Journal of High School Science

The untapped potential of the endophytic actinomyces *Streptomyces scabiei* in producing antibiotic and antitumor secondary metabolites

Dingiloglu D

Submitted: August 16, 2024, Revised: version 1, November 26, 2024, version 2, December 11, 2024
Accepted: January 6, 2025

Abstract

In an increasingly antibiotic-resistant world, the discovery of new antibiotics is essential for public health. This project used computational methods to investigate the hypothesis that strain-level investigation of the endophytic actinomyces *S. scabiei* would find a greater diversity of Biosynthetic Gene Clusters (BGC) in its genome and the presence of antitumor products. Fourteen strains from this bacterium were analyzed for strain-specific BGCs, average BGC count (sum of total BGCs/number of strains), and RIbosomally Produced and Post-translationally modified (RiPP) BGC count. It was found that the studied *S. scabiei* strains contained up to 15 strain-specific gene clusters, a BGC count of 41.6, and an average of 2.4 RiPPs. Polyketide gene clusters were discovered to be the most prominent BGC class within *S. scabiei*'s genome, on average making up 31.5% of a strain's BGC repertoire. Additionally, nine out of 14 strains contained the BGC capable of synthesizing the antitumor molecule Concanamycin A. However, it was found that a high strain-specific BGC count did not necessarily imply either a greater BGC diversity or the synthesis of a corresponding large number of unique bioactive compounds. The BGC count and/or BGC dissimilarity on the evolutionary tree alone was/were not sufficient indicator(s) for the secondary metabolite synthesizing potential of a strain.

Keywords

Antitumor, Antibiotic, Biosynthetic Gene Clusters, *Streptomyces scabiei*, RIbosomally Produced and Post-translationally modified, Endophytic actinomyces, Secondary metabolites, Polyketide synthase, Pangenome matrix, Concanamycin A

Defne Dingiloglu, Tesla STEM high school, 4301 228th Ave NE, Redmond, WA 98053, USA.
defnedingiloglu@outlook.com

Introduction

Genes found in bacteria, termed biosynthetic gene clusters (BGCs) are prolific producers of secondary metabolites that have been found to possess antibiotic and antitumor properties. The products of these BGCs contribute to 69% of all antimicrobial agents (1). The first bioactive product to be discovered was penicillin in 1928 from the genus *Penicillium*. Following this discovery, the antibiotic industry received immense support and growth, with 130 antibiotics being discovered in the remainder of the 20th century (2,3). However, bioactive product discovery decreased significantly at the start of the 21st century and it was generally thought that the natural world's antibiotic repertoire had been depleted. The challenge in finding new BGCs stems largely from the difficulty of activating these gene clusters in a laboratory setting. BGCs are activated in response to environmental stressors, such as heat, pH, nutrient availability, and osmotic pressure, and each gene cluster has a specific stressor or set of stressors that causes it to activate (4). These stressors are essential to understanding the function of a BGC, but it is difficult to predict which environmental stressor or stressors will cause which BGC to activate. Some studies predict that 75% of BGCs within a bacterial genome are not activated during *in vitro* studies (5). This dilemma has made it difficult for scientists to isolate BGC derived bioactive products from bacterial strains.

This study aimed to use genome mining to elucidate the bioactive product potential of *Streptomyces scabiei*, an endophytic actinomyces found in the roots of the common wheat plant (*Triticum Aestivum L*). Genome mining is a fairly recent practice that has been used to compare the genes of bacteria computationally, so that a single bacterium with all its BGCs activated can be isolated from nature and continuously experimented on (6-8). Genome mining has revolutionized the antibiotic industry, making access to bacteria more cost-effective, simple, and quick. Among the bacteria mined for bioactive products, *Streptomyces* is known to be the most fruitful, 55% of antibiotics discovered between 1945 and 1978 originated from the genus Streptomyces and it is the producer of almost 75% of all antibiotics (5-9). Additionally, 27% of this species belongs to the group endophytic actinomyces, meaning they are found attached to the roots of plants (10). Endophytic actinomyces are known for their potential to produce a variety of bioactive products, including those with anti-tumor properties (11). Included in this variety are alkaloids, peptides, phenolics, polyketides, and terpenoids (12). This study will focus on the RIbosomally Produced and Post translationally modified (RiPP) and polyketide synthase (PKS) classes. RiPPs are structurally diverse and have drawn attention for their unique ability to produce a multitude of products with antitumor, antiviral, and antioxidant properties (12,13). PKSs are an integral part of the production of a multitude of known antitumor products and therefore their presence in the bacterial genome may be linked with antitumor product production (14).

