component TssK (ImpJ/VasE)	443	443	443
PLF_286_00001464	1	1	T6SS component TssG (ImpH/VasB)	335	335	335
PLF_286_00001557	1	1	T6SS lysozyme-like component TssE	135	135	135
PLF_286_00001571	1	1	T6SS secretion lipoprotein TssJ (VasD)	166	166	166
PLF_286_00001834	**1**	**1**	**T6SS component TssG (ImpH/VasB)**	**341**	**341**	**341**
PLF_286_00001835	**1**	**1**	**T6SS component TssK (ImpJ/VasE)**	**443**	**443**	**443**
PLF_286_00001847	**1**	**1**	**T6SS component TssC (ImpC/VipB)**	**494**	**494**	**494**
PLF_286_00001853	**1**	**1**	**T6SS component TssF (ImpG/VasA)**	**597**	**597**	**597**
PLF_286_00001861	**1**	**1**	**T6SS component Hcp**	**165**	**165**	**165**
PLF_286_00001862	**1**	**1**	**T6SS component TssB (ImpB/VipA)**	**180**	**180**	**180**
PLF_286_00012155	1	1	T6SS Serine/threonine protein kinase (EC 2.7.11.1) PpkA	333	333	333
PLF_286_00015869	1	1	T6SS component Hcp	513	513	513
PLF_286_00016026	**1**	**1**	**T6SS lysozyme-like component TssE**	**147**	**147**	**147**
PLF_286_00037841	1	1	T6SS protein serine/threonine phosphatase PppA	242	242	242
PLF_286_00075352	**1**	**1**	**T6SS component TssM (IcmF/VasK)**	**1273**	**1273**	**1273**
PLF_286_00095010	**1**	**1**	**T6SS AAA+ chaperone ClpV (TssH)**	**846**	**846**	**846**
PLF_286_00108270	1	1	T6SS sigma-54-dependent regulator VasH	508	508	508
PLF_286_00279238	1	1	T6SS component Hcp	158	158	158
PLF_286_00279796	1	1	T6SS AAA+ chaperone ClpV (TssH)	885	885	885
PLF_286_00297984	**1**	**1**	**T6SS component TssA (ImpA)**	**362**	**362**	**362**
PLF_286_00364932	1	1	T6SS forkhead associated domain protein ImpI/VasC	397	397	397
PLF_286_00367574	1	1	T6SS PAAR-repeat protein	174	174	174
PLF_286_00373843	1	1	T6SS PAAR-repeat protein / RhaS protein	1556	1556	1556
PLF_286_00374554	1	1	T6SS PAAR-repeat protein	165	165	165
PLF_286_00382154	**1**	**1**	**T6SS outer membrane component TssL (ImpK/VasF)**	**254**	**254**	**254**
PLF_286_00436076	1	1	T6SS component TssA (ImpA)	520	520	520
PLF_286_00443528	1	1	T6SS PAAR-repeat protein / RhaS protein	1547	1547	1547

When the keyword "T6SS" was used to filter through the protein families, 34 families were identified. In bold represents the families not found in Pseudomonas B217 and the first two proteins listed are not found in either of the other two genomes.

Table 23 "Polyketide" Protein Families for Pseudomonas B228 and B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00010678	1	1	Polyketide synthase modules and related proteins	1129	1129	1129
PLF_286_00012144	1	1	Polyketide synthase modules and related proteins	3031	3031	3031
PLF_286_00014642	1	1	Polyketide synthase modules and related proteins	1370	1370	1370
PLF_286_00058876	**1**	**1**	**Polyketide synthase modules and related proteins**	**4108**	**4108**	**4108**
PLF_286_00061738	1	1	Polyketide synthase modules and related proteins	3002	3002	3002
PLF_286_00071315	**1**	**1**	**Polyketide synthase modules and related proteins**	**528**	**528**	**528**
PLF_286_00081078	1	1	Polyketide synthase modules and related proteins	4332	4332	4332
PLF_286_00298348	1	1	Polyketide synthase modules and related proteins	2136	2136	2136

When the keyword "Polyketide" was used to filter through the protein families, 8 families were identified. In bold represents the two families not found in Pseudomonas B151

Table 24 "Polyketide" Protein Families for Pseudomonas B151

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00010678	1	1	Polyketide synthase modules and related proteins	1129	1129	1129
PLF_286_00012144	1	1	Polyketide synthase modules and related proteins	3028	3028	3028
PLF_286_00014642	1	1	Polyketide synthase modules and related proteins	1370	1370	1370
PLF_286_00061738	1	1	Polyketide synthase modules and related proteins	3002	3002	3002
PLF_286_00081078	1	1	Polyketide synthase modules and related proteins	4332	4332	4332
PLF_286_00298348	1	1	Polyketide synthase modules and related proteins	2136	2136	2136

When the keyword "Polyketide" was used to filter through the protein families, 6 families were identified.

Table 25 "Phenazine" Protein Families for Pseudomonas B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00010680	1	1	Phenazine biosynthesis protein PhzF like	260	260	260

When the keyword "Phenazine" was used to filter through the protein families, 1 family was identified.

Table 26 "Phenazine" Protein Families for Pseudomonas B228 and B151

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00007397	**1**	**1**	**Phenazine biosynthesis protein PhzF like**	**293**	**293**	**293**
PLF_286_00010680	1	1	Phenazine biosynthesis protein PhzF like	260	260	260

When the keyword "Phenazine" was used to filter through the protein families, 2 families were identified. In bold represents the one family not found in Pseudomonas B217

Table 27 "PCA" Protein Families for Pseudomonas B228, B151, and B217

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00000612	1	1	Pca regulon regulatory protein PcaR	280	280	280
PLF_286_00006232	1	1	dicarboxylic acid transporter PcaT	431	431	431

When the keyword "PCA" was used to filter through the protein families, 2 families were identified.

Table 28 "Pyoverdine" Protein Families for Pseudomonas B217and B228

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001523	1	1	PvdO, pyoverdine responsive serine/threonine kinase (predicted by OlgaV)	295	295	295
PLF_286_00002468	**1**	**1**	**PvdE, pyoverdine ABC export system, fused ATPase and permease components**	**549**	**549**	**549**
PLF_286_00006551	1	1	Pyoverdine biosynthesis related protein PvdP	541	541	541
PLF_286_00012918	**1**	**1**	**Pyoverdine sidechain non-ribosomal peptide synthetase PvdI @ Siderophore biosynthesis non-ribosomal peptide synthetase modules**	**2611**	**2611**	**2611**
PLF_286_00013013	1	1	Outer membrane ferripyoverdine receptor	724	724	724
PLF_286_00019479	**1**	**1**	**Non-ribosomal peptide synthetase modules, pyoverdine @ Siderophore biosynthesis non-ribosomal peptide synthetase modules**	**1133**	**1133**	**1133**
PLF_286_00173350	1	1	Pyoverdine chromophore precursor synthetase PvdL @ Siderophore biosynthesis non-ribosomal peptide synthetase modules	4333	4333	4333

