2023

Antibiotics and Secondary Metabolites Analysis SHell (antiSMASH) as a tool to detect putative novel antibiotics

Rajiv Snape
rs5764@rit.edu

Antibiotics and **S**econdary **M**etabolites **A**nalysis **SH**ell (antiSMASH) as a tool to detect putative novel antibiotics

By: Rajiv Snape

A Thesis submitted in Partial Fulfillment of the Requirements for the Master of Science Degree in Bioinformatics

Thomas H. Gosnell School of Life Sciences in The College of Science

Rochester Institute of Technology

2023

Rochester Institute of Technology
Thomas H. Gosnell School of Life Sciences
Bioinformatics Program

To: Head, Thomas H. Gosnell School of Life Sciences

The undersigned state that Rajiv Snape, a candidate for the Master of Science degree in Bioinformatics, has submitted his thesis and has satisfactorily defended it.

This completes the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology.

Thesis committee members:

Name	**Date**
Andre O. Hudson, Ph.D. Thesis Advisor	
Gregory A. Babbitt, Ph.D.	
Michael A. Savka, Ph.D.	

Michael V. Osier, Ph.D.
Director of Bioinformatics MS Program

475-4392 (voice)
mvoscl@rit.edu

Table of Contents

1. Acknowledgements

I would like to thank my thesis advisor and committee members, Dr Andre Hudson, Dr Renata Rezende Miranda, Dr Michael Savka and Dr Gregory Babbitt for all their advice, guidance and help on crafting this project and writing this thesis. Thanks as well to my family and friends for their continued support and encouragement in helping me achieve my goals.

Lastly this research was funded by a National Institutes of Health award (R15GM144862) to AOH.

2. Abstract

In the Hudson Lab, which is focused on discovering new antibiotics, bacteria samples are taken from the environment and cultured in large quantities. Then they are tested for antibiotic resistance before they are sequenced and their secondary metabolite compounds are extracted. This is both a lengthy and expensive process that becomes more and more difficult as the number of samples one is working increases. This project assessed a different approach to rejuvenate antibiotic development with antiSMASH. antiSMASH is an online tool created by collaborators from many different institutions that uses profile Hidden Markov Models (pHMMs) to detect gene clusters which produce secondary metabolites in bacteria. The antiSMASH tool has its own repository of these "profiles" which are position specific information about an amino acid from a protein encoding gene derived from multiple sequence alignments. Once a genome is entered into antiSMASH, if these profile modules are detected and they are outputted to the user if a certain metabolite/cluster is present. Many gene clusters are known to produce metabolites with antimicrobial properties which the antiSMASH tool could potentially detect. Using this tool, the goal was to identify a potential pipeline of antibiotic discovery that would be a great improvement in time and reduce costs by using the tool as a screen of a possible viable candidate for antibiotics. In this project 30 genomes were used and fed into antiSMASH. They were broken down into positive and negative controls, known producers and unknown producers. We then looked at the tools ability to screen for antibiotics in each of those data types.

3. List of Figures

4. LIST OF TABLES

5. INTRODUCTION

Background & Relevance of this Project

The creation of a drug or any new form of medicine has long been known to be a long term, difficult and costly process, that far more often than not ends in failure. To be exact, approximately 90% of all drugs that have advanced to phase I clinical trials fail (Sun et al., 2022). This statistic doesn't even include drugs that failed in pre-clinical trials. The estimated time to develop a drug from start to finish is approximately 12 years at a cost of 2 billion dollars (Mohs & Greig, 2017).

This is especially true for antibiotics. Along with the aforementioned problems, antibiotics also face the compounded problems of rising inefficacy and a multi-decade long discovery void which has gone on for the last 30 years. These factors made antibiotic development risky due to the great costs and very little return. But what if there was an alternative method to finding antibiotics that would drastically reduce costs and speed up production time? This really isn’t a magic bullet but rather using all the tremendous advancements made in genomics and sequencing technology to our advantage. Since the advent of next generation sequencing technologies in the mid-2000s, there has been an explosion in sequencing data which has led to various databases with genomic information (Muir et al., 2016). And better yet this information is all freely available.

The use of this freely available data lies at the heart of this project. The antiSMASH software tool was created specifically to utilize this data. antiSMASH is able to take users into genomic sequence and find gene clusters that produce secondary metabolites (non-essential to

cellular life) in bacteria (Medema et al., 2011). Many of these secondary metabolites are the source of antibiotics.

I propose and examine the viability of the following workflow where antiSMASH is at the beginning of the antibiotic discovery process. A lab or group of researchers have a bacterium which they presume might produce antibiotics. They sequence its genome and use antiSMASH for genome mining. If the results look favorable or provide interesting leads, that is when the true experimentation and tests are done to prove that the bacterium has antibiotic-production capabilities. This workflow has the potential to massively reduce the time, cost and effort of discovering antibiotics as a lot of the guess work has been removed. Only targets with genomic potential are chosen via antiSMASH, which would lead to a more efficient use of time and resources.

Current crisis of antibiotic resistance

One of the most pressing challenges plaguing the scientific community is that of antibiotic resistance. Antibiotic resistance describes bacterial pathogens that are multidrug resistant and do not respond to therapeutic treatment in clinical settings (Akova, 2016). Any bacteria can exhibit antimicrobial resistance (AMR) from a particular drug but still be susceptible to others (Akova, 2016). The true danger and threat lie when these infection-causing bacteria are resistant to many antibiotics and chemotherapeutic agents (Nikaido, 2009). These bacteria are now often responsible for infections that are difficult to treat by conventional antibiotics, if not outright untreatable (Frieri et al., 2017). Take for example the infamous case of a 90-year-old Nevada woman who died in 2016 from a bacterial infection resistant to every

available antibiotic in the US (Hudson, 2021). This development coupled with a sparse drying up of antibiotic development documents a bleak picture for the future. Antibiotic resistance can be observed in two ways, microbiologically or clinically (MacGowan & Macnaughton, 2017). Microbiological resistance describes when a genetically determined resistance mechanism is present in the bacteria genome (MacGowan & Macnaughton, 2017). This mechanism can be acquired via other bacteria (i.e., horizontal gene transfer) or mutation. Antibiotic resistance can be an intrinsic feature in some bacterial species. For example, take Gram-positive organisms being resistant to colistin or *Enterobacteriaceae* to glycopeptides (MacGowan & Macnaughton, 2017)

Acquisition of resistance

Bacteria can acquire resistance in a few different ways, which all revolve around the same concept. That is, the acquisition of gene(s) that encodes for a resistance mechanism. Here are the three main ways as illustrated by Figure 1. 1) Conjugation describes the direct cell to cell contact between two bacterial cells resulting in a plasmid transfer between a susceptible cell and one that already has resistance genes (MacGowan & Macnaughton, 2017). 2) Transduction refers to the transfer of bacterial DNA with resistant genes via bacteriophage. This occurs when a bacteriophage infects a bacterium for replication. The viral DNA and host DNA with resistant genes are then integrated into new phages created inside the host cell. Once those phages leave the host cell, they now contain newly authored resistance genes (MacGowan & Macnaughton, 2017). 3) Transformation, which references the up-take of DNA with resistance from the environment by the bacterial cell (MacGowan & Macnaughton, 2017).

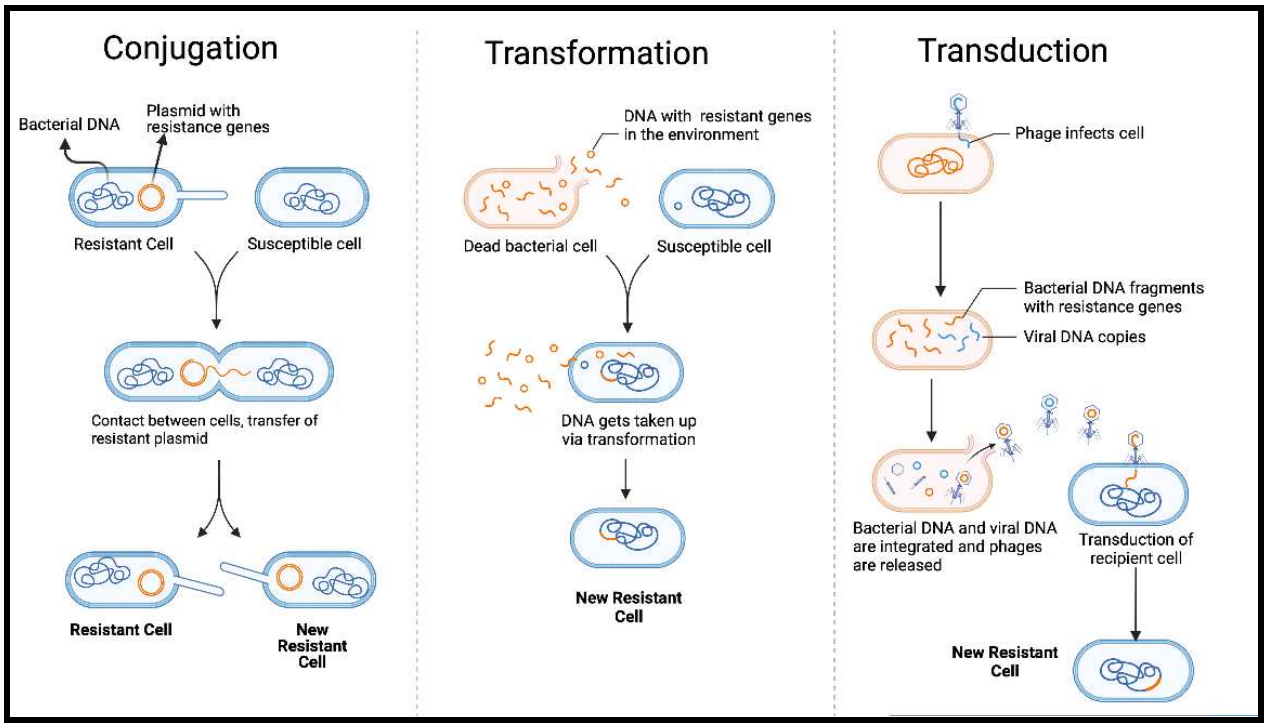


Figure 1. *Ways Bacteria Acquire Resistance genes. Overview of the three ways bacterial cells acquire resistance genes and become resistant to antibiotics. Conjugation is where a susceptible bacterial cell comes into contact with another bacterial cell with a resistant gene. They share DNA and the susceptible cell becomes resistant. Transformation is when bacteria take up DNA with resistant genes from the environment and incorporate it into the genome. Transduction is when a bacteriophage uses a bacterial host with resistant genes for replication and incorporates their DNA with the bacterial cell. Once they leave, they serve as a vector and are able to pass resistant genes to other cells. [Created with BioRender.com]*

Mechanisms of Resistance

There are a number of ways bacteria can be resistant and ward off the effects of antibiotics. One way briefly described earlier was intrinsic resistance. In this case, bacteria are able to resist and negate the action of an antibiotic simply based on their structural or functional characteristics like the absence of a specific susceptible target of an antibiotic (Blair et al., 2015). A great example of this is with the antibiotic daptomycin which is active against Gram-positive bacteria but has no effect against Gram-negative bacteria. The reason for this comes down to the difference in the cell membrane composition of these bacterial types. Gram-negative bacteria have a lower proportion of anionic phospholipids in the cell membrane than Gram-positive bacteria do. This diminishes the efficacy of daptomycin whose antimicrobial activity depends on its Ca^{2+}-mediated insertion into the cell membrane (Blair et al., 2015).

Additionally, there has been the identification of genes that are responsible for intrinsic resistance in bacteria to different antibiotic classes most notably β-lactams, fluoroquinolones and aminoglycosides (Blair et al., 2015). For example, molecular experiments performed by researcher Anne Liu et al at the University of California Los Angeles identified close to o 4,000 single-gene knockouts that when performed increase *Escherichia coli* susceptibility to antibiotics like rifampicin, triclosan, nitrofurantoin, aminoglycosides and some β-lactams (Liu et al., 2010).

Besides intrinsic resistance, bacteria can develop resistance to antibiotics through other means. One way could be through target modification and change. This is when a molecular target of a specific antibiotic undergoes changes that render the antibiotic ineffective. Often this comes down to point mutations in select genes resulting in rise in resistance (Wright, 2011). An example of this is with ciprofloxacin, a synthetic fluoroquinolone antibiotic that targets type IIA

topoisomerases. A single mutation in target genes like gyrA, where a serine is changed to an amino acid with a bulkier side-chain, often leads to fluoroquinolone resistance in those bacterial cells (Wright, 2011).

Another way is through the bacterial cell's ability to remove and eject antibiotic molecules. This is accomplished by the efflux pump. An efflux pump are composed of cell membrane transport proteins whose role is to rid the cell of toxic substances (including antibiotics) and eject them into the environment (Webber & Piddock, 2003). While many efflux pumps are narrow and allow for substrate specificity, there are some that are able to transport a wide range of structurally dissimilar substrates and are known as multidrug resistance (MDR) efflux pumps (Blair et al., 2015). While it's been shown that almost all bacteria carry the genes to code such a protein, the problem arises when they are overexpressed, as shown in Figure 2. This overexpression is possibly due to a point mutation, which leads to high levels of resistance to antibiotics (Blair et al., 2015).

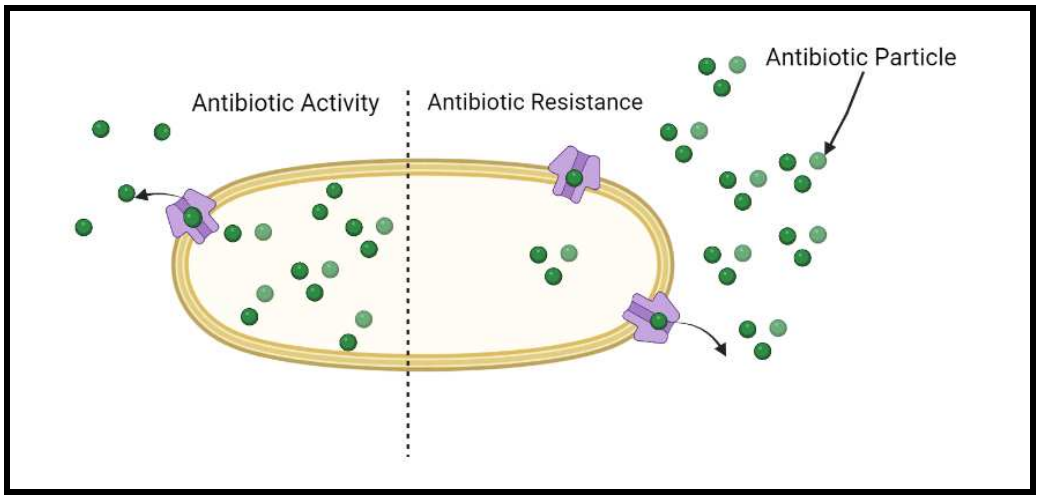


Figure 2. *Efflux pump ejection of antibiotic particles. This Figure shows the effect that efflux pump overexpression has on antibiotic resistance. The left side of the Figure shows a standard cell with an efflux pump. Even though almost all bacterial cells have an efflux it's normally not enough to prevent antibiotic activity in the cell. On the right side of the cell, overexpression of the efflux pump leads to a rapid removal of a large amount of the antibiotic molecules from the cell. [Created with BioRender.com]*

Lastly there is the mechanism of chemical or enzymatic modifications of the antibiotic. This is thought to be the most effective tool that bacteria have in counteracting antimicrobials (Wright, 2011). In processes related to this mechanism, the bacterial cell is able to directly interfere and change the chemistry or structure of antibiotic products and render them ineffective. This was first observed in the early 1940s when scientists took note of penicillin inactivating β-lactamase activity which degraded and destroyed penicillin (De Pascale & Wright, 2010). Another example of this mechanism is when the bacteria adds to chemical groups to vulnerable sites on the antibiotic. This induces antibiotic resistance by preventing antibiotics from binding to their target site/protein due steric hindrance (Blair et al., 2015).

History and Discovery of antibiotics

Ever since their inception almost a century ago (Kingston, 2000), antibiotics have proven as one of the most significant discoveries in modern medicine. For centuries prior to their discovery and production, bacterial pestilence used to ravage humanity and human civilizations. Each epoch of human history has been dominated by waves of infections such as syphilis, typhoid fever, and tuberculosis (Mohr, 2016), but the discovery of antibiotics have leveled the playing field and even changed the game in regard to human health and life span. Today, antibiotics have a much wider range of use in a variety of different fields and industries. This is a cause for concern for it is the rampant overuse of these drugs that is believed to be the cause of the great increases in the number of bacteria that are resistant to antibiotic treatment. This has a tremendous impact on all facets of modern society that regularly use and have been reliant on antibiotics.

What makes antibiotic creation and technology so intriguing and revolutionary is that it was one of the first instances where we were able to harness and genetically manipulate an already existing organism and use the resulting product in humanity's favor. The term antibiotic can be used to describe a molecule or a chemical substance created by one organism to inhibit the growth and proliferation of a microorganism (Clardy et al., 2009). Antibiotics can be derived from a number of different sources. One method through which antibiotics can be obtained is through the natural processes of biosynthesis. As previously stated, antibiotics can arise as the natural by-product of a biochemical pathway in certain bacterium or fungus. For example, Penicillin, discovered in the 1940s by Alexander Flemming, arose from what was described as “mold juice” (*Alexander Fleming Discovery and Development of Penicillin - Landmark*, n.d.).

This juice was noted for its ability to kill and neutralize other harmful bacteria like *Streptococcus*, and *Meningococcus* (*Alexander Fleming Discovery and Development of Penicillin - Landmark*, n.d.). Flemming and others knew how remarkable this discovery was and the importance of this mystery mold. Using a wide array of processes and assays like fermentation, scientists were able to finally isolate pure penicillin for use and distribution (*Alexander Fleming Discovery and Development of Penicillin - Landmark*, n.d.). This case of antibiotic production is probably one of the most famous and well-known examples of how it can be done, but technology and science has progressed a very long way since then and scientists and pharmaceutical companies now have means to mass produce a continual supply of antibiotics. Some means still have their basis in naturally occurring antibiotics while others are more synthetic. Take for example ampicillin. While its main backbone and structure is derived from natural produced/occurring penicillin, scientists were able to extrinsically alter its structure and function by adding an amino group, a benzylpenicillin molecule, creating a new class of drug with different levels of activity and effectiveness (Raynor, 1997).

How antibiotics are made

Remarkably, the basis of their creation and production has not changed much since Flemming accidentally identified and produced penicillin. Albeit the process is a whole lot more industrialized and scaled up. It all starts with finding and obtaining a desired organism that is suspected to have antimicrobial properties. The organism must be isolated for researchers to be able to generate pure cell cultures (*How Antibiotics Are Made,* n.d.). These cultures are grown and harvested in environments suitable for thriving like an agar plate or suspension flasks

containing food and other nutrients (*How Antibiotics Are Made,* n.d.). These cultures are then transferred to containers called “Seed Tanks". These containers are designed to provide the most ideal conditions for microorganism growth [1]. Key ingredients like warm water, carbohydrate rich foods like glucose or other sugars, alcohols, nitrogen sources along vital growth factors like vitamins, amino acids are all needed to make these seed tanks a hotbed of activity and perfect for harvesting bacteria (*How Antibiotics Are Made,* n.d.). The next step in the process of antibiotic manufacturing is fermentation. Very similar to the seed tank, a fermentation tank is set up containing growth media from the seed tank and provides much more room and area for growth. Additionally, acids and bases are added regularly to regulate the pH which is vital at this stage for growth. Here the bacteria can multiply and excrete their desired metabolite (*How Antibiotics Are Made,* n.d.). After a few days the maximum amount of the metabolite will have been produced. The metabolite is then purified and separated from the fermentation. The process taken to purify is very specific to the properties and chemistry of the metabolite (*How Antibiotics Are Made,* n.d.). Once purified the metabolite is then tested to see how it performs against desired bacterial targets. If successful, it is packaged and shipped as a product, though before that final step it must pass and go through rigorous quality control measures (*How Antibiotics Are Made,* n.d.). This process, summarized by Figures 3 and 4, shows antibiotic development to be a very intricate, expensive process which has contributed to research and development to be stagnant and almost abandoned (Plackett, 2020). This situation is very alarming and is happening at the worst possible time due to the rising rates of antibiotic resistance among our current drugs. The United Nation approximates 700,000 people each year die of drug resistant infections, and this could swell to a whopping 10 million people by 2050 (Plackett, 2020). Therefore, the main goal of this project is to look at another possibility to aid and fast-track antibiotic development.

There really has not been much in the way of circumventing this difficult process by leveraging the high-throughput data now available thanks to whole-genome sequencing, metabolomics and bioinformatics so far. By using antiSMASH, a web-based tool that allows for rapid genome-wide identification and analysis of secondary metabolite biosynthesis gene clusters in bacterial genomes, we looked to determine if researchers can determine whether a bacterial species has antimicrobial properties just by examining its whole genome.

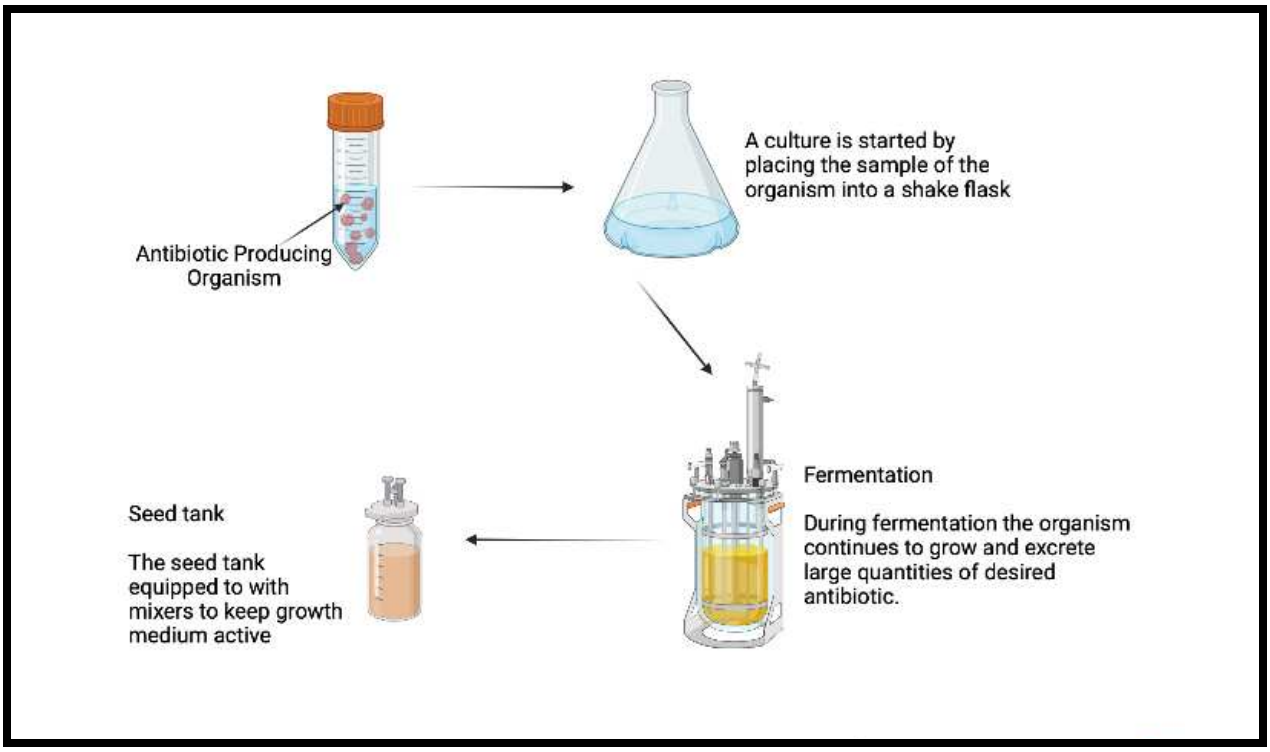


Figure 3. *Manufacturing process of antibiotics. First the desired organism is cultured and suspended in a flask with required growing nutrients. This suspension is transferred to seed flask to increase the and sustain growth before being placed in a fermentation tank where microorganism growth explodes and produces desired metabolite. [Taken from How Products are made. Antibiotics.]*

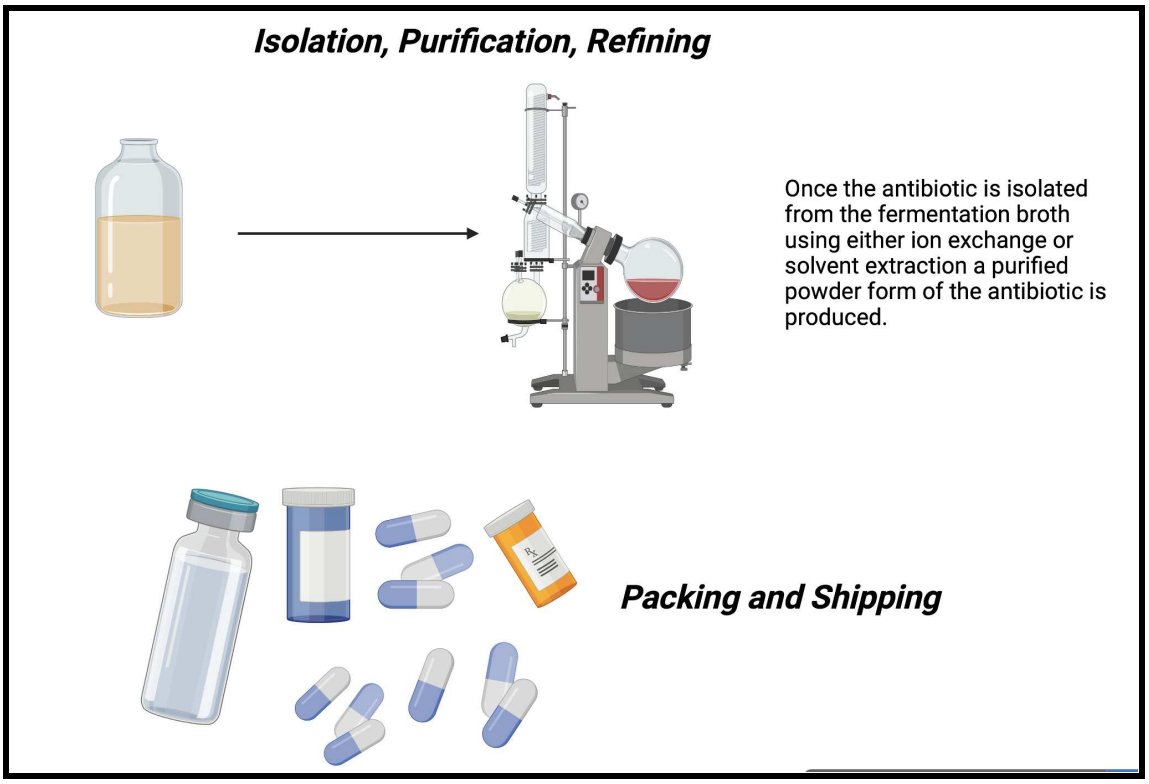


Figure 4. *Isolation and refining the metabolite product. This shows the process of isolating and refining the metabolite product before packaging and shipping if deemed to be a successful antibiotic product. Purification of the metabolite can be achieved by a number of different ways i.e., ion-exchange or solvent extraction. Once isolated tests are performed to test efficacy. If effective and pass quality assurance trials, it is a viable product ready to be shipped [Taken from How Products are made. Antibiotics]*

Types of Antibiotics

Antibiotics can largely be broken down into at least 16 different types or classes (Yim et al., 2006). What differentiates each of these classes are mainly their mechanism of action and how they go about neutralizing a microbial threat (Alanis, 2005). Table 1 demonstrates this by showing the classes and their main mechanism of action. Furthermore, the mechanisms of antibiotics can be further classified into two larger categories depending on the effect they have on pathogenic bacteria. Bacteria can be bactericidal or bacteriostatic. If an antibiotic is classified as bactericidal, it is able to kill bacteria without the help of the host's immune system (*Static or Cidal; Which Is Best? - Microbiology Nuts & Bolts*, n.d.). If an antibiotic is classified as bacteriostatic, it stops the pathogen from multiplying and allows for the host immune system to kill the bacteria (*Static or Cidal; Which Is Best? - Microbiology Nuts & Bolts*, n.d.).