Despite their untapped potential, endophytic actinomyces remain relatively understudied. An estimated 10-15% of higher plant species have been investigated, with 6% of this percentage having been screened for BGC production, even though 80% of the world's population relies on herbal medicines (15). One way to increase the utilization of this

potentially untapped source of antibiotics is to investigate strain-specific antibiotic production. Strain-specific BGCs are those that are only found in one strain of a bacterial species. Conventionally, bacterial studies have taken a representative strain from a species and tested that singular genome for the bioactive products (BP) it produces. A 2015 study revealed that "each strain of a *Streptomyces* species likely harbors at least one strain-specific biosynthetic gene cluster" (8,16,17). By mining the genes of multiple strains of the same bacterial species, it is possible to find many more BGCs than previously thought existed. This statistic is amplified in *S. scabiei* because a recent study found that *Streptomyces scabiei* is the third most common bacteria in the roots of the common wheat plants found in South Australia (18). Due to *S. scabiei* being commonly grown, it has been subject to multiple mutations and convergent evolution events, allowing unique strains with differing BGCs to develop. By investigating this bacterium, it is possible to provide insight into the potential source of antibiotics. It is hence possible that each of the different strains of *S. scabiei* may contain BGCs unique to that strain, thus providing strain specific bioactive products. Hence, the study of BGCs from a specific strain may yield bioactive products with non-overlapping structure and function distinct from other strains.

The hypothesis of this study is that strain-level investigation of the endophytic actinomyces *S. scabiei* will reveal a greater diversity of BGCs in its genome and the presence of antitumor products. BGC diversity is defined as the number of unique strain-specific BGCs found in a strain. Through computational analysis of the bacterium *S. scabiei*, it was discovered that it exhibits 0-15 unique BGCs per strain. The species displays an average of 41.643 repeated and 39.143 unrepeated BGCs, with 2.457 RIPPs per strain. In addition, on average, 31.465% of BGCs within *S. scabiei* strains are polyketide synthase BGCs, a class of gene clusters known for their ability to produce antitumor products. This finding is further supported by 9 out of 14 total strains of this bacterium producing at least one antitumor product. These results provide important insight into endophytic actinomyces strain-level diversity and current bioactive product discovery efforts, while solidifying *S. scabiei*'s potential in producing antitumor products.

Methods

Analyzing proteins within the genome

Before performing analysis using genome mining tools, the genome sequences were processed for proteins that they are known to contain. This information allowed for the visualization of what proteins the species contained and provided insight into what the genome mining analysis could be envisaged to uncover. To determine these proteins, the protein and .gff files of all *S. scabiei* species were downloaded from the NIH Genome List (Genome - NCBI – NLM). Fourteen strains (Table 1) were analyzed because the other strains were too scaffolded to be accurately studied.

Table 1. Assembly for investigated *Streptomyces scabiei* strains

Assembly	GenBank
ASM9130v1	GCF_000091305.1
ASM73869v1	GCF_000738695.1
ASM73871v1	GCF_000738715.1
ASM100540v1	GCF_001005405.1
ASM148512v1	GCF_001485125.1
ASM155021v1	GCF_001550215.1
ASM155022v1	GCF_001550225.1
ASM155023v1	GCF_001550235.1
ASM155024v1	GCF_001550245.1
ASM155029v1	GCF_001550295.1
ASM157211v1	GCF_001572115.1
ASM157968v1	GCF_001579685.1
ASM215573v1	GCF_002155735.1
ASM358481v1	GCF_003584815.1

A Python script was written to parse through the protein files and search for genomes that were common among all strains. The results were then output into an Excel file. While parsing through the protein file, any proteins that were marked as hypothetical or did not have a defined name were noted. These proteins do not yet have a clearly defined function, and it is important to take notice as further analysis may reveal what their function is. A similar script was written for the .gff files. This script searched for the name and function of proteins within the file. First it output every protein that it found and second, it identified only the proteins that were common among all genomes. Performing two outputs with the .gff file provided information into the complete protein arsenal of *S. scabiei* and allowed the protein file output to be checked against the .gff file output. A third script was written to parse through the .gff files and look for proteins with a CoDing Sequence "CDS" identifier. A CDS identifier identified the protein as a viable product, and therefore, most likely to be a part of a gene cluster. All proteins that were listed with a CDS identifier were output along with the strain name, product, and location. This list allowed for a thorough visualization of the location of the proteins on each strain in relation to each other and to provide information on which proteins may constitute a given gene cluster.