When the keyword "Pyoverdine" was used to filter through the protein families, 7 families were identified. In bold represents the families not found in Pseudomonas B151

Table 29 "Pyoverdine" Protein Families for Pseudomonas B151

Family ID	Proteins	Genomes	Description	Min AA length	Max AA length	Mean
PLF_286_00001523	1	1	PvdO, pyoverdine responsive serine/threonine kinase (predicted by OlgaV)	302	302	302
PLF_286_00006551	1	1	Pyoverdine biosynthesis related protein PvdP	539	539	539
PLF_286_00012491	**1**	**1**	**Non-ribosomal peptide synthetase modules, pyoverdine @ Siderophore biosynthesis non-ribosomal peptide synthetase modules**	**4136**	**4136**	**4136**
PLF_286_00013013	1	1	Outer membrane ferripyoverdine receptor	724	724	724
PLF_286_00016144	**1**	**1**	**Non-ribosomal peptide synthetase modules, pyoverdine @ Siderophore biosynthesis non-ribosomal peptide synthetase modules**	**4515**	**4515**	**4515**
PLF_286_00034542	**1**	**1**	**Non-ribosomal peptide synthetase modules, pyoverdine @ Siderophore biosynthesis non-ribosomal peptide synthetase modules**	**1411**	**1411**	**1411**
PLF_286_00091144	**1**	**1**	**PvdE, pyoverdine ABC export system, fused ATPase and permease components @ ABC-type siderophore export system, fused ATPase and permease components**	**553**	**553**	**553**
PLF_286_00173350	1	1	Pyoverdine chromophore precursor synthetase PvdL @ Siderophore biosynthesis non-ribosomal peptide synthetase modules	4333	4333	4333

When the keyword "Pyoverdine" was used to filter through the protein families, 8 families were identified. In bold represents the families not found in Pseudomonas B228 and B217

Secondary metabolite gene clusters and antibiotic detection

Genome mining using anti-SMASH resulted in the prediction of potential metabolic gene clusters. The type of metabolite detected, along with similar clusters are detailed in Tables 28-30.

Data from anti-SMASH showed 7 gene clusters in 15 regions, 7 gene clusters in 10 regions, and 8 gene clusters in 11 regions associated with secondary metabolite biosynthesis for isolates B228, B151, and B217 respectively (Tables 30 through 32). Isolates showed predicted BGCs that encode non-ribosomal synthases (NRPS), bacteriocins, N-γ-acetyl glutaminyl glutamine 1-amide (NAGGN), beta lactones, aryl polyenes, and siderophores. Some of these predicted metabolites showed similarity to known BGC clusters. Two of the most found BGC found in all three genomes with the most antibiotic potential were NRPS and bacteriocins. NRPS are known to produce antibiotics and other important pharmaceuticals (124). Bacteriocins are small AMPs produced by bacteria that can inhibit the growth of closely related species via destruction of target cells by pore formation and/or inhibition of cell wall synthesis (125).

ARTS created a BigSCAPE analysis on all found anti-SMASH biosynthetic gene clusters (BGC) from the genomes (Figure 16). From the analysis, the combined genomes have total of 33 BGCs in 13 families. While majority of the BGC are closely related, there are two singletons. Isolates B151 and B217 both have BGC that are not found in the other genomes. Additionally, there are two BGC families found in two of the genomes, but not all three. These shows that even though the three genomes are very closely related, they are able to inhibit different MDR *Salmonella* strains, because they are able to produce different NRPS and bacteriocins.

Table 30 Summary of anti-SMASH results for Pseudomonas B228

Region	Type of metabolite	From	To	Most similar known cluster	Similarity
2.1	Bacteriocin	185,320	196,207	-	-
2.2	NRPS	821,879	885,452	Pyoverdin	19%
4.1	Arylpolyene	306,411	350,024	APE Vf	40%
9.1	Betalactone	104,874	137,247	Fengycin	13%
10.1	NRPS-like	234,494	266,270	Fragin	37%
11.1	NAGGN	25,730	40,503	-	-
11.2	NRPS	169,111	222,112	Pyoverdin	19%
12.1	Bacteriocin	28,642	399,481	-	-
12.2	NRPS	147,307	208,673	Anikasin	100%
15.1	NRPS	1	62,077	Rimosamide	28%

The analysis detected 15 regions of the genome as potential metabolic gene clusters. Of the 15 regions, 7 contain gene clusters related to antibiotics.
NAGGN: N-γ-acetyl glutaminyl glutamine 1-amide, NRPS: non-ribosomal peptide synthases

Table 31 Summary of anti-SMASH 5.1.2 results for Pseudomonas B151

Region	Type of metabolite	From	To	Most similar known cluster	Similarity
1.1	NRPS	18,838	81,038	Rimosamide	28%
1.2	Siderophore	123,847	135,700	-	-
2.1	NRPS-like	1	31,776	Fragin	60%
2.2	Arylpolyene	365,447	409,060	APE Vf	40%
5.1	Bacteriocin	248,920	259,759	-	-
5.2	NRPS	384,255	444,911	Lokisin	71%
6.1	Betalactone	104,123	127,315	Fengycin	13%
7.1	NRPS, terpene	58,575	134,996	Pyoverdin	20%
8.1	NRPS	16,034	69,035	Pyoverdin	19%
8.2	NAGGN	211,124	225,996	-	-
10.1	Bacteriocin	228,182	239,069	-	-