Additionally, there has also been the proliferation of synthetic biology in this domain. This comes in the form of there being significant synthetic modifications in creation of some antibiotic classes. For example, tetracycline and macrolide are now largely being produced with fully synthetic platforms and completely in a laboratory setting (Mitcheltree et al., 2021). This method of antibiotic production arose as a way to drastically increase the functionally and ability of antibiotics. Scientists are faced with many challenges with the creation of natural or semi-synthetic antibiotics due to the limitations and challenges linked to the rate of drug discovery and the complexity of developing and modifying the chemical compounds found in classes of antibiotics (Mitcheltree et al., 2021). By going through the synthetic route, scientists look to alleviate these problems. This is because scientists would be able to keep structural

changes and modifications that would be much more difficult to achieve by natural based methods (Mitcheltree et al., 2021).

Table 1. *Antibiotic classes and their mechanism of action* (Yim et al., 2006).

Antibiotic classes	***Mechanism of action***
penicillins, cephalosporins, carbapenems, monobactams, daptomycin	Inhibition of cell wall synthesis
Tetracyclines, aminoglycosides, linezolid, quinupristin-dalfopristin, ketolides, macrolides, lincosamides	Inhibition of protein synthesis
Fluoroquinolones	Inhibition of DNA synthesis
Rifampin	Inhibition of RNA synthesis
Sulfonamides; trimethoprim	Competitive inhibition of folic acid synthesis Inhibition
Polymyxins	Membrane disorganizing agents

Advantage of computational methods

Due to rapid advancements in technology, computational methods are becoming more and more necessary to solve important biological problems. The discovery of viable antibiotic candidates could be one of those problems. This is the area that this project intends to investigate in order to verify and offer an alternative way of tackling this problem by utilizing the advances made in bioinformatics, statistics, etc. to aid in discovering new antibiotics. By utilizing genome screening, we intend to speed up the discovery pipeline by pre-selecting promising candidates from those results to then perform metabolite extraction. In essence, instead of starting the antibiotic discovery pipeline from the organism level, we would start at the genomic level. One of the things that makes this procedure feasible is the abundance and availability of data that we benefit from today. Technologies like whole-genome sequencing are so cheaper than ever and will continue to be affordable and available. Right now, there are databases with hundreds of thousands of genome sequences of bacteria and organisms not well known or examined. This approach also alleviates the need for the great investments or large amounts of time spent looking for a viable organism and performing the research and development to create a viable product as shown in Figure 3. Instead, we start with information and data that is abundant and free to then harness the progress made in whole genome sequencing by using antiSMASH to find gene clusters related to secondary metabolites in bacteria that have antibiotic properties.

The Tool antiSMASH

he use of antiSMASH lies at the heart of this project. antiSMASH is an online based tool that is able to detect and characterize biosynthetic loci of secondary metabolite compounds (Medema et al., 2011). The first iteration of this tool was created in 2011 precisely to leverage the advances made and diminished cost of genomic sequencing to detect gene clusters which far outstrips the rate at which laboratory researchers are able to keep up and detect gene clusters through experimental analysis (Medema et al., 2011). The tool works by first receiving a genome sequence in the form of a file or accession number. The user then has the option to decide on a number of features, downstream analysis and gene cluster types to specify in their search (Figure 5). The server then identifies gene clusters present. Then it aligns identified gene cluster regions to other known gene clusters and shows what metabolites are related/linked to those clusters (Figure 6). The NRPS (Nonribosomal peptides synthetases)/PKS (Polyketides) are examples of such gene clusters antiSMASH uses in alignment and holds particular interest to this project. PKS are a class of naturally produced compounds that host a wide variety of clinical/pharmaceutical properties such as anticancer and antibiotic activity (Gomes et al., 2013). In a similar fashion, NRPS compounds are known for their prolific therapeutic/industrial usage in the domains of antibiotics and immunosuppressant (Martínez-Núñez & López, 2016). The NRPS/PKS are gene clusters synonymous with antimicrobial activity and antiSMASH servers allow for automatic detection and annotation if present (Medema et al., 2011). The presence or absence of these two clusters will play a key role in finding out whether a bacterial genome has possible antibiotics capabilities.

Continuing on in the pipeline, the server then presents all previously available secondary metabolite gene analysis in one interactive view (Medema et al., 2011). At the core of antiSMASH is the use of the HMMer3 tool. The HMMer3 tool uses profile Hidden Markov Models (pHMM) to search for amino acid sequence translations of all the protein encoding genes (Medema et al., 2011). pHMMs are tools that have long been used as a method of determining homology (Johnson et al., 2010). They are a variant of Hidden Markov Models (HMM), specifically related to biological sequences. They are probabilistic models that have been derived from multiple sequence alignments, from which they are able to create a suitable scoring system for detecting homologous sequences. They use position-specific information from alignments to create a profile model in looking for homologous sequences (Johnson et al., 2010). The antiSMASH server utilizes a number of pHMMs, from already existing to new ones created by seed alignments (Medema et al., 2011). The overall pipeline for antiSMASH is as follows. A genome sequence is uploaded to the server and then gene cluster identification is performed with the use of pHMMs. Gene clusters are identified by locating clusters of signature gene pHMM hits spaced with <10kb from each other. Flanking accessory genes to these gene clusters are included as well. This is done by taking the last signature gene hit and extending it from 5kb - 20kb. This varies based on the type of the gene cluster. Then a slew of analyses can be performed like chemical structure prediction, ClusterBlast, secondary metabolism Clusters of Orthologous Groups (smCOG analysis) (Medema et al., 2011) (Figure 5).

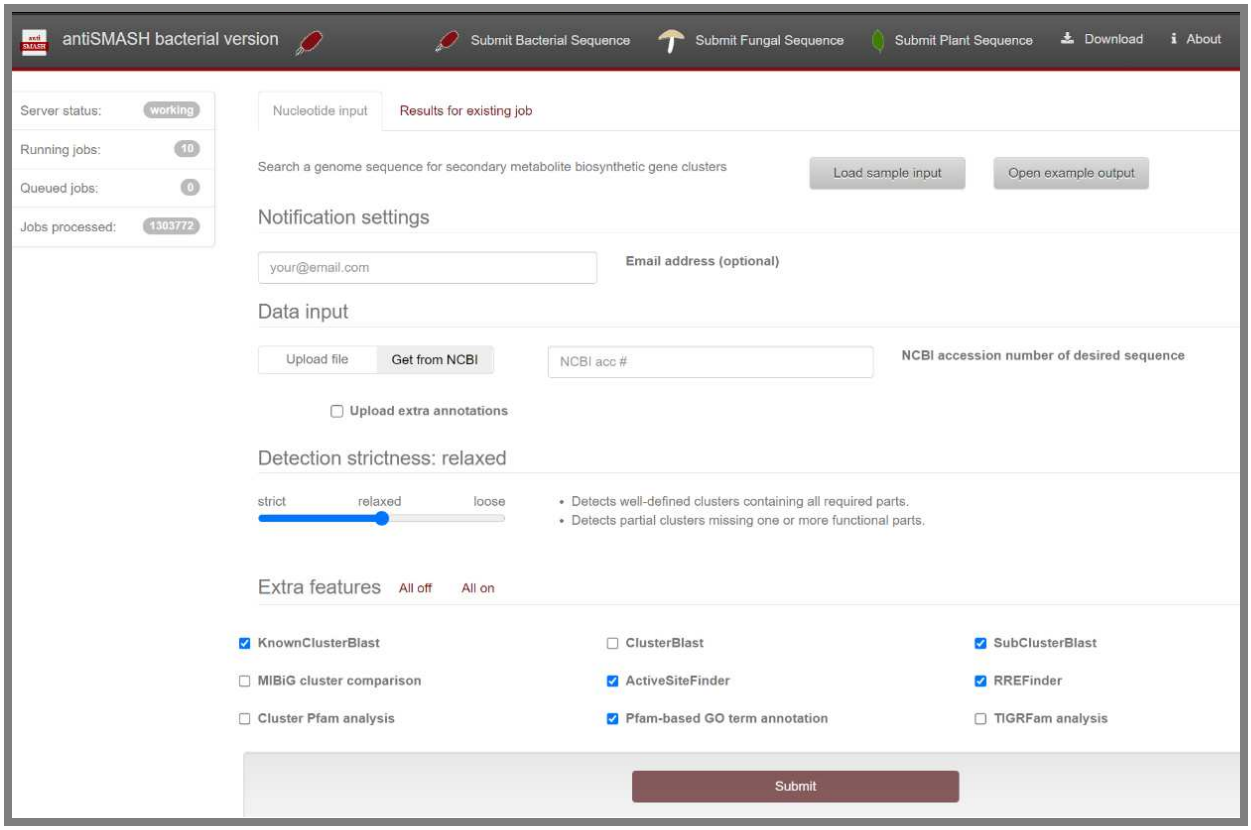
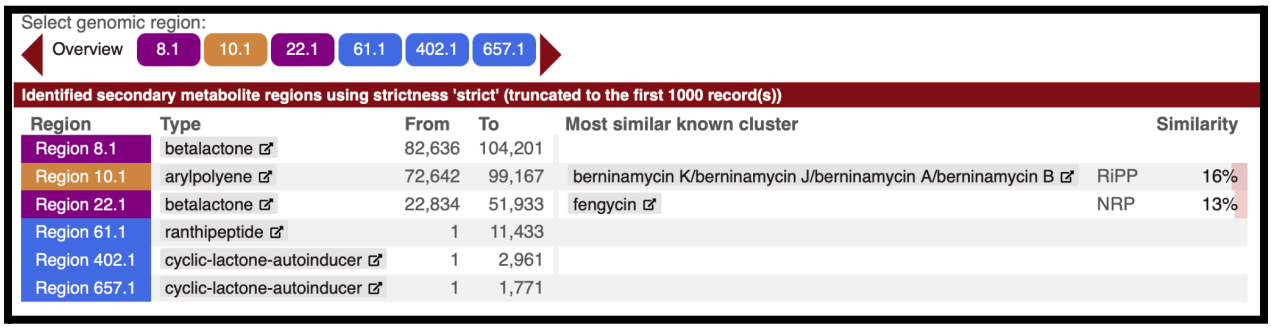


Figure 5. *A picture of the antiSMASH server shell/user interface. Genomic data can be entered in the form of a fasta file or accession number. Users can also curtail the strictness/specificity of the search, more relaxed yielding more distant hits. Extra features for further analysis shown at the bottom of the page.*



Region	Type	From	To	Most similar known cluster		Similarity
Region 8.1	betalactone	82,636	104,201			
Region 10.1	arylpolyene	72,642	99,167	berninamycin K/berninamycin J/berninamycin A/berninamycin B	RiPP	16%
Region 22.1	betalactone	22,834	51,933	fengycin	NRP	13%
Region 61.1	ranthipeptide	1	11,433			
Region 402.1	cyclic-lactone-autoinducer	1	2,961			
Region 657.1	cyclic-lactone-autoinducer	1	1,771			

Figure 6a*. An example of the initial results from antiSMASH after a bacterial genome is scanned. This is a summary of the gene clusters found and their secondary metabolites. The first*

section described as a genomic region lists the gene clusters detected and are able to be selected for closer inspection. Below that, the tool summarizes the properties of each cluster. Type indicates the type of metabolite it is predicted to produce. From & to describe location in the genome, the most similar known cluster is the prediction of that secondary metabolite. Similarly is the homology scoring of a queried cluster compared to a database cluster that produces predicted metabolites.

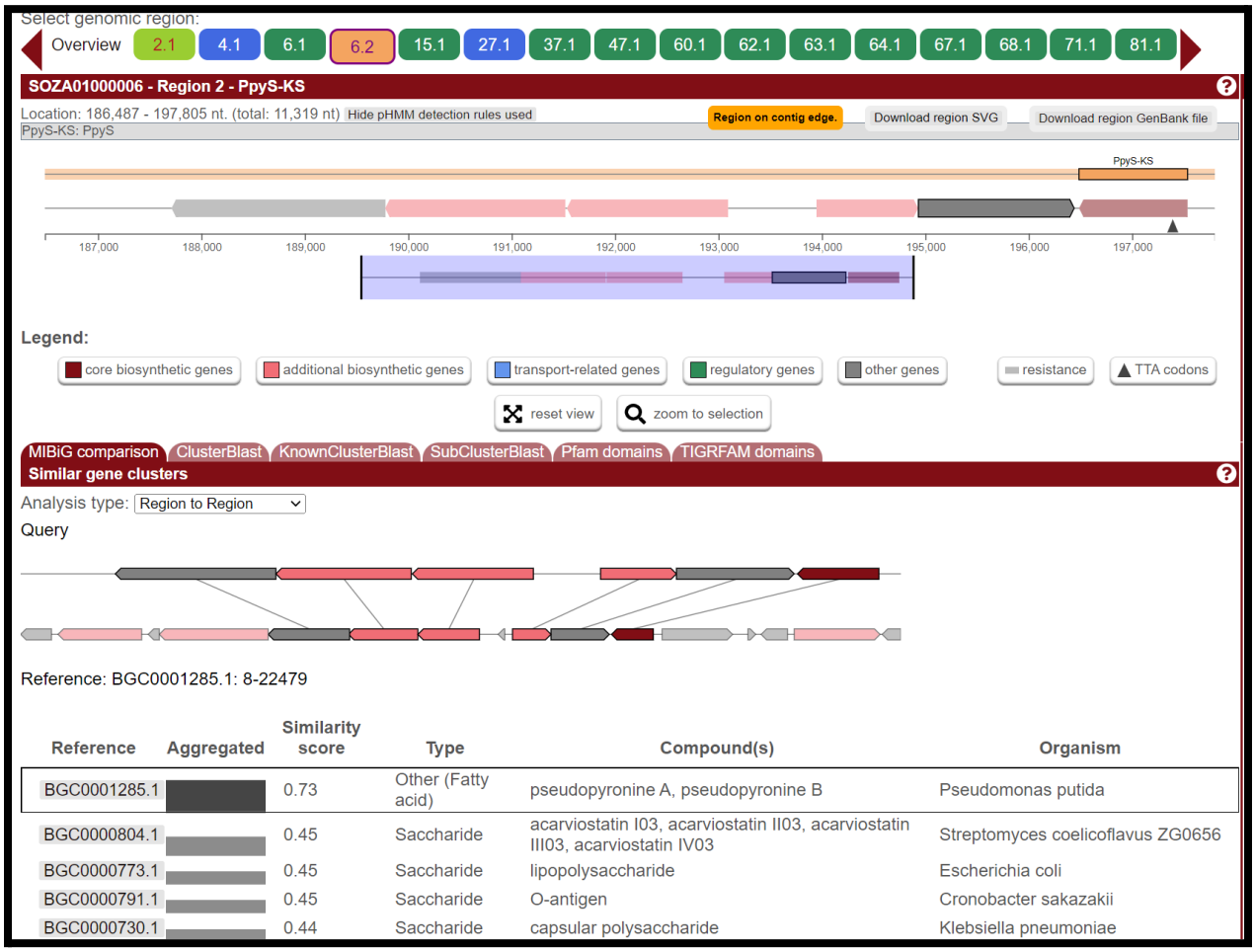


Figure 6b. *An example of a result query from an antiSMASH search. Very top regions of the genome with identified clusters are shown. Below that is a gene map displaying the location of the gene clusters, along with their type/function. Bottom half of the page shows results from the*

various analyses done by the server. Most relevant is MIBiG (Minimum Information about a Biosynthetic Gene Cluster) comparison which shows other related homologous clusters and metabolites they produce.

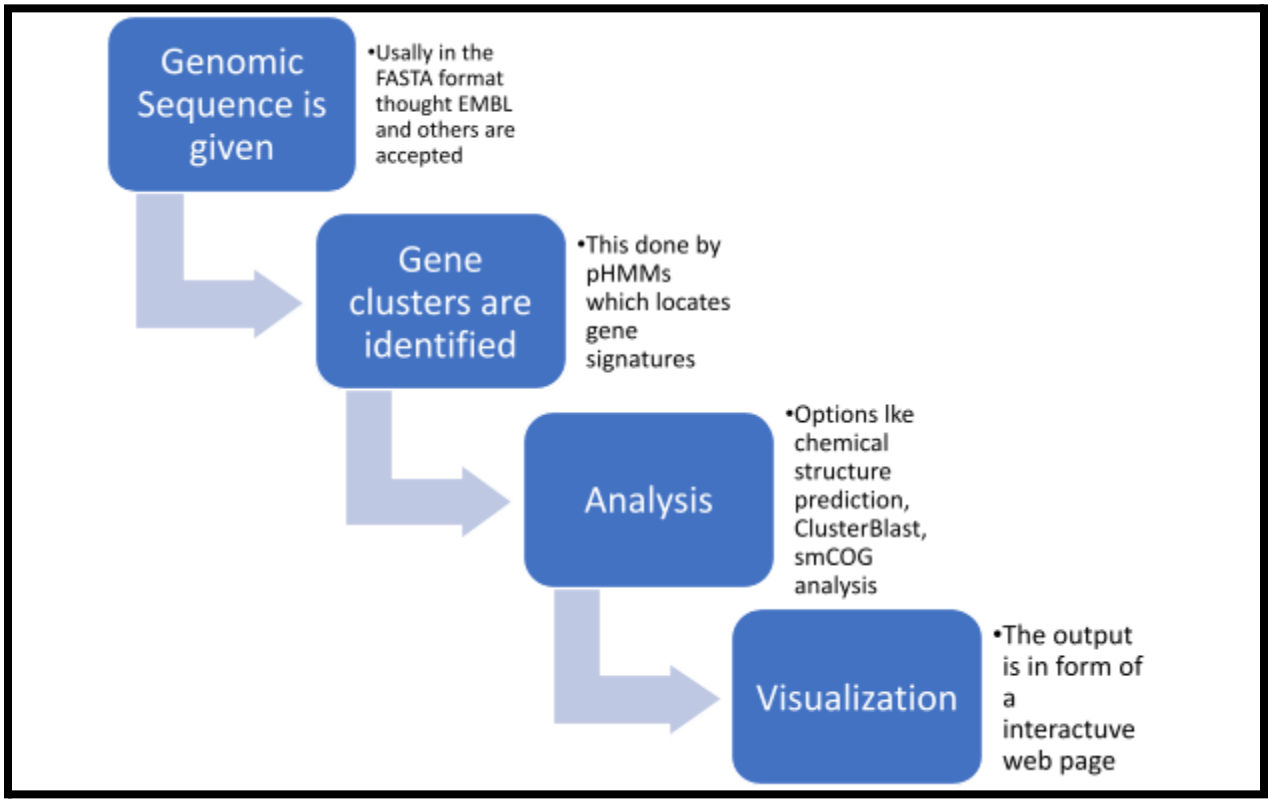


Figure 7. *Adapted from antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Figure gives an outline of the antiSMASH analysis pipeline. Genetic data is given to the server, and gene clusters are identified with pHMMs. Downstream analyses can be performed like: NRPS/PKS domain analysis and annotation, chemical structure prediction, ClusterBlast gene cluster comparative analysis, and smCOG secondary metabolism protein family analysis. The output is visualized in an interactive web page*

Hidden Markov Models & Profile Hidden Markov Models

Hidden Markov Models (HMM) are named after Russian mathematician Andrey Andreyevich Markov (Franzese & Iuliano, 2019). They are statistical models used to capture information hidden from observable sequential symbols like a nucleotide sequence (Franzese & Iuliano, 2019). In HMM, the system being modeled is assumed to be a Markov process where the objective is to determine the hidden parameters from known parameters (Franzese & Iuliano, 2019). In bioinformatics, HMM are used to model biological sequences (Franzese & Iuliano, 2019). Profile HMMs are Markov models applied to protein families (Birney, 2001). They are able to capture the evolutionary changes that have occurred in a set of related sequences by using position-specific information about how conserved each amino acid based on a multiple alignment (EMBL-EBI, n.d.).

These pHMMs work by first starting with a multiple sequence alignment (MSA). For each alignment column or consensus column of a MSA, there is a corresponding “match state” in the model. Each match state has 20 residue (amino acid) emission probabilities which model the distribution of residues allowed in the column (Eddy, 1998). Along with this there’s also an “insert” and “delete” state at each column to allow for insertion or deletion of one or more consensus residues between one column and the next (Eddy, 1998). The probability of scoring parameters in a pHMM are in the form of additive log-odds score created before alignment and scoring of a query sequence (Eddy, 1998). If the probability of the match state transitioning to another residue x is Px and the expected background frequency of x in the database is fx, then the match state would be log(Px/fx) (Eddy, 1998).

Biosynthetic Gene Clusters

Biosynthetic Gene Clusters, or simply gene clusters, are another key and important player in this project. Gene clusters can be described as a grouping of genes that participate in the same metabolic pathway (Rokas et al., 2020). These genes often reside next to each other on the same chromosome, but this is not always the case. The primary goal of these gene clusters is to create and synthesize secondary metabolites for the bacteria or fungi (Rokas et al., 2020). These secondary metabolites, often called “natural products”, are compounds an organism produces that are not needed for growth or the overall survival of the organism, but serve as important tools for signaling interactions with the environment and protection against threats (Osbourn, 2010). These secondary metabolites are versatile and often hold significant pharmacological value. Scientists have capitalized on their versatility and applied them in a variety of domains like antibiotics, antitumor/anticancer agents, insecticides, immunosuppressants and herbicides (Osbourn, 2010).

6. METHODS

The main goal of this study is to determine whether the antiSMASH tool is an effective resource in the development of antibiotics. Due to how cheap and available sequencing technology is as well as the abundance of genomic data available, genome mining presents a real opportunity to break the antibiotic discovery void. Figure 8 shows one example of a workflow using antiSMASH. It shows the standard workflow of finding an antibiotic in the Hudson Lab and how antiSMASH could disrupt and speed up this process. First is environment discovery from which the bacterium is isolated. Tests are performed to check for resistance. Then extraction is performed to get secondary metabolites of the bacteria from which antibiotic tests are performed. If successful, these samples are sequenced and annotated. With antiSMASH, this workflow would be fast tracked. After isolating the samples, the bacterial genomes would be sequenced immediately and then passed into antiSMASH to perform cluster analysis to see if the bacterial samples have the potential and capability to produce antibiotics.

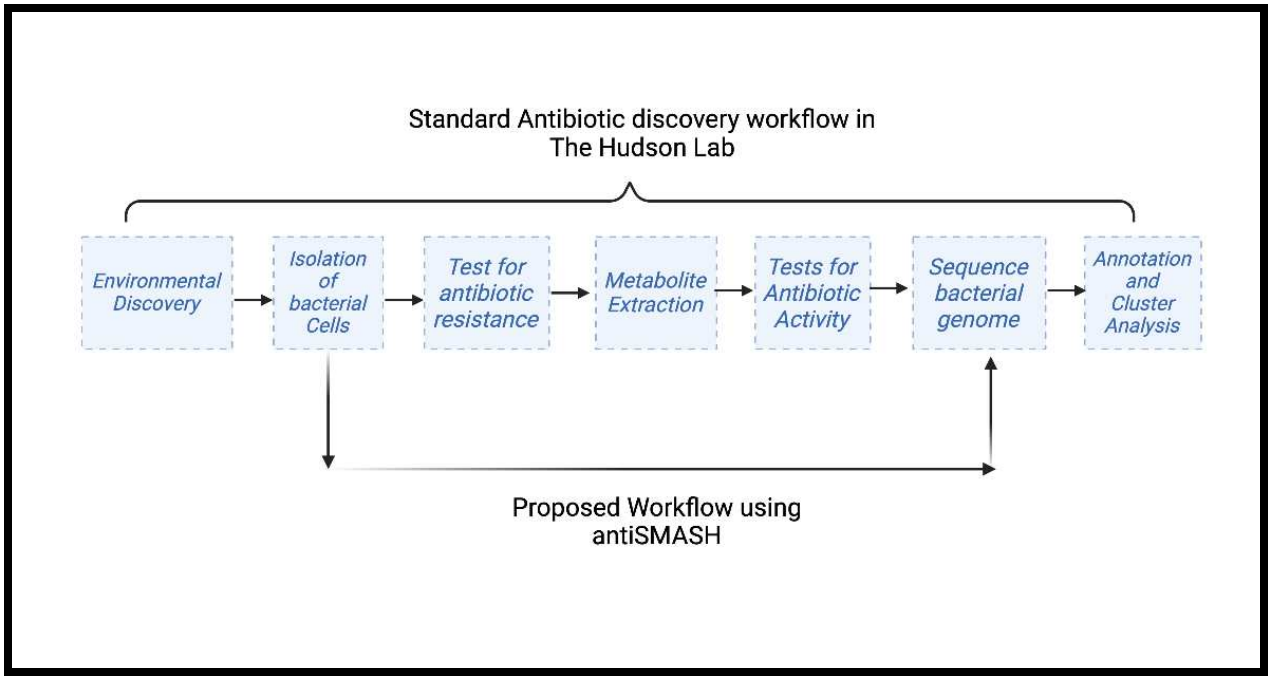


Figure 8. *Antibiotic discovery workflow in Hudson Lab. An example of the standard workflow in the Hudson Lab for discovering antibiotics. First, the bacterium is identified and retrieved from the environment. The cells are isolated and tested whether or not they are susceptible to antibiotics. The metabolites are extracted and then tested for antibiotic activity. If they show signs of activity, the bacterial genome is sequenced and annotated. antiSMASH is proposed to disrupt this process. By sequencing after isolation, researchers can perform annotation and cluster analysis to see if secondary metabolites from the bacteria can potentially produce antibiotics.*

This project utilized antiSMASH as a screener of antibiotics by using the already existing genomic data. First was the antiSMASH analysis of the positive controls from the Hudson Lab. These five (5) positive controls are the genome sequences of bacteria that have proven to produce antibiotics via the traditional workflow. These bacterial samples used were *Pseudomonas sp. RIT 623* (accession number SOZA00000000), *Exiguobacterium sp RIT 594*

(accession number QPKF00000000), *Acinetobacter sp RIT 592* (accession number QPKU00000000), *Exiguobacterium sp. RIT 452* (accession number QXJB00000000), and *Yimella sp. RIT 621* (accession number SEIP00000000). Negative controls used were in the form of the reference genomes of the following bacterial species: *Escherichia coli* (accession number NC_000913), *Streptococcus pneumoniae* (accession number NZ_CP020549), and *Clostridium perfringens* (accession number NZ_CP075979). Two randomly generated DNA sequences, called Genome_1 and Genome_2 were used as well. They were generated via The Sequence Manipulation Suite (Stothard, 2000) DNA sequence generator.. Along with this we took the genome sequence of *Streptomyces rapamycinicus NRRL 5491* and shuffled its contents and created a new genome "Shuffled_Genome.txt". This was done via script in python used for the file processing and then using file converter from "HIV Sequence Database" (*Format Conversion*, n.d.) to convert text file generated from the script to fasta format to be used in antiMSASH.These samples serve as examples of genomes that don't have clusters that produce antibiotics.