Identifying BGC count within the genome

AntiSMASH (https://antismash.secondarymetabolites.org) was run in order to find BGCs within each strain of *S. scabiei*. AntiSMASH is a database of all currently discovered BGCs that takes an input sequence and uses profile Hidden Markov Models to find matches between the query and the gene clusters in its database (19). The problem with this method is that in the case of some BGCs, a single amino acid substitution can cause an entirely new gene cluster to be synthesized. Due to Markov Models relying chiefly on amino acid substitutions to calculate similarity, the output

percentage of AntiSMASH is not always accurate. In fact, a lower percentage (70%) can sometimes be more accurate than a higher percentage (99%). For the purpose of this study, all percentages were accepted as true, and visually compared against the MIBiG database (https://mibig.secondarymetabolites.org/) for greater accuracy. AntiSMASH was run first on relaxed mode, which identified the more general classes of BGCs. It was run a second time in strict mode with the MIBiG option selected. MIBiG is an additional database that outputs a Top 10 Hit list of the most probable gene clusters for a region on *S. scabiei*. This Top 10 Hit list was visually compared with the output most similar to the gene cluster. The AntiSMASH data, including the BGCs found in each strain, the BGC's class, and the BGC's predicted product was manually entered into an Excel spreadsheet. Excel's functions were used to count the number of BGCs per strain and find the average BGC count of the species (total of BGCs in each strain/number of strains). Next, BGCs of the RiPP class were counted and the total and average for this class were also found. Finally, the BGCs in each strain were separated by class and the number of the BGCs in each class were counted in order to find the most common BGC classes within *S. scabiei*'s genome.

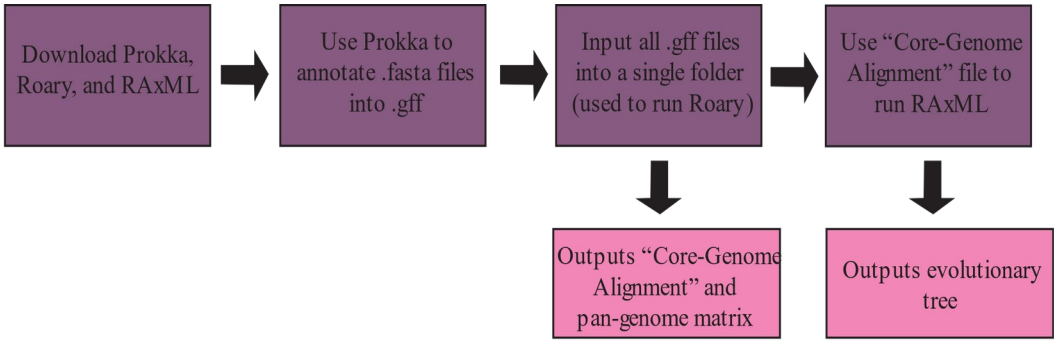

Figure 1. Prokka, Roary, and RAxML Workflow, Summarized workflow used to analyze the evolutionary history of S. scabiei. It is required to first download Prokka, Roary, and RAxML with the required additional packages. Next, Prokka is used to annotate the .fasta (nucleotide sequence) files in .gff (protein sequence) files. The .gff files are used to run Roary which outputs the "Core Gene Alignment" and a pan-genome matrix. The "Core Gene Alignment" file is used as input for RAxML, which uses this file to create an evolutionary tree of the 14 strains.

Investigating evolutionary history and homology

In order to discover more information on the evolutionary history of the 14 strains in relation to each other, homology analysis was performed (Figure 1). The .fasta files for *S. scabiei* were downloaded from the NIH Genome List. The .fasta files were then annotated into .gff files using Prokka (https://github.com/tseemann/prokka), to make sure that all proteins were labeled correctly (20). These new .gff files were then used as inputs for Roary (https://github.com/sanger-pathogens/Roary) to generate a core gene alignment (21). The core gene alignment was then used to generate the pan-genome matrix.