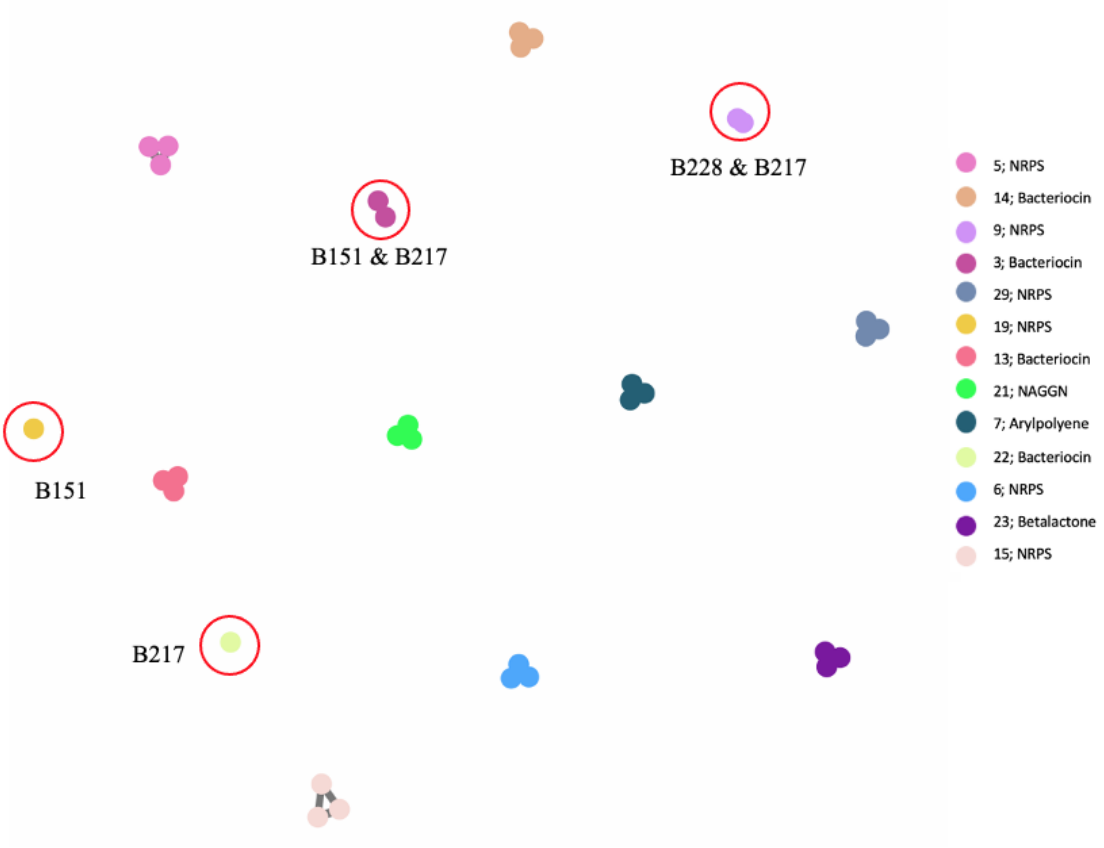
The analysis detected 10 regions of the genome as potential metabolic gene clusters. Of the 10 regions, 7 contain gene clusters related to antibiotics.
NAGGN: N-γ-acetylglutaminyl glutamine 1-amide, NRPS: non-ribosomal peptide synthases

Table 32 Summary of anti-SMASH 5.1.2 results for Pseudomonas B217

Region	Type of metabolite	From	To	Most similar known cluster	Similarity
1.1	NAGGN	719,034	733,807	-	-
1.2	NRPS	872,962	925,963	Pyoverdin	19%
2.1	Siderophore	646,869	658,722	-	-
2.2	NRPS	689,463	751,675	Rimosamide	28%
4.1	NRPS-like	1	31,777	Fragin	60%
4.2	Arylpolyene	327,123	370,751	APE Vf	40%
5.1	Bacteriocin	184,024	194,911	-	-
6.1	Betalactone	193,781	220,316	Fengycin	13%
7.1	NRPS	92,450	156,023	Pyoverdin	19%
9.1	Bacteriociin	160,046	170,960	3-thiaglutamate	70%
11.1	Bacteriociin	27,940	38,779	-	-
11.2	NRPS	153,072	224,887	Lokisin	71%

The analysis detected 11 regions of the genome as potential metabolic gene clusters. Of the 11 regions, 8 contain gene clusters related to antibiotics.
NAGGN: N-γ-acetylglutaminyl glutamine 1-amide, NRPS: non-ribosomal peptide synthases

Figure 16 Biosynthetic Genes Similarity Clustering



Conclusion

We have shown that *Pseudomonas* isolates from soil can inhibit the growth of MDR clinical *Salmonella* strains. We also know that *Pseudomonas* species are able to produce antibiotics (126). Therefore, to learn more about these strains and potentially identify the means of this inhibition ability, we sequenced the three genomes. Through genome mining and comparative genomics, we have shown that three *Pseudomonas fluorescens* species have the capability of inhibiting MDR *Salmonella* strains most likely through various BGCs, secondary metabolites, genes, and proteins that are produced during growth.

The three genomes showed high genomic similarities, but also numerous differences in the numbers of certain genes or pathways. For example, B151 showed the most inhibition capabilities, and its genomes had more proteins associated with efflux pumps and T6SS, as well as more specialty genes than the other two isolates. Future studies including knockout experiments would be beneficial to see if these genes are responsible for increased inhibition properties in comparison to the other genomes.

The program, anti-SMASH detected non-ribosomal peptide synthases (NRPS) and bacteriocins in the three genomes. Both NRPS and bacteriocins have recently been investigated as both can be used as a substitute or conjugate to current therapeutics (127,128). These secondary metabolites identified through anti-SMASH could be verified through a mass spectrometric analysis (129). Additionally, we found proteins in the genomes associated with phenazines and genes like *dxr* and *modA*, that support how these three isolates might have been able to inhibit the MDR clinical *Salmonella* strains.

Further research is needed to see if the components found throughout the genomes of these isolates would be used for novel antibiotics to help battle the rising MDR bacteria problem. For example, these antibacterial compounds produced would need to be tested with human cells to ensure there is no damage or harm caused.

CHAPTER IV

Conclusion

A world without useful antibiotics is a terrifying future that is quickly approaching as numbers of multi-drug resistant bacteria continue to increase. There is an urge to find and create new antibiotics to avoid this alarming future. As many of the antibiotics in use today have come from *Streptomyces* species, researchers have resorted back to exploration of novel antibiotics from bacteria in the soil. Through various screening, extractions and genome mining, isolation of useful antimicrobial compounds has become more accessible.

In these studies, we began our search for antibiotic producing bacteria by screening soil isolates against safe ESKAPE relatives. When bacteria are stressed, they often go into a defense mode to allow them to survive, often times producing antimicrobial compounds (2). After narrowing the isolates down to only those capable of inhibiting safe ESKAPE relatives, the isolates were screened against clinical MDR *Salmonella* strains. When screened together, the soil isolates showed inhibition abilities, supporting our first hypothesis that soil bacteria can produce antibacterial compounds that can inhibit growth of MDR *Salmonella* strains.

From safe ESKAPE relatives to MDR *Salmonella* strains, our total isolates went from over 300 to about 30. Further highlighting, how difficult it is to treat MDR infections in the hospitals. Although these ESAKPE relatives are the leading cause of nosocomial infections across the world, they are not pathogens isolated clinically. Thus, the MDR *Salmonella* strains used in this study became increasingly more difficult to

inhibit when the soil isolates were screened against them because they have more resistance genes and virulence characteristics.

Once 16s data was analyzed, we were able to conclude that the three isolates belong to *Pseudomonas* spp. In 2016, *Pseudomonas* spp. were reported to be the ideal target for SWI antibiotic discovery as they are cable of producing a very diverse range of natural metabolites (7). Thus, our study again illustrates that *Pseudomonas* spp. should be further investigated as there is still potential for novel antibiotics being discovering among various *Pseudomonas* species.

Roughly 30 isolates were found to be antibiotic producers against MDR *Salmonella* strains and were subjected to further testing. Successful extractions via ethyl acetate supported our second hypothesis that extracts from bacteria can possess antibacterial compounds. Interestingly, isolates that originally showed inhibition of all 5 MDR *Salmonella* strains, lost the ability to inhibit one or more of the MDR *Salmonella* strains after extraction. Thus, additional studies with extracts including mass spectrometry techniques could better identify and isolate specific compounds responsible for inhibition as well as therapeutic use.