Next we used 19 samples for antiSMASH analysis. 9 of these samples were known producers of antibiotics. These genomes were sourced from peer reviewed published scholarly papers. The samples were: *Streptomyces rapamycinicus NRRL 5491* (accession number: CP085193), *Streptomyces rapamycinicus SRMK07* (accession number: CP085309), *Streptomyces noursei ATCC 11455* (accession number: CP011533), *Streptomyces venezuelae ATCC 15439* (accession number: LN881739), *Streptomyces filamentosus NRRL 15998* (accession number: ABYB00000000), *Streptomyces griseus IFO 13350* (accession number: AP009493), *Streptomyces lincolnensis B48* (accession number: CP046024), *Streptomyces*

ambofaciens ATCC 23877 (accession number: CP012382), *Streptomyces atratus SCSIO ZH16* (accession number: CP027306).

The other 10 samples were randomly selected genomes of unknown characteristics. All chosen bacteria will be sourced and verified using The Integrated Microbial Genomes & Microbiomes (IMG/M) system. The IGM/M system is a public repository that houses publicly available microbial genomes, fully annotated and detailed (Chen et al., 2021). The genomes chosen were *Ignavibacterium album JCM 16511* (accession number: CP003418), *Streptomyces sp. CLI2509* (accession number: CP021118), *Streptomyces sp. WAC 01438* (accession number: CP029601), *Pseudomonas aeruginosa DSM 50071* (accession number: CP026680), *Streptomyces hawaiiensis NRRL 15010* (accession number: CP021978), *Pseudomonas aeruginosa DSM 50071* (accession number: CP012001), *Streptomyces sp. S1D4-23* (accession number: CP041613), *Fischerella sp. NIES-3754* (accession number: AP017305), *Streptomyces lydicus WYEC 108* (accession number: CP029042) and *Pseudomonas chlororaphis aurantiaca 464* (accession number: CP027742).

After being fed through antiSMASH each cluster of each genome was evaluated to see if it holds antibiotic properties. The evaluation consisted of a screening of the secondary metabolites produced to see if they were potential antimicrobials. This involved an analysis and searching of literature of said secondary metabolites and their known properties and capabilities. If found the genome was declared to be a potential candidate for antibiotics. Different analyses were performed to assess the overall difference among the genome categories like comparing average number of clusters, seeing the most abundant cluster types and metabolites, as well as comparing the ratio of the number of clusters present versus the number clusters/metabolites had antibiotic capabilities/potential.

7. RESULTS

After each genome was fed through the antiMSASH server, key features were recorded and stored in tabular form. This includes features such as Region, Type, distance/length of cluster (From & To), Most Known Cluster and Similarity (%). Region describes the cluster and a given area of the genome it is located in. To determine whether or not a specific cluster had antibiotic potential, the compound given in "Most Known Cluster" was searched for in the literature to indicate any evidence of antibiotic action or potential. If a connection was found that compound would be highlighted in yellow and the organism would be declared a candidate for being a potential antibiotic producer. This was represented in those species being bolded.

Positive Controls

Out of the 5 positive controls from the Hudson lab, antiSMASH gave back results indicating that only 2 of the 5 could produce metabolites that had antibiotic potential. This is shown in tables 2-6. These bacterial samples were the *Acinetobacter sp. RIT 592* and *Pseudomonas sp. RIT 623*. The metabolites with antibiotic connections/potential are Lankacidin C (Cai et al., 2020), Pseudopyronine A/pseudopyronine B (Bouthillette et al., 2017), Rhizomide A/rhizomide B/rhizomide C (Qi et al., 2021), Berninamycin K/berninamycin J/berninamycin A/berninamycin B (Malcolmson et al., 2013), and Fengycin (Medeot et al., 2020). NRPS or PKS related cluster types are the most abundant cluster types present in the positive controls. This is followed by terpene and betalactone cluster types.

Table 2. ***Pseudomonas sp. RIT623***

Region	Type	From	To	Most similar known cluster	Similarity (%)
2.1	NAGGN	42,676	57,575	-	-
4.1	redox-cofactor	56,992	79,139	lankacidin C	13
6.1	PpyS-KS	186,487	197,805	pseudopyronine A/pseudopyronine B	37
15.1	NRP-metallophoe, NRPS	51,137	107,419	Pf-5 pyoverdine	11
37.1	NRPS	1	37,589	MA026	12
47.1	NRP-metallophoe, NRPS	1	34,514	variobactin A/variobactin B	14
62.1	NRPS	1	14,591	Pf-5 pyoverdine	9
63.1	NRPS	1	14,095	putisolvin	63
64.1	NRPS	1	13,527	pyoverdine SMX-1	9
67.1	NRPS	1	9,413	vacidobactin A/vacidobactin B	11
68.1	NRPS	1	9,118	rhizomide A/rhizomide B/rhizomide C	100
71.1	NRPS	1	8,037	rhizomide A/rhizomide B/rhizomide C	100

Table 3. *Yimella sp. RIT 621*

Region	Type	From	To	Most similar known cluster	Similarity (%)
2.1	NAPAA, terpene	195,227	244,323	carotenoid	33
2.2	butyrolactone	322,234	328,720	-	-
12.1	LAP, thiopeptide	17,719	50,247	-	-

Table 4. *Exiguobacterium sp. RIT 452*

Region	Type	From	To	Most similar known cluster	Similarity (%)
1.1	terpene	69,899	90,726	carotenoid	33
1.2	NRPS-independent-siderophore	505,835	519,165	-	-
4.1	terpene	160,337	181,161	-	-

Table 5. ***Acinetobacter sp. RIT 592***

Region	Type	From	To	Most similar known cluster	Similarity (%)
8.1	betalactone	82,636	104,201	-	-
10.1	arylpolene	72,642	99,167	berninamycin K/berninamycin J/berninamycin A/berninamycin B	16
22.1	betalactone	22,834	51,933	fengycin	13
61.1	ranthipeptide	1	11,433	-	-
402.1	cyclic-lactone-autoinducer	1	2,961	-	-
657.1	cyclic-lactone-autoinducer	1	1,771	-	-

Table 6. *Exiguobacterium sp. RIT 594*

Region	Type	From	To	Most similar known cluster	Similarity (%)
1.1	terpene	156,066	176,513	-	-
1.2	terpene	1,802,363	1,823,190	carotenoid	33

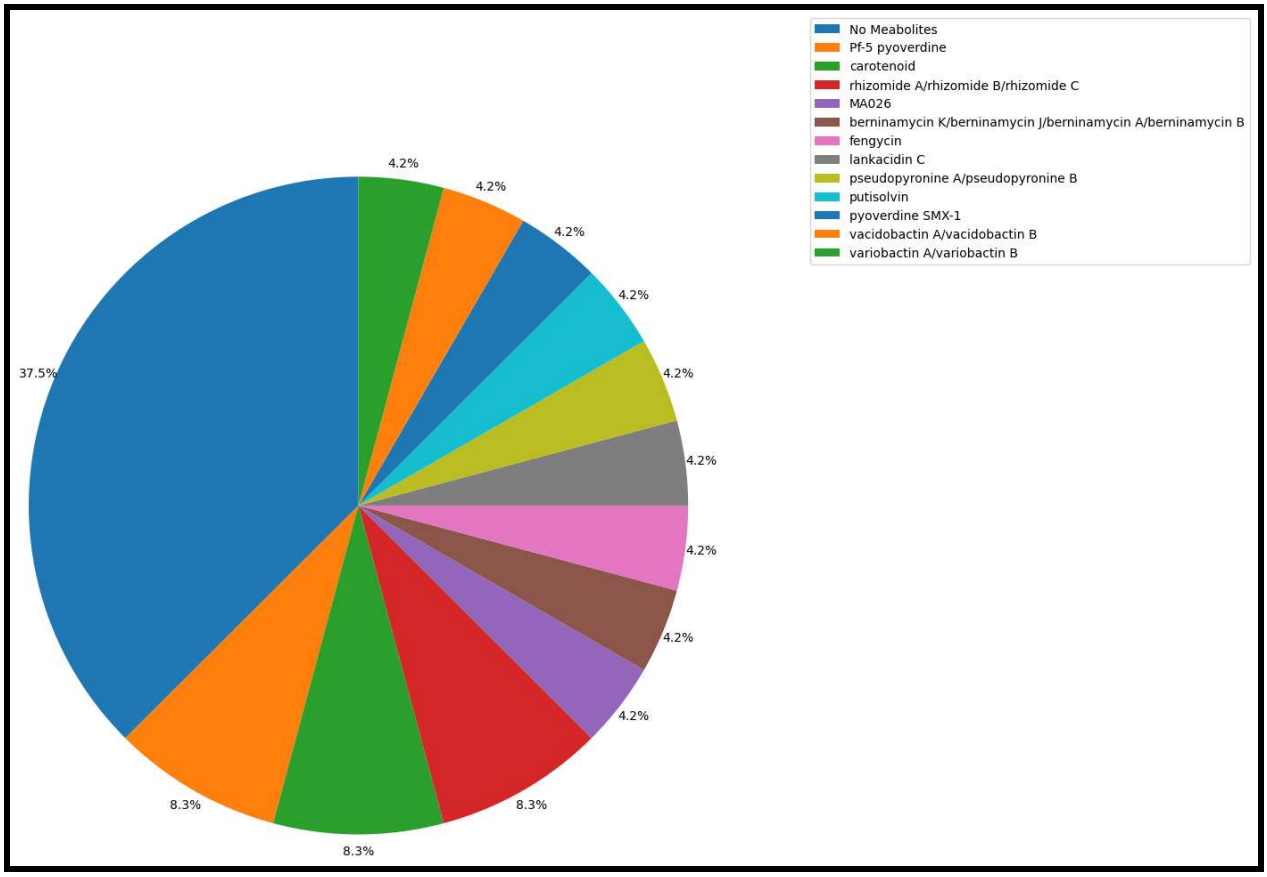


Figure 9. *A representation of the secondary metabolites produced by the positive controls data type. Vast majority of clusters, 37.5%, produced no metabolites, as predicted by antiSMASH. Following that was Pf-5 pyoverdine at 8.3%, carotenoid at 8.3% and rhizomide A/rhizomide B/rhizomide C, which does demonstrate antibiotic activity (Qi et al., 2021) at 8.3%.*

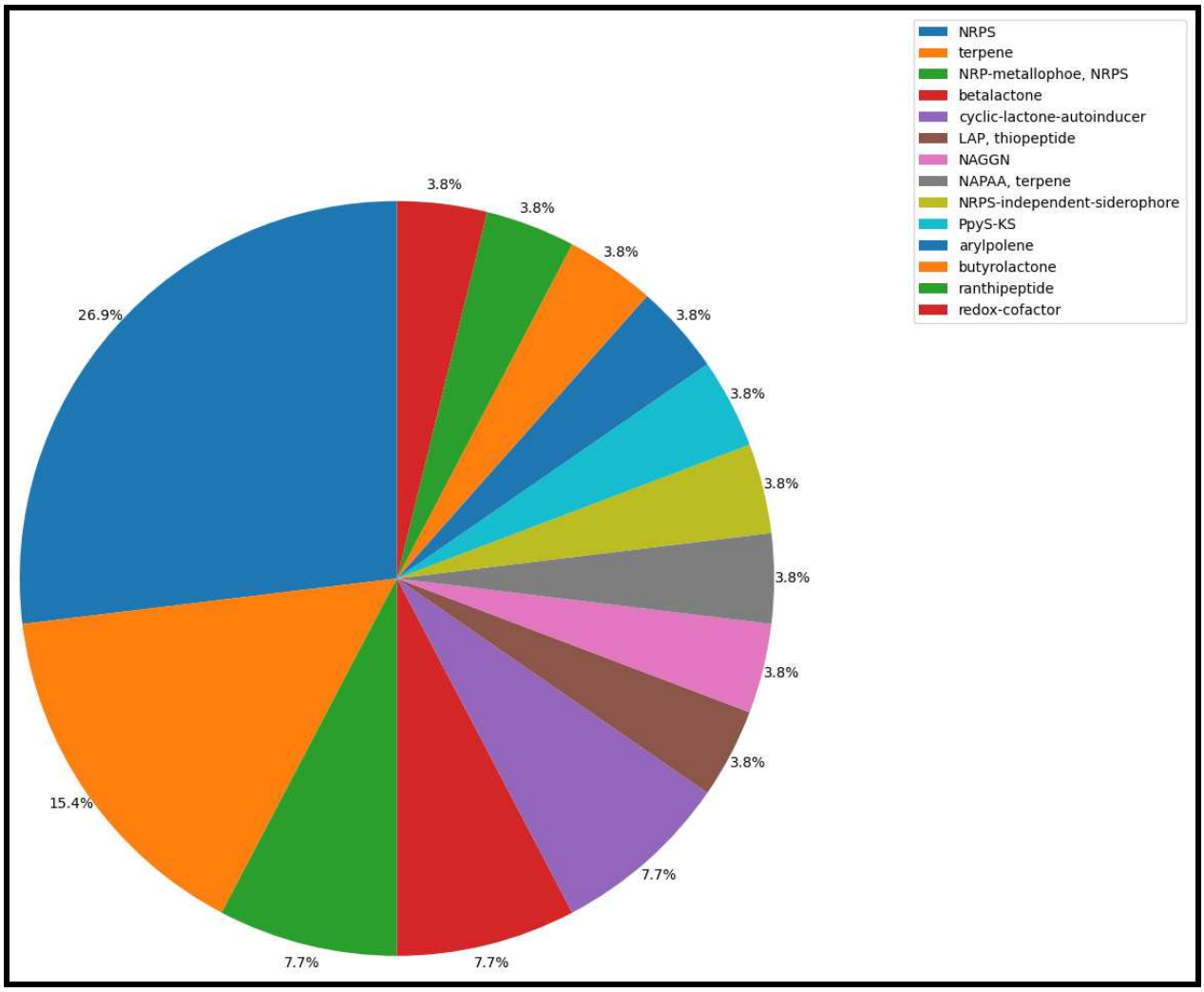


Figure 10. *The proportion of cluster types most abundant among the positive control data types. NRPS cluster types represent the majority of the gene clusters in the positive controls at 34.6%, followed by terpene cluster types at 15.4% and betalactone at 7.7%.*

Negative Controls

None of the 6 bacterial samples chosen for negative control showed any potential for antibiotic activity. The reference genomes for *Streptococcus pneumoniae*, and *Clostridium perfringens* produced results that showed clusters but no predicted metabolites for them (Tables 8 & 9). The *E. coli* reference genome showed 2 clusters producing metabolites but these had no link to antibiotics or any antimicrobial properties (Table 7). The 2 randomly generated genomes and the shuffled DNA sequence as well as the randomly shuffled sequence sample showed "no results" from the antiSMASH Tool (Figure 8).

Table 7. *E. Coli* reference genome: NC_000913

Region	Type	From	To	Most similar known cluster	Similarity
1	NRP-metallophore, NRPS	594,157	649,297	enterobactin	100%
2	thiopeptide	940,340	966,632	O-antigen	14%

Table 8. *Streptococcus pneumoniae* reference genome: NZ_CP020549

Region	Type	From	To	Most similar known cluster	Similarity
1	T3PKS	1,616,501	1,657,673	-	

Table 9. *Clostridium perfringens* reference genome: NZ_CP075979

Region	Type	From	To	Most similar known cluster	Similarity
1	cyclic-lactone-autoinducer	1,088,282	1,099,897	-	
2	cyclic-lactone-autoinducer	2,030,346	2,051,145	-	
3	ranthipeptide	2,440,643	2,462,226	-	

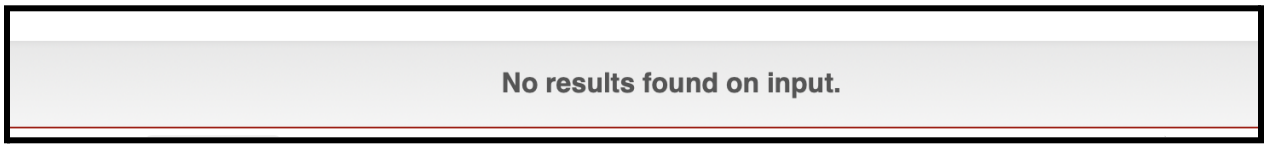


Figure 11. *A screenshot of the antiSMASH output generated by the 2 randomly generated DNA sequences and the shuffled DNA sequence. The tool was unable to identify any clusters or biological relevant data from the sequences.*

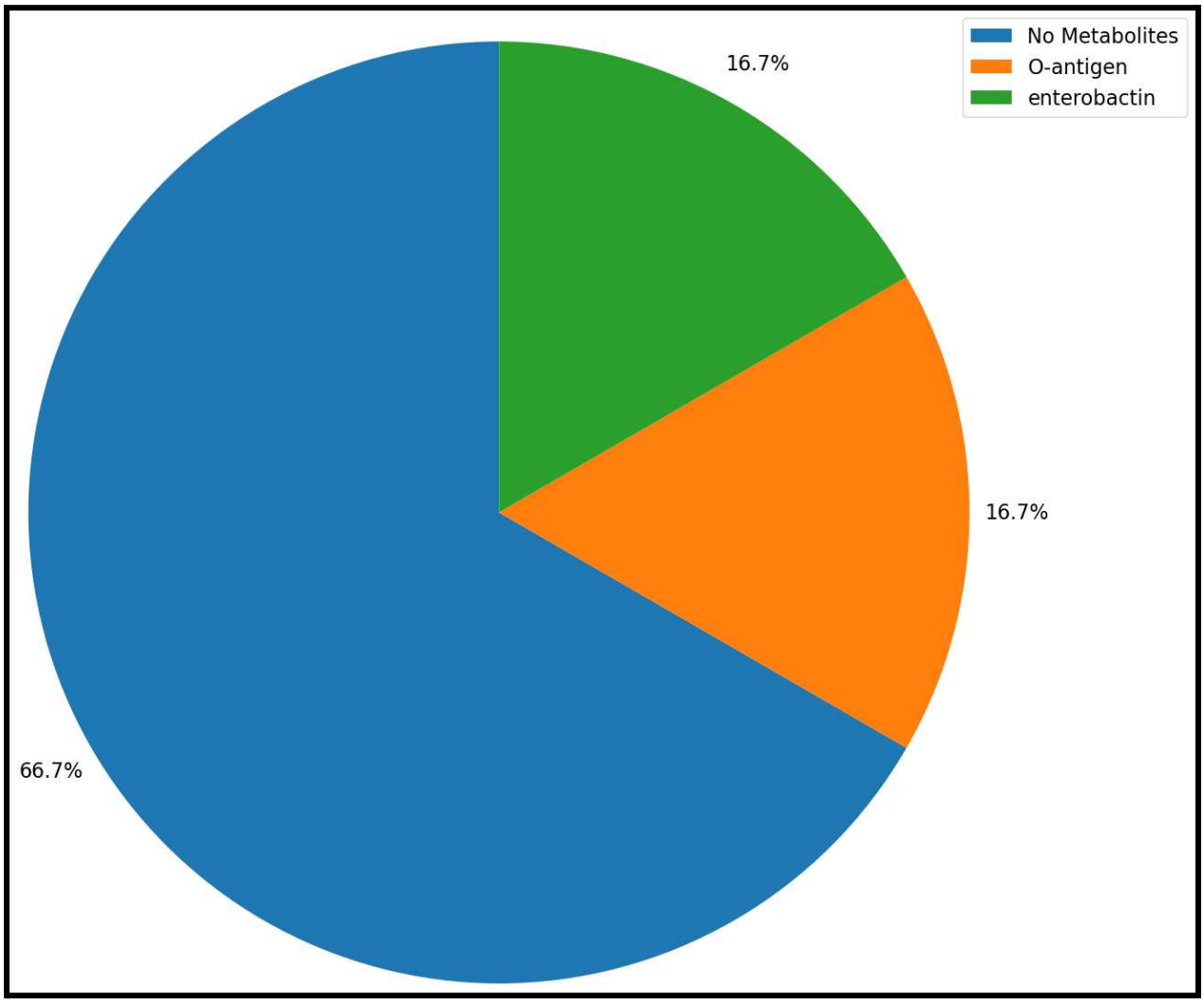


Figure 12. *This Figure shows the proportion of secondary metabolites produced by clusters of the negative controls. The overwhelming majority of clusters produced no metabolites which is to*

be expected. This was followed by O-antigen and enterobactin, neither of which had antibiotic properties.

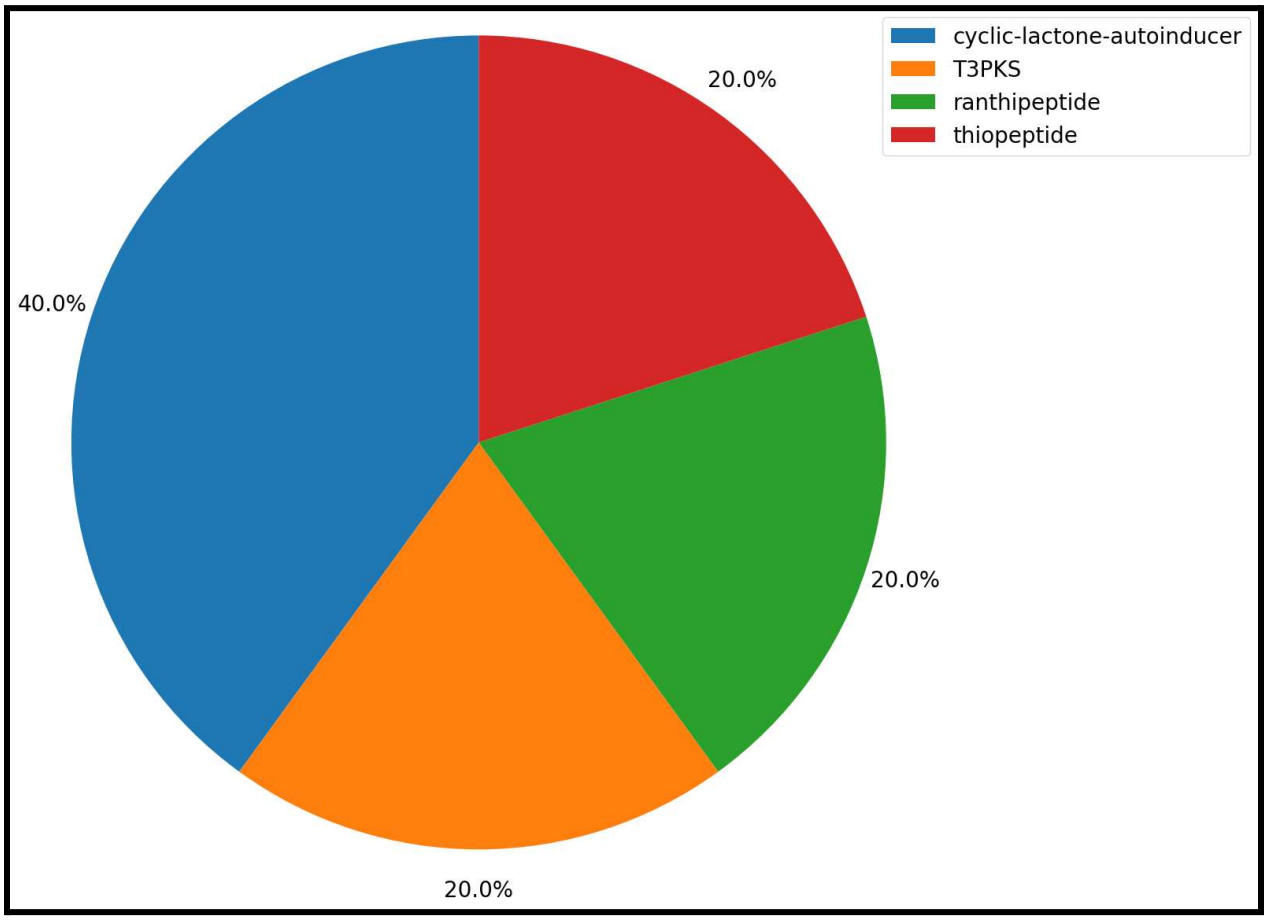


Figure 13. *Summary of cluster types present in negative control. This is a representation of the cluster types detected by antiSMASH in the negative control data type.*

Cyclic-lactone-autoinducer was the most abundant cluster type at 40%, followed by thiopeptide, T3PKS, and ranthipeptide all at 20%.