The pan-genome matrix showed which strains had the most alignment and therefore were the most similar. Finally, RAxML (https://github.com/amkozlov/raxml-ng) was used to generate an evolutionary tree with the core gene alignment file to check Roary's results (22) The evolutionary tree generated by RAxML was visualized using the Interactive Tree of Life (https://itol.embl.de/).

Results

BGC counts align with other Streptomyces species

AntiSMASH was run on two different modes, relaxed and strict. The Relaxed mode searched for moderately similar gene clusters between the database and query sequence. The Strict mode searched for similar gene clusters that contained full length protein sequences. The AntiSMASH analysis found that *S. scabiei* possessed a total of 675 BGCs among all fourteen strains on the relaxed mode and 583 BGCs on the strict mode. Average BGC counts were calculated by dividing this total BGC count by the number of strains. Average BGC counts of 48.214 and 41.643 were found on the relaxed and strict modes respectively. The number of BGCs per strain when antiSMASH was run on strict mode is shown in Figure 2.A. Only the total counts for strict are visualized because they are more precise than the relaxed run. There was an average of 1.833 RiPP BGCs on the relaxed mode and 2.357 on the strict mode (Figure 2.B). This data is considered standard for *Streptomyces* species. The variance between the results of the relaxed and strict runs was because strict antiSMASH searched for complete sequences; essentially, what the relaxed mode identified as two different gene clusters were counted as one when run using the strict mode. This was also supported by the fact that the difference between the total and unrepeated BGCs was less for the strict run than that for the relaxed run.

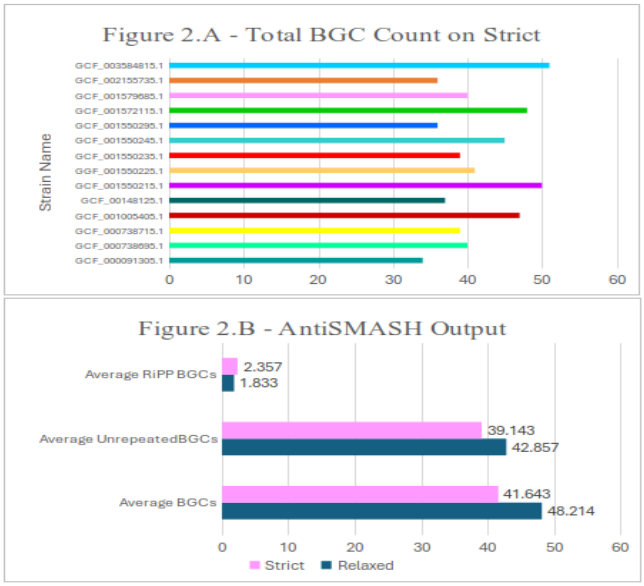

Figure 2. Output of AntiSMASH. (A) Number of BGCs in each strain based on result of strict antiSMASH run. (B) Average number of RiPP and general BGCs for both relaxed and strict antiSMASH runs.

BGCs with antitumor potential and strain-specific diversity found

Examining the BGCs found in *S. scabiei* by the class they belonged to showed that the most prominent class of BGC in *S. scabiei*'s genome were the polyketide synthases. This class of BGC constituted 31.465% of *S. scabiei*'s total BGC count with type 1 PKS found in all strains and type 2 PKS found in 12 out of 14 strains (Figure 3.A). The other common classes of BGCs in *S. scabiei* were terpenes, Non-Ribosomal Peptide synthases (NRPs), and RiPPs (specifically lanthipeptides, NI-Siderophores, and bottromycin). Previous studies have found *Streptomyces* species to contain up to 37% PKS in their genome (8, 23). The difference between the percentage found in this study and previous ones could be attributed to the fact that different species were investigated and therefore a difference in the BGCs they produced was expected due to the large diversity of the *Streptomyces* genus. For example, bottromycin was found to be prominent in *S.* scabiei but is rare among other species (8). Nine out of fourteen of the strains encoded for the antitumor BGC Concanamycin A (24). These findings supported the potential of *S. scabiei* to produce a variety of useful products, specifically with antitumor properties, due to PKSs being part of the production of antitumor products. Additionally, there were 0-15 unique BGCs per strain (Figure 3.B). Strain GCF_000738695.1 contained the maximum number of unique BGCs, suggesting that it was the most evolutionarily diverse strain among the 14 strains studied.