Lastly, we aimed to compare antibacterial producing bacteria's genomes to one another. The first step after sequencing was to identify the species of the three isolates. Through various bioinformatic programs, we have predicted that these three isolates belong to IG *Pseudomonas fluorescens.* As previously noted, *Pseudomonas* species are often hard to taxonomically classify due to it being such a diverse genus (74, 75). With the three isolates having different inhibition abilities as well as different components apart of their genomes, we predict that three genomes belong to different SG

of the IG *Pseudomonas* fluoresces or are different strains under the same SG. This is hard to conclude as just in 2020, there have been two novel isolates identified belonging to the IG *Pseudomonas* fluoresces group (77, 80).

The goal of this study was not to taxonomically identify soil isolates; however, we have shown sufficient results to conclude that the three isolates belong to *Pseudomonas fluoresces*. However, the exact subgroup and strain of each isolates is more challenging. Additional phenotypic studies would need to be conducted to further classify these isolates into their appropriate SGs. Our initial goal of this study was to identify potential antibacterial compounds, and through genome mining we have found multiple genomic properties relating to antibacterial production. Future studies isolating different members of the IG *Pseudomonas* fluoresces group could help identify potential compounds for therapeutic use, as here we have shown just three isolates with such diversity.

We have shown through comparative genomics that these genomes have similar genes, proteins, and pathways. In comparing their general genomic features, overall their totally numerical representations are comparable. However, there are more noticeable differences amongst the genomes that may account for the ability for some isolates to inhibit more MDR *Salmonella* strains than the other. As multiple different genes or proteins may have been involved in the production of antimicrobial compounds, the absence or addition of one or the other may give the isolates more of an advantage over than the others. For example, B151 was able to inhibit all 5 of the MDR *Salmonella* strains, while the other isolates were not. Isolate B151's genomes also had a few more copies of antibacterial relevant genes than the other two. This feature, might have

allowed for more of an ability to produce a different range of antibiotics and/or protection of the secreted compounds.

When looking specifically at the different features represented in the genomes, genes associated with metabolism stood out the most in comparison with the other genomes, as B151 had nearly 30 genes more than both B217 and B228. While B151 was able to inhibit all 5 MDR *Salmonella* strains, B228 and B217 were not far behind as both isolates could inhibit 3 of the strains. Interestingly enough, both isolates genomes contain more stress response, defense, and virulence genes than isolate B151. Therefore, transcriptomics data would be interesting to see which genes specifically are being upregulated and down regulated as that could help explain these inhibition properties.

As previously mentioned, B151 was able to inhibit all 5 MDR *Salmonella* strains, it also contained unique pathways that the other genomes did not. The isoflavoniod pathway found in B151 could potentially explain its successful inhibition capability. The byproducts produced in this pathway could have antibacterial activity (83). Future studies involving metabolic engineering to genetically interfere with this pathway should be done to see if after alteration, more antibacterial activity is produced (130).

There were specific proteins and genes found throughout the genomes that could potentially be responsible for antibacterial production. Proteins families associated with T6SS, ABC transporters, AMPs, phenazines, pyoverdines, and efflux pumps are all of interest as they are responsible for the production of antibiotics, as well as protection of the bacterium itself (101, 103, 105, 108, 109, 110, 131). The findings of these specific

proteins and genes further support that these isolates are capable of producing antibiotics. Comparison of these features between the three genomes shows that although these genomes are all from the same IG, there are differences among the genes that allow for differing inhibition abilities. Additional studies involving gene knockouts could support this theory. Nonetheless, these findings supported our last hypothesis that the *Pseudomonas* strains have homologous genomic features for antibiotic production.

Lastly, genome mining using anti-SMASH resulted in the prediction of potential metabolic gene clusters. These biosynthetic gene clusters are associated with secondary metabolite biosynthesis. Secondary metabolites are often mined for as they are important for the bacterium's survival ad can be made into antibiotics (98). We have shown that each other genomes have unique secondary metabolites throughout their genomes. Additional studies including isolation of these metabolites found should be done for potential therapeutic use.

This study was just the beginning to discovering novel antibiotics. More experiments will need to be conducted in order to make any final conclusions. We have already begun phase two of this study by sequencing the 5 MDR *Salmonella* strains. We are planning on comparing the data collected from the two data sets to get a better understanding on antimicrobial production and regulation. Through the data collected from the WGS of the 5 MDR *Salmonella* strains, we can exclusively look at potential drug targets in the genomes to narrow down what antibiotic compounds would work. Specifically, through the genome mining and extraction work, we could find an antibiotic that could be used clinically. Furthermore, additional studies including isolation of

antimicrobial compounds through various chemical mechanisms and different gene knockouts are needed to wrap up this study.

As a whole, we are on the road to novel antibiotic discovery. Through advances in technology in microbial genomics and metagenomics, it is becoming more possible to overcome this battle with MDR pathogens. Through SWI and researchers joining forces, there is a chance to one day slow the rate of MDR pathogens.

REFERENCES

1. Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I et al. (2015). A new antibiotic kills pathogen without detectable resistance. *Nature*, Research; 517(7535):455-459.
2. Sharma M, Manhas RK. (2019). Purification and characterization of actinomycins from Streptomyces strain M7 active against methicillin resistant Staphylococcus aureus and vancomycin resistant Enterococcus. *BMC Microbiology*;19(1).
3. Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. (2010). *Microbiology and Molecular Biology Reviews.* 74(3):417–433.
4. Lopes LD, Weisberg AJ, Davis EW, Varize CDS, Michele De C. Pereira E Silva, Chang JH, Loper JE, Andreote FD. (2019). Genomic and metabolic differences between Pseudomonas putida populations inhabiting sugarcane rhizosphere or bulk soil. *Plos One,* 14(10).
5. Zhang L, Chen W, Jiang Q, Fei Z, Xiao M. (2020). Genome analysis of plant growth-promoting rhizobacterium Pseudomonas chlororaphis subsp. aurantiaca JD37 and insights from comparison of genomics with three Pseudomonas strains. *Microbiol Res*., 237:126483.
6. Hibbing ME, Fugua C, Parsek MR, Peterson SB. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microboil,* 8(1): 15–25.
7. Davis E, Sloan T, Aurelius K, et al. (2017). Antibiotic discovery throughout the Small World Initiative: A molecular strategy to identify biosynthetic gene clusters involved in antagonistic activity. *Microbiologyopen*, 6(3).