Known Producers

Among the samples of the known producers various different trends were observed. In particular, analysis and breakdown of the cluster types present in each of these genomes. Figure 9 breaks down the spread of these cluster types. Overall the most dominant cluster types are Terpene, T1PKS related and NRPS related clusters. These cluster types account for 56.2% of the cluster types present. There was a far greater distribution in terms of the secondary metabolites present among the samples. Figure 10 shows the spread and how much more various the number of metabolites predicted were. Steffimycin D (Koyama et al., 2020) (present in *Streptomyces rapamycinicus NRRL 549, Streptomyces rapamycinicus SRMK07, Streptomyces filamentosus NRRL 15998 and Streptomyces atratus SCSIO ZH16*), Notonesomycin A (Sasaki et al., 1986) (present in *Streptomyces rapamycinicus NRRL 5491, Streptomyces rapamycinicus SRMK07, and Streptomyces filamentosus NRRL 15998*), Stenothricin (Liu et al., 2014) (present in *Streptomyces filamentosus NRRL 15998, Streptomyces lincolnensis B48, and Streptomyces atratus SCSIO ZH16*), Istamycin (Slattery et al. 2001) (present in *Streptomyces venezuelae ATCC 15439, Streptomyces griseus IFO 13350, and Streptomyces lincolnensis B48*) and Lankacidin C (Cai et al., 2020) (present in *Streptomyces rapamycinicus NRRL 5491, Streptomyces rapamycinicus SRMK07 and Streptomyces noursei ATCC 11455*) were the most commonly predicted secondary metabolites that could possess antimicrobial properties. In total they accounted for 30.8% of the predicted secondary metabolites. These secondary metabolites along with being the most abundant, also show evidence for antibiotic properties.

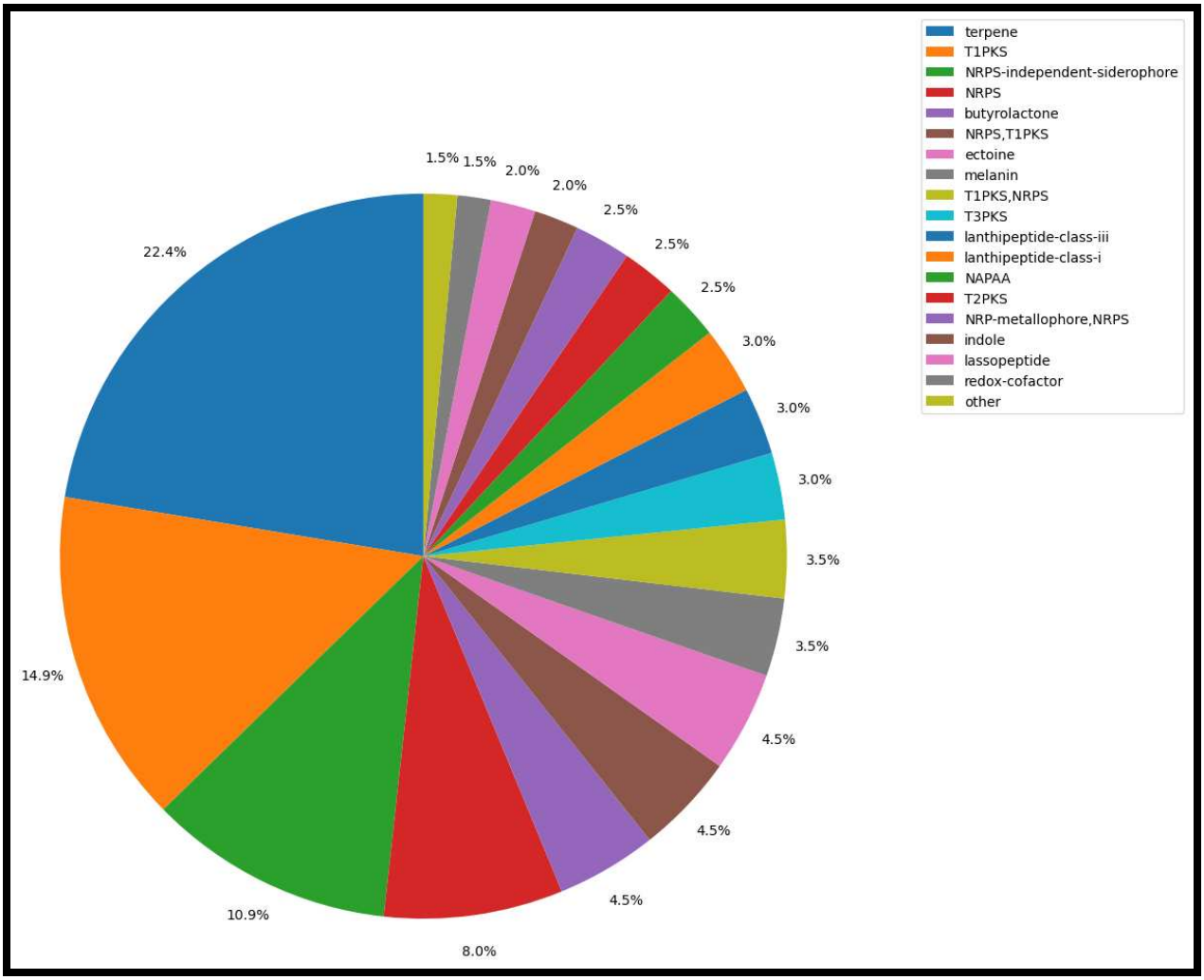


Figure 14. *Cluster types in the known producers. A graphical representation showing the proportion of the various cluster types among the "known producers". The graph was created with a cutoff to only record the frequency of cluster types that appear greater than 2 times in the data type. The largest common occurring cluster type is Terpene at 22.4.% followed by T1PKS, 14.9% and NRPS types at 10.9%.*

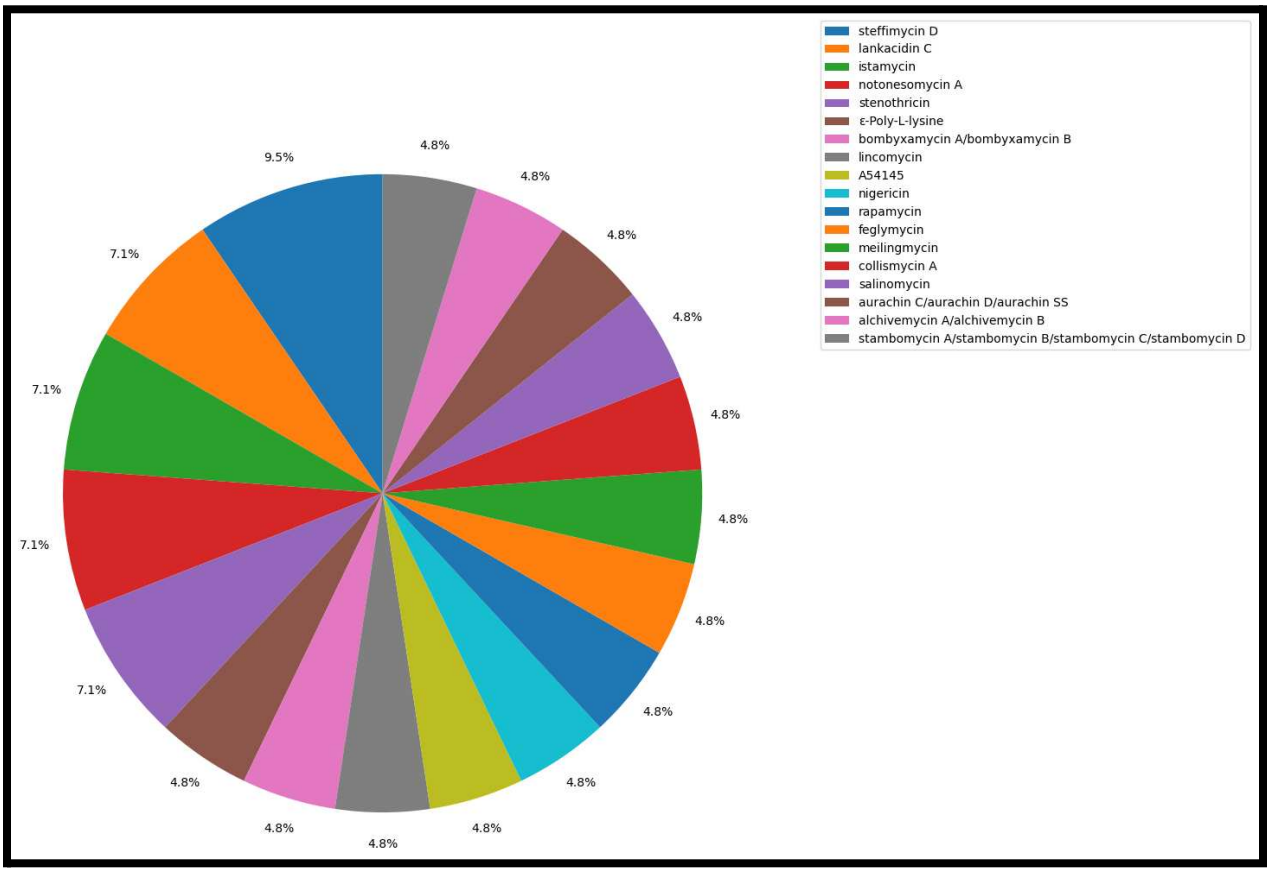


Figure 15. *Secondary metabolites produced by known producers. A representation of the proportion of secondary metabolites found in the known producers. The graph was created with a cutoff to only record the frequency of secondary metabolites that appear greater than once in the data type. Seffimycin D was the most common secondary metabolite of the known producers with 9.5% of clusters producing it. This was followed by Lankacidn C, 7.1%, Istamycin, 7.1% and Notonesomycin A at 7.1% as well.*

Table 10. ***Streptomyces rapamycinicus NRRL 5491***

Region	Type	From	To	Most similar known cluster	Similarity
1	T3PKS,NRPS,betalactone, T1PKS	1	106,876	lobophorin A	36%
2	NRP-metallophore,NRPS	490,714	549,079	coelichelin	100%
3	butyrolactone	859,144	868,463	cyphomycin	9%
4	T1PKS	1,024,465	1,112,648	azalomycin F3a	100%
5	T1PKS	1,298,486	1,437,158	nigericin	66%
6	T1PKS	1,439,806	1,483,779	salinomycin	8%
7	T1PKS	1,585,031	1,663,130	efomycin K/efomycin L	95%
8	redox-cofactor	1,690,280	1,712,398	lankacidin C	13%
9	hserlactone	1,836,607	1,857,362	heronamide A/heronamide B/heronamide C/heronamide D/heronamide E/heronamide F	8%
10	butyrolactone	1,867,930	1,878,862	-	-
11	NRPS,T1PKS	2,083,936	2,136,122	meilingmycin	4%
12	NRPS,T3PKS, other	2,194,053	2,294,639	feglymycin	84%
13	T1PKS, NRPS	2,376,523	2,596,226	thiazostatin/watasemycin A/watasemycin B/2-hydroxyphenylthiazoline enantiopyochelin/isopyochelin	86%
14	lanthipeptide-class-i	2,599,171	2,623,572	steffimycin D	16%
15	lassopeptide	2,763,243	2,785,685	-	-
16	terpene	2,825,088	2,847,488	hopene	76%
17	T2PKS	3,223,797	3,296,312	spore pigment	83%
18	T1PKS	3,440,208	3,501,621	notonesomycin A	10%
19	NRPS-independent-siderophore	3,744,643	3,755,156		
20	NRPS-independent-siderophore	4,931,246	4,942,037	legonoxamine A/desferrioxamine B/legonoxamine B	83%
21	T1PKS,NRPS	5,234,705	5,353,578	alchivemycin A/alchivemycin B	88%
22	terpene	6,279,853	6,301,507	TVA-YJ-2	9%
23	ladderane,arylpolyene,NRPS	6,595,547	6,698,033	kitacinnamycin A/kitacinnamycin B/kitacinnamycin C/kitacinnamycin	54%

				D/kitacinnamycin E/kitacinnamycin F	
24	indole	6,995,321	7,016,463	5-isoprenylindole-3-carboxylate β-D-glycosyl ester	61%
25	NRPS	7,233,050	7,275,286	ochronotic pigment	75%
26	ladderane,arylpolyene	7,325,972	7,367,065	cinnapeptin	50%
27	T1PKS	7,536,753	7,718,885	desulfoclethramycin/clethramycin	69%
28	terpene	8,222,414	8,242,984	-	-
29	ectoine	8,900,244	8,910,648	ectoine	100%
30	NRPS-independent-siderophore,T1PKS	9,081,916	9,139,461	peucechelin	20%
31	terpene	9,152,238	9,168,506	aurachin C/aurachin D/aurachin SS	20%
32	T1PKS	9,476,760	9,606,190	akaeolide	20%
33	T1PKS,nucleoside,NRPS	9,758,976	10,004,255	rapamycin	100%
34	terpene	10,130,161	10,149,631	2-methylisoborneol	100%
35	terpene	10,565,677	10,585,422	pristinol	100%
36	T1PKS,ladderane,arylpolyene	10,679,301	10,730,137	cinnapeptin	46%
37	NRPS,T1PKS	10,976,611	11,092,238	meridamycin	76%
38	T1PKS	11,099,441	11,181,894	hygrocin A/hygrocin B	93%
39	T1PKS,hglE-KS	11,336,115	11,459,137	hexacosalactone A	79%
40	NRPS	11,477,449	11,554,061	cyclofaulknamycin	8%
41	terpene	11,606,908	11,627,280	brasilicardin A	38%
42	terpene	11,765,022	11,785,930	clipibycyclene	4%
43	betalactone	11,969,269	11,998,132	Sch-47554/Sch-47555	7%
44	T1PKS	12,035,215	12,100,906	linearmycin A/linearmycin B/linearmycin C/linearmycin C	37%
45	NRPS	12,222,724	12,274,150	desulfoclethramycin/clethramycin	4%
46	T1PKS	12,289,607	12,332,156	geldanamycin	39%

Table 11. ***Streptomyces rapamycinicus SRMK07***

Region	Type	From	To	Most similar known cluster	Similarity
1	T1PKS	1	35,972	venturicidin D/venturicidin E/venturicidin F/venturicidin A	28%
2	T1PKS	144,136	282,790	nigericin	66%
3	T1PKS	285,438	329,410	salinomycin	8%
4	T1PKS	430,662	508,760	efomycin K/efomycin L	95%
5	redox-cofactor	535,910	558,028	lankacidin C	13%
6	hserlactone	682,236	702,991	heronamide A/heronamide B/heronamide C/heronamide D/heronamide E/heronamide F	8%
7	butyrolactone	713,559	724,491	-	-
8	NRPS,T1PKS	929,567	981,753	meilingmycin	4%
9	NRPS,T3PKS,other	1,039,683	1,137,114	feglymycin	84%
10	T1PKS,NRPS	1,218,996	1,438,690	rifamycin	15%
11	lanthipeptide-class-i	1,441,635	1,466,036	steffimycin D	16%
12	lassopeptide	1,605,708	1,628,150	-	-
13	terpene	1,667,552	1,689,952	hopene	76%
14	T2PKS	2,066,256	2,138,771	spore pigment	83%
15	T1PKS	2,282,668	2,344,081	notonesomycin A	10%
16	NRPS-independent-siderophore	2,587,103	2,597,615	-	-
17	NRPS-independent-siderophore	3,773,728	3,784,519	legonoxamine A/desferrioxamine B/legonoxamine B	83%
18	T1PKS,NRPS	4,077,188	4,178,506	alchivemycin A/alchivemycin B	88%
19	terpene	5,104,801	5,126,455	TVA-YJ-2	9%
20	ladderane,arylpolyene,NRPS	5,420,493	5,522,978	kitacinnamycin A/kitacinnamycin B/kitacinnamycin C/kitacinnamycin D/kitacinnamycin E/kitacinnamycin F	54%
21	indole	5,820,148	5,841,290	5-isoprenylindole-3-carboxylate β-D-glycosyl ester	61%
22	NRPS	6,057,874	6,100,110	ochronotic pigment	75%

23	ladderane,arylpolyene	6,149,892	6,192,275	cinnapeptin	50%
24	T1PKS	6,361,576	6,543,708	desulfoclethramycin/clethramycin	69%
25	terpene	7,047,234	7,067,804	-	-
26	ectoine	7,725,059	7,735,463	ectoine	100%
27	NRPS-independent-siderophore,T1PKS	7,906,733	7,964,277	peucechelin	20%
28	terpene	7,977,054	7,993,322	aurachin C/aurachin D/aurachin SS	20%
29	T1PKS	8,301,577	8,431,007	lydicamycin	28%
30	T1PKS,nucleoside,NRPS	8,583,793	8,830,075	rapamycin	100%
31	terpene	8,954,977	8,974,447	2-methylisoborneol	100%
32	terpene	9,390,245	9,411,345	pristinol	100%

Table 12. ***Streptomyces noursei ATCC 11455***

Region	Type	From	To	Most similar known cluster	Similarity
1	T3PKS,lassopeptide	86,199	139,067	A54145	5%
2	terpene	223,236	243,282	legonindolizidine A6	12%
3	T1PKS	350,589	409,828	simocyclinone D8	8%
4	terpene	584,233	603,773	-	-
5	terpene	688,481	708,120	bombyxamycin A/bombyxamycin B	3%
6	T1PKS	769,835	911,487	nystatin A1	100%
7	terpene	1,065,461	1,085,454	-	-
8	redox-cofactor	1,211,303	1,233,379	lankacidin C	20%
9	butyrolactone	1,244,054	1,254,532	merochlorin A/merochlorin B/deschloro-merochlorin A/deschloro-merochlorin B/isochloro-merochlorin B/dichloro-merochlorin B/merochlorin D/merochlorin C	4%
10	butyrolactone,NAPAA	1,488,958	1,524,741	galtamycin C/galtamycin D	16%
11	T1PKS	1,644,557	1,690,956	actinomycin D	10%
12	terpene	1,958,530	1,984,265	hopene	61%

13	lassopeptide	2,157,118	2,178,457	echoside A/echoside B/echoside C/echoside D/echoside E	11%
14	T1PKS	2,214,299	2,292,209	lankamycin	53%
15	T3PKS	2,743,304	2,782,651	naringenin	100%
16	betalactone,NRPS	2,853,457	2,914,351	ulleungmycin	80%
17	NRPS-independent-siderophore	2,922,576	2,935,753	synechobactin C9/synechobactin C11/synechobactin 13/synechobactin 14/synechobactin 16/synechobactin A/synechobactin B/synechobactin C	9%
18	CDPS	3,258,238	3,278,957	albonoursin	83%
19	NAPAA	3,889,567	3,923,700	-	-
20	linaridin	4,361,283	4,382,173	legonaridin	66%
21	thiopeptide	4,401,731	4,438,354	radamycin/globimycin	94%
22	terpene	4,524,960	4,544,791	geosmin	100%
23	lanthipeptide-class-i	6,032,920	6,057,223	neomediomycin B	7%
24	T1PKS	6,511,872	6,553,402	collismycin A	18%
25	ectoine	6,984,258	6,994,662	ectoine	100%
26	NRPS-independent-siderophore	7,075,833	7,087,638	desferrioxamine E	100%
27	lanthipeptide-class-i	7,331,458	7,356,001	-	
28	T2PKS	7,849,563	7,922,078	spore pigment	83%
29	transAT-PKS	8,267,262	8,329,672	cycloheximide	50%
30	NRPS	8,386,871	8,446,854	WS9326B/WS9326A/WS9326G/WS9326F	7%
31	terpene	8,911,916	8,931,024	neoabyssomicin/abyssomicin	6%
32	T1PKS,NRPS,T3PKS,terpene	9,018,969	9,100,074	canucin A/canucin B	42%
33	NRPS,lanthipeptide-class-iii	9,140,586	9,194,236	s56-p1	17%
34	T1PKS,terpene	9,247,044	9,361,469	tetramycin B	95%
35	lassopeptide,T3PKS	9,676,601	9,729,936	A54145	5%

Table 13. ***Streptomyces venezuelae ATCC 15439***

Region	Type	From	To	Most similar known cluster	Similarity
1	terpene	261,244	280,090	-	-
2	betalactone,NAPAA	674,746	710,001	ε-Poly-L-lysine	100%
3	T1PKS,NRPS	974,021	1,090,430	camporidine A/camporidine B	65%
4	lanthipeptide-class-iv,T1PKS ,NRPS	1,099,741	1,159,021	oxalomycin B	18%
5	lanthipeptide-class-ii,melani n	1,166,022	1,188,871	melanin	28%
6	lanthipeptide-class-iii	1,467,557	1,490,088	-	-
7	terpene	1,806,765	1,832,645	hopene	76%
8	T1PKS	2,028,496	2,072,836	formicamycins A-M	18%
9	butyrolactone	2,192,142	2,203,092	griseoviridin/fijimycin A	8%
10	NRPS-independent-sideroph ore	2,543,752	2,556,327	synechobactin C9/synechobactin C11/synechobactin 13/synechobactin 14/synechobactin 16/synechobactin A/synechobactin B/synechobactin C	9%
11	other	2,927,421	2,968,527	stambomycin A/stambomycin B/stambomycin C/stambomycin D	16%
12	melanin	3,385,785	3,396,198	istamycin	8%
13	T1PKS	3,419,066	3,491,507	mycinamicin II	61%
14	cyanobactin	4,105,357	4,127,723	-	-
15	NRPS	4,712,038	4,763,391	foxicin A/foxicin B/foxicin C/foxicin	7%
16	thiopeptide,LAP	5,202,490	5,232,532	-	-
17	NRPS-independent-sideroph ore	5,737,842	5,749,626	desferrioxamin B	100%
18	lanthipeptide-class-i	6,173,061	6,198,323	-	-
19	T2PKS,T1PKS	6,409,069	6,509,238	LL-D49194α1 (LLD)	18%

20	CDPS	6,585,513	6,606,226	malacidin A/malacidin B	5%
21	indole	7,770,196	7,793,435	rebeccamycin	25%
22	T3PKS	7,864,814	7,904,164	alkylresorcinol	100%
23	lanthipeptide-class-iv,NRPS	7,910,366	7,951,627	venezuelin	100%
24	NRP-metallophore,NRPS	7,968,001	8,025,100	peucechelin	55%
25	T3PKS,terpene	8,246,101	8,298,735	geosmin	100%
26	T2PKS,terpene	8,320,823	8,393,323	spore pigment	83%
27	ectoine	8,414,418	8,424,837	ectoine	100%
28	terpene	8,476,458	8,497,729	ebelactone	5%

Table 14. ***Streptomyces filamentosus NRRL 15998***

Region	Type	From	To	Most similar known cluster	Similarity
1	NRPS	313,359	402,438	daptomycin	100%
2	nucleoside,T1PKS	542,942	609,136	ansaseomycin A/ansaseomycin B	69%
3	terpene	750,819	769,736	steffimycin D	19%
4	NRPS	874,699	942,614	stenothricin	86%
5	ectoine	1,254,796	1,265,194	ectoine	100%
6	lanthipeptide-class-ii,lanthipeptide-class-iii	2,311,100	2,341,376	notonesomycin A	3%
7	NRPS-independent-siderophore	2,418,590	2,428,540	desferrioxamin B	100%
8	other,betalactone	3,538,559	3,579,389	FR-900098	90%
9	NRPS-independent-siderophore	4,245,409	4,253,429	desferrioxamin B	60%
10	lassopeptide	4,281,297	4,302,977	keywimysin	100%
11	lanthipeptide-class-ii	4,435,763	4,459,002	SRO15-3108	100%
12	lanthipeptide-class-iii	5,211,350	5,233,099	SRO15-2212	71%
13	terpene	5,594,782	5,613,354	-	
14	thioamide-NRP	5,623,838	5,683,186	BD-12	17%
15	NRPS-independent-siderophore	6,039,454	6,054,198	schizokinen	20%

16	T1PKS,transAT-PKS,oligosaccharide	6,090,668	6,194,339	auroramycin	79%
17	NRPS	6,206,856	6,270,735	arylomycin	100%
18	NRPS,T1PKS	6,379,148	6,428,716	collismycin A	70%
19	terpene	7,116,845	7,143,418	hopene	69%
20	NRPS,T1PKS	7,223,347	7,271,547	SGR PTMs/SGR PTM Compound b/SGR PTM Compound c/SGR PTM Compound d	100%
21	T1PKS,NRPS	7,436,778	7,485,679	valinomycin/montanastatin	13%
22	melanin	7,520,011	7,530,493	melanin	100%
23	T3PKS	7,565,513	7,606,652	tetronasin	11%
24	NRPS,T1PKS	7,691,684	7,743,872	balhimycin	8%

Table 15. ***Streptomyces griseus IFO 13350***

Region	***Type***	***From***	***To***	***Most similar known cluster***	***Similarity***
1	terpene	48,328	70,469	isorenieratene	100%
2	lanthipeptide-class-iv	149,896	172,562	-	-
3	T1PKS,NRPS	274,592	326,653	A-47934	8%
4	NRPS,T3PKS	464,992	570,745	CDA1b/CDA2a/CDA2b/CDA3a/CDA3b/CDA4a/CDA4b	7%
5	melanin	604,068	611,793	istamycin	4%
6	NRPS,LAP,T1PKS,other	641,140	792,830	C-1027	36%
7	NRPS,T1PKS	937,762	985,289	SGR PTMs/SGR PTM Compound b/SGR PTM Compound c/SGR PTM Compound d	100%
8	NRPS	1,040,889	1,086,995	nucleocidin	47%
9	terpene	1,127,348	1,153,611	hopene	69%
10	terpene	1,482,046	1,501,955	2-methylisoborneol	100%
11	linaridin	1,768,290	1,786,697	pentostatine/vidarabine	9%
12	NRPS-independent-siderophore	2,026,431	2,039,870	schizokinen	20%
13	terpene	2,452,552	2,470,816	-	-