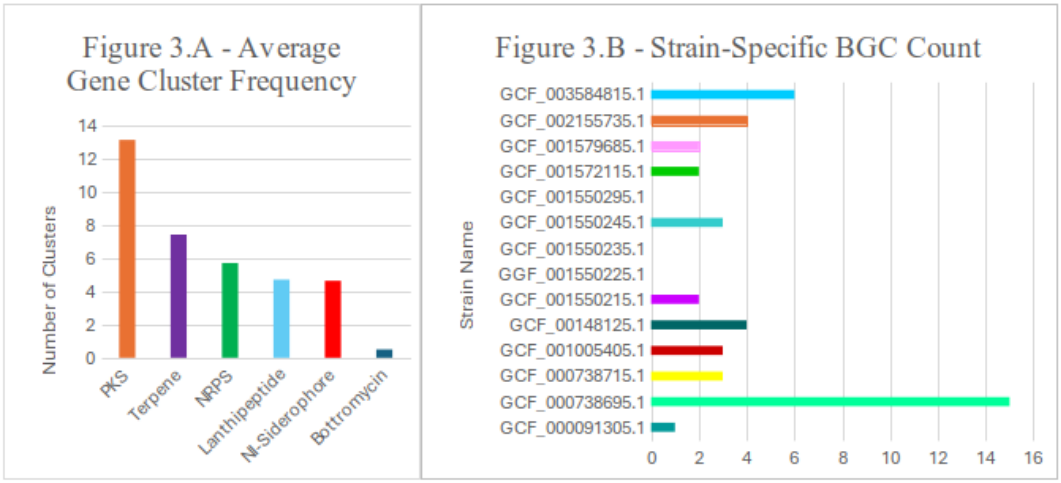

Figure 3. AntiSMASH Results, (A) All BGCs that are found within each strain based on strict AntiSMASH results. (B) Polyketide Synthase (PKS), Terpene, NRPs, and RiPPs are the most commonly found classes of BGCS. (C) Count of BGCs that are only found in one strain out of the 14 studied. GCF_000789695.1 has 15 unique BGCs, the most out of any strain.

Gene alignments reveal two unique strains

The pangenome matrix output by Roary was constructed using 100 bootstrap replicates. It contains an evolutionary tree on the left and a matrix on the right. Each row in the matrix matches a corresponding bacterium from the evolutionary tree, and each blue line within the row represents a gene cluster. An alignment in

the strains' DNA sequence is shown by blue lines from different rows aligning vertically with one another. The matrix showed that strains GCF_002155 and GCF_000738 were the most dissimilar from the other 12 strains (Figure 4). This indicated that they were the most evolutionarily distant strains of *S. scabiei*. Strains GCF_001550 and GCF_000091 were the ancestors of the remaining 12 strains as they were located on the outermost bracket and the other strains branched off them. Additionally, strains GCF_002155 and GCF_000738 showed the least gene alignments with the other strains, each blue line represented a gene cluster, while GCF_001550 and GCF_000091 possessed the most.