8. Stierle AA, Stierle DB. (2015). Bioactive Secondary Metabolites Produced by the Fungal Endophytes of Conifers. *Nat Prod Commun*., 10(10):1671-1682.
9. Van Lanen, Steven & Shen, Ben. (2006). Microbial genomics for the improvement of natural product discovery. *Current opinion in microbiology.* 9. 252-60.
10. Depoorter E, Bull MJ, Peeters C, Coenye T, Vandamme P, Mahenthiralingam E. (2016). Burkholderia: an update on taxonomy and biotechnological potential as antibiotic producers. *Appl Microbiol Biotechnol*., 100(12):5215-5229.
11. Martínez JL. (2008). Antibiotics and Antibiotic Resistance Genes in Natural Environments. *Science.* 321(5887): 365-367.
12. Nielsen K.M, Van Elsas JD, Smalla K. (2000). Transformation of *Acinetobacter* sp. Strain BD413(pFG4Δ*nptII*) with Transgenic Plant DNA in Soil Microcosms and Effects of Kanamycin on Selection of Transformants. *Applied and Environmental Microbiology*., 66 (3) 1237-1242
13. World Health Organization. (2018). High levels of antibiotic resistance found worldwide, new data shows. Bangkok. https://www.who.int/mediacentre/news/releases/2018/antibiotic-resistance-found/en/
14. World Health Organization. Office of Library and Health Literature Services. (2019).
15. Centers for Disease Control and Prevention. (2019). CDC Wonder. http://wonder.cdc.gov/.
16. Small World Initiative, Inc. (2019). http://www.smallworldinitiative.org.

17. Lynne AM, Kaldhone P, David D, White DG, Foley SL. (2009). Characterization of Antimicrobial Resistance in Salmonella enterica Serotype Heidelberg Isolated from Food Animals. *Foodborne Pathog Dis.* 6(2):207-15.

18. Foley SL, Lynne AM. (2008). Food animal-associated Salmonella challenges: pathogenicity and antimicrobial resistance. *Journal of Animal Science.*86(14 Suppl):E173-E187.

19. Lynne AM, Rhodes-Clark BS, Bliven K, Zhao S, Foley SL. (2008). Antimicrobial Resistance Genes Associated with Salmonella enterica Serovar Newport Isolates from Food Animals. *Antimicrob Agents Chemother series (1)*, 353-356.

20. Lynne AM, Dorsey LL, David DE, Foley SL. (2009). Characterisation of antibiotic resistance in host-adapted Salmonella enterica. *Int J Antimicrob Agents*;34(2):169-172.

21. Han J, David DE, Deck J, Lynne AM, Kaldhone P et al. (2011). Comparison of Salmonella enterica Serovar Heidelberg Isolates from Human Patients with Those from Animal and Food Sources. *J Clin Microbiol series*. (3), 1130-1133.

22. Butler MS, Blaskovich MA, Cooper MA. (2013). Antibiotics in the clinical pipeline in 2013. *The Journal of Antibiotics*. 66, pages571–591.

23. Gionco B, Tavares ER, de Oliveira AG, Yamada-Ogatta SF, Carmo AO et al. (2017). New Insights about Antibiotic Production by Pseudomonas aeruginosa: A Gene Expression Analysis. *Front Chem*. 5, 66.

24. Cain, C. C., Lee, D., Waldo, R. H., 3rd, Henry, A. T., Casida, E. J., Jr, Wani, M. C., Wall, M. E., Oberlies, N. H., & Falkinham, J. O., 3rd (2003). Synergistic

antimicrobial activity of metabolites produced by a nonobligate bacterial predator. *Antimicrobial agents and chemotherapy*, 47(7), 2113–2117.

25. Qing Yan, Lucas D. Lopes, Brenda T. Shaffer, Teresa A. Kidarsa, Oliver Vining, Benjamin Philmus, Chunxu Song, Virginia O. Stockwell, Jos M. Raaijmakers, Kerry L. McPhail, Fernando D. Andreote, Jeff H. Chang, Joyce E. (2018). Secondary Metabolism and Interspecific Competition Affect Accumulation of Spontaneous Mutants in the GacS-GacA Regulatory System in Pseudomonas protegens *Loper mBio*, 9 (1) e01845-17.
26. Matthijs S, Vander Wauven C, Cornu B, Ye L, Cornelis P, Thomas CM, Ongena M. (2014). Antimicrobial properties of Pseudomonas strains producing the antibiotic mupirocin. *Res Microbiol,* 165(8):695-704.
27. Park SB, Lee IA, Suh JW, Kim JG, Lee CH. (2011). Screening and identification of antimicrobial compounds from Streptomyces bottropensis suppressing rice bacterial blight. *J Microbiol Biotechnol* , 21(12):1236-1242.
28. Nandhini, S. & Sudha, S. & V, Anusha & S, Manisha. (2018). Isolation, Identification and Extraction of antimicrobial compounds produced by Streptomyces sps from terrestrial soil. *Biocatalysis and Agricultural Biotechnology*. 15, 317-321.
29. Teng Hern, Tan & Lee, Learn Han & Goh, Bey-Hing. (2020). Critical review of fermentation and extraction of anti-Vibrio compounds from Streptomyces. *Progress In Microbes & Molecular Biology*. 3(1),

30. Cho JY and Kim MS,. (2012). Antibacterial benzaldehydes produced by seaweed-derived Streptomyces atrovirens PK288-21. *Fisheries Sci*, 78(5): 1065–1073.
31. Fields FR, Lee SW, McConnell MJ. (2017). Using bacterial genomes and essential genes for the development of new antibiotics. *Biochem Pharmacol*,134:74-86.
32. Steiner, K.K., Parthasarathy, A., Wong, N.H. et al. (2020). Isolation and whole-genome sequencing of Pseudomonas sp. RIT 623, a slow-growing bacterium endowed with antibiotic properties. *BMC Res Notes* 13, 370
33. Drehe I, Simonetti E, Ruiz JA. (2018). Contribution of the Siderophores Pyoverdine and Enantio-Pyochelin to Fitness in Soil of Pseudomonas protegens Pf-5. *Curr Microbiol.*, 75(12):1560-1565.
34. Scott Quainoo, Jordy P. M. Coolen, Sacha A. F. T. van Hijum, Martijn A. Huynen, Willem J. G. Melchers, Willem van Schaik, Heiman F. L. Wertheim. (2017). Whole-Genome Sequencing of Bacterial Pathogens: The Future of Nosocomial Outbreak Analysis. *Clinical Microbiology Reviews*, 30 (4) 1015-1063.
35. Rajan, B. M., & Kannabiran, K. (2014). Extraction and Identification of Antibacterial Secondary Metabolites from Marine Streptomyces sp. VITBRK2. *International journal of molecular and cellular medicine*, 3(3), 130–137.
36. Hernandez, S., Tsang, T., Bascom-Slack, C., and Handelsman, J. (2015). The Small World Initiative Research Protocols. Forth Edition.

37. Devadass BJ, Paulraj MG, Ignacimuthu S, et al. (2015). Antimicrobial activity of soil actinomycetes isolated from Western Ghats in Tamil Nadu, India. *J Bacteriol Mycol Open Access*, 3(2):224-232.