14	lanthipeptide-class-iii	2,817,598	2,838,618	AmfS	100%
15	melanin	2,880,853	2,889,385	melanin	100%
16	T1PKS	2,908,041	2,950,375	leinamycin	5%
17	NRPS	3,029,854	3,093,422	atratumycin	13%
18	arylpolyene,ladderane,NRPS	3,756,885	3,841,258	kitacinnamycin A/kitacinnamycin B/kitacinnamycin C/kitacinnamycin D/kitacinnamycin E/kitacinnamycin F	47%
19	betalactone	4,064,789	4,092,758	divergolide A/divergolide B/divergolide C/divergolide D	6%
20	lanthipeptide-class-i	4,495,428	4,520,684	-	-
21	melanin	4,960,832	4,971,221	grixazone A	61%
22	thiopeptide,LAP	5,142,038	5,174,689	-	-
23	NRPS-independent-siderophore	5,573,822	5,585,600	desferrioxamin B	100%
24	lanthipeptide-class-iii,lanthipeptide-class-ii	5,637,061	5,668,604	-	-
25	ectoine	6,603,626	6,614,024	ectoine	100%
26	amglyccycl	6,935,101	6,958,692	streptomycin	55%
27	terpene,T1PKS	7,080,060	7,191,361	griseochelin	100%
28	T1PKS	7,271,755	7,372,510	stambomycin A/stambomycin B/stambomycin C/stambomycin D	40%
29	linaridin,T1PKS	7,562,138	7,638,703	iminimycin A/iminimycin B	100%
30	T3PKS	7,888,518	7,929,636	naringenin	100%
31	NRP-metallophore,NRPS,transAT-PKS,T1PKS	7,988,318	8,173,814	griseobactin	100%
32	terpene	8,196,619	8,236,630	carotenoid	100%
33	butyrolactone	8,268,444	8,279,388	coelimycin P1	8%
34	terpene	8,473,224	8,498,819	isorenieratene	100%

Table 16. ***Streptomyces lincolnensis B48***

Region	Type	From	To	Most similar known cluster	Similarity
1	lanthipeptide-class-iii	116,680	138,939	A-500359 A/A-500359 B	8%
2	other	295,876	334,808	lincomycin	72%
3	melanin,terpene	549,413	569,762	melanin	57%
4	NRPS	831,827	889,060	cysteoamide	100%
5	NRP-metallophore,NRPS	1,552,138	1,607,148	coelibactin	72%
6	T3PKS	1,614,875	1,655,972	flaviolin/1,3,6,8-tetrahydroxynaphthalene	100%
7	terpene	2,152,319	2,171,287	-	-
8	NAPAA	2,236,592	2,269,281	ε-Poly-L-lysine	100%
9	ectoine	2,510,170	2,520,574	ectoine	100%
10	butyrolactone	3,074,186	3,084,076	A-factor	100%
11	melanin	3,675,352	3,685,828	istamycin	5%
12	NRPS-independent-siderophore	3,783,837	3,794,804	desferrioxamin B/desferrioxamine E	83%
13	amglyccycl,butyrolactone	3,977,664	4,004,708	pyralomicin 1a	18%
14	ladderane	4,764,701	4,805,870	colabomycin E	13%
15	terpene	6,642,058	6,662,143	albaflavenone	100%
16	lanthipeptide-class-v	6,869,451	6,911,522	pristinin A3	47%
17	NRPS,T1PKS	6,949,121	7,009,823	BD-12	17%
18	NRPS-independent-siderophore	7,383,320	7,394,202	-	-
19	phosphoglycolipid	7,541,735	7,570,490	teichomycin A1	83%
20	butyrolactone,terpene	7,814,584	7,839,498	γ-butyrolactone	100%
21	NAPAA	7,874,031	7,909,906	stenothricin	13%
22	NRPS-independent-siderophore	8,074,348	8,087,135	-	-
23	T2PKS	8,206,886	8,279,347	julichrome Q3-3/julichrome Q3-5	35%
24	NRP-metallophore,NRPS	8,364,323	8,422,847	coelichelin	100%

25	terpene	8,595,995	8,620,257	hopene	92%
26	T1PKS	8,658,353	8,706,443	desulfoclethramycin/clethramycin	30%
27	terpene	9,041,364	9,063,235	mycotrienin I	11%
28	lipolanthine,lanthipeptide-class-iii	9,184,524	9,227,192	informatipeptin	42%
29	NRPS-independent-siderophore	9,251,035	9,265,733	peucechelin	25%
30	other	9,673,768	9,714,337	lincomycin	72%
31	lanthipeptide-class-iii	9,869,728	9,892,415	A-500359 A/A-500359 B	8%

Table 17. ***Streptomyces ambofaciens ATCC 23877***

Region	Type	From	To	Most similar known cluster	Similarity
1	T2PKS,butyrolactone	127,762	213,859	fluostatins M-Q	67%
2	terpene,T1PKS,NRPS	352,107	409,148	antimycin	100%
3	indole	706,084	727,211	5-dimethylallylindole-3-acetonitrile	77%
4	terpene	784,459	815,058	isorenieratene	37%
5	ectoine	2,000,972	2,011,370	ectoine	100%
6	melanin	2,873,096	2,883,623	melanin	80%
7	NRPS-independent-siderophore	2,980,926	2,991,871	desferrioxamin B/desferrioxamine E	83%
8	butyrolactone	3,469,415	3,478,485	prejadomycin/rabelomycin/gaudimycin C/gaudimycin D/UWM6/gaudimycin A	4%
9	furan	4,172,601	4,193,614	methylenomycin A	9%
10	terpene	5,303,655	5,323,095	albaflavenone	100%
11	T2PKS	5,357,993	5,430,181	spore pigment	66%
12	NRPS-independent-siderophore	5,867,788	5,878,156	-	-
13	oligosaccharide,T1PKS,NRPS	5,967,339	6,141,725	spiramycin	100%
14	terpene	6,259,807	6,280,958	geosmin	100%
15	NRPS-independent-siderophore	6,452,952	6,466,082	grincamycin	8%
16	lanthipeptide-class-iii	6,817,278	6,839,887	SapB	100%

17	terpene	6,901,409	6,927,295	hopene	92%
18	NRPS	7,197,949	7,240,502	netropsin	86%
19	terpene	7,277,164	7,296,598	-	-
20	NRP-metallophore,NRPS	7,557,873	7,616,302	coelichelin	100%
21	terpene	7,630,675	7,650,381	isorenieratene	62%
22	T1PKS	7,666,256	7,834,600	stambomycin A/stambomycin B/stambomycin C/stambomycin D	100%
23	lanthipeptide-class-i,terpene	7,931,838	7,968,985	-	-
24	T1PKS	8,007,437	8,054,819	alanylclavam/2-hydroxymethylclavam/2-formyloxymethylclavam/clavam-2-carboxylate	18%
25	T2PKS,butyrolactone	8,089,907	8,176,435	fluostatins M-Q	72%

Table 18. ***Streptomyces atratus SCSIO ZH16***

Region	***Type***	***From***	***To***	***Most similar known cluster***	***Similarity***
1	NRPS	170,926	288,305	enduracidin	14%
2	NRPS-like,NRPS,PKS-like	376,516	460,291	bonnevillamide D/bonnevillamide E	16%
3	T2PKS,terpene	690,553	763,068	spore pigment	83%
4	NAPAA	961,970	996,576	stenothricin	13%
5	phosphonate	1,070,024	1,085,463	-	-
6	T1PKS,NRPS-like,NRPS,indole,prodigiosin	1,133,399	1,261,249	ilamycin B1/ilamycin B2/ilamycin C1/ilamycin C2/ilamycin D/ilamycin E1	90%
7	terpene	1,750,130	1,770,154	steffimycin D	19%
8	ectoine	2,266,355	2,276,753	ectoine	100%
9	butyrolactone	2,769,017	2,780,069	coelimycin P1	8%
10	other,T1PKS	3,223,312	3,278,431	desulfoclethramycin/clethramycin	12%
11	NRPS-independent-siderophore	3,339,130	3,348,929	desferrioxamin B	100%
12	T3PKS	4,180,353	4,221,408	vazabitide A	17%
13	ladderane,arylpolyene,NRPS,aminocoumarin	5,612,335	5,700,782	atratumycin	89%

14	NRPS-independent-siderophore	6,722,359	6,737,076	synechobactin C9/synechobactin C11/synechobactin 13/synechobactin 14/synechobactin 16/synechobactin A/synechobactin B/synechobactin C	9%
15	terpene	7,866,656	7,890,983	hopene	84%
16	RiPP-like	7,970,276	7,981,174	griselimycin	7%
17	NAPAA	8,018,199	8,052,332	-	-
18	RiPP-like	8,079,112	8,089,909	tautomycin	6%
19	NRPS,T1PKS	8,247,955	8,366,892	montamide A/capsimycin/clifednamide A/frontalamide A/combamide A	55%
20	T1PKS,NRPS,terpene,hglE-KS	8,423,092	8,627,676	bombyxamycin A/bombyxamycin B	51%
21	terpene,NRPS-like	8,641,283	8,690,049	raimonol	90%
22	NRPS	8,772,080	8,834,071	bonnevillamide D/bonnevillamide E	13%
23	NRPS,lanthipeptide-class-ii	8,942,762	9,015,218	omnipeptin	9%
24	T1PKS	9,248,753	9,295,067	sporolide A/sporolide B	8%
25	NRPS	9,352,426	9,470,470	cadaside A/cadaside B	28%

Unknown Samples

9 of 10 the randomly selected bacterial genomes were predicted to produce secondary metabolites that had antibiotic properties or potential as indicated by the literature. The bacteria sample known as *Ignavibacterium album JCM 16511* (Table 19) was the odd one out. In this genome antiSMASH only detected 3 clusters none of which 2 of which no product was predicted and the other cluster produced a compound that doesn't appear to have any evident antibiotic properties. A large majority of the clusters, in fact the largest category predicted in these

unknown producers, were clusters that produced no metabolites. They represent approximately 48.8% of all the predicted products in the unknown producers. This is shown in Figure 16. Other represented secondary metabolites predicted in this set of genomes are Ectoine at 6.2%, Hopene at 6.2%, ε-Poly-L-lysine at 5%, and Pf-5 pyoverdine at 5%. ε-Poly-L-lysine is a poly amino acid that has shown antimicrobial activity as reported by Wang et al., 2021 and is a predicted metabolite in *Streptomyces hawaiiensis NRRL 15010, Streptomyces sp. S1D4-23, and Streptomyces lydicus WYEC 108.* The largest cluster type in the unknown producers was terpene at 24.6% as shown in Figure 17. Like with the known antibiotic producing data set the other well represented cluster types were NRPS and T1PKS related. They accounted for 20% of the cluster types in unknown genome samples (Figure 17).

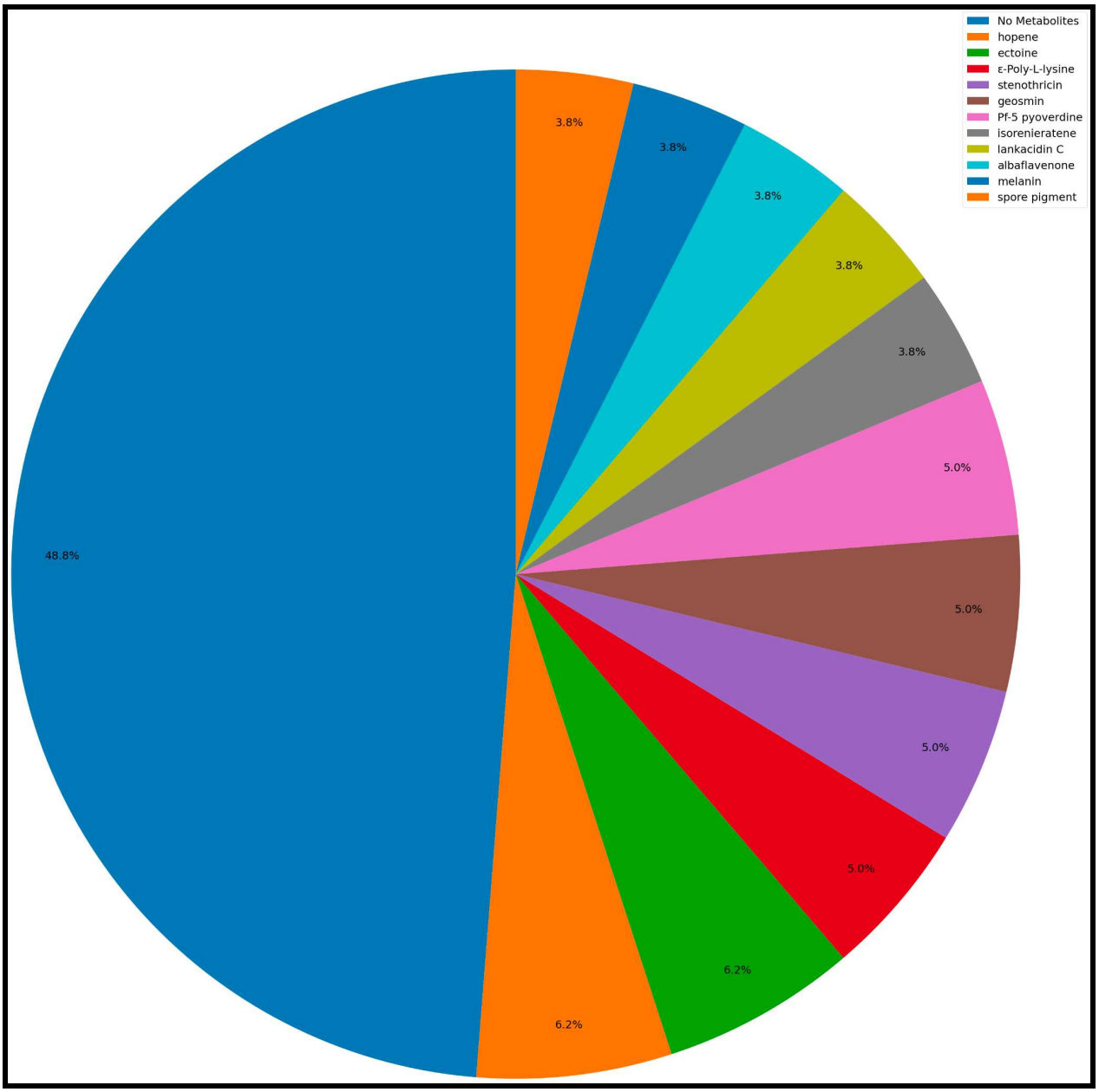


Figure 16. *Secondary metabolites produced by unknown producers. The proportion of the most common metabolites found throughout the unknown producers dataset. The graph was created with a cutoff to only record the frequency of secondary metabolites that appear greater than 2 times in the data type. Most of the clusters predicted "No metabolites" which was 48.8% of them. Terpene was the next most commonly predicted metabolite at 6.2% was ectoine, followed*

by hopene at 2.8% as well. The most common metabolite predicted with antibiotic potential was ε-Poly-L-lysine from 5.0% of clusters.

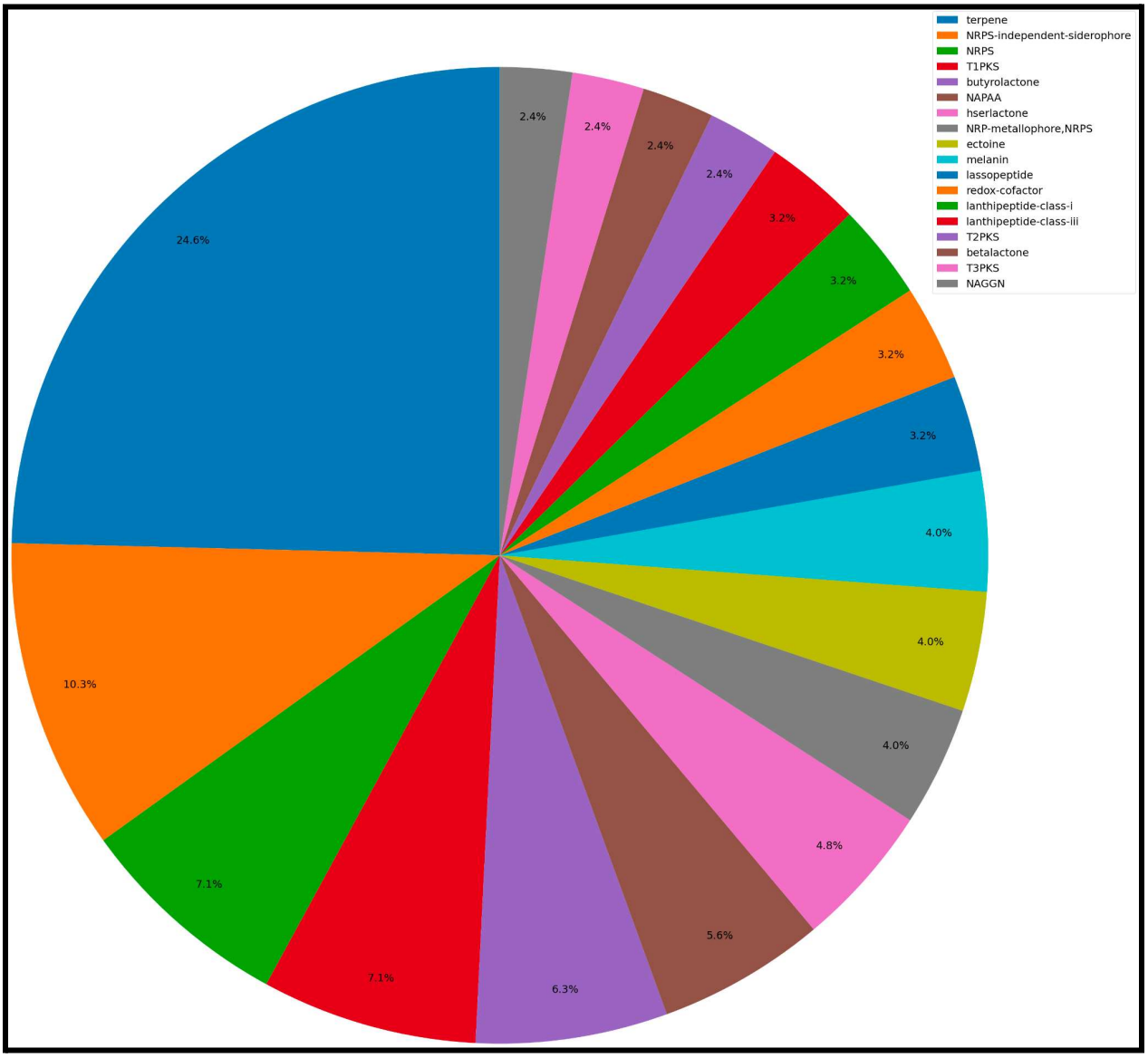


Figure 17. *Cluster types in the unknown producers. Representation of the proportion of the various cluster types present among the unknown producers. The graph was created with a cutoff to only record the frequency of cluster types that appear greater than 2 times in the data type. The largest cluster type was Terpene followed by NRPS and T1PKS cluster types.*

Table 19. Ignavibacterium album JCM 16511

Region	Type	From	To	Most similar known cluster	Similarity
1	terpene	427,519	445,374	carotenoid	28%
2	terpene	2,194,131	2,215,821	-	
3	T1PKS,NRPS	2,812,977	2,860,950	-	

Table 20. ***Streptomyces sp. CLI2509***

Region	Type	From	To	Most similar known cluster	Similarity
1	terpene	52,576	75,394	isorenieratene	85%
2	melanin	152,162	162,575	istamycin	4%
3	NRP-metallophore,NRPS	318,360	378,300	coelibactin	100%
4	NRP-metallophore,NRPS	487,013	535,712	griseobactin	53%
5	hglE-KS,T1PKS	745,185	798,303	ambactin	25%
6	terpene	854,730	880,507	hopene	76%
7	NRPS,T3PKS	1,204,878	1,269,039	A50926 A/A50926 B	16%
8	NRPS-independent-siderophore	1,703,324	1,718,806	-	-
9	terpene	2,274,493	2,293,228	julichrome Q3-3/julichrome Q3-5	21%
10	NRPS,T1PKS	2,652,862	2,701,727	SGR PTMs/SGR PTM Compound b/SGR PTM Compound c/SGR PTM Compound d	100%
11	butyrolactone	2,830,823	2,841,827	-	
12	terpene	3,077,282	3,099,762	geosmin	100%
13	T1PKS,NRPS	4,729,166	4,802,657	sporolide A/sporolide B	51%
14	ladderane,NRPS	5,343,740	5,402,778	colibrimycin	21%
15	ectoine	5,952,685	5,963,083	ectoine	100%
16	lanthipeptide-class-iii	6,567,919	6,590,612	AmfS	100%

Table 21. ***Streptomyces sp. WAC 01438***

Region	Type	From	To	Most similar known cluster	Similarity
1	NRPS	38,800	92,020	-	
2	T3PKS,terpene,NRP-metallophore,NRPS	633,376	708,172	scabichelin	100%
3	NRPS,terpene	773,920	832,223	hopene	92%
4	other,T3PKS,NRPS-independent-siderophore	1,405,124	1,447,543	diazepinomicin	55%
5	terpene	1,589,879	1,610,470	geosmin	100%
6	NRP-metallophore,NRPS	1,915,130	1,978,569	coelibactin	100%
7	NRPS-independent-siderophore	2,074,386	2,085,388	-	
8	terpene	2,557,359	2,578,219	atolypene A/atolypene B	29%
9	terpene	2,645,553	2,666,190	albaflavenone	100%
10	T1PKS,prodigiosin	3,636,105	3,677,686	marineosin A/marineosin B	95%
11	butyrolactone	4,034,784	4,045,773	coelimycin P1	8%
12	lanthipeptide-class-v	5,163,800	5,206,066	-	
13	NRPS-independent-siderophore	5,254,731	5,266,503	FW0622	62%
14	T1PKS	5,365,010	5,454,165	neomediomycin B	53%
15	ectoine	6,412,946	6,423,344	ectoine	100%
16	T1PKS	7,187,752	7,284,681	bafilomycin B1	72%
17	T3PKS,NRPS,T2PKS	7,342,610	7,524,993	spore pigment	83%
18	melanin	7,583,908	7,594,279	melanin	42%
19	terpene	7,722,638	7,744,853	carotenoid	54%
20	terpene	7,806,444	7,827,415	cyslabdan	81%

Table 22. ***Pseudomonas aeruginosa F5677***

Region	Type	From	To	Most similar known cluster	Similarity
1	NRP-metallophore,NRPS,phenazine	757,035	817,928	pyochelin	85%
2	hserlactone	1,668,715	1,688,659	-	
3	NAGGN	1,692,327	1,707,087	-	

4	NRPS	1,857,415	1,904,246	azetidomonamide A/azetidomonamide B	100%
5	NRPS,NRP-metallophore	2,930,753	3,050,943	Pf-5 pyoverdine	25%
6	redox-cofactor	3,511,701	3,533,845	lankacidin C	13%
7	hserlactone	4,221,613	4,242,218	-	
8	betalactone	4,354,966	4,383,167	thanamycin	27%
9	opine-like-metallophore	5,728,120	5,750,209	pseudopaline	100%

Table 23.***Streptomyces hawaiiensis NRRL 15010***

Region	Type	From	To	Most similar known cluster	Similarity
1	NRPS,terpene	257,697	303,726	polyoxypeptin	10%
2	LAP	334,786	357,328	streptamidine	91%
3	NAPAA	396,042	429,217	ε-Poly-L-lysine	100%
4	melanin	591,712	601,681	melanin	57%
5	T1PKS	1,234,260	1,281,301	hedamycin	9%
6	terpene	1,283,753	1,303,874	A23187	10%
7	T3PKS	1,350,817	1,390,663	flaviolin/1,3,6,8-tetrahydroxynaphthalene	100%
8	ectoine	2,097,817	2,108,215	ectoine	100%
9	terpene	2,828,484	2,847,948	streptozotocin	23%
10	NAPAA	3,057,802	3,092,795	stenothricin	13%
11	melanin	3,187,244	3,197,741	istamycin	4%
12	NRPS-independent-siderophore	3,287,256	3,298,017	desferrioxamin B/desferrioxamine E	83%
13	phenazine	4,060,437	4,080,928	endophenazine A/endophenazine B	44%
14	butyrolactone	4,441,748	4,452,782	lactonamycin	5%
15	NRPS	4,731,577	4,775,566	-	
16	T2PKS	4,809,151	4,880,065	vazabitide A	26%
17	T2PKS	4,989,918	5,062,349	gilvocarcin V	88%
18	terpene	5,648,756	5,672,930	isorenieratene	100%
19	terpene	5,819,551	5,839,612	albaflavenone	100%

20	NRPS-independent-siderophore	6,482,226	6,493,148	-	
21	terpene	6,824,981	6,846,944	geosmin	100%
22	NRPS-independent-siderophore	7,080,485	7,093,618	-	
23	ladderane,NRPS	7,148,027	7,211,204	acyldepsipeptide 1	73%
24	lanthipeptide-class-i	7,263,664	7,288,770	-	
25	terpene	7,700,473	7,725,131	hopene	92%
26	NRP-metallophore,NRPS	7,810,531	7,868,857	coelichelin	100%
27	T1PKS	8,068,147	8,113,229	thiotetroamide	58%
28	T1PKS	8,118,850	8,166,394	-	
29	lanthipeptide-class-iii	8,277,872	8,300,520	informatipeptin	85%
30	T3PKS	8,535,316	8,576,503	clipibycyclene	13%
31	terpene	8,595,930	8,617,258	ebelactone	5%

Table 24. ***Pseudomonas aeruginosa DSM 50071***

Region	***Type***	***From***	***To***	***Most similar known cluster***	***Similarity***
1	NRP-metallophore,NRPS,phenazine	725,806	786,485	pyochelin	92%
2	hserlactone	1,588,426	1,608,370	-	
3	NAGGN	1,611,924	1,626,684	-	
4	NRPS	1,736,397	1,783,244	azetidomonamide A/azetidomonamide B	100%
5	NRPS,NRP-metallophore	2,738,840	2,870,816	Pf-5 pyoverdine	24%
6	NRPS	2,922,350	2,974,630	L-2-amino-4-methoxy-trans-3-butenoic acid	100%
7	redox-cofactor	3,326,364	3,348,508	lankacidin C	13%
8	thiopeptide	3,393,173	3,426,176	oxalomycin B	6%
9	phenazine	3,428,356	3,449,368	pyocyanine	100%
10	hserlactone	3,941,080	3,961,691	-	
11	betalactone	4,178,885	4,206,730	thanamycin	27%
12	opine-like-metallophore	5,462,538	5,484,627	pseudopaline	100%