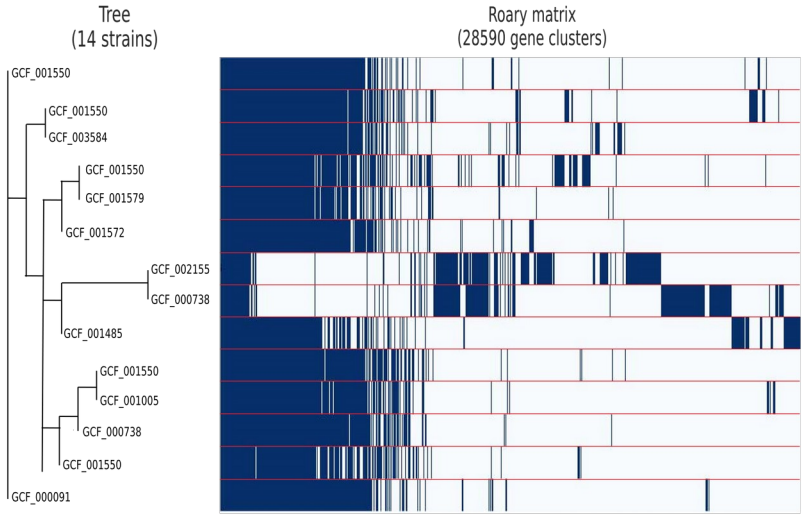

Figure 4. Pan-Genome Matrix, Evolutionary tree and pangenome matrix. Each blue line represents a gene cluster. Where the lines align on top of each other represents an alignment in the DNA sequence of the bacterial strains. Therefore, more alignments indicate a greater similarity.

Discussion

The hypothesis that strain-level investigation of the endophytic actinomyces *S. scabiei* would reveal a greater diversity of BGCs in its genome and the presence of antitumor products was not supported. *S. scabiei* contains the standard BGC counts and diversity for *Streptomyces* species. Nine out of 14 strains produced the antitumor product Concanamycin A. The presence of polyketide synthases in the *Streptomyces* has been well established and the number of PKS BGCs found in *S. scabiei* is not more than other studied *Streptomyces* species. Previous studies have found up to 80 BGCs within a single *Streptomyces* species (8). This shows that *S. scabiei* does not have a greater BGC number propensity than other species and, in fact, the number of BGCs its genome contains is standard for *Streptomyces*. These findings support the ability of endophytic actinomycin to contribute to antitumor product development efforts.

The hypothesis of this study was based on an implicit inference that *S. scabiei* would be a prolific producer of bioactive compounds. The *Streptomyces* species has long been recognized for its ability to produce a large variety of

products that are useful in human medicine (23). A 2003 study found that *S. scabiei* is the 3rd most common bacteria found in the roots of the common wheat plant, *Triticum aestivum L.*, which is widely grown (18). This study also found that many strains that match *S. scabiei*'s genome except that they do not produce the toxin nec1, meaning that there are other varieties of this species evolving in the wheat plant's roots. Due to *Streptomyces'* known potential, *S. scabiei*'s high bioavailability, and its shown rapid evolution, it was inferred that *S. scabiei* would be able to produce a larger variety of BGCs and bioactive products compared to previously studied *Streptomyces* species.

It is important to note that the results may have been affected by the scaffolded nature of *S. scabiei*'s nucleotide sequence. There was only one complete scaffold and the other 13 strains' nucleotide sequences were divided into thousands of scaffolds. This could have left nucleotide gaps between the scaffolds, causing the BGC output to be altered. Additionally, the outputs of the pan-genome matrix and Figure 3.B differed. Strain GCF_000738695.1 was both the most dissimilar strain in the pan-genome matrix and possessed the greatest number of unique products. However, although strain GCF_002155 was seen as dissimilar on the evolutionary tree, it possessed the standard number of unique products. This implies that strain-specific BGC counts do not necessarily mean that an exclusive study of GCF_000738695 will yield the greatest number of bioactive compounds. Based on these results, it can be concluded that BGC count alone and/or BGC dissimilarity on the evolutionary tree is/are not a sufficient indicator for the BGC potential of a strain. However, there is evidence that "*Streptomyces* strains with a greater number of BGCs might also contain a greater number of cryptic silent BGCs" (23). Therefore, further study of GCF_000738695.1 may reveal hidden BGCs, but it is not clear whether the extent to which the number of unique products this strain contains influences its BGC diversity (the number of different BGCs that are present in the strain). In fact, studying the other 13 strains simultaneously may yield many more compounds than studying GCF_000738695 alone, because none of them are assumed to be subsets of GCF_000738695. Previous studies have found that the presence of strain-specific BGCs in individual strains of a bacterial species is "indicative of recent gene loss and horizontal gene transfer" (8). This also shows that it would be more beneficial to study all strains of *S. scabiei*, rather than just GCF_000738695, because the diversity of one strain may very well depend on the BGC production of another strain(s).