38. Valot, B., Guyeux, C., Rolland, J. Y., Mazouzi, K., Bertrand, X., & Hocquet, D. (2015). What It Takes to Be a Pseudomonas aeruginosa? The Core Genome of the Opportunistic Pathogen Updated. *PloS one*, 10(5).

39. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. (2019). AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*., 47:81–7

40. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 45(D1):D535-D542.

41. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19:455-477.

42. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365.

43. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.,* 32:D431-D433.

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics,* 25:25.

45. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.,* 44:D457-462.

46. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H. (2016). PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front Microbiol*., 7:118.

47. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*., 33:5691-5702.

48. Saier Jr MH, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. (2015). The transporter classification database (TCDB): recent advances. *Nucleic Acids Res*., 44:D372-D379.

49. Mao C, Abraham D, Wattam AR, Wilson MJ, Shukla M, Yoo HS, Sobral BW. (2015). Curation, integration and visualization of bacterial virulence factors in PATRIC. *Bioinformatics,* 31:252-258.

50. Chen L, Zheng D, Liu B, Yang J, Jin Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res*., 44:D694-D697.

51. Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, Huang L, Guo Y, Han L, Zheng C. (2009). Update of TTD: therapeutic target database. *Nucleic Acids Res.,* 38:D787-D791.

52. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*., 42:D1091-1097.

53. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L. (2013). The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy*, 57:3348-3357.

54. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res*, 45:D535-D542.

55. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology*, 17:132.

56. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H. (2016). PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front Microbiol*, 7:118.

57. Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.,* 32:1792-1797.

58. Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30:1312-1313.

59. Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic biology,* 57:758-771.

60. Ondov BD, Treangen TJ, Melsted P, et al. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol,* 17: 1–14.

61. Lagesen K, Hallin P. (2017). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res. Oxford Univ Press,* 35: 3100–3108.

62. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,*10: 421.

63. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics.14: 60.

64. Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol., 32: 2798–2800.

65. Farris JS. (1972) Estimating phylogenetic trees from distance matrices. *Am Nat.* 106: 645–667.

66. Kreft L, Botzki A, Coppens F, Vandepoele K, Van Bel M. (2017). PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics,* 33: 2946–2947.

67. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. (2014). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci*. 9: 2.

68. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. (2012). Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol*., 50(4):1355-61.

69. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O. J (2012). Benchmarking of methods for genomic taxonomy. Species Finder. *Clin Microbiol*., 52(5):1529-39.

70. Saputra, D., Rasmussen, S., Larsen, M.V. et al. (2015). Reads2Type: a web application for rapid microbial taxonomy identification. *BMC Bioinformatics*, 16, 398

71. Mungan,M.D., Alanjary,M., Blin,K., Weber,T., Medema,M.H. and Ziemert,N. (2020) ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res.*,10,1093

72. Alanjary,M., Kronmiller,B., Adamek,M., Blin,K., Weber,T., Huson,D., Philmus,B. and Ziemert,N. (2017) The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res*.,10.

73. Lalucat J, Mulet M, Gomila M, García-Valdés E. (2020). Genomics in Bacterial Taxonomy: Impact on the Genus Pseudomonas. *Genes (Basel)*,29;11(2):139.

74. Peix A, Ramírez-Bahena MH, Velázquez E. (2018). The current status on the taxonomy of Pseudomonas revisited: An update. *Infect Genet Evol*., 57:106-116.

75. Lalucat J, Mulet M, Gomila M, García-Valdés E. (2019). Genomics in Bacterial Taxonomy: Impact on the Genus Pseudomonas. *Genes (Basel).* 29;11(1):139.

76. Garrido-Sanz, D., Meier-Kolthoff, J. P., Göker, M., Martín, M., Rivilla, R., & Redondo-Nieto, M. (2016). Genomic and Genetic Diversity within the Pseudomonas fluorescens Complex. *PloS one*, 11(2).

77. Jia, Jiayuan & Wang, Xiaoqiang & Deng, Peng & Ma, Lin & Baird, Sonya & Li, Xiangdong & Lu, Shien. (2020). Pseudomonas glycinae sp. nov. isolated from the soybean rhizosphere. *Microbiology Open*, 9.

78. Mulet M, Gomila M, Scotta C, Sánchez D, Lalucat J, García-Valdés E (2012) Concordance between whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry and multilocus sequence analysis approaches in species discrimination within the genus Pseudomonas. *Syst Appl Microbiol*, 35:455–464

79. Chang, D., Rhee, M., Kim, J. et al. (2016). Pseudomonas kribbensis sp. nov., isolated from garden soils in Daejeon, Korea. *Antonie van Leeuwenhoek*, 109, 1433–1446

80. Morimoto Y, Uwabe K, Tohya M, Hiramatsu K, Kirikae T, Baba T. (2020). Pseudomonas atagosis sp. nov., and Pseudomonas akappagea sp. nov., New Soil

Bacteria Isolated from Samples on the Volcanic Island Izu Oshima, Tokyo. *Curr Microbiol*., 77(8):1909-1915.

81. Bobay, L. M., & Ochman, H. (2017). The Evolution of Bacterial Genome Architecture. *Frontiers in genetics*, 8, 72.
82. Rokas A, Payne G, Fedorova ND, Baker SE, Machida M, Yu J, Georgianna DR, Dean RA, Bhatnagar D, Cleveland TE, Wortman JR, Maiti R, Joardar V, Amedeo P, Denning DW, Nierman WC. (2007). What can comparative genomics tell us about species concepts in the genus Aspergillus? *Stud Mycol.*, 59:11-7.
83. Mukne, A. P., Viswanathan, V., & Phadatare, A. G. (2011). Structure pre-requisites for isoflavones as effective antibacterial agents. *Pharmacognosy reviews*, 5(9), 13–18.
84. Tercero JA, Espinosa JC, Lacalle RA, Jiménez A. (1996). The biosynthetic pathway of the aminonucleoside antibiotic puromycin, as deduced from the molecular analysis of the pur cluster of Streptomyces alboniger. *J Biol Chem.*, 271(3):1579-90.
85. Pickens, L. B., & Tang, Y. (2009). Decoding and engineering tetracycline biosynthesis. *Metabolic engineering*, 11(2), 69–75.
86. Steffensky, M., Mühlenweg, A., Wang, Z. X., Li, S. M., & Heide, L. (2000). Identification of the novobiocin biosynthetic gene cluster of Streptomyces spheroides NCIB 11891. *Antimicrobial agents and chemotherapy*, 44(5), 1214–1222.

87. Jia, N., Ding, M. Z., Luo, H., Gao, F., & Yuan, Y. J. (2017). Complete genome sequencing and antibiotics biosynthesis pathways analysis of Streptomyces lydicus 103. *Scientific reports*, 7, 44786.

88. Solecka, J., Rajnisz, A. & Laudy, A. (2009). A novel isoquinoline alkaloid, DD-carboxypeptidase inhibitor, with antibacterial activity isolated from Streptomyces sp. 8812. Part I: Taxonomy, fermentation, isolation and biological activities. *J Antibiot*, 62, 575–580.