Table 25. ***Streptomyces sp. S1D4-23***

Region	Type	From	To	Most similar known cluster	Similarity
1	terpene	35,035	53,906	BE-14106	10%
2	butyrolactone	844,406	854,354	-	
3	NRPS,betalactone	865,791	909,233	diisonitrile antibiotic SF2768	66%
4	lanthipeptide-class-ii,terpene	1,188,614	1,222,223	2-methylisoborneol	100%
5	NAPAA	1,324,600	1,358,505	stenothricin	13%
6	NRPS,nucleoside	1,711,450	1,763,634	detoxin P1/detoxin P2/detoxin P3	75%
7	guanidinotides	1,825,247	1,848,778	-	
8	melanin	1,955,172	1,965,564	melanin	57%
9	NAPAA	2,004,693	2,036,244	ε-Poly-L-lysine	100%
10	aminopolycarboxylic-acid	2,175,856	2,189,368	EDHA	88%
11	T1PKS	2,407,612	2,461,088	foxicin A/foxicin B/foxicin C/foxicin	24%
12	T3PKS	2,688,325	2,729,389	flaviolin/1,3,6,8-tetrahydroxynaphthalene	100%
13	NAPAA	3,319,993	3,352,156	ε-Poly-L-lysine	100%
14	ectoine	3,599,026	3,609,430	ectoine	100%
15	NRPS-independent-siderophore	4,972,385	4,983,185	desferrioxamin B/desferrioxamine E	83%
16	T1PKS	5,022,707	5,064,844	armeniaspirol A/armeniaspirol B/armeniaspirol C	13%
17	lassopeptide	5,603,813	5,626,381	siamycin I	100%
18	T2PKS,butyrolactone	6,289,965	6,360,977	prejadomycin/rabelomycin/gaudimycin C/gaudimycin D/UWM6/gaudimycin A	33%
19	terpene	7,776,087	7,795,720	albaflavenone	100%
20	NAPAA	7,871,485	7,907,312	stenothricin	13%
21	NRPS,arylpolyene,ladderane,other	8,059,332	8,139,296	pepticinnamin E	100%
22	NRPS-independent-siderophore	8,566,980	8,581,212	synechobactin C9/synechobactin C11/synechobactin 13/synechobactin 14/synechobactin 16/synechobactin	9%

				A/synechobactin B/synechobactin C	
23	terpene	8,847,425	8,867,812	geosmin	100%
24	T1PKS,hglE-KS	9,032,711	9,082,042	hexacosalactone A	11%
25	NRPS-independent-siderophore	9,246,559	9,259,686	-	
26	NRP-metallophore,NRPS	9,277,275	9,328,692	myxochelin B/myxochelin N/myxochelin O/myxochelin P/myxochelin Q/myxochelin A	25%
27	lanthipeptide-class-i	9,529,539	9,554,589	-	
28	redox-cofactor	9,763,383	9,786,176	-	
29	terpene	10,143,824	10,168,314	hopene	92%
30	indole,other,T2PKS	10,194,208	10,310,520	spore pigment	83%
31	lanthipeptide-class-iii	10,801,754	10,824,402	AmfS	80%
32	T1PKS	11,268,264	11,316,234	herboxidiene	42%
33	terpene	11,529,810	11,550,763	rustmicin	10%
34	butyrolactone	11,689,830	11,700,705	-	

Table 26. ***Fischerella sp. NIES-3754***

Region	Type	From	To	Most similar known cluster	Similarity
1	lassopeptide	741,982	761,668	kijanimicin	4%
2	crocagin	1,208,591	1,238,046	O-antigen	14%
3	hglE-KS	1,754,440	1,806,495	-	
4	lassopeptide	1,894,634	1,917,098	-	
5	NRPS,T1PKS	2,080,593	2,171,143	hapalosin	40%
6	terpene	3,072,838	3,094,764	-	
7	T1PKS,hglE-KS	3,854,249	3,905,107	heterocyst glycolipids	85%
8	terpene	4,330,127	4,350,116	-	
9	terpene	4,660,992	4,681,924	-	
10	lanthipeptide-class-v	4,820,221	4,862,587	-	

11	NRPS	4,972,017	5,013,663	puwainaphycin A/puwainaphycin B/puwainaphycin C/puwainaphycin D	50%
12	hglE-KS,resorcinol	5,644,062	5,698,477	heterocyst glycolipids	57%

Table 27. **Streptomyces lydicus WYEC 108**

Region	Type	From	To	Most similar known cluster	Similarity
1	nucleoside,NAPAA,butyrolactone	17,169	68,999	ε-Poly-L-lysine	100%
2	lassopeptide	384,099	405,946	-	
3	NRPS,LAP,thiopeptide,CDPS	430,915	498,640	muraymycin C1	23%
4	T1PKS,oligosaccharide,trans AT-PKS	594,119	828,983	caniferolide A/caniferolide B/caniferolide C/caniferolide D	61%
5	T3PKS,lanthipeptide-class-iii	863,875	906,994	SapB	100%
6	NAPAA	1,006,025	1,041,631	stenothricin	13%
7	T2PKS	1,118,464	1,190,979	spore pigment	83%
8	NRP-metallophore,NRPS,melanin,T1PKS	1,210,846	1,388,016	lydicamycin	96%
9	NRPS-independent-siderophore	1,527,062	1,539,513	peucechelin	25%
10	blactam	1,615,118	1,638,619	valclavam/(-)-2-(2-hydroxyethyl) clavam	71%
11	NRPS-independent-siderophore	2,221,748	2,230,631	desferrioxamine E	100%
12	ectoine	2,319,228	2,329,644	ectoine	100%
13	T1PKS	3,279,881	3,324,607	cyclofaulknamycin	16%
14	lanthipeptide-class-iii	3,938,390	3,961,008	pentalenolactone	15%
15	lanthipeptide-class-i	4,039,967	4,062,984	-	
16	terpene	4,520,042	4,537,633	-	
17	terpene	4,956,617	4,978,225	salinomycin	6%
18	thioamitides,LAP,thiopeptide,NRPS,lassopeptide	5,134,934	5,207,338	ulleungdin	100%
19	terpene	5,848,082	5,867,980	notonesomycin A	5%
20	NRPS-independent-siderophore	6,747,644	6,762,338	schizokinen	20%
21	butyrolactone	6,948,186	6,958,456	-	

22	butyrolactone	7,432,679	7,440,099	-	
23	indole	7,469,816	7,491,138	CDA1b/CDA2a/CDA2b/CDA3a/ CDA3b/CDA4a/CDA4b	15%
24	terpene	7,808,551	7,834,001	hopene	69%
25	hglE-KS,T1PKS	7,953,819	8,005,747	hexacosalactone A	9%
26	aminopolycarboxylic-acid	8,104,695	8,118,145	EDHA	88%
27	lanthipeptide-class-i	8,163,025	8,187,573	-	
28	butyrolactone	8,556,760	8,567,743	merochlorin A/merochlorin B/deschloro-merochlorin A/deschloro-merochlorin B/isochloro-merochlorin B/dichloro-merochlorin B/merochlorin D/merochlorin C	4%
29	NRPS,other	8,632,920	8,687,263	antipain	100%
30	terpene	8,720,529	8,746,092	isorenieratene	100%
31	amglyccycl	8,945,434	8,969,380	streptomycin	55%

Table 28. ***Pseudomonas chlororaphis aurantiaca 464***

Region	***Type***	***From***	***To***	***Most similar known cluster***	***Similarity***
1	arylpolyene	510,505	554,125	APE Vf	45%
2	hserlactone	2,834,491	2,854,226	-	
3	NRPS	3,250,875	3,298,199	JBIR-06	16%
4	other	3,574,128	3,615,210	pyrrolnitrin	100%
5	NRPS-independent-siderophore	3,628,660	3,647,629	schizokinen	20%
6	NRPS	4,050,259	4,118,608	MA026	12%
7	betalactone	4,475,345	4,498,594	fengycin	13%
8	NRP-metallophore,NRPS,resorcinol,ranthipeptide	4,770,459	4,868,244	Pf-5 pyoverdine	33%
9	NRPS	4,911,687	4,964,703	Pf-5 pyoverdine	18%
10	NAGGN	5,096,081	5,110,917	-	
11	hserlactone	5,790,162	5,808,414	-	
12	hserlactone,phenazine	5,862,427	5,885,214	endophenazine A/endophenazine B	38%
13	redox-cofactor	6,356,371	6,378,539	lankacidin C	13%

Overall, another trend in the data observed can be seen in the number of clusters detected for each category of bacteria as well as the number of that could potentially produce antimicrobial compounds. Figures 13 and 14 show this. This category of known genomes in general has the highest number of gene clusters detected with an average of 31 clusters detected. This was followed by the unknown category with 18 clusters and then the positive control with 5 clusters. The negative control showed the least amount of clusters detected with approximately 2 on average across the category (Figure 13). This order was also reflected in the number of clusters with antibiotic potential. The negative control had zero clusters detected with antibiotic potential, but the "known producer" category had on average ~9 clusters per genome that had antibiotic potential. Positive control had ~ 1.4 while the unknown producers had ~4.4. Additionally I looked at the ratio or the proportion of the number of clusters with antibiotic potential to the number of clusters detected in each genome (Figure 14). The category type with the highest ratio was again "the known producers" with a ratio of 0.31 or ~31% of the clusters detected showing some sort of antibiotic potential or activity as described in the literature. This was followed closely by positive control with a ratio of 0.24 or ~ 24% and "unknown producers" with a ratio of 0.27 or 27% of clusters detected showing potential for antibiotic activity. The negative control genomes did not chart with 0% having antibiotic activity.

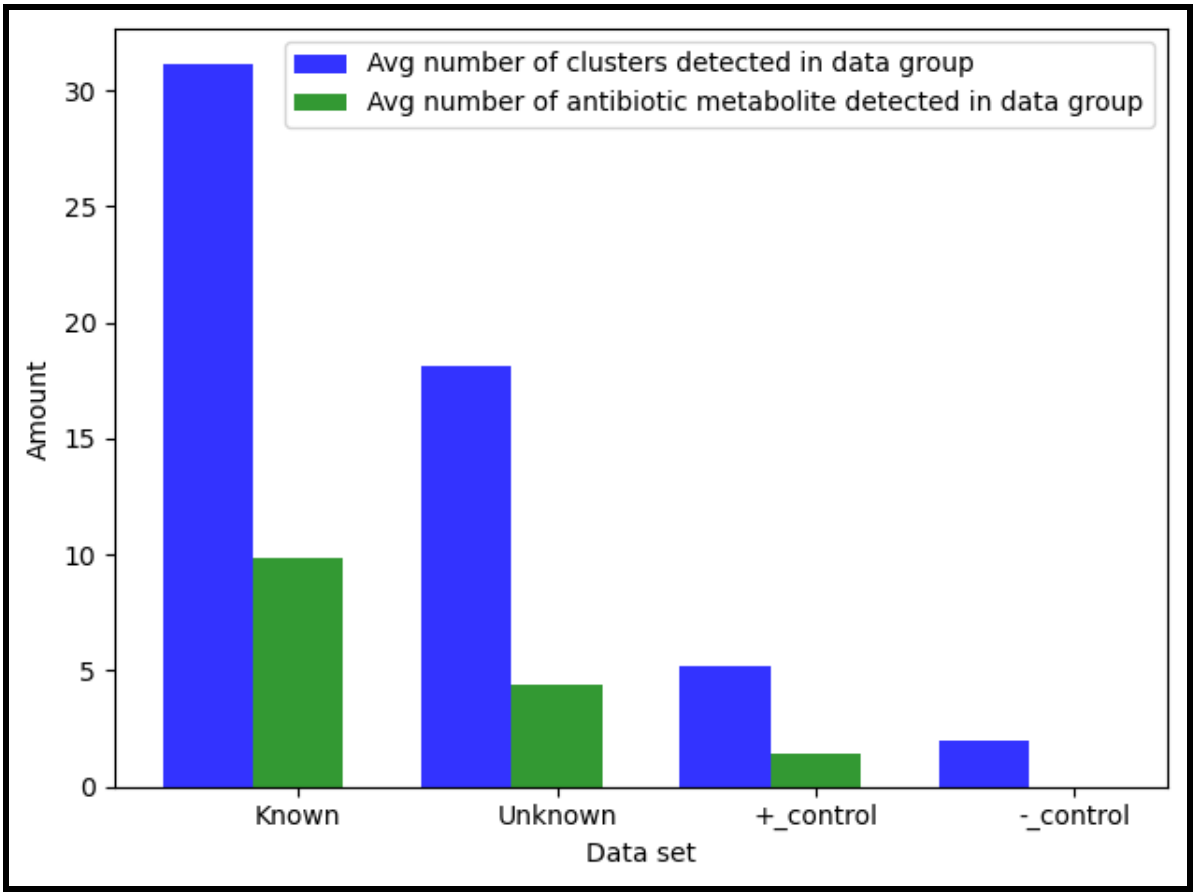


Figure 18. *Number of clusters detected for data type. A representation showing the number of clusters detected for each category of bacteria as well as the number of those clusters that could potentially produce antimicrobial compounds. Known genomes had the highest number of gene clusters detected. Unknown category and then the positive control followed next respectively and lastly negative control. This order was also reflected in the number of clusters with antibiotic potential. The negative control had zero clusters detected with antibiotic potential. ** The values in this figure are Non-Significant*

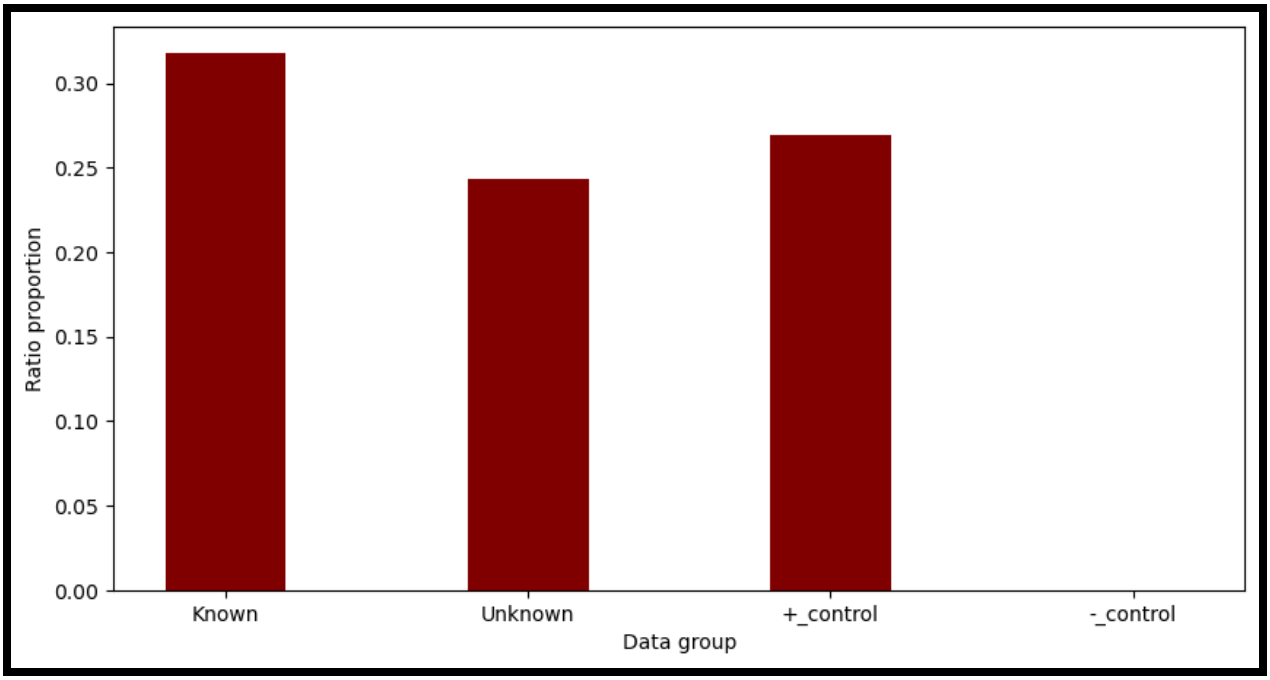


Figure 19. *Ratio of the number of antibiotic clusters. This shows the ratio of the number of clusters with antibiotic potential vs the number of clusters detected in each genome. The category type with the highest ratio was again "the known producers". This was followed closely by positive control and "unknown producers". The negative control genomes did not chart. ** The values in this figure are Non-Significant*

8. DISCUSSION

The results from antiSMASH showed interesting results starting with the positive control group. antiSMASH detected secondary metabolites with antibacterial activity in only 2 of the 5 bacterial genomes as seen in the results section. This was surprising at first considering that all 5 bacterial samples used in the positive control were isolated, sequenced and proved to be antibiotic producers from various assays done in the Hudson Lab. The genomes predicted to have antibacterial activity was *Pseudomonas sp. RIT 623* which as predicted to produce Lankacidin C (Cai et al., 2020), Pseudopyronine A/pseudopyronine B (Bouthillette et al., 2017), Rhizomide A/rhizomide B/rhizomide C (Qi et al., 2021) and *Acinetobacter sp. RIT 592* which was predicted to produce Berninamycin K/berninamycin J/berninamycin A/berninamycin B (Malcolmson et al., 2013), and Fengycin (Medeot et al., 2020). These antiSMASH results of *Pseudomonas sp. RIT 623* and *Acinetobacter sp. RIT 592* are both supported by disc-diffusion inhibitory assays in Figure 22 (Steiner et al., 2020) as well as Figure 21 (Hudson, n.d.) which show a zone of inhibition disc surrounding increasing volumes of each bacterial strain. It is important to note that the compounds from *Pseudomonas sp. RIT 623* and *Acinetobacter sp. RIT 592* have not been isolated so it isn't possible to verify antiSMASH in this scenario other than that it detected secondary metabolites that have antimicrobial activity which does align with the disc assays. Discrepancy arises when looking at the other samples in the positive control. Figures 20, 23 and 24 show disc assays of *Yimella sp. RIT 621, Exiguobacterium sp. RIT 452, and Exiguobacterium sp. RIT 594,* respectively, demonstrating antibacterial activity. This could be down to a few reasons. First could be that the compounds produced by *Yimella sp. RIT 621* (Figure 20), *Exiguobacterium sp. RIT 452* (Figure 23), and *Exiguobacterium sp. RIT 594* (Figure

24) could be novel and does not exist in the MIBiG database where antiSMASH searches for cluster matches. Another reason could be that antiSMASH was not able to detect and accurately characterize the clusters and the metabolites they produce due to a lack of comprehensive or up to date moldues to access the wide/diverse make-up of these gene clusters.

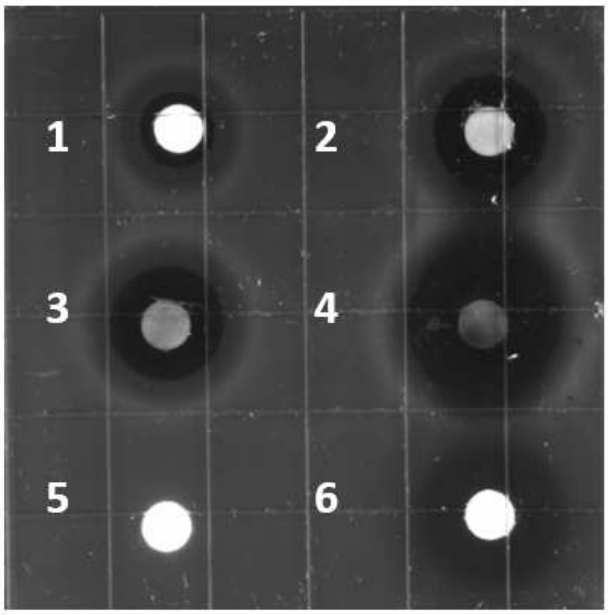


Figure 20. *Yimella sp. RIT 621 produces antibiotic activity against Escherichia coli ATCC 25922. (A) Disc-diffusion inhibitory assay with increasing volumes of a 683 mg/mL solution of Yimella sp. RIT 621 spent TSB medium extract applied to disks 1-4: 10 µL (1), 20 µL (2), 30 µL (3), 40 µL (4); 40 µL of methanol (5) and 10 µL of a 10 mg/mL solution of tetracycline (6) were used as negative and positive control, respectively [Taken from Steiner et al., 2020]*

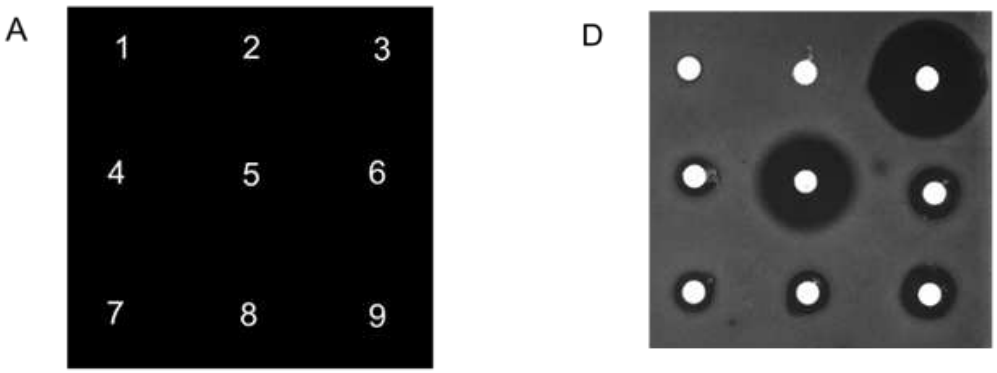


Figure 21. *Disc diffusion assay of Acinetobacter sp. RIT 592 shows antibiotic activity against a B. subtilis BGSC 168 medium. A) Temple, D) B. subtilis BGSC: 1) sterile un-inoculated LB medium, 2) Methanol, 3) Tetracycline (10 mg/ml), 4) RIT 594, 5) RIT 594 stimulated by 20% sterile filtered supernatant of RIT 592 6) RIT 594 stimulated by 20% sterile filtered supernatant of E. coli ATCC 25922, 7) RIT 592 8) RIT 592 stimulated by 20% sterile filtered supernatant of RIT 594, 9) RIT 594 stimulated by 20% sterile filtered supernatant of P. aeruginosa ATCC 27853. [Figure taken from Hudson, n.d]*

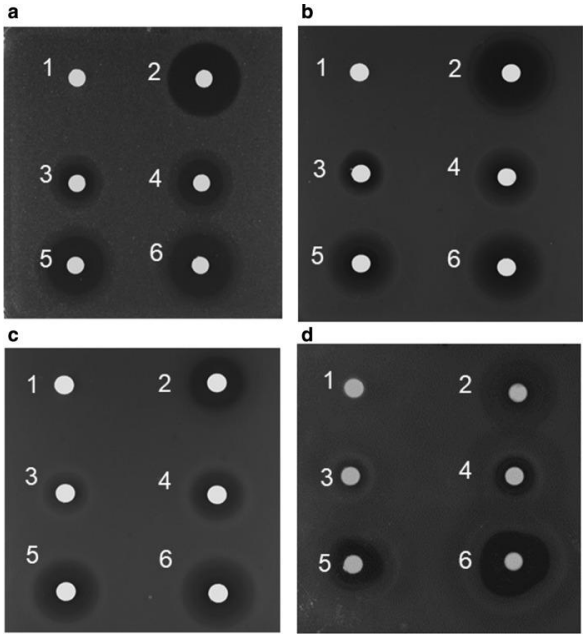


Figure 22. *(A-D) Disk diffusion assay of Pseudomonas sp. RIT 623. Against a) Bacillus subtilis BGSC 168, b) Staphylococcus aureus ATCC 25923, c) Escherichia coli ATCC 25922, d) Pseudomonas aeruginosa ATCC 27853. (1) Methanol, 20 µL; (2) Tetracycline, 20 µL (10 mg/mL); and (3) 10 µL, (4) 20 µL, (5) 40 µL, and (6) 60 µL of Pseudomonas sp. RIT 623 extract dissolved in 100% methanol, respectively. [Taken from (Steiner et al., 2020)]*

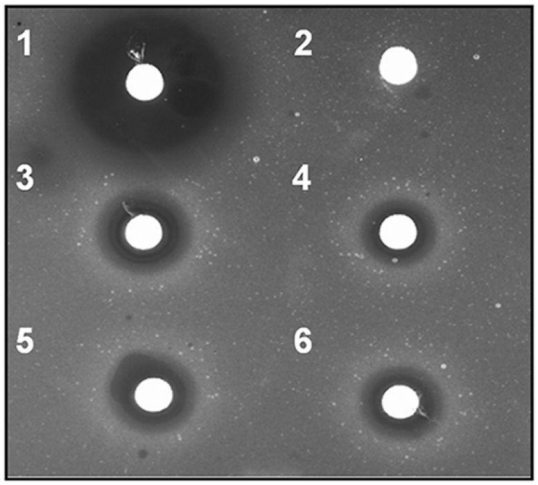


Figure 23. *Disk diffusion assay using ethyl acetate extract of spent medium of Exiguobacterium sp. RIT 452 against Escherichia coli ATCC 25922. (1) Tetracycline, 20 µl (10 mg/ml); (2) methanol, 20 µl; and (3, 4, 5, and 6) 25 µl, 10 µl, 20 µl, and 15 µl of RIT452 extract, respectively. [Taken from (Parthasarathy et al., 2018)]*

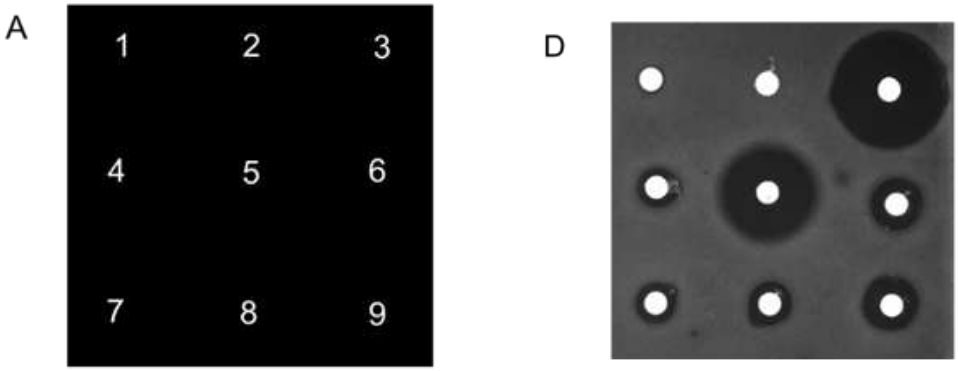


Figure 24. *Disc diffusion assay of Acinetobacter sp. RIT 594 shows antibiotic activity against a B. subtilis BGSC 168 medium. A) Temple, D) B. subtilis BGSC: 1) sterile un-inoculated LB medium, 2) Methanol, 3) Tetracycline (10 mg/ml), 4) RIT 594, 5) RIT 594 stimulated by 20% sterile filtered supernatant of RIT 592 6) RIT 594 stimulated by 20% sterile filtered supernatant of E. coli ATCC 25922, 7) RIT 592 8) RIT 592 stimulated by 20% sterile filtered supernatant of RIT 594, 9) RIT 594 stimulated by 20% sterile filtered supernatant of P. aeruginosa ATCC 27853. [Figure taken from Hudson, n.d]*

The results from the negative control were more straightforward and what was expected. None of the 6 samples used showed any clusters that had potential antimicrobial activity. In fact a majority of the clusters detected were predicted to produce no metabolites as represented in Figure 12. In fact, out of the 3 reference genomes used (Tables 7-9) only the *E. coli* reference genome had predicted compounds, enterobactin, and O-antigen. This result would be anticipated as these samples are the reference genomes of very common bacterial species that are not generally known to produce antibiotics. The results for the other test cases for the negative control were also consistent. The randomly generated genomes produced "no result found" output (Figure 11) which indicates the tools ability to properly assess and detect real and viable genetic content in sequence files. This is also the same for the shuffled sequence of the *Streptomyces rapamycinicus NRRL 5491* genome. The shuffle process as you would assume would randomize and destroy the genetic contents contained within the file and the antiSMASH tool was able to detect this and output a no result output as shown in Figure 11.