Future studies could include searching for the function of all discovered BGCs. Finding the chemical structure and function of BGCs is the most complicated part of the genome mining process. These gene clusters often have synthesis pathways that are difficult to trace from start to finish, because of this, the function of many BGCs from the antiSMASH output is unknown. One way to accomplish this could be to run a BLAST search to find the functions of the individual proteins within the gene clusters and using this information to find the function of the BGC. To more accurately ascertain the RiPP repertoire of *S. scabiei*, the leader peptide of each gene cluster can be analyzed. The leader peptide is a protein in the RiPP BGC that remains unchanged throughout

the gene cluster's entire synthesis, acting as an identifier for the gene cluster. Analyzing this protein could reveal more RiPP gene clusters within the *S. scabiei* genome and a more accurate identification of the product(s) of the RiPP gene cluster.

An additional future step would be to identify/analyze cryptic BGCs, those whose products are "hidden" due to low expression under laboratory conditions, which have also hindered antibiotic discovery efforts (25). Cryptic BGCs lower the perceived production of a strain, therefore, the number of BGCs that can be uncrypticized can serve as an additional measure of the biosynthetic potential of a bacterium. A study investigating *Streptomyces griseus* found that " 2- to 3-bp lesions [insertion-deletion lesions, INDELs] between −10 promoter regions and AdpA operator sites can cause substantial differences in transcription strength and biosynthetic output" (26). Based on these findings, investigating the indels surrounding the transcription site of a particular BGC may reveal insight into its productivity. Understanding the productivity level of a BGC can allow a decision to be made as to whether the gene cluster should be researched further. If a BGC with a desired product is found, indels can artificially be copied to increase production of the BGC in multiple strains of a bacterium.

These analyses will contribute to ongoing cancer research as well as reveal more information about the stressor-gene cluster relationship in bacteria. By continuing to employ genome mining techniques on endophytic actinomyces bacteria, it may be possible to uncover more producers of antitumor products. Ultimately, this research will provide important insight into antibiotic and antitumor product research.

Conclusion

The BGCs of fourteen individual strains of *Streptomyces scabiei* were investigated to determine if the number of BGC counts on an individual strain, combined with that strain's greater diversity on the evolutionary tree would enable that strain to yield a proportionally larger number of bioactive products such as antibiotics and antitumor compounds. This was found not to be the case. From a drug discovery viewpoint, it is therefore more productive to investigate more strains of a particular species – rather than focus on the most diverse strain with the largest number of BGCs – in order to increase the probability of finding secondary metabolites that can function as antibiotics and/or antitumor compounds.

Acknowledgments

I would like to thank my mentor, Dr. Manoj Pratim Samanta and Mrs. Kate Allender.

References

1. Park CJ, Andam CP. Within-Species Genomic Variation and Variable Patterns of Recombination in the Tetracycline Producer *Streptomyces rimosus*. Frontiers in microbiology, 10: 552, 2019. https://doi.org/10.3389/fmicb.2019.00552

2. Sepkowitz KA. One Hundred Years of Salvarsan. New England Journal of Medicine, 365 (4): 291–293, 2011. https://doi.org/10.1056/NEJMp1105345

3. Powers JH. Antimicrobial drug development – the past, the present, and the future. Clinical Microbiology and Infection, 10 (s4): 23–31, 2004. https://doi.org/10.1111/j.1465-0691.2004.1007.x

4. Hernandez DJ, David AS, Menges ES, Searcy CA, Afkhami ME. Environmental stress destabilizes microbial networks. ISME J, 15: 1722–1734, 2021. https://doi.org/10.1038/s41396-020-00882-x

5. Hutchings MI, Truman, AW, Wilkinson B. Antibiotics: past, present and future. Current Opinion in Microbiology, 51: 72–80, 2019. https://doi.org/10.1016/j.mib.2019.10.008

6. Rutledge P, Challis G. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol, 13: 509–523, 2015. https://doi.org/10.1038/nrmicro3496

7. Zhu JW, Zhang SJ, Wang WG, Jiang H. Strategies for Discovering New Antibiotics from Bacteria in the Post-Genomic Era. Curr Microbiol, 77: 3213–3223, 2020. https://doi.org/10.1007/s00284-020-02197-8

8. Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. Sci Rep, 10: 2003, 2020. https://doi.org/10.1038/s41598-020-58904-9

9. Arulprakasam KR, Dhanasekaran D. Genome mining of biosynthetic gene clusters intended for secondary metabolites conservation in actinobacteria. Microbial Pathogenesis, 161, 2012. https://doi.org/10.1016/j.micpath.2021.105252