89. UniProt Consortium European Bioinformatics Institute Protein Information ResourceSIB Swiss Institute of Bioinformatics. (n.d.). UniProt Consortium. https://www.uniprot.org/keywords/KW-0759

90. Tanaka JC, Silva CC, Oliveira AJ, Nakamura CV, Dias Filho BP. (2006). Antibacterial activity of indole alkaloids from Aspidosperma ramiflorum. *Braz J Med Biol Res*., 39(3):387-91.

91. Al-Janabi A. A. (2011). Potential activity of the purine compounds caffeine and aminophylline on bacteria. Journal of global infectious diseases, 3(2), 133–137.

92. Martín JF, Casqueiro J, Kosalková K, Marcos AT, Gutiérrez S. (1999). Penicillin and cephalosporin biosynthesis: mechanism of carbon catabolite regulation of penicillin production. *Antonie Van Leeuwenhoek*, 75(1-2):21-31.

93. KEGG PATHWAY: beta-Lactam resistance - Pseudomonas aeruginosa PAO1. Genome.jp. https://www.genome.jp/kegg-bin/show_pathway?pae01501

94. Vogt T. (2010). Phenylpropanoid biosynthesis. *Mol Plant*. 2010 Jan;3(1):2-20.

95. Basri, D. F., Luoi, C. K., Azmi, A. M., & Latip, J. (2012). Evaluation of the Combined Effects of Stilbenoid from Shorea gibbosa and Vancomycin against

Methicillin-Resistant Staphylococcus aureus (MRSA*). Pharmaceuticals (Basel, Switzerland),* 5(9), 1032–1043.

96. Choo W.S. (2018) Betalains: Application in Functional Foods. In: Mérillon JM., Ramawat K. (eds) Bioactive Molecules in Food. Reference Series in Phytochemistry. Springer, Cham.
97. Sharma V, Chitranshi N, Agarwal AK. (2014). Significance and biological importance of pyrimidine in the microbial world. *Int J Med Chem.* 2014.
98. Sánchez S, Demain AL. (2011). The biochemical basis/secondary metabolites. In: Moo-Young Murray, editor. Comprehensive Biotechnoogy. Ontario: Elsevier; 1(2), 155–167
99. Sharma, V., Chitranshi, N., Agarwal, A.K., (2014). "Significance and Biological Importance of Pyrimidine in the Microbial World", *International Journal of Medicinal Chemistry*, 2014, 31.
100. Sivapragasam, S., & Grove, A. (2019). The Link between Purine Metabolism and Production of Antibiotics in Streptomyces. *Antibiotics (Basel, Switzerland),* 8(2), 76.
101. Delmar, J. A., Su, C. C., & Yu, E. W. (2014). Bacterial multidrug efflux transporters. *Annual review of biophysics*, 43, 93–117.
102. Tsuchiya H, Doki S, Takemoto M, Ikuta T, Higuchi T, Fukui K, Usuda Y, Tabuchi E, Nagatoishi S, Tsumoto K, Nishizawa T, Ito K, Dohmae N, Ishitani R, Nureki O. (2016). Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature*, 534(7607):417-20.

103. Sinha, R., & Shukla, P. (2019). Antimicrobial Peptides: Recent Insights on Biotechnological Interventions and Future Perspectives. *Protein and peptide letters*, 26(2), 79–87.

104. Qiaomei Cheng, James T. (2002). Substrate Specificity of the AmpG Permease Required for Recycling of Cell Wall Anhydro-Muropeptides *Park Journal of Bacteriology,* 184 (23) 6434-6436.

105. Méndez C, Salas JA. (2001). The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res Microbiol*., 152(3-4):341-50.

106. Putman, M., van Veen, H. W., & Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiology and molecular biology reviews*: *MMBR*, 64(4), 672–693.

107. de Castro AP, Fernandes Gda R, Franco OL. (2014). Insights into novel antimicrobial compounds and antibiotic resistance genes from soil metagenomes. *Front Microbiol*, 18;5:489.

108. Cascales, E., & Cambillau, C. (2012). Structural biology of type VI secretion systems. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences*, 367(1592), 1102–1111.

109. Wulf Blankenfeldt, Alexandre P. Kuzin, Tatiana Skarina, Yuriy Korniyenko, Liang Tong, Peter Bayer, Petra Janning, Linda S. Thomashow, Dmitri V. Mavrodi. (2004). Structure and function of the phenazine biosynthetic protein PhzF from Pseudomonas fluorescens. *Proceedings of the National Academy of Science*, 101 (47) 16431-16436.

110. Calcott, M. J., Owen, J. G., Lamont, I. L., & Ackerley, D. F. (2014). Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of Pseudomonas aeruginosa. Applied and environmental microbiology, 80(18), 5723–5731.

111. Li, H., Luo, Y. F., Williams, B. J., Blackwell, T. S., & Xie, C. M. (2012). Structure and function of OprD protein in Pseudomonas aeruginosa: from antibiotic resistance to novel therapies. *International journal of medical microbiology: IJMM*, 302(2), 63–68.

112. Mahizan, N. A., Yang, S. K., Moo, C. L., Song, A. A., Chong, C. M., Chong, C. W., Abushelaibi, A., Lim, S. E., & Lai, K. S. (2019). Terpene Derivatives as a Potential Agent against Antimicrobial Resistance (AMR) Pathogens. *Molecules (Basel, Switzerland)*, 24(14), 2631.

113. BioCyc pathway/genome database collection. Biocyc.org. https://biocyc.org

114. Méndez C, Salas JA. (2001). The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res Microbiol.*, 152(3-4):341-50.

115. UniProt. Uniprot.org. https://www.uniprot.org

116. Ferenci, T., & Phan, K. (2015). How Porin Heterogeneity and Trade-Offs Affect the Antibiotic Susceptibility of Gram-Negative Bacteria. Genes, 6(4), 1113–1124.

117. Lee, K. Y., & Mooney, D. J. (2012). Alginate: properties and biomedical applications. *Progress in polymer science*, 37(1), 106–126.

118. Andrew G. McArthur GDW. The comprehensive antibiotic resistance database. Mcmaster.ca. https://card.mcmaster.ca/ontology/36672

119. Sharma, A., Gupta, V. K., & Pathania, R. (2019). Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *The Indian journal of medical research,* 149(2), 129–145.

120. Carriel, D., Simon Garcia, P., Castelli, F., Lamourette, P., Fenaille, F., Brochier-Armanet, C., Elsen, S., & Gutsche, I. (2018). A Novel Subfamily of Bacterial AAT-Fold Basic Amino Acid Decarboxylases and Functional Characterization of Its First Representative: Pseudomonas aeruginosa LdcA. *Genome biology and* evolution, 10(11), 3058–3075.