The resulting data from the known producers category was also fairly consistent with what was expected. With the known producers, we already know what antibiotic compounds they produce. We can easily verify the accuracy of antiSMASH and the secondary metabolites it predicts. Please see Table 29 for breakdown of each known producer, their antibiotic and whether that antibiotic was detected by antiSMASH.

Table 29. *Cross-checking of antiSMASH in known producers*

Known Producing bacterial species	***Known antibiotic***	***Predicted by antiSMASH?***
Streptomyces rapamycinicus NRRL 5491	Rapamycin	Yes
Streptomyces rapamycinicus SRMK07	Rapamycin	Yes
Streptomyces noursei ATCC 11455	Nystatin A1	Yes
Streptomyces venezuelae ATCC 15439	Methymycin and pikromycin	No
Streptomyces filamentosus NRRL 15998	Napsamycin and mureidomycin	No
Streptomyces griseus IFO 13350	Streptomycin	Yes
Streptomyces lincolnensis B48	Lincomycin	Yes
Streptomyces ambofaciens ATCC 23877	Spiramycin	Yes
Streptomyces atratus SCSIO ZH16	Atratumycin	Yes

Normally cross checking of antiSMASH tools would not be done when dealing with an unknown or newly sequenced bacterial species but since the antibiotics of known producers are already documented it is easier to see the performance of antiSMASH. Overall, the tool performed fairly well. It was able to predict the known antibiotic produced in 7 of the 9 bacterial genome sequences. The 2 species in which the predictions did not accurately predict their known antibiotics were *Streptomyces filamentosus NRRL 15998* and *Streptomyces venezuelae ATCC 15439*. While antiSMASH did predict other secondary metabolites with potential antibiotic activity, this test shows that the tool can possibly miss or fail to predict some secondary metabolites. In samples where there aren't many antimicrobial compounds predicted by

antiSMASH, not detecting one could be the difference between further experimentation and finding a new antibiotic producer or a missed opportunity on a viable sample.

The analysis of randomly chosen unknown producers gives us a chance to simulate or create a mock environment of where a lab would sequence an unknown bacterial sample then feed it though antiSMASH for secondary metabolite screening and analysis. As stated in the results section all but one of the selected genomes were predicted to have secondary metabolites with antimicrobial activity. The lone genomic sample left out was *Ignavibacterium album JCM 16511* (Table 19). This could be explained somewhat by the fact that the genera *Ignavibacterium* are not known generally to be producers of antibiotics at least as compared to other bacterial genera. Take for example the genera *Streptomyces* which is known to be a prolific producer of antibiotics (Watve et al., 2001). All the *Streptomyces* genomes were predicted to produce antimicrobial compounds with a minimum of at least 2 predicted in each genome and contain an average of ~9.6 antimicrobial secondary metabolites. This could also be due to the fact as well that Streptomyces have the largest number of clusters detected. This point is reflected in Figure 18 where it shows the average number of gene clusters detected and the average number of those clusters which were predicted to produce antimicrobial compounds in each category type. In general more clusters show more potential for antibiotic compounds. This could be good criteria for a lab when evaluating the antimicrobial properties of an unknown genome sample as more predicted antimicrobial secondary samples in a genome could lower the chance of a false positive. Take for example *Fischerella sp. NIES-3754* (Table 26) in the unknown samples. antiSMASH did predict it would produce an antibiotic, kijanimicin (Nakayama et al., 1987), it was the only one. If there were multiple, perhaps researchers would be more confident in this

sample's antimicrobial potential. This was explored in Figure 19 which shows the proportion of total clusters which were predicted to produce a secondary metabolite with antimicrobial properties. The positive control and the known producers had a higher ratio of clusters detected being antibiotic producing which supports the idea of more antimicrobial metabolites predicted increases the odds that a specific sample is an antibiotic producer.

Another form of analysis was to look at the trends in cluster types and secondary metabolites produced in each sample category. This would provide some insights into the predominant characteristics of the metabolites that make up in each sample category. As stated in the introduction, cluster types like NRPS and PKS are known to be producers of secondary metabolites with a wide range of pharmacological properties including antibiotics. If researchers were to perform gene cluster analysis and notice a trend in these cluster types being common it could serve as an indicator that a sample produces antibiotics. In the positive control Figure 10 shows a summary of the most common cluster types. Since we know from Figures 20-24 that these positive controls do have antimicrobial activity we would expect a larger proportion of the cluster to be NRPS and PKS related. This turns out to be the case as Figure 10 shows that NRPS cluster types in particular represent the largest share of gene clusters present from the positive controls. This trend is reversed in the negative controls as shown in Figure 13, where NRPS cluster types make up a smaller fraction of the cluster types and the largest share belongs to Cyclic-lactone-autoinducer. This trend also continues in the known producers and unknown producers. NRPS and PKS (T1PKS, T2PKS, and T3PKS) represent the largest portion of clusters at 49.8% percent of clusters and in the unknown producers they represent 33.3%. NRPS and PKS cluster types being the largest portion of genomes makes sense since both the sample categories

were predicted to be composed almost entirely of antibiotic producers. It is important to note though that NRPS and PKS presence doesn't directly translate to antibiotics, since there were other clusters types that were predicted to produce antimicrobial compounds, but their value and importance lie in their known ability to produce these compounds and the overwhelming knowledge and literature to support the pharmacological properties. If a bacterial sample has a large proportion or majority of their clusters being of NRPS or PKS types then one can assume or infer that the organism has a lot of medicinal potential. But this assumption will have to be supported with other information as discussed before.

9. CONCLUSION

With the rates of antibiotic resistance increasing and antibiotic development very low, a change is needed to ensure bacterial infections do not cause the pandemics and devastation as they once did in the past. With the many new scientific and technological advances that have been made throughout the years many are turning to computational and bioinformatics methods to tackle these pressing problems. With its ability to detect gene clusters and secondary metabolites of bacterial genome sequences, antiSMASH could aid in this area.

This project focused on using antiSMASH as a screening tool for antibiotics and evaluated the platform as an aide in the detection to find antibiotics in-silico before confirmation with wet-lab experimental approaches. This has the potential to save time and resources since it would afford researchers to be more selective. Overall based on its performance and analysis antiSMASH would be valuable to any lab or research focused on discovering antibiotics with a few caveats. First seen in the positive control, the tool can miss secondary metabolites and not detect those with antibiotic properties. This could be due to the novelty of the compound or it simply not being in their database. This was shown to be the case as well when performing cross checking with the known producers and perhaps this could be a point of improvement. With the unknown samples is where we saw the potential value of antiSMASH in where it predicted secondary metabolites with antibiotics properties in 9 of the 10 samples. Even though there is no way to cross check with unknowns, looking at trends in the abundance of certain cluster types or proportion of clusters predicted to produce antibiotics could give clues and reassurance as to whether a bacterial isolate is a potential antibiotic producer.

Appendix A: Code for shuffling genome

```
from Bio import SeqIO

from Bio.Seq import Seq

import random


for seq_record in SeqIO.parse("sequence-3.fasta", "fasta"):

    temp_file = seq_record.seq

temp_file = str(temp_file)

new_file = ''.join(random.sample(temp_file,len(temp_file)))

file = open("Shuffled_genome.txt", "w")

file.write(new_file)

file.close()

#Fasta Conversion: "https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html"
```

Appendix B: Code for graphs and Figures

```
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np

Known_producers = pd.read_csv('Known Producers - Sheet1.csv')
print(Known_producers)
df = Known_producers.value_counts().rename_axis('unique_values').reset_index(name='counts')
print(df)
df["unique_values"]
df = df[df['counts'] > 2]
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.1, labeldistance=0.03,
radius=2.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
plt.show()
"_____"
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np

Known_producers_metabolites = pd.read_csv('Known Producers - Metabolites - Sheet1.csv')
print(Known_producers_metabolites)
```

```
df =
Known_producers_metabolites.value_counts().rename_axis('unique_values').reset_index(name='
counts')
df
df["unique_values"]
df = df[df['counts'] > 1]
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.1, labeldistance=0.03,
radius=2.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
plt.show()
```

*** Positive Controls ***

```
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np

Positive_Controls = pd.read_csv('Positive Controls.csv')
Positive_Controls
df = Positive_Controls.value_counts().rename_axis('unique_values').reset_index(name='counts')
df
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.05, labeldistance=0, radius=2.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
plt.show()
Positive_Metabolites = pd.read_csv('Positive Control - Cluster Types - Sheet1.csv')
```

```
Positive_Metabolites
df = Positive_Metabolites.value_counts().rename_axis('unique_values').reset_index(name='counts')
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.05, labeldistance=0, radius=2.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
plt.show()

*** Negative Controls ***

import pandas as pd
from matplotlib import pyplot as plt
import numpy as np

Negative_Controls = pd.read_csv('Negative Controls.csv')
Negative_Controls
df = Negative_Controls.value_counts().rename_axis('unique_values').reset_index(name='counts')
df
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.1, labeldistance=3, radius=3.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
plt.rc('font', size=20)
plt.show()
Negative_Cluster = pd.read_csv('Negative Control - Cluster Types - Sheet1.csv')
Negative_Cluster
df = Negative_Cluster.value_counts().rename_axis('unique_values').reset_index(name='counts')
df
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=1.1, labeldistance=3, radius=3.5)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(1.85, 2))
```

```
plt.rc('font', size=20)
plt.show()
```

*** Unknown Genomes ***

```
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np

Unknown_Genomes = pd.read_csv('Unknown Genomes.csv')
Unknown_Genomes
df = Unknown_Genomes.value_counts().rename_axis('unique_values').reset_index(name='counts')
df
df = df[df['counts'] > 2]
plt.rcParams.update({'font.size': 23})
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=0.9, labeldistance=0.03, radius=10)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(4, 5))
plt.show()
Unknown_Genome_types = pd.read_csv("Unknwon Genomes -  Cluster Types - Sheet1.csv")
df = Unknown_Genome_types.value_counts().rename_axis('unique_values').reset_index(name='counts')
df
df = df[df['counts'] > 2]
plt.rcParams.update({'font.size': 23})
plt.pie(df["counts"],autopct='%1.1f%%', startangle=90, pctdistance=0.9, labeldistance=0.03, radius=10)
plt.legend(labels=df["unique_values"], loc="upper left", bbox_to_anchor=(4, 5))
plt.show()
```

References

Adams, R. A., Leon, G., Miller, N. M., Reyes, S. P., Thantrong, C. H., Thokkadam, A. M., Lemma, A. S., Sivaloganathan, D. M., Wan, X., & Brynildsen, M. P. (2021). Rifamycin antibiotics and the mechanisms of their failure. *The Journal of Antibiotics*, *74*(11), Article 11. https://doi.org/10.1038/s41429-021-00462-x

Akova, M. (2016). Epidemiology of antimicrobial resistance in bloodstream infections. *Virulence*, *7*(3), 252–266. https://doi.org/10.1080/21505594.2016.1159366

Alanis, A. J. (2005). Resistance to Antibiotics: Are We in the Post-Antibiotic Era? *Archives of Medical Research*, *36*(6), 697–705. https://doi.org/10.1016/j.arcmed.2005.06.009

Alexander Fleming Discovery and Development of Penicillin—Landmark. (n.d.). American Chemical Society. Retrieved February 27, 2023, from https://www.acs.org/education/whatischemistry/landmarks/flemingpenicillin.html

Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y., & Kato, F. (2003). Organization of the biosynthetic gene cluster for the polyketide macrolide mycinamicin in Micromonospora griseorubida. *FEMS Microbiology Letters*, *218*(1), 135–141. https://doi.org/10.1111/j.1574-6968.2003.tb11509.x

Arisetti, N., Fuchs, H. L. S., Coetzee, J., Orozco, M., Ruppelt, D., Bauer, A., Heimann, D., Kuhnert, E., Bhamidimarri, S. P., Bafna, J. A., Hinkelmann, B., Eckel, K., Sieber, S. A., Müller, P. P., Herrmann, J., Müller, R., Winterhalter, M., Steinem, C., & Brönstrup, M. (2021). Total synthesis and mechanism of action of the antibiotic armeniaspirol A. *Chemical Science*, *12*(48), 16023–16034. https://doi.org/10.1039/D1SC04290D

Armstrong, H., Bording-Jorgensen, M., Chan, R., & Wine, E. (2019). Nigericin Promotes NLRP3-Independent Bacterial Killing in Macrophages. *Frontiers in Immunology*, *10*. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02296

Asano, T., & Adachi, Y. (2006). Effects of Griseoviridin and Viridogrisein against Swine Dysentery in Experimental Infection by Using Mice and Pigs. *Journal of Veterinary Medical Science*, *68*(6), 555–560. https://doi.org/10.1292/jvms.68.555

Belousoff, M. J., Shapira, T., Bashan, A., Zimmerman, E., Rozenberg, H., Arakawa, K., Kinashi, H., & Yonath, A. (2011). Crystal structure of the synergistic antibiotic pair, lankamycin and lankacidin, in complex with the large ribosomal subunit. *Proceedings of the National Academy of Sciences*, *108*(7), 2717–2722. https://doi.org/10.1073/pnas.1019406108

Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism | PNAS. (n.d.). Retrieved March 31, 2023, from https://www.pnas.org/doi/full/10.1073/pnas.0708750105

Birney, E. (2001). Hidden Markov models in biological sequence analysis. *IBM Journal of Research and Development*, *45*(3.4), 449–454. https://doi.org/10.1147/rd.453.0449

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*(1), 42–51. https://doi.org/10.1038/nrmicro3380

Boeck, L. D., Papiska, H. R., Wetzel, R. W., Mynderse, J. S., Fukuda, D. S., Mertz, F. P., & Berry, D. M. (1990). A54145, A NEW LIPOPEPTIDE ANTIBIOTIC COMPLEX: DISCOVERY, TAXONOMY, FERMENTATION AND HPLC. *The Journal of Antibiotics*, *43*(6), 587–593. https://doi.org/10.7164/antibiotics.43.587

Borghi, A., Coronelli, C., Faniuolo, L., Allievi, G., Pallanza, R., & Gallo, G. G. (1984). Teichomycins, new antibiotics from Actinoplanes teichomyceticus nov. Sp. IV. Separation and characterization of the components of teichomycin (teicoplanin). *The Journal of Antibiotics*, *37*(6), 615–620. https://doi.org/10.7164/antibiotics.37.615

Bouthillette, L. M., Darcey, C. A., Handy, T. E., Seaton, S. C., & Wolfe, A. L. (2017). Isolation of the antibiotic pseudopyronine B and SAR evaluation of C3/C6 alkyl analogs. *Bioorganic & Medicinal Chemistry Letters*, *27*(12), 2762–2765. https://doi.org/10.1016/j.bmcl.2017.04.067

Bowyer, J. E., LC. de los Santos, E., Styles, K. M., Fullwood, A., Corre, C., & Bates, D. G. (2017). Modeling the architecture of the regulatory system controlling methylenomycin production in Streptomyces coelicolor. *Journal of Biological Engineering*, *11*(1), 30. https://doi.org/10.1186/s13036-017-0071-6

Brautaset, T., Sekurova, O. N., Sletta, H., Ellingsen, T. E., Strøm, A. R., Valla, S., & Zotchev, S. B. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces noursei ATCC 11455: Analysis of the gene cluster and deduction of the biosynthetic pathway. *Chemistry & Biology*, *7*(6), 395–403. https://doi.org/10.1016/S1074-5521(00)00120-4

Cai, L., Yao, Y., Yeon, S. K., & Seiple, I. B. (2020). Modular approaches to lankacidin antibiotics. *Journal of the American Chemical Society*, *142*(35), 15116–15126. https://doi.org/10.1021/jacs.0c06648

Charan, R. D., Schlingmann, G., Janso, J., Bernan, V., Feng, X., & Carter, G. T. (2004). Diazepinomicin, a New Antimicrobial Alkaloid from a Marine Micromonospora sp. *Journal of Natural Products*, *67*(8), 1431–1433. https://doi.org/10.1021/np040042r

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R., Roux, S., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2021). The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Research*, *49*(D1), D751–D763. https://doi.org/10.1093/nar/gkaa939

Clardy, J., Fischbach, M. A., & Currie, C. R. (2009). The natural history of antibiotics. *Current Biology*, *19*(11), R437–R441. https://doi.org/10.1016/j.cub.2009.04.001

Cobongela, S. Z. Z., Makatini, M. M., Mdluli, P. S., & Sibuyi, N. R. S. (2022). Acyldepsipeptide Analogues: A Future Generation Antibiotics for Tuberculosis Treatment. *Pharmaceutics*, *14*(9), 1956. https://doi.org/10.3390/pharmaceutics14091956

Conly, J. M., & Johnston, B. L. (2005). Where are all the new antibiotics? The new antibiotic paradox. *Canadian Journal of Infectious Diseases and Medical Microbiology*, *16*, 159–160. https://doi.org/10.1155/2005/892058

De Pascale, G., & Wright, G. D. (2010). Antibiotic Resistance by Enzyme Inactivation: From Mechanisms to Solutions. *ChemBioChem*, *11*(10), 1325–1334. https://doi.org/10.1002/cbic.201000067

Deboer, C., Meulman, P. A., Wnuk, R. J., & Peterson, D. H. (1970). GELDANAMYCIN, A NEW ANTIBIOTIC. *The Journal of Antibiotics*, *23*(9), 442–447. https://doi.org/10.7164/antibiotics.23.442

Dong, L., Shen, Y., Hou, X.-F., Li, W.-J., & Tang, G.-L. (2019). Discovery of Druggability-Improved Analogues by Investigation of the LL-D49194α1 Biosynthetic

Pathway. *Organic Letters*, *21*(7), 2322–2325. https://doi.org/10.1021/acs.orglett.9b00610

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763. https://doi.org/10.1093/bioinformatics/14.9.755

EMBL-EBI. (n.d.). *What are profile hidden Markov models? | Pfam*. Retrieved March 3, 2023, from https://www.ebi.ac.uk/training/online/courses/pfam-creating-protein-families/what-are-profile-hidden-markov-models-hmms/

Flatman, R. H., Howells, A. J., Heide, L., Fiedler, H.-P., & Maxwell, A. (2005). Simocyclinone D8, an Inhibitor of DNA Gyrase with a Novel Mode of Action. *Antimicrobial Agents and Chemotherapy*, *49*(3), 1093–1100. https://doi.org/10.1128/AAC.49.3.1093-1100.2005

Flatt, P. M., Wu, X., Perry, S., & Mahmud, T. (2013). Genetic Insights Into Pyralomicin Biosynthesis in Nonomuraea spiralis IMC A-0156. *Journal of Natural Products*, *76*(5), 939–946. https://doi.org/10.1021/np400159a

Format Conversion. (n.d.). Retrieved April 14, 2023, from https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html

Franzese, M., & Iuliano, A. (2019). Hidden Markov Models. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 753–762). Academic Press. https://doi.org/10.1016/B978-0-12-809633-8.20488-3

Fredenhagen, A., & Séquin, U. (1985). THE PHOTODEACTIVATION OF HEDAMYCIN, AN ANTITUMOR ANTIBIOTIC OF THE PLURAMYCIN TYPE. *The Journal of Antibiotics*, *38*(2), 236–241. https://doi.org/10.7164/antibiotics.38.236

Frieri, M., Kumar, K., & Boutin, A. (2017). Antibiotic resistance. *Journal of Infection and Public Health*, *10*(4), 369–378. https://doi.org/10.1016/j.jiph.2016.08.007

Fukumoto, A., Kim, Y.-P., Hanaki, H., Shiomi, K., Tomoda, H., & Ōmura, S. (2008). Cyslabdan, a New Potentiator of Imipenem Activity against Methicillin-resistant Staphylococcus aureus, Produced by Streptomyces sp. K04-0144. *The Journal of Antibiotics*, *61*(1), Article 1. https://doi.org/10.1038/ja.2008.102

Gomes, E. S., Schuch, V., & Lemos, E. G. de M. (2013). Biotechnology of polyketides: New breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Brazilian Journal of Microbiology*, *44*, 1007–1034. https://doi.org/10.1590/S1517-83822013000400002

Gonsior, M., Mühlenweg, A., Tietzmann, M., Rausch, S., Poch, A., & Süssmuth, R. D. (2015). Biosynthesis of the Peptide Antibiotic Feglymycin by a Linear Nonribosomal Peptide Synthetase Mechanism. *ChemBioChem*, *16*(18), 2610–2614. https://doi.org/10.1002/cbic.201500432

Goodsell, D. (n.d.). *PDB101: Molecule of the Month: Actinomycin*. RCSB: PDB-101. Retrieved March 27, 2023, from http://pdb101.rcsb.org/motm/160

Grant, M. A., Baron, R. M., Macias, A. A., Layne, M. D., Perrella, M. A., & Rigby, A. C. (2009). NETROPSIN IMPROVES SURVIVAL FROM ENDOTOXEMIA BY DISRUPTING HMGA1-BINDING TO THE NOS2 PROMOTER. *The Biochemical Journal*, *418*(1), 103. https://doi.org/10.1042/BJ20081427

Greule, A., Marolt, M., Deubel, D., Peintner, I., Zhang, S., Jessen-Trefzer, C., De Ford, C., Burschel, S., Li, S.-M., Friedrich, T., Merfort, I., Lüdeke, S., Bisel, P., Müller, M., Paululat, T., & Bechthold, A. (2017). Wide Distribution of Foxicin Biosynthetic Gene Clusters in Streptomyces Strains – An Unusual Secondary Metabolite with Various Properties. *Frontiers in Microbiology*, *8*. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00221

Hassan, H. M., & Fridovich, I. (1980). Mechanism of the antibiotic action pyocyanine. *Journal of Bacteriology*, *141*(1), 156–163.