10. Lacey HJ, Rutledge PJ. Recently Discovered Secondary Metabolites from *Streptomyces* Species. Molecules (Basel, Switzerland), 27(3): 887, 2022. https://doi.org/10.3390/molecules27030887

11. Zotchev SB. Unlocking the potential of bacterial endophytes from medicinal plants for drug discovery. Microbial Biotechnology, 17(2): e14382, 2024. https://doi.org/10.1111/1751-7915.14382

12. Anavadiya B, Chouhan S, Saraf M, Goswami D. Exploring endophytic actinomycetes: A rich reservoir of diverse antimicrobial compounds for combatting global antimicrobial resistance. The Microbe, 4, 2024. https://doi.org/10.1016/j.microb.2024.100110

13. Aggarwal E, Chauhan S, Sareen D. Thiopeptides encoding biosynthetic gene clusters mined from bacterial genomes. Journal of Biosciences, 46(2), 2021. https://doi.org/10.1007/s12038-021-00158-2

14. Gao Q, Deng S, Jiang T. Recent developments in the identification and biosynthesis of antitumor drugs derived from microorganisms. Engineering Microbiology, 2(4), 2022. https://doi.org/10.1016/j.engmic.2022.100047

15. Alvin A, Miller KI, Neilan BA. Exploring the potential of endophytes from medicinal plants as sources of antimycobacterial compounds. Microbiological Research, 169(7): 483-495, 2014. https://doi.org/10.1016/j.micres.2013.12.009

16. Seipke RF. Strain-Level Diversity of Secondary Metabolism in *Streptomyces albus*. PLoS ONE 10(1): e0116457, 2015. https://doi.org/10.1371/journal.pone.0116457

17. Vicente CM, Thibessard A, Lorenzi J.N, Benhadj M, Hôtel L, Gacemi-Kirane D, *et al*. Comparative Genomics among Closely Related *Streptomyces* Strains Revealed Specialized Metabolite Biosynthetic Gene Cluster Diversity. Antibiotics (Basel, Switzerland), 7(4): 86, 2018. https://doi.org/10.3390/antibiotics7040086

18. Coombs JT, Franco CMM. Isolation and Identification of Actinobacteria from Surface-Sterilized Wheat Roots. Applied and Environmental Microbiology, 69(9): 5603–5608, 2003. https://doi.org/10.1128/aem.69.9.5603-5608.2003

19. Blin K, Shaw S, Augustijn HE, Retix ZL, Biermann F, Alanjary M, *et al*. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res, 51(W1): W46-W50, 2023. https://doi.org/10.1093/nar/gkad344

20. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics, 30(14): 2068-2069, 2014. https://doi.org/10.1093/bioinformatics/btu153

21. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, *et al*. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics, 31(22): 3691-3693, 2015. https://doi.org/10.1093/bioinformatics/btv421

22. Kozlov AM, Darriba D, Flouri R, Morel B, Stamatakis A. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics, 35(21): 4453-4455, 2019. https://doi.org/10.1093/bioinformatics/btz305

23. Chung YH, Kim H, Ji CH, Je HW, Lee D, Shim SH, *et al*. Comparative Genomics Reveals a Remarkable Biosynthetic Potential of the *Streptomyces* Phylogenetic Lineage Associated with

Rugose-Ornamented Spores. mSystems, 6, 2021. https://doi.org/10.1128/msystems.00489-21

24. Kiyoshima T, Yoshida H, Wada H, Nagata K, Fujiwara H, Kihara M, *et al*. Chemoresistance to concanamycin A1 in human oral squamous cell carcinoma is attenuated by an HDAC inhibitor partly via suppression of Bcl-2 expression. PLoS One, 8(11): e80998, 2013. https://doi.org/10.1371/journal.pone.0080998

25. Hoskisson PA, Seipke RF. Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. mBio, 11, 2020. https://doi.org/10.1128/mbio.02642-20

26. Qi Y, Nepal KK, Blodgett JAV. A comparative metabologenomic approach reveals mechanistic insights into *Streptomyces* antibiotic crypticity. PNAS, 118(31): e2103515118, 2021. https://doi.org/10.1073/pnas.2103515118