121. Lizewski SE, Schurr JR, Jackson DW, Frisk A, Carterson AJ, Schurr MJ. (2004). Identification of AlgR-regulated genes in Pseudomonas aeruginosa by use of microarray analysis. *J Bacteriol*., 186(17):5672-84.

122. Cunliffe, H. E., Merriman, T. R., & Lamont, I. L. (1995). Cloning and characterization of pvdS, a gene required for pyoverdine synthesis in Pseudomonas aeruginosa: PvdS is probably an alternative sigma factor. *Journal of bacteriology*, 177(10), 2744–2750.

123. de Kievit, T. R., & Lam, J. S. (1997). Isolation and characterization of two genes, waaC (rfaC) and waaF (rfaF), involved in Pseudomonas aeruginosa serotype O5 inner-core biosynthesis. *Journal of bacteriology,* 179(11), 3451–3457.

124. Marahiel MA. (2009). Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J Pept Sci*., 15(12):799-807.

125. Hassan M, Kjos M, Nes IF, Diep DB, Lotfipour F. (2012). Natural antimicrobial peptides from bacteria: characteristics and potential applications to fight against antibiotic resistance. *J Appl Microbiol*.,113(4):723-36.

126. Matthijs S, Vander Wauven C, Cornu B, Ye L, Cornelis P, Thomas CM, Ongena M. (2014). Antimicrobial properties of Pseudomonas strains producing the antibiotic mupirocin. *Res Microbiol*.,165(8):695-704.

127. Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., & Thomas, M. G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Molecular pharmaceutics*, 5(2), 191–211.

128. Meade E, Slattery MA, Garvey M. (2020). Bacteriocins, Potent Antimicrobial Peptides and the Fight against Multi Drug Resistant Species: Resistance Is Futile? *Antibiotics (Basel)*, 9(1):32.

129. Hu, Dini & Chen, Yan & Sun, Chenghang & Jin, Tao & Fan, Guangyi & Liao, Qiwen & Mok, Kai Meng & Lee, Ming-Yuen. (2018). Genome guided investigation of antibiotics producing actinomycetales strain isolated from a Macau mangrove ecosystem. *Scientific Reports.*

130. Pickens LB, Tang Y, Chooi YH. (2011). Metabolic engineering for the production of natural products. *Annu Rev Chem Biomol Eng*., 2:211-36.

131. Xu G, Kong L, Gong R, Xu L, Gao Y, Jiang M, Cai YS, Hong K, Hu Y, Liu P, Deng Z, Price NPJ, Chen W. (2018). Coordinated Biosynthesis of the Purine Nucleoside Antibiotics Aristeromycin and Coformycin in Actinomycetes. *Appl Environ Microbiol*., 84(22)

VITA

Caitlyn Mary Gaffney

CURRENT POSITION

Graduate Research Assistant
Department of Biological Sciences,
Sam Houston State University, Huntsville, TX 77340,
August 2018-present

EDUCATION:

2018-2020 Master of Science
Department of Biological Sciences
Sam Houston State University
Huntsville, Texas
Advisors: Aaron Lynne, Ph.D. and Jeremy Bechelli, Ph.D.

2014-2018 Bachelor of Science
Department of Chemistry
Sam Houston State University
Huntsville, Texas

RESEARCH EXPERIENCE

Thesis: Identification of Bacteria and their Antimicrobial Properties Against 5 Multi-Drug Resistant *Salmonella* Strains from Soil Samples with Comparative Genomics

Confocal laser scanning microscopy, transmission electron microscopy, cell culture, PCR, DNA extraction, BSL-2 pathogens

PROFESSIONAL EXPERIENCE

Graduate Assistant. Department of Biological Sciences, Sam Houston State University, Huntsville, TX 77380, 2018-present.

Undergraduate Teaching Assistant, Department of Chemistry, Sam Houston State University, Huntsville, TX 77380, 2016.

COURSES TAUGHT

General Microbiology Lab (Head TA)
Introductory Applied Microbiology Lab (Head TA)
General Zoology Lab
Organic Chemistry II Lab

ACADEMIC/PROFESSIONAL AWARDS

Graduate Assistantship Scholarship: $2000.00
Graduate School Travel Award: $500.00
Graduate Studies Scholarship: $1,000.00
COAS Scholarship: $1,000.00

RESEARCH PRESENTATIONS

Gaffney C, Onyoni F, Bechelli J, Lynne A. Identification of Bacteria and their antimicrobial properties against 5 multi-drug Resistant Salmonella strains from soil samples in the Piney Woods. Texas Academy of Sciences. Nacogdoches, TX. February 2020

Onyoni F, **Gaffney C**, Bechelli J, Lynne A. Drug Resistant Environmental Enterobacteriaceae. Texas Academy of Sciences. Nacogdoches, TX. February 2020

Gaffney C, Onyoni F, Bechelli J, Lynne A. Identification of Bacteria and their antimicrobial properties against 5 multi-drug Resistant Salmonella strains from soil samples in the Piney Woods. American Society for Microbiology, Texas Branch Meeting. San Antonio, TX. November 2019.

Johnson M, Chacon J, **Gaffney C**, Bechelli J, Trad T. Water dispersible cobalt ferrite nanoparticles by silane ligand exchange of carboxylate-functionalized hydrophobic nanoprecursors, and their anti-bacterial activity. American Chemical Society, Southwest and Rocky Mountain Regional Meeting. El. Paso, TX. November 2019.

MEMBERSHIP IN SOCIETIES

Texas Academy of Sciences, 2018-present

American Society for Microbiology (ASM), 2018- present.

American Society for Microbiology (ASM) Texas Branch, 2018-present

Biological Science Graduate Student Organization (BSGSO), 2018-present

PUBLICATIONS IN PRESS

Johnson M, Chacon J, **Gaffney C**, Bechelli J, Trad T. Non-hydrolytic synthesis of caprylate capped cobalt ferrite nanoparticles and their application against *Erwinia*

carotovora and *Stenotrophomonas maltophilia.* **Accepted** in RSC's Journal of Materials Chemistry B.

PUBLICATIONS IN PROGRESS

Gaffney C, Bechelli J, Lynne A. Identification and Antimicrobial Properties of *Pseudomonas* from Soil. Letters of Applied Microbiology.

Gaffney C, Bechelli J, Lynne A. Identification and Antimicrobial Properties of *Flavobacterium* from Soil. Letters of Applied Microbiology.

Gaffney C, Bechelli J, Lynne A. Draft Genome Sequences of 5 *Pseudomonas* species isolated from Piney Woods of Texas. Microbiology Resource Announcements.

Owen S**,** Grado L, **Gaffney C**, Woodson M, Balaraman R, Bechelli J. Colorado Tick Fever Virus Induces Apoptosis and Caspase-3/7 in Human Endothelial Cells. Ticks and Tick-borne Diseases.

Gaffney C, Owen S, Balaraman R, Bechelli J. Electron Microscopic Observations of Colorado Tick Fever Virus in THP-1 cells. Journal of General Virology