Hayakawa, Y., Kanamaru, N., Shimazu, A., & Seto, H. (1991). LYDICAMYCIN, A NEW ANTIBIOTIC OF A NOVEL SKELETAL TYPE I. TAXONOMY, FERMENTATION, ISOLATION AND BIOLOGICAL ACTIVITY. *The Journal of Antibiotics*, *44*(3), 282–287. https://doi.org/10.7164/antibiotics.44.282

He, J., Sundararajan, A., Devitt, N. P., Schilkey, F. D., Ramaraj, T., & Melançon, C. E. (2016). Complete Genome Sequence of Streptomyces venezuelae ATCC 15439, Producer of the Methymycin/Pikromycin Family of Macrolide Antibiotics, Using PacBio Technology. *Genome Announcements*, *4*(3), e00337-16. https://doi.org/10.1128/genomeA.00337-16

Heidary, M., Khosravi, A. D., Khoshnood, S., Nasiri, M. J., Soleimani, S., & Goudarzi, M. (2018). Daptomycin. *Journal of Antimicrobial Chemotherapy*, *73*(1), 1–11. https://doi.org/10.1093/jac/dkx349

Holzgrabe, U. (2015). New Griselimycins for Treatment of Tuberculosis. *Chemistry & Biology*, *22*(8), 981–982. https://doi.org/10.1016/j.chembiol.2015.08.002

Hover, B. M., Kim, S.-H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., Maniko, J., Estrela, A. B., Molina, H., Park, S., Perlin, D. S., & Brady, S. F. (2018). Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature Microbiology*, *3*(4), Article 4. https://doi.org/10.1038/s41564-018-0110-1

How antibiotic is made—Material, history, used, processing, components, composition, structure, procedure, steps. (n.d.). Retrieved February 28, 2023, from http://www.madehow.com/Volume-4/Antibiotic.html

Huang, S., Liu, Y., Liu, W.-Q., Neubauer, P., & Li, J. (2021). The Nonribosomal Peptide Valinomycin: From Discovery to Bioactivity and Biosynthesis. *Microorganisms*, *9*(4), 780. https://doi.org/10.3390/microorganisms9040780

Hudson, A. (n.d.). *R15 Proposal 2021—592-594 strains.docx*. Google Docs. Retrieved April 21, 2023, from https://docs.google.com/document/u/0/d/1bY1JwR58LJIGag_VhBkeKFf08Ps0NEW8/edit?usp=gmail_attachment_preview&usp=embed_facebook

Igarashi, Y., Kim, Y., In, Y., Ishida, T., Kan, Y., Fujita, T., Iwashita, T., Tabata, H., Onaka, H., & Furumai, T. (2010). Alchivemycin A, a Bioactive Polycyclic Polyketide with an Unprecedented Skeleton from Streptomyces sp. *Organic Letters*, *12*(15), 3402–3405. https://doi.org/10.1021/ol1012982

Iscla, I., Wray, R., Wei, S., Posner, B., & Blount, P. (2014). Streptomycin potency is dependent on MscL channel expression. *Nature Communications*, *5*(1), Article 1. https://doi.org/10.1038/ncomms5891

Ji, X., Dong, Y., Ling, C., Zhou, Z., Li, Q., & Ju, J. (2020). Elucidation of the Tailoring Steps in Julichrome Biosynthesis by Marine Gastropod Mollusk-Associated Streptomyces sampsonii SCSIO 054. *Organic Letters*, *22*(17), 6927–6931. https://doi.org/10.1021/acs.orglett.0c02469

Jiang, L., Wang, L., Zhang, J., Liu, H., Hong, B., Tan, H., & Niu, G. (2015). Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in Streptomyces roseosporus NRRL 15998. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep14111

Jo, H.-G., Adidjaja, J. J., Kim, D.-K., Park, B.-S., Lee, N., Cho, B.-K., Kim, H. U., & Oh, M.-K. (2022). Comparative genomic analysis of Streptomyces rapamycinicus NRRL 5491 and its mutant overproducing rapamycin. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-14199-6

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, *11*(1), 431. https://doi.org/10.1186/1471-2105-11-431

Jørgensen, H., Degnes, K. F., Sletta, H., Fjærvik, E., Dikiy, A., Herfindal, L., Bruheim, P., Klinkenberg, G., Bredholt, H., Nygård, G., Døskeland, S. O., Ellingsen, T. E., & Zotchev, S. B. (2009). Biosynthesis of Macrolactam BE-14106 Involves Two Distinct PKS Systems and Amino Acid Processing Enzymes for Generation of the Aminoacyl Starter Unit. *Chemistry & Biology*, *16*(10), 1109–1121. https://doi.org/10.1016/j.chembiol.2009.09.014

Kara, M., Asano, K., Kawamoto, I., Takiouchi, T., Katsumata, S., Takahashi, K.-I., & Nakano, H. (1989). LEINAMYCIN, A NEW ANTITUMOR ANTIBIOTIC FROM

STREPTOMYCES; PRODUCING ORGANISM, FERMENTATION AND ISOLATION. *The Journal of Antibiotics*, *42*(12), 1768–1774. https://doi.org/10.7164/antibiotics.42.1768

Kharel, M. K., Pahari, P., Lian, H., & Rohr, J. (2010). Enzymatic Total Synthesis of Rabelomycin, an Angucycline Group Antibiotic. *Organic Letters*, *12*(12), 2814–2817. https://doi.org/10.1021/ol1009009

Kingston, W. (2000). Antibiotics, invention and innovation. *Research Policy*, *29*(6), 679–710. https://doi.org/10.1016/S0048-7333(99)00045-1

Kodani, S., Sato, K., Hemmi, H., & Ohnish-Kameyama, M. (2014). Isolation and structural determination of a new hydrophobic peptide venepeptide from Streptomyces venezuelae. *The Journal of Antibiotics*, *67*(12), Article 12. https://doi.org/10.1038/ja.2014.81

Koyama, N., Shigeno, S., Kanamoto, A., & Tomoda, H. (2020). Steffimycin E, a new anti-mycobacterial agent against Mycobacterium avium complex, produced by Streptomyces sp. OPMA02852. *The Journal of Antibiotics*, *73*(8), Article 8. https://doi.org/10.1038/s41429-020-0290-9

Kunze, B., Höfle, G., & Reichenbach, H. (1987). THE AURACHINS, NEW QUINOLINE ANTIBIOTICS FROM MYXOBACTERIA: PRODUCTION, PHYSICO-CHEMICAL AND BIOLOGICAL PROPERTIES. *The Journal of Antibiotics*, *40*(3), 258–265. https://doi.org/10.7164/antibiotics.40.258

Kwun, M. J., & Hong, H.-J. (2014). Genome Sequence of Streptomyces toyocaensis NRRL 15009, Producer of the Glycopeptide Antibiotic A47934. *Genome Announcements*, *2*(4), e00749-14. https://doi.org/10.1128/genomeA.00749-14

Lim, J., Chintalapudi, V., Gudmundsson, H. G., Tran, M., Bernasconi, A., Blanco, A., Song, L., Challis, G. L., & Anderson, E. A. (2021). Synthesis of the C1–C27 Fragment of Stambomycin D Validates Modular Polyketide Synthase-Based Stereochemical Assignments. *Organic Letters*, *23*(19), 7439–7444. https://doi.org/10.1021/acs.orglett.1c02650

Lim, Y. H., Wong, F. T., Yeo, W. L., Ching, K. C., Lim, Y. W., Heng, E., Chen, S., Tsai, D.-J., Lauderdale, T.-L., Shia, K.-S., Ho, Y. S., Hoon, S., Ang, E. L., Zhang, M. M., & Zhao, H. (2018). Auroramycin: A Potent Antibiotic from Streptomyces roseosporus by CRISPR-Cas9 Activation. *Chembiochem: A European Journal of Chemical Biology*. https://doi.org/10.1002/cbic.201800266

Liu, A., Tran, L., Becket, E., Lee, K., Chinn, L., Park, E., Tran, K., & Miller, J. H. (2010). Antibiotic Sensitivity Profiles Determined with an Escherichia coli Gene Knockout Collection: Generating an Antibiotic Bar Code. *Antimicrobial Agents and Chemotherapy*, *54*(4), 1393–1403. https://doi.org/10.1128/AAC.00906-09

Liu, W.-T., Lamsa, A., Wong, W. R., Boudreau, P. D., Kersten, R., Peng, Y., Moree, W. J., Duggan, B. M., Moore, B. S., Gerwick, W. H., Linington, R. G., Pogliano, K., & Dorrestein, P. C. (2014a). MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in Streptomyces roseosporus. *The Journal of Antibiotics*, *67*(1), Article 1. https://doi.org/10.1038/ja.2013.99

Liu, W.-T., Lamsa, A., Wong, W. R., Boudreau, P. D., Kersten, R., Peng, Y., Moree, W. J., Duggan, B. M., Moore, B. S., Gerwick, W. H., Linington, R. G., Pogliano, K., & Dorrestein, P. C. (2014b). MS/MS-based networking and peptidogenomics guided

genome mining revealed the stenothricin gene cluster in Streptomyces roseosporus. *The Journal of Antibiotics*, *67*(1), Article 1. https://doi.org/10.1038/ja.2013.99

Ma, J., Huang, H., Xie, Y., Liu, Z., Zhao, J., Zhang, C., Jia, Y., Zhang, Y., Zhang, H., Zhang, T., & Ju, J. (2017). Biosynthesis of ilamycins featuring unusual building blocks and engineered production of enhanced anti-tuberculosis agents. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-00419-5

MacGowan, A., & Macnaughton, E. (2017). Antibiotic resistance. *Medicine*, *45*(10), 622–628. https://doi.org/10.1016/j.mpmed.2017.07.006

Malcolmson, S. J., Young, T. S., Ruby, J. G., Skewes-Cox, P., & Walsh, C. T. (2013). The posttranslational modification cascade to the thiopeptide berninamycin generates linear forms and altered macrocyclic scaffolds. *Proceedings of the National Academy of Sciences*, *110*(21), 8483–8488. https://doi.org/10.1073/pnas.1307111110

Mann, A., Nehra, K., Rana, J. S., & Dahiya, T. (2021). Antibiotic resistance in agriculture: Perspectives on upcoming strategies to overcome upsurge in resistance. *Current Research in Microbial Sciences*, *2*, 100030. https://doi.org/10.1016/j.crmicr.2021.100030

Marquis, R. E. (1965). Nature of the Bactericidal Action of Antimycin A for Bacillus megaterium. *Journal of Bacteriology*, *89*(6), 1453–1459. https://doi.org/10.1128/jb.89.6.1453-1459.1965

Martínez-Núñez, M. A., & López, V. E. L. y. (2016). Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes*, *4*(1), 13. https://doi.org/10.1186/s40508-016-0057-6

Matsumoto, N., Tsuchida, T., Nakamura, H., Sawa, R., Takahashi, Y., Naganawa, H., Iinuma, H., Sawa, T., Takeuchi, T., & Shiro, M. (1999). Lactonamycin, a new antimicrobial antibiotic produced by Streptomyces rishiriensis MJ773-88K4. II. Structure determination. *The Journal of Antibiotics*, *52*(3), 276–280. https://doi.org/10.7164/antibiotics.52.276

Mauvezin, C., & Neufeld, T. P. (2015). Bafilomycin A1 disrupts autophagic flux by inhibiting both V-ATPase-dependent acidification and Ca-P60A/SERCA-dependent autophagosome-lysosome fusion. *Autophagy*, *11*(8), 1437–1438. https://doi.org/10.1080/15548627.2015.1066957

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., & Breitling, R. (2011). antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, *39*(suppl_2), W339–W346. https://doi.org/10.1093/nar/gkr466

Medeot, D. B., Fernandez, M., Morales, G. M., & Jofré, E. (2020). Fengycins From Bacillus amyloliquefaciens MEP218 Exhibit Antibacterial Activity by Producing Alterations on the Cell Surface of the Pathogens Xanthomonas axonopodis pv. Vesicatoria and Pseudomonas aeruginosa PA01. *Frontiers in Microbiology*, *10*. https://www.frontiersin.org/articles/10.3389/fmicb.2019.03107

Mitcheltree, M. J., Pisipati, A., Syroegin, E. A., Silvestre, K. J., Klepacki, D., Mason, J. D., Terwilliger, D. W., Testolin, G., Pote, A. R., Wu, K. J. Y., Ladley, R. P., Chatman, K., Mankin, A. S., Polikanov, Y. S., & Myers, A. G. (2021). A synthetic antibiotic class

overcoming bacterial multidrug resistance. *Nature*, *599*(7885), Article 7885. https://doi.org/10.1038/s41586-021-04045-6

Mohr, K. I. (2016). History of Antibiotics Research. In M. Stadler & P. Dersch (Eds.), *How to Overcome the Antibiotic Crisis: Facts, Challenges, Technologies and Future Perspectives* (pp. 237–272). Springer International Publishing. https://doi.org/10.1007/82_2016_499

Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimer's & Dementia : Translational Research & Clinical Interventions*, *3*(4), 651. https://doi.org/10.1016/j.trci.2017.10.005

Montoya, J. G., Laessig, K., Fazeli, M. S., Siliman, G., Yoon, S. S., Drake-Shanahan, E., Zhu, C., Akbary, A., & McLeod, R. (2021). A fresh look at the role of spiramycin in preventing a neglected disease: Meta-analyses of observational studies. *European Journal of Medical Research*, *26*(1), 143. https://doi.org/10.1186/s40001-021-00606-7

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., Zhang, J., Weinstock, G. M., Isaacs, F., Rozowsky, J., & Gerstein, M. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biology*, *17*(1), 53. https://doi.org/10.1186/s13059-016-0917-0

Nakayama, H., Takatsu, T., Abe, Y., Shimazu, A., Furihata, K., Ikeda, K., Furihata, K., Seto, H., & Ōtake, N. (1987). Rustmicin, a New Macrolide Antibiotic Active Against Wheat Stem Rust Fungus. *Agricultural and Biological Chemistry*, *51*(3), 853–859. https://doi.org/10.1080/00021369.1987.10868081

Newbold, C. J., Wallace, R. J., Watt, N. D., & Richardson, A. J. (1988). Effect of the novel ionophore tetronasin (ICI 139603) on ruminal microorganisms. *Applied and*

Environmental Microbiology, *54*(2), 544–547. https://doi.org/10.1128/aem.54.2.544-547.1988

Nikaido, H. (2009). Multidrug Resistance in Bacteria. *Annual Review of Biochemistry*, *78*, 119–146. https://doi.org/10.1146/annurev.biochem.78.082907.145923

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., & Horinouchi, S. (2008). Genome Sequence of the Streptomycin-Producing Microorganism Streptomyces griseus IFO 13350. *Journal of Bacteriology*, *190*(11), 4050–4060. https://doi.org/10.1128/JB.00204-08

Osbourn, A. (2010). Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. *Trends in Genetics*, *26*(10), 449–457. https://doi.org/10.1016/j.tig.2010.07.001

Pang, B., Liao, R., Tang, Z., Guo, S., Wu, Z., & Liu, W. (2021). Caerulomycin and collismycin antibiotics share a trans-acting flavoprotein-dependent assembly line for 2,2'-bipyridine formation. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-23475-4

Parthasarathy, A., Wong, N. H., Cavanaugh, N. T., Steiner, K. K., Wengert, P. C., Savka, M. A., & Hudson, A. O. (2018). Whole-Genome Sequencing and Annotation of Exiguobacterium sp. RIT 452, an Antibiotic-Producing Strain Isolated from a Pond Located on the Campus of the Rochester Institute of Technology. *Microbiology Resource Announcements*, *7*(17), e01341-18. https://doi.org/10.1128/MRA.01341-18

Plackett, B. (2020). Why big pharma has abandoned antibiotics. *Nature*, *586*(7830), S50–S52. https://doi.org/10.1038/d41586-020-02884-3

Pommerehne, K., Walisko, J., Ebersbach, A., & Krull, R. (2019). The antitumor antibiotic rebeccamycin—Challenges and advanced approaches in production processes. *Applied Microbiology and Biotechnology*, *103*(9), 3627–3636. https://doi.org/10.1007/s00253-019-09741-y

Qi, D., Zou, L., Zhou, D., Zhang, M., Wei, Y., Zhang, L., Xie, J., & Wang, W. (2021). Identification and Antifungal Mechanism of a Novel Actinobacterium Streptomyces huiliensis sp. Nov. Against Fusarium oxysporum f. Sp. Cubense Tropical Race 4 of Banana. *Frontiers in Microbiology*, *12*. https://www.frontiersin.org/articles/10.3389/fmicb.2021.722661

Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., Wilkinson, B., & Hutchings, M. I. (2017). Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. *Chemical Science*, *8*(4), 3218–3227. https://doi.org/10.1039/C6SC04265A

Quaderer, R., Omura, S., Ikeda, H., & Cane, D. E. (2006). Pentalenolactone Biosynthesis. Molecular Cloning and Assignment of Biochemical Function to PtlI, a Cytochrome P450 of Streptomyces avermitilis. *Journal of the American Chemical Society*, *128*(40), 13036–13037. https://doi.org/10.1021/ja0639214

Radics, L., Incze, M., Dornberger, K., & Thrum, H. (1982). Tetramycin B, a new polyene macrolide antibiotic: The structure of tetramycins A and B as studied by high-field NMR spectroscopy. *Tetrahedron*, *38*(1), 183–189. https://doi.org/10.1016/0040-4020(82)85064-3

Raynor, B. D. (1997). Penicillin and ampicillin. *Primary Care Update for OB/GYNS*, *4*(4), 147–152. https://doi.org/10.1016/S1068-607X(97)00012-7

Reference. (n.d.). Retrieved May 11, 2023, from https://www.bioinformatics.org/sms2/reference.html

Röhl, F., Rabenhorst, J., & Zähner, H. (1987). Biological properties and mode of action of clavams. *Archives of Microbiology*, *147*(4), 315–320. https://doi.org/10.1007/BF00406126

Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A., & Oberlies, N. H. (2020). Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Natural Product Reports*, *37*(7), 868–878. https://doi.org/10.1039/C9NP00045C

Rothe, M. L., Li, J., Garibay, E., Moore, B. S., & McKinnie, S. M. K. (2019). Synthesis, bioactivity, and enzymatic modification of antibacterial thiotetromycin derivatives. *Organic & Biomolecular Chemistry*, *17*(13), 3416–3423. https://doi.org/10.1039/C8OB03109F

Sadaka, C., Ellsworth, E., Hansen, P. R., Ewin, R., Damborg, P., & Watts, J. L. (2018). Review on Abyssomicins: Inhibitors of the Chorismate Pathway and Folate Biosynthesis. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, *23*(6), 1371. https://doi.org/10.3390/molecules23061371

Sakoulas, G., Nam, S.-J., Loesgen, S., Fenical, W., Jensen, P. R., Nizet, V., & Hensler, M. (2012). Novel Bacterial Metabolite Merochlorin A Demonstrates in vitro Activity against Multi-Drug Resistant Methicillin-Resistant Staphylococcus aureus. *PLOS ONE*, *7*(1), e29439. https://doi.org/10.1371/journal.pone.0029439

Sasaki, T., Furihata, K., Shimazu, A., Seto, H., Iwata, M., Watanabe, T., & Otake, N. (1986). A NOVEL MACROLIDE ANTIBIOTIC, NOTONESOMYCIN A. *The Journal of Antibiotics*, *39*(4), 502–509. https://doi.org/10.7164/antibiotics.39.502

Shi, P., Li, Y., Zhu, J., Shen, Y., & Wang, H. (2021). Targeted Discovery of the Polyene Macrolide Hexacosalactone A from Streptomyces by Reporter-Guided Selection of Fermentation Media. *Journal of Natural Products*, *84*(7), 1924–1929. https://doi.org/10.1021/acs.jnatprod.1c00144

Shin, Y.-H. (2019). Bombyxamycins A and B, Cytotoxic Macrocyclic Lactams from an Intestinal Bacterium of the Silkworm Bombyx mori. *Organic Letters*, *21*(6), 1804–1808. https://doi.org/10.1021/acs.orglett.9b00384

Slattery, M., Rajbhandari, I., & Wesson, K. (2001). Competition-mediated antibiotic induction in the marine bacterium Streptomyces tenjimariensis. *Microbial Ecology*, *41*(2), 90–96. https://doi.org/10.1007/s002480000084

Smith, P. A., & Romesberg, F. E. (2012). Mechanism of Action of the Arylomycin Antibiotics and Effects of Signal Peptidase I Inhibition. *Antimicrobial Agents and Chemotherapy*, *56*(10), 5054–5060. https://doi.org/10.1128/AAC.00785-12

Son, S., Hong, Y.-S., Jang, M., Heo, K. T., Lee, B., Jang, J.-P., Kim, J.-W., Ryoo, I.-J., Kim, W.-G., Ko, S.-K., Kim, B. Y., Jang, J.-H., & Ahn, J. S. (2017). Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a Streptomyces Species. *Journal of Natural Products*, *80*(11), 3025–3031. https://doi.org/10.1021/acs.jnatprod.7b00660

Sr, N., Mv, P., S, C., Ek, V., Kr, D., J, B., Bn, G., & M, L. (1994). Balhimycin, a new glycopeptide antibiotic produced by Amycolatopsis sp. Y-86,21022. Taxonomy, production, isolation and biological activity. *The Journal of Antibiotics*, *47*(3). https://doi.org/10.7164/antibiotics.47.334

Static or cidal; which is best? - Microbiology Nuts & Bolts. (n.d.). Retrieved February 28, 2023, from http://www.microbiologynutsandbolts.co.uk/the-bug-blog/static-or-cidal-which-is-best

Steiner, K. K., Parthasarathy, A., Wong, N. H., Cavanaugh, N. T., Chu, J., & Hudson, A. O. (2020a). Isolation and whole-genome sequencing of Pseudomonas sp. RIT 623, a slow-growing bacterium endowed with antibiotic properties. *BMC Research Notes*, *13*(1), 370. https://doi.org/10.1186/s13104-020-05216-w

Steiner, K. K., Parthasarathy, A., Wong, N. H., Cavanaugh, N. T., Chu, J., & Hudson, A. O. (2020b). Isolation and whole-genome sequencing of Pseudomonas sp. RIT 623, a slow-growing bacterium endowed with antibiotic properties. *BMC Research Notes*, *13*(1), 370. https://doi.org/10.1186/s13104-020-05216-w

Stubbendieck, R. M., Brock, D. J., Pellois, J.-P., Gill, J. J., & Straight, P. D. (2018). Linearmycins are lytic membrane-targeting antibiotics. *The Journal of Antibiotics*, *71*(3), Article 3. https://doi.org/10.1038/s41429-017-0005-z

Sun, C., Yang, Z., Zhang, C., Liu, Z., He, J., Liu, Q., Zhang, T., Ju, J., & Ma, J. (2019). Genome Mining of Streptomyces atratus SCSIO ZH16: Discovery of Atratumycin and Identification of Its Biosynthetic Gene Cluster. *Organic Letters*, *21*(5), 1453–1457. https://doi.org/10.1021/acs.orglett.9b00208

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, *12*(7), 3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Sun, Y., Zhou, X., Tu, G., & Deng, Z. (2003). Identification of a gene cluster encoding meilingmycin biosynthesis among multiple polyketide synthase contigs isolated from

Streptomyces nanchangensis NS3226. *Archives of Microbiology*, *180*(2), 101–107. https://doi.org/10.1007/s00203-003-0564-1

Terwilliger, D. W., & Trauner, D. (2018). Selective Synthesis of Divergolide I. *Journal of the American Chemical Society*, *140*(8), 2748–2751. https://doi.org/10.1021/jacs.7b13092

Thibessard, A., Haas, D., Gerbaud, C., Aigle, B., Lautru, S., Pernodet, J.-L., & Leblond, P. (2015). Complete genome sequence of Streptomyces ambofaciens ATCC 23877, the spiramycin producer. *Journal of Biotechnology*, *214*, 117–118. https://doi.org/10.1016/j.jbiotec.2015.09.020

Tripathi, R. K., & Gottlieb, D. (1969). Mechanism of Action of the Antifungal Antibiotic Pyrrolnitrin. *Journal of Bacteriology*, *100*(1), 310–318. https://doi.org/10.1128/jb.100.1.310-318.1969

van der Voort, M., Meijer, H., Schmidt, Y., Watrous, J., Dekkers, E., Mendes, R., Dorrestein, P., Gross, H., & Raaijmakers, J. (2015). Genome mining and metabolic profiling of the rhizosphere bacterium Pseudomonas sp. SH-C52 for antimicrobial compounds. *Frontiers in Microbiology*, *6*. https://www.frontiersin.org/articles/10.3389/fmicb.2015.00693

Wang, L., Zhang, C., Zhang, J., Rao, Z., Xu, X., Mao, Z., & Chen, X. (2021). Epsilon-poly-L-lysine: Recent Advances in Biomanufacturing and Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://www.frontiersin.org/articles/10.3389/fbioe.2021.748976

Wang, R., Kong, F., Wu, H., Hou, B., Kang, Y., Cao, Y., Duan, S., Ye, J., & Zhang, H. (2020). Complete genome sequence of high-yield strain S. lincolnensis B48 and

identification of crucial mutations contributing to lincomycin overproduction. *Synthetic and Systems Biotechnology*, *5*(2), 37–48. https://doi.org/10.1016/j.synbio.2020.03.001

Watve, M. G., Tickoo, R., Jog, M. M., & Bhole, B. D. (2001). How many antibiotics are produced by the genus Streptomyces? *Archives of Microbiology*, *176*(5), 386–390. https://doi.org/10.1007/s002030100345

Webber, M. A., & Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, *51*(1), 9–11. https://doi.org/10.1093/jac/dkg050

Wiegmann, D., Koppermann, S., Wirth, M., Niro, G., Leyerer, K., & Ducho, C. (2016). Muraymycin nucleoside-peptide antibiotics: Uridine-derived natural products as lead structures for the development of novel antibacterial agents. *Beilstein Journal of Organic Chemistry*, *12*, 769–795. https://doi.org/10.3762/bjoc.12.77

Wright, G. D. (2011). Molecular mechanisms of antibiotic resistance. *Chemical Communications*, *47*(14), 4055. https://doi.org/10.1039/c0cc05111j

Wu, C., Shang, Z., Lemetre, C., Ternei, M. A., & Brady, S. F. (2019). Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug resistant bacteria. *Journal of the American Chemical Society*, *141*(9), 3910–3919. https://doi.org/10.1021/jacs.8b12087

Wu, C., van Wezel, G. P., & Hae Choi, Y. (2015). Identification of novel endophenaside antibiotics produced by Kitasatospora sp. MBT66. *The Journal of Antibiotics*, *68*(7), Article 7. https://doi.org/10.1038/ja.2015.14

Xu, L., Xu, X., Yuan, G., Wang, Y., Qu, Y., & Liu, E. (2018). Mechanism of Azalomycin F_{5a} against Methicillin-Resistant*Staphylococcus aureus*. *BioMed Research International*, *2018*, e6942452. https://doi.org/10.1155/2018/6942452

Xu, Y., & Tan, D. S. (2019). Total Synthesis of the Bacterial Diisonitrile Chalkophore SF2768. *Organic Letters*, *21*(21), 8731–8735. https://doi.org/10.1021/acs.orglett.9b03348

Yi, W., Newaz, A. W., Yong, K., Ma, M., Lian, X.-Y., & Zhang, Z. (2022). New Hygrocins K–U and Streptophenylpropanamide A and Bioactive Compounds from the Marine-Associated Streptomyces sp. ZZ1956. *Antibiotics*, *11*(11), Article 11. https://doi.org/10.3390/antibiotics11111455

Yim, G., Huimi Wang, H., & Davies, J. (2006). The truth about antibiotics. *International Journal of Medical Microbiology*, *296*(2), 163–170. https://doi.org/10.1016/j.ijmm.2006.01.039

Yin, X., & Zabriskie, T. M. (2006). The enduracidin biosynthetic gene cluster from Streptomyces fungicidicus. *Microbiology*, *152*(10), 2969–2983. https://doi.org/10.1099/mic.0.29043-0

Zabolotneva, A. A., Shatova, O. P., Sadova, A. A., Shestopalov, A. V., & Roumiantsev, S. A. (2022). An Overview of Alkylresorcinols Biological Properties and Effects. *Journal of Nutrition and Metabolism*, *2022*, 4667607. https://doi.org/10.1155/2022/4667607

Zhao, Q., Wang, M., Xu, D., Zhang, Q., & Liu, W. (2015). Metabolic coupling of two small-molecule thiols programs the biosynthesis of lincomycin A. *Nature*, *518*(7537), Article 7537. https://doi.org/10.1038/nature14137

Zheng, D., Ding, N., Jiang, Y., Zhang, J., Ma, J., Chen, X., Liu, J., Han, L., & Huang, X. (2016). Albaflavenoid, a new tricyclic sesquiterpenoid from Streptomyces violascens. *The Journal of Antibiotics*, *69*(10), Article 10. https://doi.org/10.1038/ja.2016.12

Zhou, S., Wang, F., Wong, E. T., Fonkem, E., Hsieh, T.-C., Wu, J. M., & Wu, E. (2013). Salinomycin: A Novel Anti-Cancer Agent with Known Anti-Coccidial Activities. *Current Medicinal Chemistry*, *20*(33), 4095. https://doi.org/10.2174/15672050113109990199

Zhu, X. M., Hackl, S., Thaker, M. N., Kalan, L., Weber, C., Urgast, D. S., Krupp, E. M., Brewer, A., Vanner, S., Szawiola, A., Yim, G., Feldmann, J., Bechthold, A., Wright, G. D., & Zechel, D. L. (2015). Biosynthesis of the Fluorinated Natural Product Nucleocidin in Streptomyces calvus Is Dependent on the bldA-Specified Leu-tRNAUUA Molecule. *ChemBioChem*, *16*(17), 2498–2506. https://doi.org/10.1002/cbic.201500402